

سلسلة إصدارات مركز التميز في اللغة العربية

جامعة محمد بن زايد للعلوم الإنسانية

2023 - 2

كتاب

المؤتمر الدولي الأول

اللسانيات الحاسوبية واللغة العربية

التصورات والتطبيقات والتطلعات

(أبوظبي 25 - 26 أكتوبر 2022م)

كتاب البحوث المنشورة

إصدارات محكمة

اسم الكتاب: المؤتمر الدولي الأول للسانيات الحاسوبية واللغة العربية
التصورات والتطبيقات والتطلعات
تأليف: مجموعة مؤلفين
الطبعة الأولى
سنة النشر: 2023
رقم الإيداع: MC-10-01-8545092
الترقيم الدولي: 978-9948-783-96-1

جميع الحقوق محفوظة

لا يجوز نشر أي جزء من هذا الكتاب أو تخزين مادته بطريقة
الاسترجاع أو نقله على أي نحو أو بأية طريقة سواء كانت
إلكترونية أو ميكانيكية أو خلاف ذلك إلا بموافقة مركز
التميز في اللغة العربية - جامعة محمد بن زايد للعلوم الإنسانية -
على هذا كتابة ومقدماً.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

المحتويات

كلمة معالي الشيخ/ عبد الله بن بيه

في مؤتمر اللسانيات الحاسوبية واللغة العربية يومي 25 و 26 أكتوبر 2022م..... 7

- 1- أ. د. عمر مهديوي: نحو تطوير برمجيات حرة ومفتوحة المصدر في خدمة تطبيقات المعالجة الآلية للغة العربية..... 13
- 2- د. أسامة عبد الوهاب ريس: مستقبلات اللغة العربية في فضاء الويب الثالث (WEB 3) والعوالم الافتراضية (الميتافيرس)..... 53
- 3- د. فاطمة الغنام: تمثيل النصوص العربية وتصنيفها بناء على نهج ثنائي متواليات الحروف الهجائية..... 87
- 4- أ.د. محسن عبد الرازق رشوان & د. نعيم محمد عبد الغني: التحقق من النطق الصحيح باستخدام تقنية الحاسوب..... 125
- 5- د. سندس كرونة: إشكالات المعالجة الآلية للغة العربية وآفاق تطويرها بنك المشجرات العربية أنموذجا..... 161
- 6- د. أفراح التميمي: استكشاف أداء نموذج التميمي المحسن للتوسيم النحوي الآلي على سور من القرآن الكريم..... 205
- 7- د. أكرم محمد زكي & د. بسام عركوك: التصنيف الموضوعي للنص العربي باستخدام التصنيف غير المتوازن مواضيع القرآن الكريم أنموذجا..... 233

- 8- أ.د. صديق جوهر: مزلق الترجمة الآلية- دراسة تطبيقية على نصوص مختارة 265
- 9- د. حسام الدين سمير عبد العال محمد: نَمْدَجَةُ التَّمَثَلَاتِ الصَّرْفِيَّةِ العربية وتعليميتها بواسطة الحاسوب في ضوء تحليلٍ لِأَمْثَلَةٍ نَصِيَّةٍ تطبيقية 307
- 10- م. إلهام المفشى & د. أميمة الدكاك: تصنيف صوتيمات اللغة العربية باستخدام نموذج طرف إلى طرف في سبيل كشف أخطاء النطق 429
- 11- د. عيسى صالح الحمادي: إطار عام مرجعي لتوظيف التقنية في تعليم اللغة العربية وتعلّمها 457
- 12- د. محمد بن عبد العزيز العميري: توصيف العربية حاسوبياً لغير الناطقين بها - إشكالات وتحديات - 525
- 13- د. بدرية براك بنيه العنزي: تصور مقترح لبناء معجم رقمي للمتصاحبات اللفظية (فكرة تطبيق ذكي) 569
- 14- أ.د. بلقاسم الجطاري: رقمنة اللغة العربية وتحديات الانخراط في عالم الذكاء الاصطناعي 591
- 15- د. شائق عبد الرحمان: التعابير المسكوكة في اللغة العربية من التوصيف اللساني إلى المعالجة الحاسوبية 607
- 16- د. محمد الغليظ: الأصوات العربية" وتعليمها للناطقين بغيرها 643

كلمة معالي الشيخ / عبد الله بن بيه

في مؤتمر اللسانيات الحاسوبية واللغة العربية

يومي 25 و26 أكتوبر 2022م

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

الحمد لله رب العالمين، والصلاة والسلام على سيدنا محمد النبي الخاتم،
وعلى إخوانه من الأنبياء والمرسلين.

معالي الدكتور حمدان مسلم المزروعى رئيس مجلس أمناء الجامعة.

أصحاب المعالي والسعادة، أصحاب السماحة والفضيلة.

أيها المشاركون كل باسمه وجميل اسمه، السلام عليكم ورحمة الله وبركاته.

وبعد

يسرني الحضور معكم اليوم مرحباً بضيوف جامعتنا الكرام مثنيا على جهود
إدارة الجامعة في تنظيم هذه الفعالية الهامة ودعوة هذه النخبة المميزة من أعلام
الفكر واللغة وعلماء الحاسوب.

إن موضوع مؤتمركم هذا هو اللغة العربية في جانب من جوانبها وزاوية من
زواياها تتعلق بالتجديد، إن لم يكن في الأساس فهو تجديد في الوسائل، وذلك
يندرج في رؤية دولتنا، دولة الإمارات العربية المتحدة، وقيادتنا الرشيدة في تشجيع
التجديد والابتكار في شتى الميادين ومختلف الفنون.

إن اللغة العربية جديرة بأن تنال حظها وتحظى بنصيبها في اهتمام جامعاتنا
ومؤسساتنا العلمية وبخاصة تلك التي تتمتع بطموح وسعي للتميز كمثال

جامعتنا، جامعة محمد بن زايد للعلوم الإنسانية، ولذلك كان هذا المؤتمر بعنوان: "مؤتمر اللسانيات الحاسوبية واللغة العربية". وهو موضوع له جدته وجديته لما يمثل من سعي لتمكين اللغة العربية من الولوج في مجالاتٍ ظلت حصرًا على بعض اللغات العالمية.

وإذ أرحب بكم باسم جامعتنا وأشارككم هذا الاجتماع العلمي الرفيع فأود أن أقول كلمة وجيزة عن هذه اللغة العربية القديمة الحديثة. إن تاريخ اللغات وبخاصة السامية يصعب تحديده بزمان أكيد إلا أن اللغة العربية حسب الروايات التاريخية ترجع إلى حوالي أربعة آلاف سنة إذ أن المؤرخين يذكرون أن سيدنا إسماعيل عليه السلام تعلم اللغة من جرههم وهي من القبائل العربية البائدة من أصول قحطانية كجديس وطسم وعاد وثمود. هذه هي الروايات التي يذكر المؤرخون العرب. ولا توجد نصوص أدبية ترجع إلى ذلك العهد بل إن بعض الروايات التي لا يمكن الحكم بصحتها تنسب شعراً إلى أبي البشرية آدم باللغة العربية.

تغيرت البلاد ومن عليها فوجه الأرض مغرب قبيح

إلا أن الله علم آدم الأسماء كما جاء في القرآن الكريم فإذا كانت الأسماء هي اللغات كما يروى عن ابن عباس فلا شك أن العربية كانت من بينها. مهما يكن من أمر فالأشعار التي وصلت إلينا ترجع إلى أقل من ألفي سنة كأشعار الأفوه الأودي وغيره.

ولقد صارت العربية بفضل القرآن الكريم لغةً عبادَةٍ، وهي أيضاً وسيلة فهم الشريعة، هذه الوظيفة التعبدية للغة العربية (على خلاف عن أبي حنيفة الذي رأى قراءة القرآن باللغة الدارية)، وهذه الميزة ربما لا يشارك فيها من اللغات الحية إلا العبرية القديمة. وأما النصارى الكاثوليك فكانوا يتعبدون باليونانية ثم

اللاتينية، باعتبار أن الأولى هي التي ترجم إليها العهد الجديد من النسخة الأصلية الضائعة التي لا يعرف أكانت باللغة السريانية أم بالعبرية؟ وقد سمح الفاتيكان في المجمع المسكوني لسنة 1965 بأداء الطقوس بلغة كل قوم.

أما خارج الوظيفة الدينية، فإن اللغة العربية كسائر اللغات في أداءٍ وَظِيفَتَي التَّعْبِيرِ والاتِّصَالِ، التعبير عن الأفكار والتواصل مع الآخرين، Communication. ولذلك فإن العربية محترمة ومعتبرة، وهو احترام واعتبار تشاركها فيه عند بعض العلماء كلُّ اللغات لأن الحرمة للحروف كما يقولون. كما أن اللغة العربية هي لغة الضاد ولغة المثنى أي صيغة التثنية وهي ما دل على اثنين وأغنى عن المتعاطفين بزيادة ألف ونون في حالة الرفع ، أو ياء ونون في حالتي النصب والجر، على أن اللغة العربية -فيما علمنا- احتفظت بألفاظ قليلة تأتي فيها صيغة التثنية.

أيها الحضور الكريم:

إن اللغة العربية لغة تنتهي للمستقبل بخوارزمياته وعملياته المعقدة. وإن نظرة على تاريخ تطورها توضح هذا المنحى. فقد كان أحد روادها الأوائل وهو الخليل بن أحمد يتمتع بعقلية رياضية. يذكر أنه توفي وهو يبحث في معادلات رياضية دقيقة. من خلال ذلك اخترع العروض ووضع أول قاموس للغة العربية "كتاب العين" وكانت منهجيته أن يرتب الأحرف حسب ترتيبها في المخارج متدرجة من حلقية إلى شجرية إلى ذلقية إلى شفوية. أو ما سماه الشاطبي موازين الحروف.

وَهَاكَ مَوَازِينَ الْحُرُوفِ وَمَا حَكَى ... جَهَابُذَةُ النُّقَادِ فِيهَا مُحَصَّلًا

وَلَا رَيْبَةَ فِي عَيْنَيْنِ وَلَا رَيْبًا ... وَعِنْدَ صَلِيلِ الزَّيْفِ يَصْدُقُ الْإِتِّبَالُ

فيأخذ جذراً وينظر تقلباته والإمكانات المتاحة لتصاريف هذا الجذر وبناء على هذه الإمكانيات يضع قائمة للأصول المستعملة في مقابل الأصول المهملة. على سبيل المثال ج ذ ب هو جذر ثلاثي -والأقل من الثلاثي كما تعرفون لا يكون أصلاً.

وليس أدنى من ثلاثي يرى *** قابل تصريف سوى ما غيرا

كما يقول ابن مالك، هذا الثلاثي له تصاريف مختلفة وإن شئت فقل تقلبات فهذه التصاريف بناء عليها وعلى ما سمع من العرب فيها نضع قائمة الأصول المستعملة بعضها يكون كل هيئة منه أصلاً وبعضها تكون فرعاً. والفرق بين الأصل والفرع أن الأصل يتصرف والفرع لا يتصرف.

والجذر "ض ر ب"، يمكن أن نأخذ منها "ضرب" وهو مستعمل و"ضبر" و"برض" و"رضب" و"برض" و"بضر" وكلها أصول مستعملة. ومن برض قول ذي الرمة:

"رعت بارض الهى جميماً وبسرةً وصمعاء حتى أنفتها نصالها".

والمراد أن اللغة العربية بدء تدوينها بصفة شبه رياضية مذ عني بها الخليل ومن قبله قام أبو الأسود الدؤلي بوضع الشكل على الحروف كما أسس الخليل العروض كذلك بأشكال شبه موسيقية. وقد اهتم بتصاريف اللغة، والصرف يختلف عن النحو فهو يهتم ب"كينونة اللفظ" أما النحو فيهتم بأواخر الكلم وعلاقته بالعوامل الخارجية. فالحروف وأشباهاها لا تتصرف بمعنى أن هيئاتها لا تتقلب:

حرف وشبهه من الصرف بري *** وما سواهما بتصريف حري

وإذا كانت اللغة العربية بدأت هكذا فهي إذاً قابلة للتطوير والتطور. ويمكن أن نعتبر أن الأجيال التي تعاقبت اهتمت بالنحو على حساب الصرف وأن ابن جني وشيخه أبا علي الفارسي لم يتابعهما الناس في عنايتهما بما يسمى بالاشتقاق الكبير أو الأكبر. وهو ما يدعى بالقلب المكاني وهذا معنى الاشتقاق الكبير، أما الاشتقاق الأكبر فهو قلب حرف إلى حرف آخر فالأحرف تتعاقب وتتعاور.

فهل يمكن الآن أن نعود إلى أصول هذه الصناعة عند الخليل وعند ابن جني

وشيخه لنفيد منها في الحاسوبية علماً أن الأبجدية العربية وإن اشتملت على ثمانية أو تسعة وعشرين حرفاً فيمكن رجوعها إلى عدد أقل عندما نعرف أن أربعة أحرف تمثل حرفاً واحداً الباء والتاء والثاء والياء هي في الحقيقة حرف واحد إذا عريناه من النقاط التي دخلت متأخراً في رسم القرآن الكريم، وكذلك الجيم والحاء والخاء حرف واحد. وحتى التشكيل فإن الفتحة والكسرة كما يقول أهل رسم القرآن عبارة عن ألف مبطوحة في الأعلى والأسفل والضمة عبارة عن واو مطموسة الرأس.

فإذا كانت اللغة العربية من مبتدئ تدوينها دائرة على منطق دقيق واستقراء عميق فأحرى أن تكون بذلك أقرب وأيسر في البرمجة والحوسبة التي مدارها كذلك على المنطق الرياضي. والمراد هنا كيف نستطيع من خلال إمكانات الحاسوب البرمجية ومتاحات الذكاء الصناعي استيعاب اللغة العربية بمجازاتها واستعاراتها وكنياتها وظاهر دلالاتها...ذلك ما نأمل أن يسهم هذا المؤتمر فيه من خلال التعاون والتكامل بين علماء اللغة والمختصين في التقنية والشراكات بين المؤسسات الأكاديمية والقطاع الخاص.

هذه أفكار متناثرة لكن المختصين والخبراء في مؤتمرنا هذا بإمكانهم أن يجدوا فيها أو في الرجوع إلى أصلها ما ينجدهم في البرمجة ويسعدهم في الحوسبة.

وختاماً، أشكركم جميعاً وأدعو الله أن يبقي بلدنا هذا بلد خير ونماء وعطاء ونفع للناس، وأن يحفظ قيادتنا الرشيدة من كل ضر وباس، وأن يصلح أعمالنا وأن يجعل اجتماعنا اجتماع خير وأن يكمله بالنجاح إنه ولي ذلك والقادر عليه.

والسلام عليكم ورحمة الله وبركاته

**نحو تطوير برمجيات حرة ومفتوحة المصدر
في خدمة تطبيقات المعالجة الآلية
للغة العربية**

**Towards the Development of Free and Open-
Source Software in the Service of Automatic
Processing Applications for the Arabic
Language**

**أ.د. عمر مهديوي
جامعة مولاي إسماعيل
مكناس- المملكة المغربية**

الملخص

يندرج هذا البحث في إطار برمجيات معالجة اللغات الطبيعية عامة، ومعالجة اللغة العربية خاصة، ذلك أن المعالجة الحاسوبية للغات تعدُّ مجالاً معرفياً بينياً، يقع في مفترق الطرق بين اللسانيات، والمعلومات، والذكاء الاصطناعي، والعلوم المعرفية. وتنحصر أهدافه في بلورة برامج وتقنيات حاسوبية متطورة جداً، الغاية منها معالجة المستويات اللسانية للغات البشرية تحليلاً وتوليداً.

وسعيًا منا في تضيق الفجوة الرقمية بين العربية، واللغات الأجنبية في مجال الحوسبة اللغوية، فقد خصّصنا هذا البحث للتعريف بالبرمجيات الحرة ومفتوحة المصدر من جهة، ومن جهة أخرى، الوقوف عند دورها باعتبارها آلية لدعم المحتوى الرقمي للغة العربية وتعزيزه على الشبكة، وبالتالي تبيان وظيفتها في خدمة التطبيقات الحاسوبية للغة العربية عامة.

الكلمات المفتاحية: برمجيات حرة - مفتوحة المصدر - المحتوى الرقمي - التطبيقات الحاسوبية - اللغة العربية.

Abstract

This research falls within the framework of natural language processing software in general, and Arabic language processing in particular, because computer processing of languages is an interdisciplinary field, located at the crossroads between linguistics, informatics, artificial intelligence, and cognitive science. Its objectives are limited to the crystallization of highly advanced computer programs and techniques, the purpose of which is to address the linguistic levels of human languages, analyzing and generating.

In our endeavor to narrow the digital gap between Arabic and foreign languages in the field of linguistic computing, we have devoted this research to defining free and open-source software on the one hand, and on the other hand standing at its role as a mechanism to support and enhance the digital content of the Arabic language on the net, and thus demonstrate its functionality in serving Computational Arabic language applications in general.

Keywords: free software - open source - digital content - computer applications - Arabic language.

المقدمة

إن الباعث الأساس إلى الاهتمام بالبرمجيات الحرة، ومفتوحة المصدر في علاقتها باللغة العربية، يعود إلى ما يمكن أن تقوم به من دور رئيس في دعم المحتوى الرقمي للغة العربية وتعزيزه على الشبكة من جهة، ومن جهة أخرى فهي آلية ذات فاعلية كبيرة يمكن أن تفيد التطبيقات الحاسوبية للغة الضاد؛ علمًا بأن النهوض باللغة العربية وترقيتها تنظيرًا وممارسةً، وتطوير أدوات معالجتها حاسوبيًا، أصبح من الاستراتيجيات العلمية، والحضارية، والتقنية، التي لا غنى عنها في ظل مجتمع المعرفة؛ إنها حاجة ملحة تملحها طبيعة اللغة العربية، التي تنوق إلى مواكبة التقدم التقني الذي وصلت إليه لغات العالم المتقدم، وهي حاجة ملحة أيضًا ترتبط بتعطش أبناء لغة الضاد إلى تطوير برمجيات بالحرف العربي، ترقى إلى مستوى مثيلاتها في اللغات المتقدمة تقنيًا ومعرفيًا، كالإنجليزية مثلاً، مما سيفيد المستخدم العربي في البحث عن المعلومات في النصوص والتنقيب عنها، وإجراء الترجمة الآلية من وإلى اللغة العربية، وإعداد المعاجم، والقواميس الإلكترونية (أحادية، وثنائية، ومتعددة اللغة)، وغيرها من التطبيقات الحاسوبية التي تعرف تطورًا سريعًا لا مثيل له.

منهج البحث:

نظرًا لطبيعة الدراسة التي تجمع بين النظري والتطبيقي، فقد ارتأينا اعتماد المنهج الوصفي التحليلي في معالجة إشكالية البحث الرئيسة، وذلك بوصف اللغة العربية، وتوصيف قضاياها اللسانية الحاسوبية، مما يتصل بالمستويات اللسانية: الأصوات، والصرف، والتركيب، وتطبيق المنهج التحليلي من خلال تحليل التطبيقات الحاسوبية والبحث عن مدى ملاءمتها للغة العربية، ويظهر

ذلك جلياً من خلال تحليل البرمجيات المفتوحة المصدر، وتبيان أثرها على المحتوى الرقمي العربي على الشبكة، وقد كان منطلقنا في ذلك برمجية نوج باعتبار ما تتيحه من إمكانيات وصفية وتصنيفية في حوسبة اللغة العربية.

الإشكالية الرئيسية:

تتمحور إشكالية البحث الرئيسة في الأسئلة الفرعية التالية:

- ما المقصود بالبرمجيات؟
- ما أنواعها؟
- ما الفرق بين البرمجيات المغلقة والبرمجيات الحرة ومفتوحة المصدر؟
- ما خدماتها بالنسبة للمعالجة الآلية للغة العربية؟
- ما دورها في تعزيز المحتوى الرقمي العربي على الشبكة؟

2- البرمجيات الحرة ومفتوحة المصدر: رؤية في المفهوم والمصطلح

من الصَّعب بمكان إيجاد دلالة محدّدة وواضحة لمصطلحين متداخلين هما: البرمجيات الحرة ومفتوحة المصدر وذلك نظراً لعدم توفر مفردتين في اللغة الإنجليزية تعبران عن مفهومي الحرية والمجانية، إذ إنّ كلمة واحدة مثل (free) تستطيع التعبير عن هذين المفهومين المختلفين كلية⁽¹⁾. ومع ذلك سندسعى من خلال بعض الأبحاث المتوافرة لدينا تقريب تلك المصطلحات إلى القارئ العربي⁽²⁾، عسانا بذلك نتمكّن من ردم - على الأقل - أو تقليص فجوة البرمجيات

(1) طويلة، أنس، "البرمجيات الحرة ومفتوحة المصدر واللغة العربية"، المجلة العربية للعلوم والمعلومات، ع6، 2005، ص144.

http://www.freesoft.jo/people/kefah/freesoftware_whitepaper_arabic.pdf

(2) مهديوي، عمر، وزايد محمد، معجم تقنيات المعلومات: فرنسي إنجليزي عربي، في طور النشر.

بين العربية، ولغات متقدمة تقنيًا ومعرفيًا، لها باع طويل في مجالات الإنتاج، والتطوير، والتنفيذ المعلوماتي. وحسبنا، اليوم، أن العربية بدأت تتلمس طريقها الصحيح في مجال الحوسبة؛ نظرًا لما راكمته في هذا المجال خلال السنوات الأخيرة من ابتكارات، وبرامج، أقل ما يقال عنها إنها أسهمت ولو بنسبة ضئيلة في تعزيز موقع اللغة العربية في خريطة البحث الحاسوبي المعاصر.

وعلى هذا الأساس تُعرّف البرمجيات الحرة بأنها: "برمجيات لخدمة الناس، ولإدارة الحاسوب، يتم تطويرها بطرق جديدة، من قبيل عشرات الآلاف من المطوّرين المحترفين، ومئات الآلاف من ضابطي الجودة، والموثقين، والملايين من المستخدمين حول العالم"⁽¹⁾. وتختلف البرمجيات الحرة عن البرمجيات مفتوحة المصدر، في كون هذه الأخيرة لا تضمن - بالضرورة - الحريات الأربع الأساس، وسميت البرمجيات حرة باعتبار مبدأ الحرية، الذي يضمن للجميع حرية استخدامها، وتوزيعها، والحصول على أصول برمجياتها، أو ما يعرف بالنصوص المصدرية، أو الشيفرة المصدرية⁽²⁾.

وتذهب بعض الدراسات⁽³⁾ إلى أن بداية الاهتمام بالبرمجيات الحرة والمفتوحة، تعود إلى عام 1984، حينما شرع ريتشارد ستالمان⁽⁴⁾ في تطوير نظام

(1) كفاح، عيسى، مقدمة في البرمجيات الحرة، 28 أبريل 2005، ص 5.
رابط الاطلاع والتحميل:

http://www.freesoft.jo/people/kefah/freesoftware_whitepaper_arabic.pdf

(2) كفاح، عيسى، مقدمة في البرمجيات الحرة، 28 أبريل 2005، ص 5.

(3) طوبلة، أنس، البرمجيات الحرة ومفتوحة المصدر واللغة العربية، المجلة العربية للعلوم والمعلومات، ع6، 2005، ص144.

(4) ريتشارد ستالمان (Richard Stallman) مشرف مختبر الحاسوب في جامعة ماساشوستس (MIT) للتكنولوجيا.

التشغيل GNU، قصد إتاحتها على الشبكة للجميع وبدون مقابل مادي، وذلك تعبيراً عن سخطة الكبير على تحكم بعض الشركات التجارية الكبرى في مفاتيح البرمجيات المغلقة.

وتحقيقاً لهذا الغرض أسس ريتشارد في عام 1985 مؤسسة البرمجيات الحرة Free Software Foundation، التي قامت بتطوير نظام تشغيل حرّ ومتكامل، لكن هذا الأخير ظل بعيد المنال، نتيجة كثير من الصعوبات التقنية⁽¹⁾. وقد كان هذا الوضع حافزاً للعديد من المطوّرين في العالم، وفي مقدمتهم لينوس تورفالدز (Linux Torvalds) الفنلندي الجنسية، الذي يرجع إليه الفضل في تطوير نظام تشغيل حرّ يدعى لينوكس⁽²⁾ Linux (نسبة إلى اسمه)، وذلك عام 1991. وبصدور النسخة الأولى من هذا النظام، ازدادت القناعة لدى مطوري البرمجيات، بأن لينكس هو الخطوة المنتظرة في الاتجاه الصحيح لإتمام العمل الذي بدأه ريتشاردز، لأنها استطاعت التغلب على الكثير من المشاكل التقنية سابقاً. وبموجب ذلك تم دمج البرمجيتين في عمل واحد، ضمن ما أصبح يعرف حينذاك بنظام التشغيل غنو/ لينكس (GNU /Linux)، والذي سيعرف فيما بعد نموّاً متسارعاً في فترة قياسية، أبهرت الجميع وطرحت الكثير من التساؤلات والتعقيدات.

وإذا كان الأمر كذلك في البيئة الغربية، فمتى تم الربط بين العربية والبرمجيات الحرة ومفتوحة المصدر؟

على الرغم من الانتشار السريع الذي يعرفه مجال استخدام البرمجيات الحرة ومفتوحة المصدر في جميع أنحاء المعمور، إلا أن الإقبال على هذه التقنية

(1) طويلة، أنس، مرجع سابق، ص144.

(2) جنو ليوس يونكس نسبة إلى مطوره ليونكس وهو نظام حر وجنو اسم لحيوان من جنس الجواميس يعيش في أمريكا الشمالية.

في العالم العربي ما يزال ضيقاً ولم يرق إلى المستوى المطلوب، علماً بأن التعاطي مع هذا النوع من البرمجيات لم يحدث إلا في فترة متأخرة؛ وذلك لعدة أسباب: سياسية، ومادية، ولغوية، وتقنية.⁽¹⁾ لا حاجة لنا اليوم في سردها.

إن لمشاريع المصادر العربية المفتوحة على الصعيد الحاسوبي، أو البرمجي، دوراً كبيراً في توفير الأنظمة المفتوحة التي تقدم برامج داعمة للعربية على الشبكة، أو لحوسبتها من حيث معجمها، أو نحوها، أو صرفها، أو الترجمة منها وإليها، إلخ. وقد انطلقت هذه الفكرة بشكل فعال في عام 2001م، وعُرفت باسم عيون العرب، وهو مشروع عربي يهدف إلى دعم اللغة العربية في منصات المصادر المفتوحة والمجانية، في كل ما يتعلق بقضايا الخطوط، وقراءة النصوص في الملفات والواجهات العربية، كما يوفّر للعالم العربي بيئة تفاعلية مفتوحة عبر موقعها في كل ما يتعلق بحوسبة اللغة العربية، والبرمجيات. كما أسهمت أيضاً في بناء مشاريع حاسوبية عديدة، تتعلق أساساً بالترجمة الآلية والمعاجم، ومشاريع أخرى مرتبطة بتطوير برمجيات حققت نتائج قيمة في مجال إنشاء برامج مفتوحة المصدر لدعم العربية، وبخاصة القرآن الكريم، والتطبيقات الإسلامية، والمدقق النحوي العربي، وبدأ التعويل على المصادر المفتوحة لخدمة العربية عند كل من جنو لينيكس المصرية، ومجتمع موزيلا العربي، ومجموعة مستخدمي لينكس بالأردن، والكويت، وغيرها من الدول العربية⁽²⁾. وهكذا اتسعت دائرة

(1) طويلة، أنس، "البرمجيات الحرة ومفتوحة المصدر واللغة العربية"، المجلة العربية للعلوم والمعلومات، 6ع، 2005، ص145.

(2) المجبول، سلطان، الجهود العربية في حوسبة العربية، مقالة منشورة في صفحة الجزيرة، 25 يونيو 2006:

<http://www.al-jazirah.com/2016/20160625/cm25.htm>

الاستفادة من فلسفة البرمجيات الحرة ومفتوحة المصدر في البلاد العربية، سواء من طرف الأفراد، أو المؤسسات، أو المنظمات. ويكفي هنا الإشارة إلى مشروع ألكسو الرامي إلى النهوض باللغة العربية من أجل التوجه إلى مجتمع المعرفة، والذي يهدف إلى تطوير برمجيات مفتوحة المصدر للغة العربية تهم الصَّرف والمعجم والتركيب والتَّشكيل والدَّلالة.

والجدير بالذكر هنا أن جامعات غربية في أمريكا، وأوروبا، وآسيا، قد قطعت مراحل متقدمة في مجال المعالجة الآلية للغات الطبيعية، وراكت أبحاثاً ودراسات متميزة كمّاً وكيفاً، كان لها أثرها الواضح على المحتوى الرقمي لهذه اللغات. في حين أن إسهام الجامعات، والمعاهد العربية، في هذا الشأن ما يزال دون المستوى المطلوب، ولم يصل إلى مستوى تغطية التَّوصيف الشَّامل لمستويات العربية: صوتاً، وصرفاً، ونحواً، ودلالةً، ومعجماً؛ ولعلَّ أسباب هذه الفجوة الرقمية -على مستوى الحوسبة اللغوية- كثيرة ترجع في نظرنا إلى أمرين:

الأول: يتمثل في غياب استراتيجية علمية واضحة المعالم، يمكن تطبيقها في الوطن العربي، من شأنها أن تضع خيطاً رابطاً بين المقدمات والنتائج.

والثاني: يتجلى في غياب التنسيق، والتعاون بين الجامعات العربية في مجال المعالجة الآلية للغة العربية وغيرها.

3- برمجيات المعالجة الآلية للغة العربية من التنظير إلى التطبيق

إنَّ تطوير البرمجيات بنوعها الحر، أو المغلق، قصد معالجة اللغة العربية حاسوبياً أضحى مطلباً ملحاً أكثر من أي وقت مضى، إن أردنا للغتنا العربية اللِّحاق بركب اللُّغات المتقدمة في هذا الباب، وتحقيق نهضة معرفية في الحياة العامة للفرد والجماعة في البلاد العربية، هذا مع العلم أن هذا المطلب هو ديدن الأمم في عصر مجتمع المعرفة الذي نعيش فيه. وهكذا فأني تراجع أو تخاذل في

تحقيق هذا المبتغى من شأنه أن يقذف بنا - نحن العرب- إلى أرصفة الطريق السيار للمعلومات، أو أننا سنظل خارج خريطة التقانة المعرفية التي هي السبيل لكل تقدم اقتصادي واجتماعي وثقافي وعلمي.

وفيما يلي عرض لنموذج من البرمجيات المفتوحة المصدر التي استخدمت في معالجة اللغات الطبيعية عامة، ومعالجة اللغة العربية خاصة.

1.3- برمجة نوج: الخصائص، الخوارزميات والتطبيقات

إن برمجة نوج في الأصل هي حصيلة المسيرة العلمية الطويلة للباحث ماكس سلبيرشتاين (1993-2005) التي قضاها في التنظير، والتجريب موجهاً إياها لبناء معالجات آلية للغة الفرنسية. وقد طبقت هذه البرمجية في معالجة العديد من أنظمة اللغات الطبيعية، وحقت نتائج باهرة في هذا المجال، ولاسيما في التطبيق على اللغة الفرنسية، لكن في المجال العربي يبدو أن الاستفادة من هذه البرمجية ما يزال في بداية الطريق، إذ عمل عدد من الباحثين العرب في تطوير تطبيقات حاسوبية للغة الضاد، أذكر منهم على سبيل الذكر: سليم مصفار (2008) الذي استخدم نوج في تطوير معالج صر في نحوي للغة العربية⁽¹⁾.

يُعدُّ نظام نوج (Nooj) مثله مثل إنتكس (Intex)، فهو عبارة عن بيئة معلوماتية متخصصة في المعالجة اللسانية للغات الطبيعية، وأداة في النمذجة الصورية وإجراء التطبيقات الحاسوبية. وقد تمت صياغة هذا النظام بناء على النظام المعلوماتي إنتكس (Intex) بهدف تبسيط عملية الولوج إلى المعلومات والنفوذ إليها واستغلالها في بلورة تطبيقات حاسوبية للغات الطبيعية.

(1) يمكن الإشارة هنا إلى محاولات أخرى للدكتور الحناش محمد، ود. مهديوي عمر، ود. غازي عز الدين، ود. علي بولعلام وآخرين.

وهكذا فقد حصر ماكس سلبيرشتاين (1993) حدود مقارنة Intex في المبادئ التالية:

أولاً- يقوم نظام Intex على خمسة معاجم صورية.

ثانياً- لا يمكن للرسوم البيانية (Les Meta-graphes) أن تستعمل مع المعاجم التركيبية.

ثالثاً- إن الفصل بين معاجم (DELAS) و (DELACF) غير ممكن، لسبب بسيط هو أن عددًا من المصطلحات الموصوفة في (DELAS) قد تطرأ عليها تغييرات إملائية، وتغييرات أخرى في (DELACF).

يتضح من هذا الكلام أن نظام Intex غير قادر على حوسبة إعراب المركبات آلياً. ولعل هذا السرّ الكامن وراء البحث عن نظام معلوماتي آخر أكثر دقة وصرامة. إذ تمكن المهندسون المهتمون بحوسبة اللغات من بلورة نظام إلكتروني باعتباره محرّكاً لسانياً، يشغل ضمن منظومة معاجم ديلا (DELA) ألا وهو نظام نوج، وهو يعمل في إطار ثلاث طبقات من المعاجم: (DELA و DELAS و DELAC) التي تشغل في تناسق تامّ. كما يحتوي النظام أيضاً على أنحاء إلكترونية محلّية، فضلاً عن كونه مزوّداً بالأساليب الخوارزمية التي تقوم بالصيانة، والموارد المعجمية الصرفية والتركيبية التي تسهم في رفع اللبس، وخاصة مع المداخل المركّبة.

تعالج هذه المعاجم إذن بناءً على معطيات جداول المعجم-نحو Lexique- (Grammaire)، ويرجع الفضل في ذلك إلى النظام الإعرابي الذي يعالج النص المعجمي بفعل العمليات التّصريفية التي يتيحها المحرّك الإعرابي للبيئة التي لا تعدو أن تكون مجرد روزنامة من المحوّلّات والرّسوم.

وبهذا يعالج النظام الآلي النصوص صرفيّاً، وتركيبيّاً، ثم يحلّل المتواليات

لقد أظهرت هذه العملية المتبعة في معالجة اللغات الأجنبية نجاعتها، وبالتحديد في اللغة الفرنسية. ولكن الأمر بالنسبة للعربية ليس سهلاً للغاية. إذ إن تناول المركّبات المؤلفة من أكثر من لفظين قد يطرح مشاكل جمّة على مستوى المعالجة، فعندما يشرع البرنامج الآلي في التعرّف على المنطقة المعتمدة، فإنه قد يحصل أن يقع خلط بينها وبين منطقة ثابتة أخرى في النّص نفسه، مما يعرّض البرنامج للارتباك. وفي هذه الحالة يتم اختيار البنية الأولى؛ لأنها تحتوي على المنطقة المعتمدة التي يبحث عنها البرمجي ضمن البنية المتلازمة، وإذا حصل أن تداخلت منطقتان معتمتان مع بعضهما البعض، فإنه يُطبق حينئذ المعيار الصُّوري الذي بموجبه يحلّل النّص بناء على مقارنة التعبيرين. وبذلك يهتدي إلى إقصاء البنيات التي لا تتوفر على منطقة معتمدة تلائم الأصناف الصورية. ولعل هذا خيار من الخيارات الناجعة التي يتم اللجوء إليها في مثل هذه الحالات اللُّغوية المعقدة. وعليه يمكن اختصار خصائص منصة نوج فيما يلي:

- نوج منصبة لسانية ملائمة للعديد من أنظمة الاستغلال بإصداراتها المختلفة.
- تعتمد عليها مجموعة كبيرة من المراكز ومختبرات البحث والتطوير.

24

المدونة. كما تستطيع أن تحلل المدونات المرتكزة على المعاجم الإلكترونية ديلا (DELA)، وعلى الأنحاء المحليّة التي تُعرفُ انتشارًا كبيرًا في لغات عديدة، وخاصة في مراكز البحث المتخصصة، حيث يتم تكوين فرق عمل بحثية لتطوير معالجة اللغات الطبيعية، ولاسيما الأنحاء الإلكترونية للغات، وتطبيقات برمجية صناعية مختلفة. وهذه الأخيرة طوّرت كثيرًا بناء النظام، علاوة على المناهج الأخرى التي تقوم بتصفية قواعد المعطيات التي تتلاءم ومفردات النص المزمع تحليله مع النّحويات الإلكترونية المحليّة قبل الشّروع في تحليلها.

لقد كان طموح واضعي هذا النظام إذن كبيرًا، يتوق إلى جعله المعيار الصوري العام لكل اللغات الطبيعية، ومن ضمنها اللغة العربية. لهذا جهزوه بمحلّ معجمي، عبارة عن قالب يستطيع أن يعالج جميع أصناف المتواليات اللسانية المعقّدة كالّتعبيرات المركّبة أو الأساليب. والغرض الأساس من ذلك هو المساعدة على بناء نظام للترجمة الآلية يتمكن من ترجمة تعبيرات اللغات الطبيعية، والمطورة بواسطة بيئة نوج، والتعليم، والمعالجة الآلية للغات الطبيعية، في سائر مستوياتها، ومن ثم تحقيق المطالب الآتية:

- 1- أن يعمل كمدقق إملائي ونحوي في المتواليات خارج السياق النصي / المدونة.
- 2- أن يعمل كبرنامج للتحليل المعجمي استجابة لأهداف محدّدة، مثل: التعرّف على أقل عدد ممكن من الأخطاء الواردة أثناء القراءة الآلية.
- 3- يتأكد البرنامج من صحّة الكلمة من خلال قاعدة المعطيات المصنفة في النظام.
- 4- يشغل بصفة أساس مع محوّلّات الأوضاع النهائية التي تقوم بدور رئيس، وهو أمن اللبس الذي يعدّ تحدّيًا كبيرًا في معالجة اللغات الطبيعية آليًا؛ وبهذا الأسلوب يمكن توصيف النّحويات المحليّة التي تنتسب إليها المتوالية.
- 5- يوظف برنامج التوافق الآلي بهدف إيجاد البنيات الصحيحة (المورفو-تركيبية)

في سياقاتها النصية الأصلية؛ وذلك قصد تحقيق المطابقة، والتحقق منها، وهذا يطابق إلى حد ما الشاهد في المعجم الورقي. وعلى مستوى النظام الآلي يركز مبدئيًا على برمجي يحتوي على جميع المتواليات الموجودة في النص، فيقوم بالبحث في نصوص المدونة ليُطابق النتائج التي توصل إليها بالتحليل، والمعالجة، وما هو موجود في المدونة الأصلية مرفقة بحزمة من النُحويات الإلكترونية التي تساعد على إقامة هذا التطابق.

6- يقوم بمعالجة نصوص ثابتة، وأخرى دينامية، وهذا ما يميز البيئة كونها مفتوحة المصدر، وتفاعلية في آن واحد، من حيث مقارنتها بالمعطيات الجديدة، ولذلك يتطلب الأمر تطبيق مفاهيم لغوية، وكيانات اسمية، وعبارات عقلانية؛ لتنظيم، وفرز، وتصدير هذه النتائج، ببناء تقرير إحصائي اعتمادًا على أرقام دقيقة عن البنيات المعالجة.

وتستغل بيئة التطوير هذه بالموافق وهي عبارة عن أداة معلوماتية تسمح بالكشف عن المدونة باستخدام بعض الأبنية الخاصة ومقاماتها، ومن ثم تنجز التطبيقات انطلاقًا من تشغيل الموافق الآلي، وذلك باختيار المقولات الموجهة للمعالجة، حيث يقوم النظام باستخراج المصطلحات المناسبة من المدونة، وذلك بتصنيفها وتشخيصها سواء أكانت مصطلحات، أو مفردات، أو كيانات جغرافية. وإذا أردنا رصد ألفاظ تتعلق بحقل "النبات" مثلاً في نص ما، فإننا نستدعي الموثق الآلي، الذي يقوم بجرد جميع المصطلحات، والمركبات، والألفاظ، التي تتعلق بالمجال موضوع الدراسة، ثم نقوم بتفعيل العبارات العقلانية، إذ يعمل الأوتومات على رصد كل ما له انتماء أو علاقات دلالية بالكلمة.

سهولة الاستخدام: تتجلى سهولة استخدام المنصة في معالجة المدونة النصية وبناء القواعد النحوية في بضع دقائق. وتتأكد هذه الخاصية في كون

المعاجم من نوع (ديلا) أصبحت مشكلة في معجم واحد، حيث يتم معالجة كل اللغات الطبيعية باعتماد مقاييس تشفيرية موحدة، أو الشفرة الموحدة (UNICODE) القابلة للتصفح انطلاقاً من صفحات الإنترنت، أو وثائق (XML) وميكروسوفت (Microsoft). وقد استخدمت لغة (SQL) كلغة للطلبات المهيكلية لتنظيم وإدارة قاعدة البيانات المخزنة والمرتبة ترتيباً محكماً في صيغة جداول تحتوي على حقول ذات خصائص فردية مترابطة ومهيكلية حسب نموذج مفهوماتي متلائم مع التطبيقات اللغوية.

وبما أن البرمجي مفتوح المصدر وقابل للتحميل، فهو مبني بلغة (PHP) المتوافقة مع لغة الويب (HTML) التي تيسر عملية التصفح، ولغة بناء قواعد البيانات (SQL) لإنجاز أي تطبيق كيفما كان نوعه أو حجمه، وعلى جميع منصات المنظومات الحاسوبية المعروفة لونيكس أو أونيكس أو ويندوز.

كل هذه المزايا والخصائص تساعد البرمجي على معالجة اللغة الطبيعية، وبناء المعجم الحاسوبي المُبْنِي في حلة أنيقة وذكية، من شأنها أن تستجيب لمتطلبات المستعمل اللغوية، والحاسوبية، والمصطلحية، والتوثيقية، وغيرها، بدءاً من استرجاع المعلومات وانتهاءً بمرحلة الإضافة عليها إن أراد المستخدم للنظام تحليل نص ما.

وأما من حيث بنية البرمجي ونسقيته، فهو يعطي للمستعمل إمكانية الاسترجاع والبحث عن المعلومات، والحصول على الإجابات المباشرة بدون ترميز أو اختصار، وأغلب استعمالات البرمجي تنحصر في بناء، أو تشغيل واستغلال معاجم تستجيب لضرورتين:

الأولى- الضَّرورة اللُّغوية، المتمثلة في الحفاظ على تناسق اللغة، وتواترها بشكل شفاف وجلي.

الثانية- الضرورة الحاسوبية المتمثلة في تكثيف عمليات النظام وتركيزها سرعةً وإنجازًا.

يحتوي (نوج) إذن باعتباره محرِّكًا لغويًّا على معاجم من نوع ديلا (DELA)، بحيث يستطيع معالجة كل الأشكال اللغوية، بواسطة جداوله المعجمية التركيبية، بما فيها المصطلحات التي تدخل في معجم ديلاك (DELAC).

4- البرمجيات المفتوحة في خدمة تطبيقات المعالجة الآلية للغة العربية

إن تطوير الأدوات اللسانية الموجَّهة لخدمة الأهداف والمقاصد البيداغوجية يُعدُّ قطب الرِّحى بالنسبة للسانيات الحاسوبية وهندسة اللغات على عدة مستويات: 1- ترجع أهمية النظام الصرفي في اللغة العربية باعتباره دراسة لبنية الكلمة وشكلها، إلى كون اللغة العربية ذات خصوصية إعرابية من جهة، ومن جهة أخرى، فهي لغة اشتقاقية وغنية من حيث المتغيِّرات المورفيمية وتأليفاتها المتعددة، التي تجعل المتكلم ينتج بواسطتها صيغًا ذات دلالات متعددة. لهذا فإن أقسام الكلم بوصفها أشكالًا مكونة من روابط، وحروف الجر، والتعريف، واللواصق، وأقسام أخرى في حالات سابقة للبنية أو لاحقة لها، متعددة، بحيث يمكن أن تأتي مع كلمة واحدة، وقد تصل إلى أكثر من مائة شكل مختلف.

وانطلاقًا من خاصية الإنتاجية الصرفية العالية، لا بد من البحث في استراتيجية خاصة للغة العربية، تسمح ببناء معجم صرفي آلي موسوعي يضم جميع الجذور والمداخل المناسبة للكلمات السياقية، وهذا ما قام به بعض الباحثين أمثال: د. محمد الحناش، ود. عمر مهديوي؛ وذلك بتطوير نظرية لسانية تمكِّن من توصيف معجم آلي للغة العربية يتضمَّن المداخل المفردة، والبسيطة، والمركَّبة المعربة، والمتلازمة.

وعليه، فإذا تعاملنا مع هذا التعدّد في الأشكال الصرفية المؤلفة مع الجذر في اللغة العربية، الذي يعطي-لا محالة- كثرة في المداخل المعجمية، فلو أردنا معالجة كل ذلك آلياً، فلن نكون إلا أمام مسألة تقنية معلوماتية صرفة، لن تكون ذات استعداد كافٍ لاستيعاب كل هذا الكمّ الهائل من المداخل الصرفية، إلا إذا قمنا بتطوير الأدوات المعلوماتية المتاحة من جهة، ومن ثمّ ضرورة الاستفادة من الأدوات اللسانية الصورية ومفاهيمها الأساس من جهة ثانية، والتي يجب أن تنسجم انسجماً تاماً مع المفاهيم المعلوماتية المتقدمة ومبادئها، كالتحليل، والتوليد الآليين.

وإذا كانت الغاية من وراء هذا كله هي بناء نظام آلي يعالج الكلمات المعربة، وأشكالها التصريفية المطّردة، فإن ثمة صعوبات تُطرح مع الأشكال الشاذة، أو المهملة. لذلك تم التفكير في بناء نظام ذي منهجية عالية الإنتاجية؛ كأن تتخذ اللغة العربية في شموليتها بجميع متونها واستعمالاتها، وهذا من شأنه أن يطور قدرة المحلّل الصرفي العربي الذي يحتاج في نظرنا إلى الكثير من المراجعة والتدقيق، في ضوء ما تتيحه اللغويات الرقمية، والهندسة اللغوية⁽¹⁾، وقد أدى ذلك إلى بناء نظام عام للتعليم، يستطيع أن يجعل من قواعد المعرفة الصرفية قاعدة للمساءلة وفق خوارزميات تقام لهذا الغرض.

هذا، ولقد أدى استخدام الحاسوب في نقل المعلومات من المجال الطبيعي للمسار التعليمي إلى وضع هذه المعلومات في قالب يسير ومشوّق، وذلك ببرمجة المادة اللغوية المراد تلقينها، مصحوبة بالتعليقات الصوتية، والخلفية مكونة من

(1) مهديوي عمر، خوارزميات صرفية للغة العربية من التحليل إلى التوليد، أشغال مؤتمر اللغة العربية في عصر المعلوماتية، نوفمبر، 2006، سوريا، موقع مجمع اللغة العربية بدمشق:

<http://www.arabacademy.gov.sy/uploads/conferences/conference5/21.pdf>

الصور، والرسوم، واللقطات المتحركة للأفلام المصورة، بالإضافة إلى تصميم وإخراج الاختبار، ووسيلة قياس مدى استيعاب الطلاب للمادة، عبر مجموعة من المراحل، والروائز المخصصة لهذا الغرض، وهي بصفة عامة تتمثل في أسلوب المحاضر، وأسلوب المحاكاة، والتدريب والممارسة وغيرها.

وفيما يلي عرض مختصر لبعض نماذج أوتومات الأوضاع النهائية المستخدمة في نظام (Nool)⁽¹⁾ والموجهة لمعالجة اللغات الطبيعية من قبيل اللغتين الفرنسية والعربية:

أولاً- المُركَّبات والمتلازمات في اللغتين العربية والفرنسية

لا يمكن للبرامج المعدة للتعامل مع المعجم الآلي أن تتعرف على جميع مكونات النص المكتوب، ما لم يتضمن هذا المعجم جميع أنواع الوحدات الاسمية المركبة، والمتلازمة في اللغة العربية، نحو:

علامة المرور- أحد عشر- جواز السفر....

وهي مداخل غير قابلة للتجزئة، لأنها تتشكل في أغلبها من مداخل متلازمة، وبالتالي تتطلب وصفاً صرفياً، وآلياً، من نوع خاص. إذ الأمر يدعو إلى الجمع بين نوعين من الوصف:

أولاً- الاهتمام بالوضع المورفولوجي للعناصر التي يتألف منها الاسم المركب.

ثانياً- وصف وضعه الجديد في التأليفية التي أنشأها منه نظام العربية، خاصة وأن الاسم المركب يجمع في الغالب بين فصيلتين من الأسماء الموجودة في

(1) قمنا باستخراج مجموعة من الأوتومات التي تعالج المتواليات المركبة والمتلازمات في اللغة الفرنسية، وقد كان الهدف تقديمها ضمن هذا الملحق لاستكمال ما قد يكون من نقص في تصور هذه التقنية مقارنة على البنيات الشبيهة في اللغة العربية.

معجم المفردات البسيطة: الاسم والصفة، أو بين الاسم والحرف، أو بين الاسم والظرف، إلخ.

هذا الجمع، يعني أن الاسم المركب يتألف من فصيلة صرفية أيضاً، وهو ما يتطلب معالجة خاصة، أصبح معها لازم وضع معاجم تعتمد نظاماً متراصاً ومُبْنِياً من التراكيب الثابتة، والصمَاء التي تقف في وجه كل تأليف، أو تحويل لإحدى عناصرها، مما يطرح مشكلة لا تقل أهمية عما سبق، وتتعلق أساساً بعملية الترتيب في المعجم، لأن المتواليات المركبة هي متواليات لسانية تتألف من أكثر من عنصر اسمي بسيط، كما تحتوي على أكثر من بياض واحد، ذلك أن دماغ الإنسان مزود على المستوى الصوري، بقدرات لسانية معقدة تركز من حيث الإنجاز على مبدأ التوزيع التكاملي بين العناصر الاسمية والحرفية والفعلية؛ بينما يتميز الثاني باستحالة التأليف، وخاصة التعقيم الدلالي، وعدم الاستجابة للقوانين المنظمة للشكل الأول.

واعتمادنا على منهجية (المعجم- نحو) تأتي من زاوية أساس، وهي إعادة النظر في مفهوم المفردة التي قد تتضمن أكثر من متوالية لسانية، متضمنة لأكثر من منطقتين فارغتين، لكنها غير دالة على أكثر من مفردة واحدة؛ ثم إن اضطراب مفهوم المفردة جعل المعجميين العرب لا يكثرثون بوضع المعجم الخاص بالمفردات المركبة، والتعبيرات المتلازمة، رغم كثرتها، وإنتاجية بنياتها في نظام العربية، لتتأمل المثال التالي:

يضرب س. (أخماساً) 1 في (أسداس) 2

تتأسس المفردة /المتوالية: (أخماساً في أسداس) على أكثر من بياضين، ويمكن لأي برنامج توليد للمفردات اللغوية أن يدخل في الاعتبار أكثر من بياضين، بينما تبقى البياضات الأخرى بدون أي دور توليدي في عملية التعرف

عليها⁽¹⁾. إذ إنَّ هذا النوع من المركبات الاسمية لا يقبل التجزئة إلى مكونات صغرى دالة، تنفرد كل واحدة منها بביاضها في الاستعمال العادي للغة العربية، بل ولا تسمح حتى باستبدالها بمتوالية لسانية بسيطة معادلة لها دلاليًا، فهذا النوع من المركبات الاسمية يبقى هكذا منطقة سوداء معتمة لا يمكن النفاذ إليها، إلا بالإحصاء، والتصنيف حتى تُقدَّم إلى المعجم الآلي بشكل يسمح له بتوظيفها في العمليات التي ننتظر منه أن ينجزها. غير أن هذا لا يجب أن يفهم منه أن كل التعبيرات المتلازمة قائمة على المتواليات الاسمية المركبة، فهناك العديد منها تقوم على عناصر اسمية بسيطة، نحو:

ضرب س. (السكة)⁽²⁾

وقع س. (في الفخ)

شد س. (الرحال) 1 (إلى مظ) 2

فالمفردة المركبة إذن، هي متوالية لسانية اسمية مؤلفة من أكثر من مدخل معجمي بسيط أي من مداخل معجم المفردات البسيطة، وهي تتألف تارة من عنصر مشتق، وآخر جامد، وتارة من عنصرين جامدين. إن المفردة المركبة إذن لا تستجيب للتغيرات الصرفية أو حتى التوليد بظاهرة التوسيم مثلا⁽³⁾:

<س س> = شذر مذر

(1) للتعرف الآلي على المدخل المعجمي لا بد من بناء خوارزميات دقيقة بمساعدة أوتومات الأوضاع النهائية وتقنيات أخرى للتمثيل يشغل ضمن نظام المعجم الرقبي.

(2) الحناش، محمد، "المعاجم الآلية للغة العربية: بناء قاعدة المعطيات"، مج التواصل اللساني، م4، ع1، ص100-101.

(3) El Hannache Mohammed, Lexique grammaire de l'arabe classe des verbes qualitatifs, (suite) in Linguistica communicatio, V.1, N.2.1989. pp.31-41.

<س س>: = حيص بيص

يصعب مع هذه الأمثلة تحليل هذه المركبات تحليلًا (مورفو- تركيبًا)، بل وحتى تلك المبنية على المصدر، أو اسم الفاعل حيث يصعب كذلك العودة بها إلى البنية الأصلية. وذلك مثل:

مقود السيارة.

هناك فقدان للعلاقة الدلالية بين هذه المركبات، وهو ما قد يكون أصلًا لها. ذلكم ما سوف نسلط عليه الضوء لاحقًا عند التعرف الآلي على المفردات المركبة والتعبيرات المتلازمة.

هذا ولقد عُولِجَت هذه التراكيب بصفة معمّقة في اللغة الفرنسية⁽¹⁾ وكذا في بعض المحاولات في اللغة العربية⁽²⁾. وقد خُلصَ هؤلاء إلى أن هذه التراكيب تتصف بدرجة معينة من التلازم والتضام، حيث إنّ منها المتلازمة جزئيًا ومنها المتلازمة كليًا.

ولما كانت المتلازمة كليًا، تتميز بعتمة منطقتها الثابتة، إذ غالبًا ما يصعب ربطها بجملته، وفي هذه الحالة ستصبح مدخلًا مستقلًا بنفسها رغم وجود بياضات، أو عناصر مشتقة، أو جامدة، منطقتها ترفض أي استبدال كان، لأن

(1) وقد حصل هذا مع مجموعة من الباحثين نذكر من بينهم:

- ماكس سليبرشتاين (Max SILBERZTEIN) 1993.

- موريس غروس (M.GROSS) (1988).

- إيريك لابورط (PORTE) (1988).

- ج.لابيل (LABELLE. J 1988) (1988).

- م. كونيينا (CONENNA.M) (1985) ..

(2) نذكر منهم أحمد أبو سعد (1987) والحناش (1992) و (1993) و (2002).

عناصرها صامدة في وجه أي تغيير تركيبي.

هذا، وتتميز المفردة المركبة بوقوعها في الغالب في موقع المفعول، وهي بذلك تبتعد عن منطقة التطابق مع الفعل. ولقد ذهب محمد الحناش في هذا الإطار⁽¹⁾، إلى إدراج الأسماء المركبة الحملية المستقلة في هذا الباب والتي لا تُعَرَّفُ أصلاً حملياً مثل:

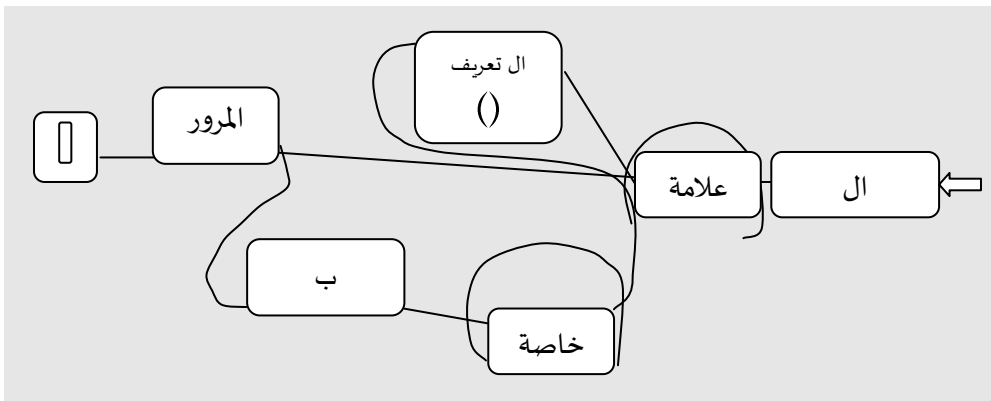
أثار زيد حفيظة عمر.

إن التعبير المتلازم لا يترك عنصراً يندمج في مكوناته مثلاً:

أخماساً في أسداس، لكن *أخماساً × في أسداس

خلافًا للمتلازمة جزئياً مثل: علامة المرور، العلامة الخاصة بالمرور.

نمثل لذلك بأوتومات الأوضاع النهائية التالي:



رسم 1: يمثل تثنية وجمع بعض أصناف المفردات المركبة

(1) El Hannache, M, Lexique grammaire de l'arabe: classe des verbes qualitatifs, in *Linguistica communicatio*, V.1, N.1, 1989. pp.9-18.

نخلص مما سبق إلى أن هذه المركبات غير متجانسة في تكوينها وفصيلتها التركيبية؛ لذلك وجب التعامل معها بتصنيفها، ووصفها، ثم معالجتها وفق طرق خاصة، فالمعجم يجب أيضًا أن يشمل كل هذه الأنواع من المركبات بدون استثناء. وإذا ما قمنا بدراسة هذه التعبيرات وفق إطار (المعجم-نحو) وخاصة تقنيات التصنيف، والتعرف على مناطقها المعتمدة (بحيث لا تختلط بمعجم المأثورات والأمثال) فإننا سنجد أنها تضمن الأسماء المتلازمة مع الفعل: أسماء بسيطة وأخرى مركبة منها:

- الاسم والضمير (قضت لطيفة نحيها).
- الاسم المتجانس (وجهًا لوجه).
- الاسم المتخالف (ذهابًا وإيابًا).
- المركب الإضافي (رد الاعتبار).
- المركب الحرفي (بمنأى عن كذا).
- ومتلازم صفة (أصاب عمرَ مرضٍ عضال⁽¹⁾)

إن هذا الوضع يتطلب وصفًا صرفيًا آليًا من نوع خاص؛ لأنها بنيات تختلف عن باقي المفردات البسيطة، لذلك فإن استدعاء مفاهيم وأدوات إجرائية جديدة، تعالج البنية الصرفية للعناصر التي يتألف منها المركب، قد تكون من المفروض موجودة في معجم المفردات البسيطة والمعربة (ينظر معجم DELAS على سبيل المثال⁽²⁾)، غير أنها تختلف من حيث كونها تجمع بين المعجم

(1) الحناش، محمد، نظرية حاسوب لسانية لبناء معاجم الآلية للغة العربية (محاولة في التأصيل)، سلسلة خاصة عن استخدام اللغة العربية في تقنية المعلومات، مجلة التواصل اللساني، 1992، ص 71.

(2) Silberstein Max, Dictionnaires électroniques et analyse automatique de textes: le système INTEX, Masson Ed. Paris.1993.

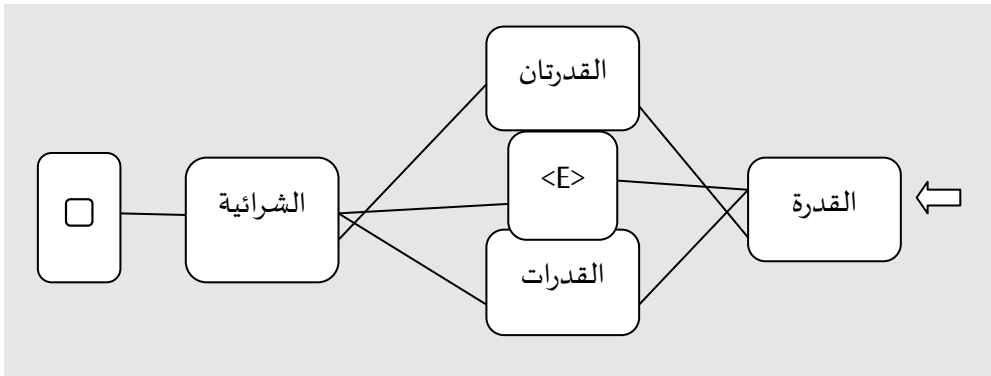
والتركيب، رغم أن الجانب الصرفي سيصبح لاحقاً في وضع ثانوي بعد التركيب. هذا، وإن غنى المتن المعجمي العربي، ووجود الزخم الهائل من هذه التعبيرات التي أغنته بشكل واضح، هي غالباً مداخل مولدة عبر الترجمة، والمثاقفة. إذ نجد أفاضاً مركبة قد انتقلت إلى العربية انطلاقاً من لغات أجنبية عديدة، منها: المصطلحات العلمية؛ بسبب المتطلبات التكنولوجية والعلمية والثقافية، إضافة إلى المركبات الأصلية في العربية المؤلفة من صفة واسم، حرف واسم، اسم وصفة، حرف واسم وحرف، اسم وحرف واسم، اسم واسم، وأشكال أخرى. كل ذلك يشكل قاعدة معطيات معجمية ضخمة يجب أن نستفيد منها في بناء المعجم الحاسوبي العربي المنشود.

وصفياً، يظهر -إلى حدود الآن- أن هذا النوع من البنيات اللسانية، باستثناء بعض المجهودات القليلة، لم يلق اهتماماً رغم إنتاجيته الكبيرة في قدرة المتكلم اللغوية، والمعجمية بخاصة. وتحتاج هذه البنيات آلياً، إلى بناء قواعد بيانات خاصة بها⁽¹⁾. وهذا ما يستدعي مزيداً من المعالجة المعلوماتية بهدف تزويد التطبيقات الحاسوبية المتعددة للترجمة، وتعليم اللغات والنشر، والتدقيق الإملائي، وذلك ببناء معجم يستجيب للتقنيات والأساليب المتبعة في المعاجم الأجنبية من هذا الحجم، (انظر مثلاً التعبيرات المتلازمة في المعجم الرقمي للغة الفرنسية (DELAC) إذ نجد حوالي 20000 تعبير متلازم في قاعدة بيانات اللغة الفرنسية)⁽²⁾.

(1) الحناش، محمد، قاعدة البيانات الخاصة بالتعابير المسكوكة والمركبة (1992) و نظام للترجمة المتعددة اللغات الذي أنتجته شركة (CIMOS) للمزيد يرجى الاطلاع على الموقع الالكتروني للشركة (www.cimos.com).

(2) Silberztein Max, op.cit., (1993).

ولما كانت هذه التعبيرات خاضعة في جميع اللغات الطبيعية لنفس عمليات التصنيف بحسب رؤية (المعجم-نحو) وطريقة الأجروميات التي وضعها الحناش محمد⁽¹⁾، باقتراح قواعد خاصة للمعالجة المعجمية، والنحوية المستفادة من المحللات التركيبية، والصرفية، والدلالية، وتراعي خصوصية، ونحو كل صنف، أولاً. وثانياً: التعامل معها بطريقة الأوتومات وأنواعه، ولاسيما أوتومات الإدماج التي تسمح بعض عقدها، حسب أنواع المركبات، إما بإدماج عنصر ما بين مكوناته أو تصريف أحد عناصره، كأن ننثي ونجمع رأس التعبير التالي: القدرتان والقدرات على التوالي في: القدرة الشرائية. فأوتومات الأوضاع النهائية يستطيع تمثيل هذه التراكيب آلياً وفي مراحل إنتاجية معجمية مختلفة (حالات إعرابية محلية)، مما تحتاج إلى خوارزميات متميزة كما في الأوتومات الممثل بالرسم التالي:

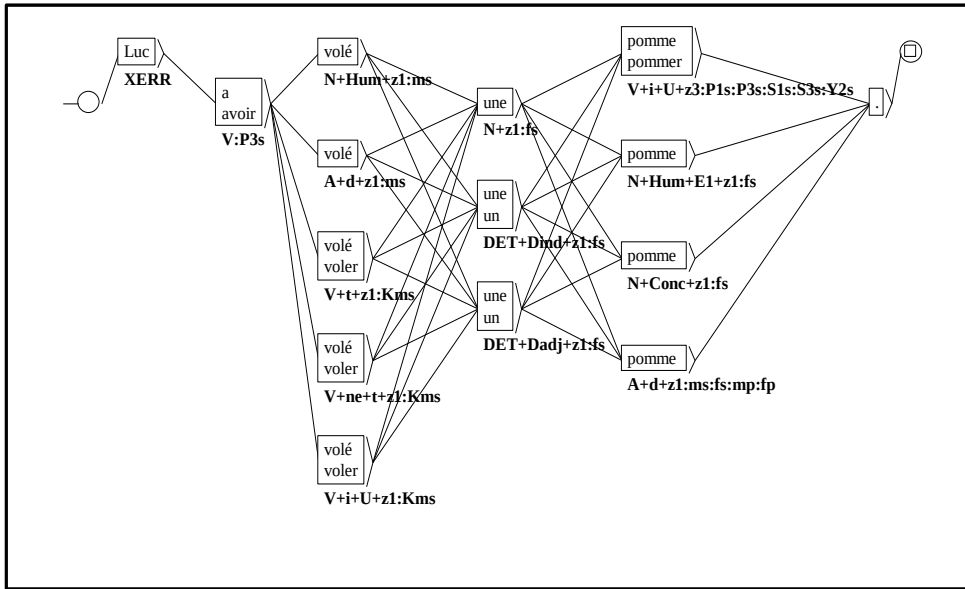


رسم 2: يمثل ثنائية وجمع بعض أصناف المفردات المركبة

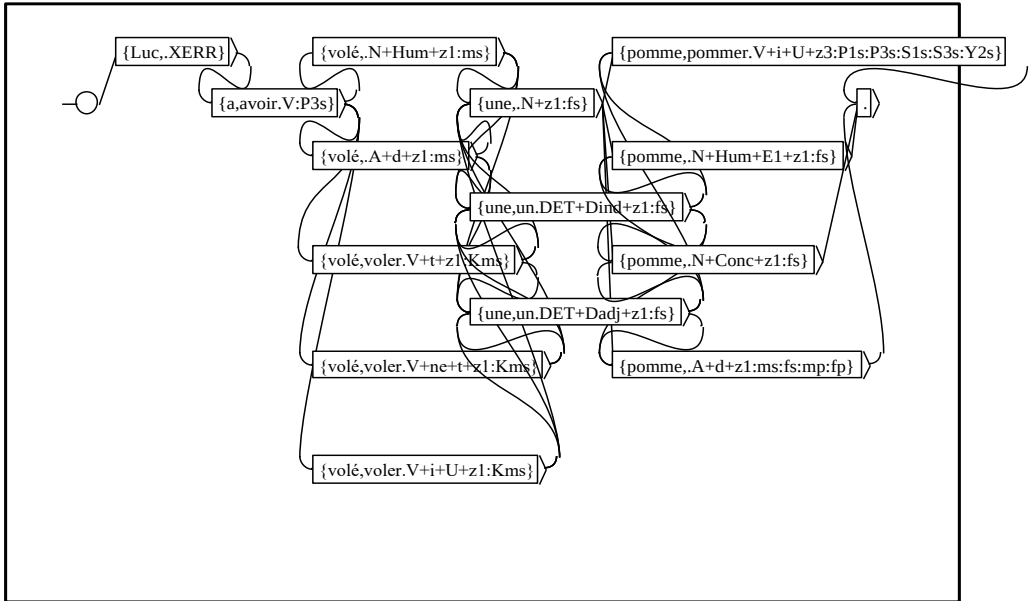
(1) الحناش، محمد، "برنامج لساني حاسوبي للتعرف الآلي على التعبيرات المسكوكة في اللغة العربية"، مجلة التواصل اللساني، مج 3، 1996، ص 86-88.

هذا وإذا كان المعجم العربي الآلي⁽¹⁾ الذي ننشده يأخذ بعين الاعتبار المنهجية الفنية، والتقنية التي أثرت بشكل كبير في تحديد المدخل المعجمي المتضمن للبسيط والمركب، وإذا كان أيضاً يستجيب للغة العقلانية المعلوماتية، فإن بناء معجم رقمي تعليمي للغة العربية، لابد أن يركز على منظومة من المعاجم التي تعمل في تناسق تامٍّ ومكتملٍ ومُبنينٍ يتمكن مُستعمله من التعامل معه بشكل تفاعلي ذكي. وفيما يلي عرض لتطبيق أتومات الحالات المحدودة في معالجة اللغة الفرنسية أولاً، وعلى اللغة العربية ثانياً.

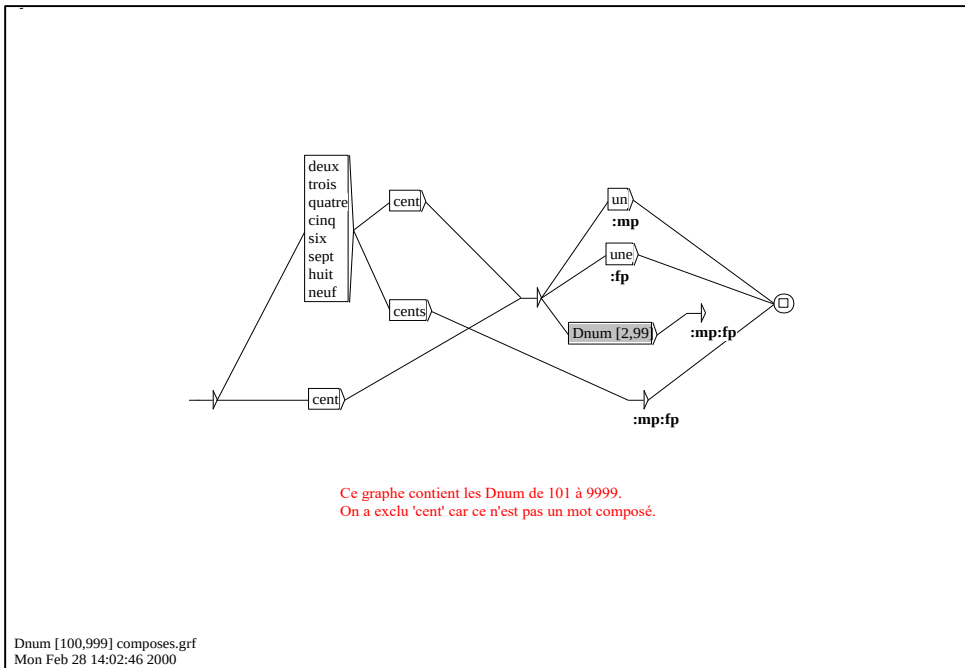
أولاً- معالجة اللغة الفرنسية باستخدام تقنية أتومات الحالات المحدودة:

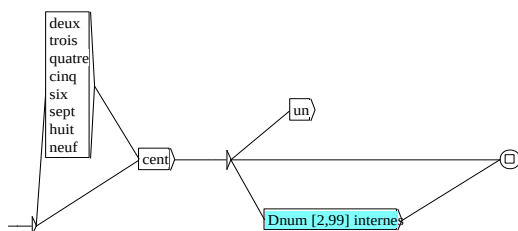


(1) مهديوي، عمر، الهندسة اللغوية وتطبيقاتها في اللغة العربية، (2017)

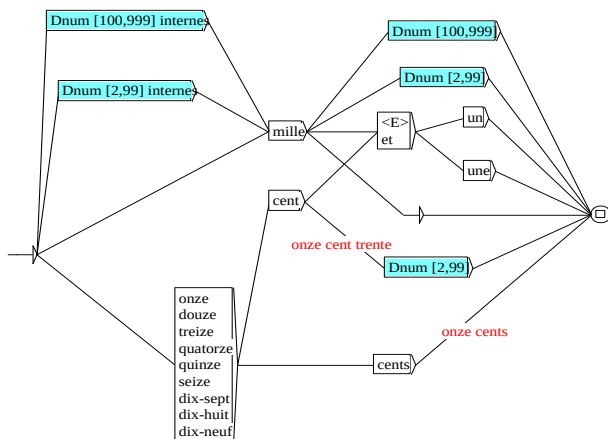


تمثل الرسوم التالية الأعداد المركبة في اللغة الفرنسية:

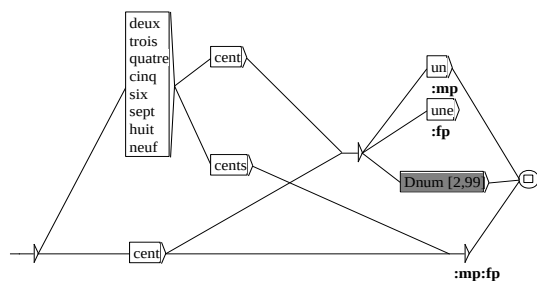




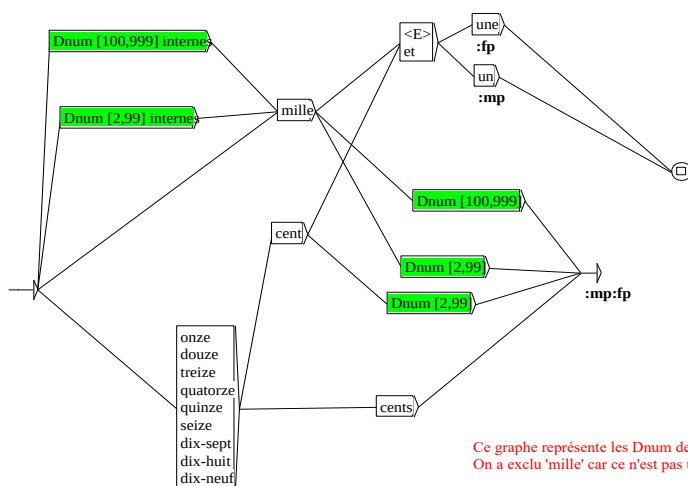
Dnum [100,999] internes.grf
Tue Aug 10 15:30:24 1999



Dnum [1000,999999].grf
Tue Nov 19 10:46:50 2002

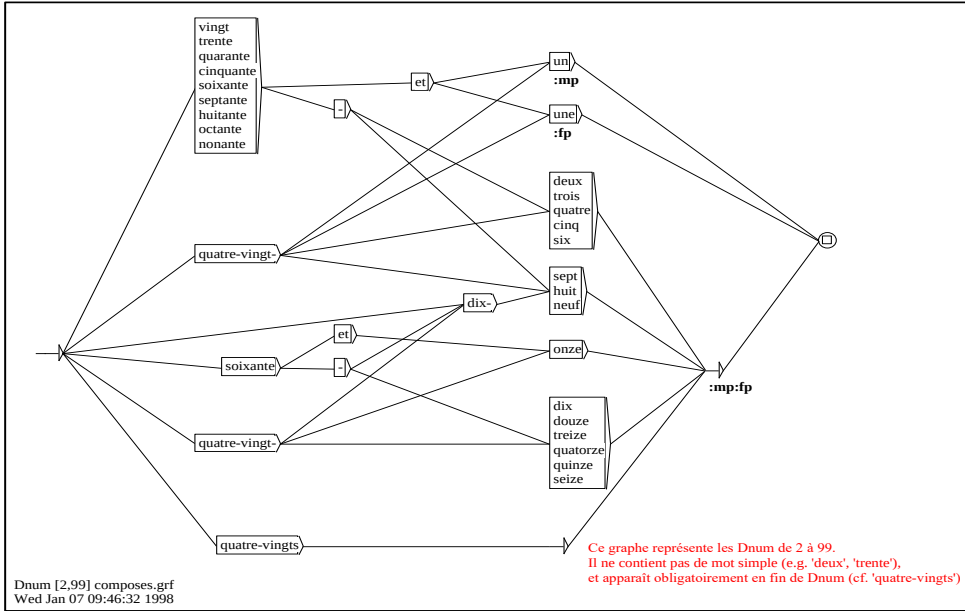


Dnum [100,999].grf
Mon Jan 05 15:29:48 1998

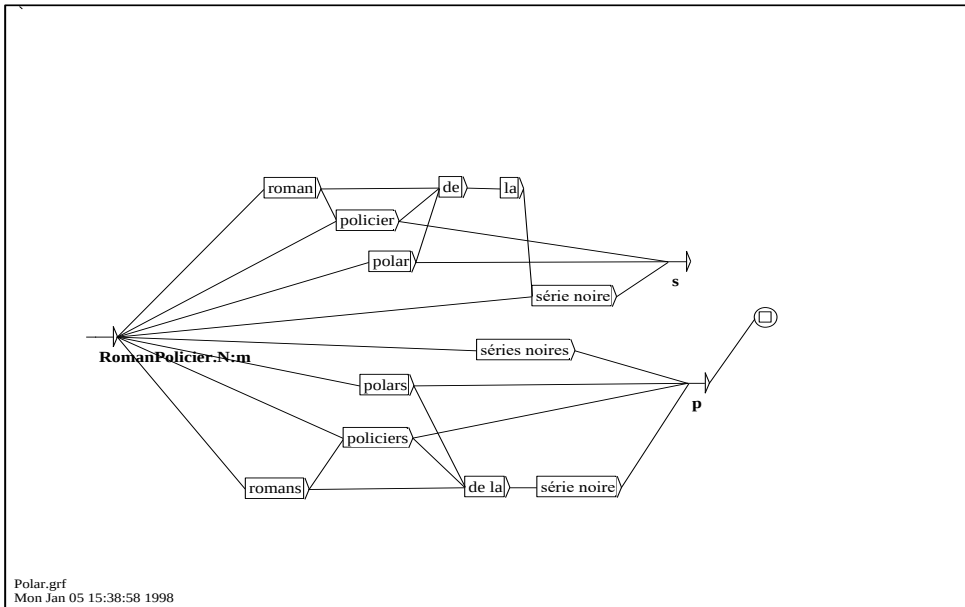


Ce graphe représente les Dnum de 1001 à 999 999.
On a exclu 'mille' car ce n'est pas un mot composé.

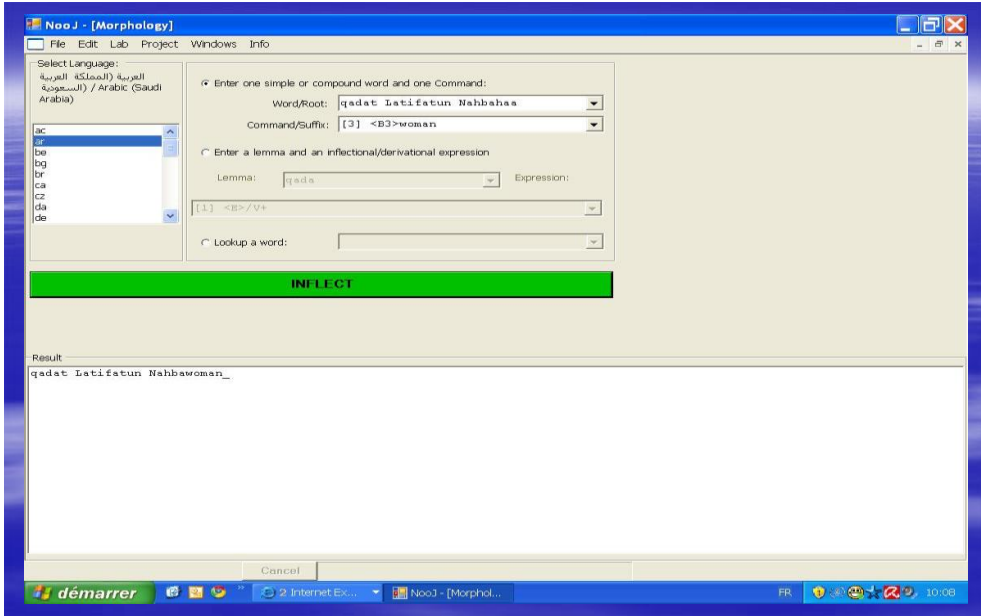
Dnum [1000,999999] composes.grf
Mon Feb 28 14:03:16 2000



يمثل الرسم، أدناه، أوتومات تحليل البنية المركبة في اللغة الفرنسية كما في
العبارة التالية: (Roman policier de la série noire)



ثانياً. التصريف الآلي والتحليل المعجمي للنصوص العربية باستخدام نافذة البيئة مفتوحة المصدر نوج FLX



لقد تخصصت معظم المعاجم العربية الورقية والرقمية في تصنيف جميع مداخل اللغة العربية البسيطة، فنظرت إليها من وجهة نظر نحوية صرفة، حيث إنَّها تميز بين الأفعال اللازمة والأفعال المتعدّية، أو التي تتعدى بحرف أو بدونه، كما صنفها إلى أفعال عادية، وأفعال قلوب وأفعال ناقصة، وهي تصنيفات تحتاج إلى إعادة نظر، في ضوء ما تتيحه لسانيات المنصات أولاً، وفي ضوء تقنيات المعالجة الهندسية للغات الطبيعية ثانياً.

وأما التصنيف المعتمد في بناء هذا المعجم فهو مستمد أساساً من أعمال كل من بلاندين كورتواز، وماكس سلبرشتاين، الذي يصنف المقولات النحوية بناءً على تغيراتها التصريفية؛ بغرض التعرف عليها، وتوليدها آلياً، وهي: الاسم،

والفعل، والظروف، والصفات. وهكذا يمكن أن يتولد من اسم مفرد مثلاً عدد من الأشكال المعربة حسب النوع، والعدد، والجنس، وفي السياق نفسه سار الحناش محمد، على المنوال نفسه، إذ اعتمد تصنيفاً يراعي كل خصوصية للمقولات النحوية في نظام اللغة العربية، سواء أكانت فعلاً، أم اسماً، أم حرفاً، وربطها بالتغيرات التصريفية الممكنة في النص المعجمي (أقسام الكلم)، وذلك بمراعاة الخصوصية النحوية للمدخل المعجمي (المفردة)، وربطها بالمستوى المورفولوجي المناسب لها (الصيغة والاشتقاق) على النمط التالي:

ف: الفعل في الزمن الماضي للمفرد الغائب المذكر

س: للمفرد المذكر المؤنث

ص: الصفة وسائر المشتقات بجميع أنواعها

ظ: للظروف بجميع أنواعها

ح: للحروف والأدوات بجميع أنواعها

يعتبر المصدر مفردة بسيطة؛ لذلك يستحسن إدراج رمزه ضمن معجم المفردات البسيطة المعربة؛ لتبقى بذلك مكونات خمسة يتأسس عليها نظام أقسام الكلم للغة العربية هي:

أ- تسنين الفعل: يعتبر الفعل⁽¹⁾ من أصعب المقولات ضبطاً في النص المعجمي، نظراً لتشعب علائقه بسائر أقسام الكلم، وخاصة المصدر منها، وسائر أنواع

(1) للفعل أدوار توزيعية في العمليات التركيبية وقد أكدت هذه الفرضية نظرية المعجم- نحو، كما أن له أبنية وصيغ صرفية تراوح مكانها بين الصحة والاعتلال، علماً بأن الصيغة المنتهية (infinitive) غير موجودة في اللغة العربية على غرار اللغات الإلصاقية مثلاً، مما يجعلها مقولة نحوية صعبة التناول في الوصف والتفسير، وهذا ما انطلق منه النحاة سواء منهم المتقدمين أو المتأخرين، انظر على سبيل المثال لا الحصر، سيبويه في الكتاب، ج. 1. تحقيق عبد السلام هارون. بيروت 1988.

الصفة. وبالتالي فقد تم تقسيم الأفعال إلى مجموعات تصنيفية، يتمثل فيها أقصى حد ممكن من الاطراد في الانتماء إلى الصنف المناسب. وهذه الأصناف تحدد من خلال الحركات التي يأخذها الفعل في المضارع، وحسب الصحة والاعتلال، وحسب الزيادة أو عدمها، وهذه كلها تنتظم كما يلي⁽¹⁾:

(1) الثلاثي (الثلاثي الصحيح): فَعَلَ: (ف [رقم]) فعل ← يفعل

(ف [رقم]) فعل ← يفعل

(2) الرباعي (الرباعي الصحيح): فَعَّلَ: (ف [رقم]) فعلل ← يفعلل.

(3) الخماسي: تفاعل: (ف [رقم]) تفاعل ← يتفاعل.

(4) أنواع أخرى: (ف [رقم]) كل فعد جامد لا يصرف، مثل: بئس ونعم.

ب- الأسماء البسيطة⁽²⁾: إذا كانت الأفعال تستجيب للترقيم المطرد اعتماداً على نظام الصيغ التصنيفية، فإن ذلك سيتعذر مع الأسماء التي لا تعرف أي انتظام تصنيفي يمكن اعتماده في إقامة تصنيف معجمي من نفس القبيل. ولهذا تتبع طرق أخرى لتصنيفها، وهي كما يلي:

الصنف 1: ويشمل سائر الأسماء التي جاءت في المذكر فقط، مثل: رجل (مذكر) ولكن: رجلة (مؤنث).

الصنف 2: ويشمل الأسماء التي لها مؤنث فقط، مثل: امرأة (مؤنث) لكن: امرأ (مذكر).

(1) سيبويه، مرجع مذكور، ص. 89.

(2) إذا كانت اللغة الفرنسية لغة إلصاقية تصنف بعض أقسام الكلم ضمن قاعدة بيانات الأسماء كالصيغة المنتهية للفعل، فإنه في اللغة العربية لا تدرج هذه الصيغة مباشرة مع الاسم بل مع أصناف أخرى، وهذا ما يتطلب الدقة في التصنيف.

الصنف3: ويشمل ما له مذكر ومؤنث بشكل مطرد بالتاء المربوطة، مثل: معلم، معلمة.

الصنف4: ويشمل كل ما ليس له جمع من لفظه ولا من غير لفظه، مثل: هواء، حب.

الصنف5: ويشمل كل ما ليس له مفرد من لفظه، مثل: جيش.

الصنف6: ويشمل الضمائر: المتصل منها، والمنفصل، وأسماء الإشارة، واسم الموصول، وأدوات التعريف، وكل ماله علاقة بالاسم البسيط.

ج- الصفات والمصادر: مراعاة للتلاؤم التوزيعي الحاصل في تصنيف الأسماء البسيطة، والصفات والمصادر وخاصة التطابق. تم التصنيف بالطريقة التالية وهي:

الصنف1: وهو مخصص للصفات الخاصة بالمذكر، مثل: عُضَال، في مرض عُضَال، والتي ليس لها مؤنث من لفظها.

الصنف2: فقد تم تخصيصه للصفات التي ليس لها مذكر من لفظها، مثل: حامل، للمرأة الحامل.

الصنف3: فقد تم تخصيصه للصفات التي يشترك فيها المذكر والمؤنث، مثل: جميل ← جميلة

الصنف4: فقد تم تخصيصه للصفات المستعملة في بعض التخصصات العلمية، مثل: أحمر قاني.

د- الظروف: من المفروض أن يهتم المعجم الآلي بالظروف(1) بشكل مستقل عن

(1) تدخل الظروف في باب المعجم الآلي للمفردات المركبة، مع أن بعضها يخضع للتحويلات التي تجرى
==

بقية أقسام الكلم في نظام اللغة العربية، فهي تشكل مداخل مستقلة عن الصفات، والأسماء، والأفعال، هذا بالرغم من قلة عددها مقارنة مع أقسام الكلم الأخرى.

هـ- الحروف والأدوات: لقد تم تخصيص هذا الصنف من المداخل والأدوات في اللغة العربية، ويعنى بها حروف الجر، والأدوات التي تقوم بدور الربط بين الكلمات، دون أن تكون هي نفسها ذات دلالة مرجعية، وهذه المداخل تدرجها كتب النحو تحت باب الحروف، أو حروف المعاني.

و- معجم المفردات البسيطة المعربة (DEAMSF):

تتمثل استعمالات المعاجم الإلكترونية في تقديم معلومات دقيقة وشاملة عن المفردات التي يتألف منها نص ما، وذلك بهدف معالجتها آلياً. وبالإضافة إلى مهمة التعرف على أقسام الكلم يجب إعطاء معلومات أساسية مرفقة بكل مفردة معربة⁽¹⁾. فبالنسبة للأفعال: تحدد طريقة تصريفها ونسبتها إلى الضمائر ومختلف المورفيمات المتعلقة بالجنس، والعدد، والزمن، أما بالنسبة للأسماء: تحدد طريقة تصغيرها، ونسبتها، وعلامتها الإعرابية الممكنة مع كل واحد منها، إلخ. وبذلك سوف يتم إرفاق كل نوع من المداخل المعجمية التي يتضمنها معجم

==

على الظروف المركبة وخاصة في عملية التوسيم، وقد أعد ماكس سلبيرشتاين (M. Silberztein) جداول المعجم التركيبي خاص بها بحيث يمكن أن نسمها بقيم الإيجاب والسلب حسب المتغيرات التي تطرأ عليها نحو: صباح مساء، التي يمكن أن نعرفها: الصباح والمساء ، لغة أجنبية: لغات أجنبية. للمزيد حول توسيم الظروف والإحالة والحذف والتعويض ينظر على سبيل الذكر:

Silberztein, Dictionnaires électroniques et analyse automatique de textes: le système INTEX, Masson Ed. Paris, (1993) p.106

(1) زايد، محمد، دراسة في المعاجم الفرنسية الحاسوبية، الاجتماع الثاني لخبراء المعجم الحاسوبي للغة العربية، مدينة الملك عبد العزيز للعلوم والتقنية، أبريل 2008، ص 1-28.

المفردات البسيطة بهذه المعلومات. وبعبارة أخرى، فإن كل شكل من أشكال هذا المعجم يجب أن تعطى عنه جميع الاحتمالات التصريفية، والصرفية، التي يوجد عليها في اللغة العربية، وهذا ما يشكل موضوع القسم الأخير من قاعدة المعطيات، والتي يطلق عليها: معجم المفردات البسيطة المعربة، على النمط التالي:

الشكل أ = جذر ب، صرف ج

حيث يقصد بالشكل المدخل المعجمي (المفردة البسيطة)، وبالجذر الصيغة التي يظهر بها في المعجم نفسه، وبالصرف مختلف الأوجه التصريفية، والصرفية، التي تَلَحَقُ المدخل البسيط، وهو ما يقصد به بالإعراب في هذا المعجم. ونحن إذ نبتغي هنا إعطاء فكرة عن معجم المفردات البسيطة بأشكالها المعربة، لأننا نفترض تمامًا بأن بناء معجم المفردات المركبة يستند في جزء هام من بياناته إلى المتواليات البسيطة التي قد تكون عنصرًا من مكوناته.

5- الخاتمة:

سعينا في هذا البحث إلى إبراز أهمية البرمجيات الحرة ومفتوحة المصدر، في تنفيذ تطبيقات حاسوبية للغة العربية، من قبيل إنجاز الترجمة الآلية وبناء المعاجم، والقواميس الإلكترونية، ومحركات البحث، وتعليم اللغة، وحفظ المعلومات، واسترجاعها، وبناء المحلل النحوي، والمحلل الصرفي، والمشكل الآلي، وغيرها؛ وكل ذلك موجه بالأساس لتعزيز المحتوى الرقمي العربي على الشبكة، ودعمه؛ لأن منظومة العربية، اليوم، تعاني من عدة فجوات، ومن بينها فجوة المحتوى الإلكتروني، إذ توجد هوة كبيرة فاصلة بين محتوى اللغات المتقدمة تقنيًا، ومعرفيًا، كاللغة الإنجليزية، ومحتوى لغات ضعيفة، ومنها العربية.

ودعنا نؤكد في خاتمة الكلام النتائج التي تم التوصل إليها بعد هذه الرحلة العلمية في تفاصيل موضوع الدراسة:

- 1- إنَّ حوسبة اللغة العربية هي المدخل السليم لحل الكثير من أزماتنا. اجتماعية كانت، أم اقتصادية، أم ثقافية، أم حضارية.
- 2- إنَّ بناء المشروع اللساني الحاسوبي العربي يتطلب فريقاً من المتخصصين (مهندسين، ولسانيين ورياضيين، ومناطقة...)، يكون فيه دور الريادة للساني أثناء المعالجة حتى نقطع الطريق أمام تدخلات غير المتخصصين، والهواة الذين نهبوا المال العربي، وأسرفوه في مشاريع تجارية ونفعية لا تعود على لغة الضاد بظائل.
- 3- إنَّ تطوير برمجيات حرة ومفتوحة المصدر للغة العربية، مطلب معرفي وتقني بامتياز من أجل دعم المحتوى العربي وتعزيزه على الشبكة.

توصيات البحث:

- إنَّ حوسبة العربية اليوم تعد من الأولويات الرئيسة التي ينبغي للجامعات والمعاهد العربية الانكباب عليها، وتشجيع الباحثين عليها.
- إنَّ تأسيس قسم اللسانيات الحاسوبية في كليات العلوم الإنسانية أضحي المطلب الرئيس لتطوير الهندسة البيداغوجية لعلوم اللغة العربية، والمناهج اللسانية الحديثة.
- إنَّ الاستثمار في مجال المعرفة اللسانية الحاسوبية، هو خيار من الخيارات الاستراتيجية التي ينبغي للقطاع العام والخاص الاسهام فيها بحظ وافر؛ من أجل بناء مجتمع معرفة عربي .

قائمة المراجع

أولا- المراجع العربية:

- 1- الحناش، محمد، "اللغة العربية والحاسوب: قراءة سريعة في الهندسة اللسانية"، مجلة التواصل اللساني، مج 9، 2003.
- 2- سيبيوه أبو بشر عمرو بن عثمان، الكتاب، تحقيق عبد السلام هارون، بيروت، ج1، 1988.
- 3- طويلة، أنس، "البرمجيات الحرة ومفتوحة المصدر واللغة العربية"، المجلة العربية للعلوم والمعلومات، ع6، 2005.
- 4- كفاح، عيسى، مقدمة في البرمجيات الحرة، 28 أبريل 2005، رابط الاطلاع والتحميل:
http://www.freesoft.jo/people/kefah/freesoftware_whitepaper_arabic.pdf
- 5- المجيلول، سلطان، الجهود العربية في حوسبة العربية، مقالة منشورة في صفحة الجزيرة، 25 يونيو 2006:
<http://www.al-jazirah.com/2016/20160625/cm25.htm>
- 6- مجهودات ألكسو في المعالجة الآلية للغة العربية، منشورات المنظمة العربية للتربية والثقافة والعلوم، رابط الاطلاع والتحميل:
<http://www.alecso.org/site/sectors/sciences/93-alecso/projects-alecso.html>
- 7- مهديوي، عمر، الهندسة اللغوية وتطبيقاتها في اللغة العربية، نور نشر، 2017.
- 8- مهديوي، عمر وآخرون، اللسانيات الحاسوبية واللغة العربية، دار كنوز المعرفة، 2018.

9- مهديوي، عمر العتاد اللساني الحاسوبي لرقمنة المعجم التاريخي للغة العربية، ضمن كتاب المعجم التاريخي للغة العربية: رؤى وملامح، الرياض، المملكة العربية السعودية، 2016.

10- مهديوي، عمر، خوارزميات صرفية للغة العربية، مؤتمر اللغة العربية في عصر المعلوماتية، نوفمبر 2006 رابط الاطلاع والتحميل:

<http://www.arabacademy.gov.sy/uploads/conferences/conference5/21.pdf>

11- مهديوي، عمر، وغازي عز الدين، وعلي بولعلام، مشروع إعداد موارد لسانية للغة العربية ضمن كتاب اللسانيات وتحليل الخطاب، منشورات مختبر الخطاب وتكامل المعارف، الكلية المتعددة التخصصات الرشيدية، 2016.

12- نبيل، علي، اللغة العربية والحاسوب، مجلة عالم الفكر، ع3، 1987.

ثانيا- المراجع الأجنبية:

- 1- El Hannach Mohammed. (1989) Lexique grammaire de l'arabe: classe des verbes qualitatifs, in *Linguistica communicatio*, Imprimerie Annajah El-Jadida, Casablanca V.1, N.1, 1989. pp.9-18.
- 2- El Hannach Mohammed (1989), Lexique grammaire de l'arabe classe des verbes qualitatifs, (suite) in *Linguistica communicatio*, Imprimerie Annajah El-Jadida, Casablanca V.1, N.2. pp.31-41.
- 3- Mehdioui Omar. (2016) Les expressions Multi-mots Arabes: Traitement morphologique syntaxique et sémantique, Formation sur NooJ et ses applications sur les langues naturelles. Faculté des sciences, Kenitra ,7 Nov. 2016.

- 4- Silberztein Max. (2000) INTEX Manual. Manuel d'utilisation du logiciel INTEX, disponible sur le site <http://intex.univ-fcomte.fr> (en anglais, 230 pages).
- 5- Silberztein Max. (1993) Dictionnaires électroniques et analyse automatique de textes: le système INTEX, Masson Ed. Paris, p.106.
- 6- Silberztein Max, Nooj Manual. Manuel du logiciel Nooj et support de cours, disponible sur le site <http://www.nooj4nlp.net> .
- 7- Slim. Mesfar. (2008) Analyse Morphologique-syntaxique automatique et reconnaissance des entités nommées en Arabe Standard, Thèse de doctorat en informatique nov. 2008, Université de France Comté, France.



**مستقبلات اللغة العربية
في فضاء الويب الثالث (3 WEB)
والعوالم الافتراضية (الميتافيرس)**

**Arabic Language Futuristics within the WEB
3.0 and the Metaverse**

د. أسامة عبد الوهاب ريس
مجمع اللغة العربية بالخرطوم

الملخص

منذ بداية العقد الثالث من القرن الحادي والعشرين، تطورت الفضاءات الرقمية، واندمجت العديد من التقانات الرائدة من الذكاء الاصطناعي، والتكنولوجيا المكانية، وتقنيات سلاسل الكتل، والحوسبة السحابية؛ لتشكيل هذه الفضاءات الرقمية الجديدة، والتي بدورها صنعت ما أصبح يعرف بالويب الثالث، وعوالم افتراضية واعدة أصبحت تسمى بالميثافيرس. حيث أفضت هذه التطورات إلى إضافة فضاءات وأبعاد جديدة للحضارة الإنسانية، ونشأ جيل جديد من الشباب (الجيل زد)، انغمس في هذه الفضاءات والأبعاد؛ مما أدى لابتعاده عن الحياة التقليدية. بدأت هذه الثورة الجديدة بالألعاب، ثم تطورت في أزمة الكورونا العالمية، بحيث خدمت قضية العمل والتعليم عن بعد، والتجارة الإلكترونية، وأصبحت الآن تتسع لممارسة جميع النشاطات البشرية. بذلك أفضى هذا التحول إلى نقلة حضارية نوعية مما يتطلب تدخلات غير تقليدية، وصناعة تموضع مستقبلي مواكب. كما يتطلب إجراء نقلة مفاهيمية في جل المعارف.

يغطي طيف البحث عدة موضوعات، تبدأ من الحوار بين الإنسان، والآلة لتصل حتى تطبيقات الذكاء الاصطناعي في قضايا اللغة العربية والإدراك الآلي. يسعى هذا البحث إلى تعميق المعرفة حول طرق بناء تفاعل أكثر ذكاءً، بين الإنسان، والآلة، باستخدام اللغة الطبيعية. وقد وسعت نطاقه التطورات الأخيرة في معالجة اللغة الطبيعية. علاوة على ذلك، أصبح من الممكن الآن ليس فقط الاستفادة من التفاعل السلبي المقيد بين الإنسان، والآلة، بل تم تطوير قنوات ذكية أكثر تفاعلية، حيث يكون للذكاء البشري، والذكاء الاصطناعي

(Artificial Intelligence AI) فرصة للتفاعل.

تقدم الورقة نموذجًا إطارياً لتفاعل اللغة العربية في هذه الأبعاد، وستعمل الورقة على استكشاف مستقبلات اللغة العربية في هذا الإطار، مع النظر في الآفاق والتحديات المصاحبة. وتقدم الورقة تعريفاً بمجموعة من المعارف المستحدثة في مجال التفاعل المحوسب باللغة العربية، والإدراك الآلي، والتفاعل بين الإنسان، والآلة، وتحديات الويب الثالث، والميتافيرس. ثم تبحر في الآفاق المستقبلية في حوسبة اللغة العربية في الفضاء الثالث، والعوالم الافتراضية، ومن ثم تقدم نموذجاً لصناعة إطار، ومنصة؛ لتطوير برمجيات قادرة على إنتاج، وفهم اللغة في صورتها الصوتية، والمرئية (الكلام والكتابة) في الفضاء السيبراني بتطوراتها المتسارعة. وتقدم توصيات بمجالات البحوث المستقبلية التي انفتحت فيما بين التخصصات اللغوية والرقمية.

الكلمات المفتاحية: العوالم الافتراضية (ميتافيرس)، اللغويات الحاسوبية العربية، إدراك اللغة، الويب الثالث، نماذج اللغة، الذكاء الاصطناعي، معالجة اللغة الطبيعية، العالم الدلالي.

Abstract

Since the beginning of the third decade of the twenty-first century, cyberspaces have evolved, and many leading technologies from artificial intelligence, spatial technology, block chain technologies, and cloud computing have merged to form these new cyberspaces. That in turn created what has been known as Web 3.0, and the promising virtual worlds have been coined to Metaverse. These developments led to the addition of new spaces and dimensions to human civilization, and a new generation of young people arose immersed in these spaces and dimensions. That led to the distance and departure of this Z generation from traditional life. This new revolution began with games, then developed in the global Corona crisis, serving the cause of work, distance education, and electronic commerce, and has now expanded to the practice of all human activities. This transformation led to a qualitative civilized shift, which requires non-traditional interventions and a future positioning process. It also requires a conceptual paradigm shift in knowledge.

The spectrum of the research covers several topics, starting from the dialogue between man and machine up to the applications of artificial intelligence in issues of the Arabic language and machine perception. This research seeks to deepen knowledge about ways to build smarter human-machine interaction using natural language. Its scope has been extended by recent developments in natural language processing. Moreover, it is now possible not only to benefit from restricted passive human-machine interaction, but more interactive

smart channels have been developed where human intelligence and artificial intelligence (AI) have a chance to interact.

The paper presents a framework model for the interaction of the Arabic language in these dimensions. The paper will explore the future of the Arabic language in this context while considering the prospects and accompanying challenges. The paper introduces a set of concepts in the field of computerized interaction in the Arabic language, automatic perception, human-machine interaction, and the challenges of the third web and metaverse. It then navigates the future horizons of computing the Arabic language in WEB 3.0 and virtual worlds, then presents a model for creating a framework and a platform for developing solutions capable of producing and understanding the language in its audio and visual forms (speech and writing) in cyberspace with its rapid developments. It provides recommendations for future research areas that have opened up between the linguistic and digital disciplines.

Keywords: Metaverse, Arabic Computational Linguistics Language Perception, WEB 3.0, Language Models, Artificial Intelligence, Natural Language Processing, Semantic Verse.

المقدمة

بدأ استخدام مصطلح الميتافيرس (Metaverse) في أوائل التسعينيات، باعتبارها كلمة مركبة جديدة في الخيال العلمي، وهي منحوتة من الكلمتين (Meta) و (Universe)، وترجم حرفيًا بـ(ما وراء العالم)، وتعرب بـ(ميتافيرس)، وقد ارتأينا أيضًا استخدام المصطلح الجديد (العوالم الافتراضية) مع الميتافيرس؛ ليعمق فهم المعنى، مثل استخدام كلمة الهاتف مع الكلمة المُعَرَّبة (تليفون) ولا حرج، أصبح المصطلح الآن ونحن في بدايات العقد الثالث من القرن الحادي والعشرين، كحقبة جديدة لتطور الإنترنت، يعني منظور افتراضي للإنترنت، كعالم افتراضي واحد غامر ومتجانس عالميًا تم تطويره عبر استخدام تكنولوجيا الواقع الافتراضي (Virtual Reality)، والواقع المعزز (Augmented Reality)، وفي الاستخدام العامي، هو عبارة عن شبكة من عوالم افتراضية ثلاثية الأبعاد تركز على التواصل الاجتماعي.

دفعت التطورات الجديدة في تكنولوجيا المعلومات، الانترنت لحقبة جديدة أصبحت تعرف بالويب الثالثة (3.0 WEB)، مدعومة بالطفرة الهائلة في تطور التطبيقات المكانية (Spatial)، والدلالية (Semantic)، وتكنولوجيا الواقع الافتراضي (Virtual Reality)، والواقع المعزز (Augmented Reality)، وانبثق المصطلح الجديد ليدمج مفاهيم مثل: اللامركزية، وتقنيات سلاسل الكتل



(Blockchain)، والاقتصاد القائم على الرمز المميز (Token-Based)، مع التطورات السابقة.

سيلج هذه العوالم أولاً الجيل الجديد، وسيواجه بتحدٍ حضاري وجودي من ناحية لغته، وثقافته، وقيمه، ودينه، وأولويات مجتمعه، وسيشهد عالمنا تحديات جوهرية في مسألة التفاعل، والحضور الرقمي، من حيث نوعية المحتوى، ووسائل التواصل الذي يعمل بأنماط جديدة غيّرت المفهوم التقليدي من ذهنية البث، إلى ذهنية التفاعل، ومن ذهنية المحتوى الثابت، إلى ذهنية المحتوى المتجدد، ومن المحتوى أحادي النوع، إلى المحتوى المتعدد المتنوع، ومن سيطرة مؤسسات الإعلام الضخمة إلى نفوذ الفرد المتمكن، ومن المركزية إلى الفردية؛ ولذا تغيرت الموازين (ر الصين 2015) (ريس 2020 و 2021). وبرغم التطور الذي شهدته تكنولوجيا الإعلام الجديد، إلا أنّها لم تُلغ وسائل الاتصال القديمة، ولكن طورتها بل غيرتها بشكل ضخم، وأدت إلى اندماج وسائل الإعلام المختلفة، والتي كانت في الماضي وسائل مستقلة لا علاقة لكل منها بالأخرى، بشكلٍ أُلغيت معه تلك الحدود الفاصلة بين تلك الوسائل، وفي إطار هذا التغيير انفتحت فرص جديدة، وفي الوقت نفسه طاقات وتحديات جديدة، يعيش في هذا الفضاء الجديد حوالي 400 مليون عربي في حدود الأوطان، و حوالي 50 مليون في المهجر ينغمسون فيه حيث تنشأ أعمالهم وتمضي دراستهم، وتتشكل قناعاتهم، وأفكارهم أساساً في هذا الفضاء.

من منظور لغوي يجب تطوير مقاربة لغوية عربية جديدة لريادة هذه الفضاءات الجديدة، تتناول أهم مجالات العمل في تكنولوجيا الواقع الافتراضي والمعزز، والذكاء الاصطناعي، والويب 3، وتطوير قنوات ذكية أكثر تفاعلية باللغة العربية، حيث يكون بالإمكان التواصل الذكي بين الإنسان العربي، والحاسوب،

وتوجد فرصة للتفاعل بين الذكاء البشري، والذكاء الاصطناعي.

يتبع هذا البحث منهجية التفكير التصميمي للبحوث النوعية (Design Thinking for Qualitative Research)، وهي منهجية بازغة أُعْتُمدت في إم آي تي، وستانفورد، والعديد من المؤسسات البحثية والتكنولوجية الرائدة، لمعالجة القضايا المعقدة ضعيفة التعريف، وذلك بتفكيك مفردات المسألة، ومن ثمَّ بناء تصميم للحل، يبدأ بتطوير المفهوم كنواة صلبة، ثم البناء التجريبي في مراحل وصولاً للحلول المقبولة. باتت هذه المنهجية من المقاربات التي تم تطويرها حديثاً، وتستخدم بفعالية في بحوث الفضاء السيبراني.

اللغة العربية في العوالم الافتراضية (ميتافيرس)

كما سبقت الإشارة، فإن الميتافيرس: هو حقبة جديدة لتطور الإنترنت، يعني منظور افتراضي للإنترنت، كعالم افتراضي واحد، غامر، متجانس، عالمياً، وتم تطويره عبر استخدام تكنولوجيا الواقع الافتراضي، والمعزز، والتطبيقات المكانية، والدلالية، والذكاء الاصطناعي، وهو شبكة من عوالم افتراضية ثلاثية الأبعاد، يدمج مفاهيم مثل: اللامركزية، وتقنيات سلاسل الكتل، والاقتصاد القائم على الرمز المميز. مع التطورات السابقة تركز على التواصل الاجتماعي، وقد حددت الشركات العالمية العملاقة وجهتها؛ بالتركيز على الاستثمار في العوالم البديلة، مما يقدم إشارة واضحة للتحول العالمي المستقبلي نحو هذه العوالم الافتراضية. وستعج العوالم الافتراضية بالشخص الحقيقين والأفاتارات (Avatars)، والبشر الرقميين (Digital Humans)، والذين هم إصدارات ثلاثية الأبعاد، من روبوتات المحادثة الموجودة في الميتافيرس، في غالبيتها تحكي هيئة، وسلوك البشر، وهي ليست نسخاً من شخص آخر، بل بدلاً من ذلك، فهي أشبه بشخصيات غير قابلة للعب، تدعم بالذكاء الاصطناعي في لعبة فيديو، ويمكنها

التفاعل والاستجابة لأفعالك، في عالم الواقع الافتراضي. يتم بناء البشر الرقميين بالكامل، باستخدام تقنية الذكاء الاصطناعي، وهم ضروريون للمشهد في الميتافيرس. هناك عدد لا يحصى من التطبيقات للبشر الرقميين، وقد استثمرت شركات عديدة بالفعل في هذا الاتجاه. وقد تطورت تطبيقات البشر الرقميين في العام 2022 بشكل هائل بحيث أصبحت الشخصيات المنتجة تقدم شيئاً كبيراً مع البشر الحقيقيين.

إحدى الطرق الأساسية التي يستخدم بها البشر الرقميين الذكاء الاصطناعي، هي معالجة اللغة، حيث يساعد الذكاء الاصطناعي في تفكيك نصوص اللغات الطبيعية، مثل اللغة العربية، وتحويلها إلى صيغ يمكن قراءتها آلياً، وإجراء التحليل والوصول إلى استجابة، ثم تحويل النتائج مرة أخرى إلى اللغة العربية، وإرسالها إلى المستخدم، وعُرفت هذه الخاصية بالوصول متعدد اللغات (Multilingual Accessibility). تستغرق هذه العملية بأكملها جزءاً من الثانية، تماماً مثل المحادثة الحقيقية. واعتماداً على تدريب الذكاء الاصطناعي، يمكن تحويل النتائج إلى أي لغة بحيث يمكن للمستخدمين من جميع أنحاء العالم الوصول إلى الميتافيرس. بذلك يتم توسع عالم الواقع الافتراضي على نطاق واسع، وعندما يتم تغذية محرك الذكاء الاصطناعي بالبيانات التاريخية، ومخازن المعرفة؛ فإنه يتعلم من المخرجات السابقة، ويحاول ابتكار بياناته الخاصة، وستتحسن مخرجات الذكاء الاصطناعي في كل مرة، مع مدخلات جديدة، وردود فعل بشرية، بالإضافة إلى تعزيز التعلم الآلي. في النهاية، سيكون الذكاء الاصطناعي قادراً على أداء المهمة، وتقديم مخرجات، مثل البشر تقريباً. تقوم شركات عديدة الآن بتدريب الذكاء الاصطناعي على إنشاء عوالم افتراضية كاملة. وسيكون هذا الاختراق مفيداً في قيادة قابلية التوسع للميتافيرس، حيث يمكن إضافة عوالم جديدة دون تدخل البشر.

اللغة هي الوجه المنطوق للفكر (سليمان 2003)، وهي وعاء الفكر (هيجل) واللغة هي هيكل الفكر (المطيري)، وقد قال أحمد فائق "تَمَكَّنُ اللغةُ الإنسانَ من تحقيق استقلال الفكر عن العالم المادي، وتصبح الرموز اللغوية موضوع التفكير بدلا من الأشياء، ويتمكن الإنسان من التفكير المُجَرَّد في غيبة الأشياء" (فائق 2003)، وإدراك اللغة هو العملية التي يتم من خلالها استقبال مفردات لغة معينة وتفسيرها وفهمها. في أعتاب التحولات التكنولوجية الكبيرة في سابق الأيام، وما استتبع من ضرورة إحداث إزاحة منهجية (Paradigm Shift) في اللسانيات.

تحاول الورقة التعريف بمجموعة من المعارف الأساسية في مجال التفاعل المحوسب باللغة العربية، ثم تبحر في الآفاق المستقبلية في حوسبة اللغة العربية. ترتبط موضوعات النظرية اللسانية الحديثة للغة العربية وتكنولوجيا اللغة العربية بموضوعات المدونات، والمحتويات الرقمية، والمعاجم، والقواميس الالكترونية، والإحصاء اللغوي، كما ترتبط بقضايا الدلالة -من منظور رقمي- والنحو والصرف الآليين، والإنتاج والتمييز الآلي للنصوص، وإنتاج وتمييز الكلام، والتعرُّف البصري على المحارف، والمخطوطات، والترجمة الآلية، واسترجاع المعلومات، وتنقيب الآراء، مثل بصمة الكاتب، وتطبيقات أخرى.

يغطي طيف البحث في المجال الحوار بين الإنسان، والآلة، إلى الإدراك الآلي، حيث يسعى هذا الجزء إلى تعميق المعرفة حول طرق بناء تفاعل أكثر ذكاءً بين الإنسان، والآلة، باستخدام اللغة الطبيعية. وقد وَسَّعَتْ نطاقه التطورات الأخيرة في معالجة اللغة الطبيعية (ريس 2019 ج). علاوة على ذلك، أصبح من الممكن الآن ليس فقط الاستفادة من التفاعل السلبي المقيد بين الإنسان، والآلة، بل تم تطوير قنوات ذكية أكثر تفاعلية، حيث يكون للذكاء البشري، والذكاء

الاصطناعي، فرصة للتفاعل. كما سيساعد الذكاء الاصطناعي أيضًا في التفاعلات بين الإنسان والحاسوب. عندما ترتدي سماعة رأس متطورة للواقع الافتراضي، والتي تدعم الذكاء الاصطناعي، بحيث ستكون مستشعراتها قادرة على قراءة أنماطك الكهربائية، والعضلية، والتنبؤ بها؛ لتعرف بالضبط كيف تريد التحرك داخل الميتافيرس، ويمكن أن يساعد الذكاء الاصطناعي في إعادة إنشاء إحساس حقيقي باللمس في الواقع الافتراضي. كما يمكن أن يساعدك أيضًا في التنقل المُمكن صوتيًا، بحيث يمكنك التفاعل مع الكائنات الافتراضية، دون الحاجة إلى استخدام أدوات التحكم اليدوية. وسيمثل ذلك فتحًا كبيرًا لذوي الحاجات الخاصة، أصحاب الهمم. تجسير الفجوة الرقمية في ما بين الإنسان والحاسوب، يجسر فجوة أخرى في ما بين الذين يتخاطبون بلغات مختلفة، فتصبح الميتافيرس مترجمات وسيطة بين اللغات. الضجة الكبيرة التي صاحبت تدشين التشغيل التجريبي في نوفمبر 2022 لتطبيقات حركة الذكاء المفتوح (Open AI) لمنتجاتها حول نماذج اللغات الذكية (ChatGPT) (محول الدردشة التوليدي المدرب مسبقًا)، والتي تتفاعل في الحوار بطريقة ذكية بَشَرَت بزوغ الجيل الجديد من أدوات حوسبة اللغات المبنية على الذكاء الصناعي (OpenAI 2023).

تشكل قضية مواكبة اللغة العربية للتكنولوجيا، أو التقانة إشكالًا ذهنيًا مجتمعيًا وفرديًا؛ لأنه في كثير من المجتمعات يُنظرُ إلى اللغة العربية، ويتم ربطها مع الواقع العربي، فتتغير المسألة إلى مواكبة المجتمع العربي للتقانة، ومثل ذلك واقع المدرسة العربية؛ إذا كانت بيئتها، أو أساتذتها، أو وسائلها، قاصرة، وتقارن مع مدرسة أخرى تُدرّس باللغة الانجليزية في بيئة متقدمة، نظيفة، بها حواسيب، ووسائل التقنية الحديثة. فلا تعد هذه إشكالات موضوعية، بل هي إشكالات ظرفية، ولا بد أن يتم التعامل معها بجدية، ولا تهمل، فعند إهمالها تستوطن وتزيد.

قضية اللغة العربية ومواكبة التطور المعرفي، وضحت تمامًا في عصر النهضة الإسلامية الأولى، خاصة في عهد الخلافة العباسية، حيث انتهت الدولة في خلفائها إلى أهمية العلوم، وبدأت أولاً عملية النقل والترجمة التي اضطلع بها بعض الأعاجم، مثل السريانيين، أو غيرهم، كابن اسحق، فترجموا، ونقلوا، عددًا من الكتب الإغريقية، والرومانية القديمة، ومن بعدهم جاء الجيل الثاني من المترجمين الذين ترجموا بلغةً فصيحَةً، بعدها جاء جيل التأليف الذي ارتبط بالنهضة العلمية. نلاحظ أن هذه القضية هي قضية فكرية، فاللغة هي مرآة للفكر، فإذا ضعف الفكر في أمة؛ يصيب اللغة والمحتوى ضمورًا أكبر. وعندما يكون الإنتاج المعرفي كثيفًا تنتشر اللغة وتزدهر، وعندما يصيب الأمة انحطاط؛ تضمحل اللغة واستخداماتها ومحتواها، واللغة في مواكبتها ما هي إلا نحن، إذا أنتجنا فكريًا فستزدهر لغتنا كذلك.

شكّل علماء اللغة العربية زيادة مبكرة، في ما أصبح يعرف اليوم باللسانيات المحوسبة، فكانت ابتكارات الفراهيدي، وسيبويه، سابقه بقرون لوضع شكل علم التشكيل - المورفولوجيا- وكانت ابتكارات الثعالبي، سابقة للشبكات الدلالية والاونتولوجي. وعمل علماء في العصر الجديد في إثراء اللسانيات العربية المحوسبة، فنجد هذا الفرع الآن في حالة تطور مستمر. تطور في الحوسبة، علم يسمى تكنولوجيا اللغة، وهو علم، ومساق معرفي، وتجارب، ومعلومات تكنولوجية، وهذا الجانب الديناميكي للغة العربية تجهله اللسانيات البنوية التقليدية؛ لأنها تركز كل اهتمامها على تشخيص الوحدات في ذاتها، وبالاعتماد على تقابل الصفات الذاتية التي تميزها عن غيرها.

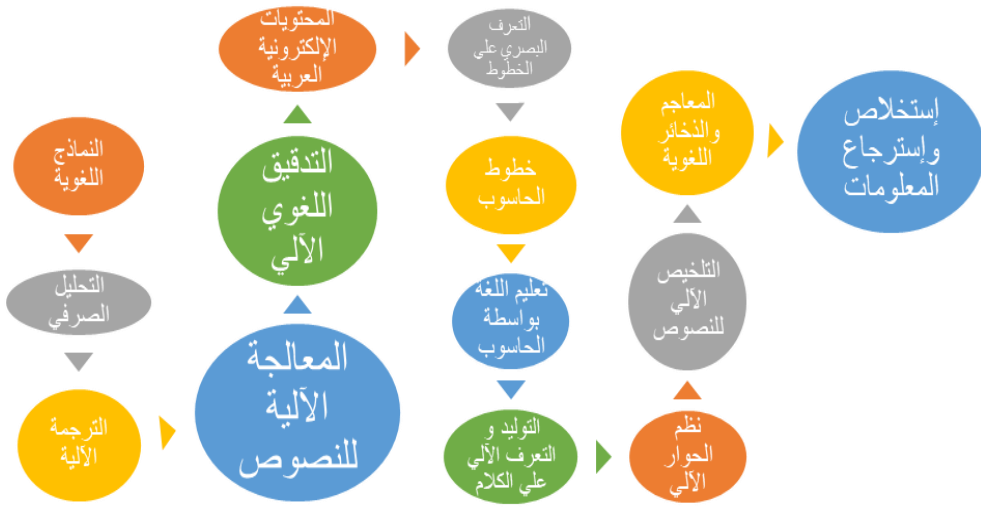
تتميز اللغة العربية بخواص فريدة، فهي لغة اشتقاقية، متسقة، ولينة في انسياقها لقوانين الصرف، وانضباط اللغة، وانسياقها للقواعد، وارتباطها بالحساب. وهي لغة اشتقاقية من الجذر بحيث يمكن تكوين كلمات كثيرة من

جذر واحد، وفي علم اللغة، تصنف من اللغات الانصهارية، وتنفرد في قيامها على تشكيل رياضي محكم، ينبني على مكونين هما: الجذر والوزن. وهي لغة متكاملة، مرتبطة مسالكها: كتابةً، وتركيباً، وصرفاً، وصوتاً، ودلالةً، وترتيباً، وتداولاً؛ فتتسق هذه القضايا، وتتكامل مع بعضها، وتترابط الأصوات، مع الأحرف، مع التراكيب، مع الدلالات. وهي لغة عالمية يتعامل معها بصورة أساسية أكثر من 500 مليون، وبصورة جزئية أكثر من مليار وربع مليار إنسان.

تكنولوجيا اللغة العربية

تطور حديثاً فرع للسانيات المحوسبة (Computational Linguistics)، يتناول قضايا استخدامات الحاسوب في معالجة القضايا اللغوية. وفلسفياً يمكن أن نشير لهذا الفرع، بأنه يتناول القضية المشتركة بين الأرقام والحروف. تطورت بحوث وتطبيقات اللسانيات المحوسبة في أواخر القرن العشرين، مع ازدياد كثافة استخدام الحاسوب في مناحي الحياة المختلفة، بواسطة أعداد أكبر من البشر، تتحدث لغات عديدة. كما كان لتطور الإنترنت ووسائل التواصل الإلكتروني دور كبير، في ولوج قطاعات بشرية كبيرة في دائرة التعامل اليومي مع الحاسوب. وتعالج قضايا التطبيقات اللسانية المحوسبة؛ لإنتاج، وتطوير نظم تعالج قضايا اللغة العربية، وفق الممارسات الهندسية.

تمت معالجة قضية التفاعل بين الإنسان والحاسوب، باللغة الطبيعية في سلسلة من الأبحاث، التي بدأت في وقت مبكر في مساق معالجة اللغات الطبيعية، وهو مجال متداخل بين تخصصات الحاسوب واللغويات، ويعنى بدراسة وتطبيق معالجة الحاسوب للغات الطبيعية. وقد مثل التَّعَرُّف الضوئي على المحارف والمخطوطات Optical Character Recognition أحد المنجزات المهمة في هذا المجال، وبواسطته يتم التعرف على المخطوطات المصورة.



كما يُمَثَّل التعرف والإنتاج الآلي للأصوات أحد الإنجازات المبكرة في المجال. إدماج هذا التطور في التعرف على الخطوط، والأصوات وإنتاجها، مع التطور في برمجيات حوسبة التعرف، وإنتاج الجمل والكلمات يمثل التحدي الآني في هذا المجال. يوجد تحدٍّ علمي تكنولوجي أوسع، في إدماج الإطار السابق في إطار أوسع يشمل وسائل التفاعل الحالية مع الميتافيرس.

وقد بدأت بحوث هذا المساق مبكرًا، بوصفها تطبيقات لبحوث الحوسبة في تطبيقات الإحصاء اللغوي، ومعالجة نظم الحوار الآلي، بواسطة الذكاء الاصطناعي، وهناك مساق متقارب لمعالجة اللغة الطبيعية وهو مساق اللسانيات الحاسوبية (Computational Linguistics)، والذي انبثق من بحوث اللسانيات. ويتكامل ويتداخل التخصصان في هذه الأيام حتي أصبحا مترادفين.

تتناول نظريات اللغويات المحوسبة قضايا ارتباط اللغويات والإدراك، من منظور شكلي Formal Perspective منهجي قابل لحوسبة حاجات الإنسان؛

لتوليد، وفهم اللغة بالاستعانة بالمعارف اللغوية المتاحة. وقد وصلت هذه النظريات في أيامنا الحالية درجة عالية من النضوج، والثراء، والتعقيد ما يفوق قدرة العقل البشري المجرد من الإحاطة بها وإدارتها؛ لذا قامت التطبيقات المحوسبة بتطوير برامج حاسوبية تستوعب النماذج الشكلية Formal Model للمعرفة اللغوية الإنسانية. وتركز تطبيقات اللغويات المحوسبة على المخرجات العلمية لنمذجة استخدامات المعرفة اللغوية الإنسانية (Cole et al 1997). ويُعرّف إطارُ المناهج والأدوات والطرائق المستخدمة في تطبيقات اللغويات المحوسبة، بالهندسة اللغوية (language engineering)، أو تكنولوجيا اللغات الإنسانية، أو تكنولوجيا اللغات الإنسانية (Human Language Technology)، وتفاعل الإنسان مع الآلات باللغة الطبيعية atural Language-Human Computer Interaction هو خط بحث بدأ مُبَكِّرًا في أبحاث وممارسات الحوسبة، ولطالما كان التفاعل بين الإنسان والآلة هدفًا وأملًا. وقد شكلت التطورات الأخيرة هذا الخط مع طرق واعدة لمستقبل هذا التفاعل، ومع لمس الأتمتة كل جزء من حياتنا الحديثة، أصبحت الحاجة إلى تفاعل أكثر ذكاءً بين الإنسان، والآلة، أكثر أهمية من أي وقت مضى. تراوحت تطبيقات التفاعل بين الإنسان والحاسوب، بين الترجمة الآلية، التدقيق اللغوي الآلي، المعالجة الآلية للنصوص، النماذج اللغوية، التحليل الصرفي، المدونات والمحتويات الرقمية، التَّعرُّف البصري على الخطوط اليدوية، خطوط الحاسوب، تعليم اللغة بواسطة الحاسوب، التوليد والتعرف الآلي على الكلام، نظم الحوار الآلي، التلخيص الآلي للنصوص، المعاجم والذخائر اللغوية، استخلاص واسترجاع المعلومات. مع الأتمتة التي تمس كل جزء من حياتنا الحديثة، أصبحت الحاجة إلى تفاعل أكثر ذكاءً بين البشر والآلات، أكثر حيوية من أي وقت مضى. وقد شكلت هذا المساق التطوراتُ

الأخيرة، والتقدم التكنولوجي مع طرق واعدة لمستقبله. فعملت على توسيع نطاقه، علاوة على ذلك، أصبح من الممكن الآن ليس فقط الاستفادة من التفاعل السلبي المقيد بين الإنسان والآلة، ولكن تم تطوير قنوات ذكية وأكثر تفاعلية، حيث أصبح لدى ذكاء الإنسان، والذكاء الآلي، فرصة للتفاعل. الكتابة اليدوية، واللمس، والصوت، والعلامات البصرية، كلها وسائل للتعبير عن اللغات الطبيعية، ويمكن أن تشكل واجهة مشتركة بين الآلات والإنسان. ويجري تطوير واجهات جديدة، مثل، استخدام تحسس موجات الدماغ. وتعد الأجهزة القابلة للارتداء، والسيارات ذاتية القيادة، والأجهزة الطبية، والمساعدون الشخصيون الرقميون، والهواتف الذكية، وآلات البيع، والأكشاك الرقمية، والروبوتات، أمثلة كثيرة على مجالات تطبيق الخطوط الأمامية. قد تؤدي النتائج الأكثر ثراءً إلى إنتاج آلات ذات عاطفة وإدراك، ويحمل هذا الخط من البحث وعودًا لسد الفجوة الرقمية للمواطنين الرقميين، والأطفال، وذوي الاحتياجات الخاصة.

المشهد اللغوي في الفضاء الرقمي

ظاهرة القرن الحادي والعشرين بحق هي وسائط التواصل الاجتماعي، والتي اجتاحت العالم في أعقاب التطور الهائل في تكنولوجيا الإنترنت، وتكنولوجيا الهواتف الجواله واندماجهما في تطبيقات الهواتف الذكية، واندياح ثقافة واستخدام هذه التكنولوجيا لدى العامة، حيث يقدر عدد المستخدمين العرب بأكثر من ثلاثمئة مليون مستخدم، ينتشرون في العالم. تشكل هذه الكتلة البشرية بيئة خصبة للتواصل والتفاعل والإثراء الثقافي. وقد استخدمت هذه الجموع وسائط التواصل الاجتماعي بفعالية لمختلف الأغراض. كل هذه الوسائط التي جذبت الملايين من العوام والخواص، وردمت فجوة الأمية، وأتاحت فضاءً حرًا للتفاعل، ودخل فيه الجميع بدون ترتيب، أو استئذان، أو تصنيف. لذا تجد

الجيد، والغث، والصحيح، والخطأ. وذلك مما يعظم الحاجة لوجود المثقفين، والعلماء؛ لتشكيل فضاء معرفي إنساني زاهر.

تتنوع الاستخدامات العامة على الوسائط والتطبيقات الرقمية تنوعاً كبيراً، فنجد استخدامات التواصل الرقمي الاجتماعي، والسياسي والمهني، وتأنس الأسر والأصدقاء، من الأعمار المختلفة: من الشيب، والشباب، وحتى الأطفال، ومن النساء والرجال، فبينما نجد خليطاً من لغة فصيحة، بسيطة، وعامية قريبة من الفصحى، في الاستخدامات المهنية، والسياسية، نجد لغة جديدة عامية مليئة بالرموز وألفاظ التورية، والألفاظ الأجنبية، في تأنس الشباب. ونجد أيضاً انتشار النشر بالعامية للشعر، والنكات، والحكايات في كل هذه الأوساط بصورة راقية وإيجابية. كما نجد التغريد علي تويتر، يشغل حيزاً مقدراً من الاهتمامات. المدونات، والمنديات مجال آخر ثرّ، فيه كثير من المساهمات العامة والفصيحة (ريس 2019- أ).

ودعا(رصين 2015) لاستخدام الوسائط لتمكين اللغة العربية، مدافعاً بأن استخدام الوسائط المتعددة يفيد اللغة العربية، والمهتمين بها، ويساعد المعلم، والتلميذ على خلق فضاء أوسع لتمكين اللغة من الاستخدام اليومي، ونَبّه على أنّ هذا التمكين مشروط بسلامة الوسائط من الأخطاء، أما إذا لم تخلُ الوسيلة من الأخطاء فتتحول إلى عقبة كؤود في طريق تطور وتحسين لغة المتلقين، وتناول مجموعة الوسائط التي تستخدم في التعليم (التعليم بالنص المكتوب وهو النمط التقليدي، والتعليم بالصورة الثابتة، والصورة المتحركة، والتعليم بالصوت، والتعليم بالفيديو، وهو الأكثر شيوعاً بين أنواع التعليم بالوسائط المتعددة. وأوضح (نهاد الموسي 2015) "إن الفصيحة والعامية تحتلان موقعين مركزيين في المشهد اللغوي، الذي شبهه ببرج بابل، ذلك أن العربية في المشهد اللغوي، تتمظهر في تجليات شتى، فهي عربية فصحي في المصحف المرتل، وعربية فصيحة بالفعل في مثل الدراما التاريخية، والتقارير الوثائقية، والنشرات الإخبارية، وعربية

فصيحة بالقوة في البحوث، والمؤلفات والدوريات، والصحف، وعربية فصيحة محكية، يحاولها ويلتزمها متخصصون ومثقفون، وعربية شبه فصيحة، تجري بها ألسنة مراسلي بعض الفضائيات، وعربية وسطى، وهي مزاج من العامية المكتسبة، والفصحى المتعلّمة، ولهجات عامية محكية، ومتداولة في سياقاتها المحلية، بل ممتدة في بعض فضاء الإعلام، والأفلام العربية، والدراما الاجتماعية، والشعر الشعبي، (أو النبطي)، وفيض الأغاني الشبابية، ولهجات عامية مكتوبة في حواشي بعض القنوات الفضائية، والإعلانات التجارية، وعربية عامية مهجنة بالهندية أو الأردنية، وعربية عامية مهجنة بالفرنسية وغير ذلك، لافتًا إلى أن الفصحى والعامية تمثّلان أبرز ما يتناظر في هذا المشهد اللغوي من وجوه التداول بالعربية، وتمثّلان الظاهرة الموسومة بالازدواجية. وأن الازدواجية في العربية تمثّل مظهرًا حاسمًا من مظاهر التحول الذي جرى على العربية، وأن صفوة الرأي في العلاقة بين الفصحى والعامية، أنها قائمة على التقابل الثنائي المتمايز بآمر العُرف الاجتماعي، وأن امتداد بعض الفصحى في العامية وتسلسل بعض العامي في الفصيحة، مرجعه إلى ما يكتنف الخطاب اللغوي من تباين المقاصد، وهو يفيء في نهاية التحليل إلى جدل الثقافي والاقتصادي.

ونافلة القول أرى وجوب حوسبة اللهجات العربية العامية لتتكامل، وتتفاعل مع اللغة العربية الفصيحة والعامية في الوسائط الرقمية إيجابيًا، وتتواجد بشكل متزن. وقد فتحت وسائط التواصل الاجتماعي طاقة واسعة للتعبير والتفاعل، والتعليم، وهي مساحة لا يمكن تجاوزها، وقد أعادت عامة الناس إلى حظيرة الثقافة المشافهة والمرئية، فلا بد من تعاضد الجهود لإثراء هذه الدائرة باللغة الرصينة، والمحتوي الجيد. وأدعو لابتدأ مشروعات الرقمنة، والحوسبة، لإثراء المحتوى العربي، ولبلورة العامية في قوالب علمية، ودعمها بأدوات رقمية وتطبيقات وسائط ذكية.

التمثيل الرقمي للغة العربية

اللغة هي تعبير عن الفكر، ولها وجهان: (الكلام و الكتابة) وأتى التمثيل الرقمي؛
ليعبر عن كل من الكلام، والكتابة، وذلك علي مستويات عديدة أهمها:

■ الحرف العربي

■ المفردات

■ الجمل

حوسبة الحرف

كان التحدي الأول هو ترميز الحروف العربية برموز يمكن ترجمتها إلى لغة الآلة، واللغات العليا للحاسوب، وكانت المحاولات الأولى في السبعينات، إلى إن تم الاتفاق علي رموز ال ASMO 449E للتعبير عن الحروف العربية مع وجود رموز أخرى. ثم ابتداء ال Unicode في أواخر القرن الماضي، وبه صفحات الرموز العربية، بالإضافة إلى رموز عالمية اللغات الموجودة علي البسيطة، ويتكون نظام الترميز العالمي الجديد من حرفين رقمين Byte، أو من ستة عشر لبنة رقمية Bit. وكان تحدٍ آخر انفردت به اللغة العربية، وهو تشكيل الحروف، الذي يشكل عائقًا إضافيًا أمام ترتيب الحروف العربية، ومن التطبيقات المهمة على مستوى الحرف، قضية الخطوط وهي التعبير الصوري للحروف والكتابة العربية، وتنفرد بخصائص كثيرة منها: كتابتها من اليمين إلى الشمال، واتصال خط حروفها، وتعدد أشكال خط الحرف، حسب موقعه، وقد شكّل هذا التفرد تحديًا كبيرًا في الحوسبة.

حوسبة الجذر

ترجع الأفعال الثلاثية إلى ستة أبواب معلومة هي: (ضرب، نصر،....)، وتتكون هذه الجذور من تباديل معلومة للحروف العربية مشكّلة. وتشكل هذه الجذور

لبنات الكلمات العربية، وهنالك تصنيف قديم يرجع الكلمات للجذور، والحوسبة هنا عملية مقابلة للأفعال والمصادر إلى أبوابها المتفق عليها (ريس 1992 و 2019 ب) (الهيئة 1993) (الخطيب 1985).

حوسبة الكلمة

أولى المجالات التطبيقية الناجحة للحوسبة علي مستوى الكلمة، كانت مجهودات شركة صخر في مطلع التسعينات، حيث استطاعوا انتاج محلل صرفي، ومولد لكلمات عربية، يعمل بدقة وكفاءة عالية، وتشكل المحللات الصرفية التطبيق الأهم لنظريات علاقات اللغات، والآلات وبالذات نظرية الأوتوماتا : Automata Theory حيث تُقسَّم عمليات حوسبة الكلمات إلى مستويين فرعيين: الأول: وهو توليد الكلمات من الجذور بقواعد الاشتقاق المعروفة؛ وذلك للتعبير عن الدلالات المعنوية بأوزان معلومة، ومثال ذلك: التعبير عن دلالة الآلة بوزن فاعول، من الجذر (حسب) لتصبح (حاسوب)، لتدل علي آلة تستخدم في الحساب.

أما المستوى العربي الثاني: فهو الكواسع والبوادي (اللواحق والصدور) (Meduna 2000) (Sudkamp 1997) (Watt 1985).

حوسبة اشباه الجمل

تصنف الكلمات العربية إلى: فعل، واسم، وحرف، وقامت قواعد النحو والصرف علي ربط الكلمات لإنتاج جمل وأشباه جمل، تعبر عن معطيات إدراكية جزئية، ولا يوجد في اللغة العربية فصل واضح بين قواعد إنتاج أشباه الجمل، والجمل، أو تعريف لمفهوم شبه الجملة. ولكن القصد من إدراج هذا التصنيف هنا هو التعبير عن المقاطع غير الكاملة، والتي تفيد معنى جزئياً مثل: النسبة، والتأنيث والتذكير، والجمع والتثنية، مثل: كتابان، و كتابي، وكتب. يتم استخدام

أساليب حوسبة الجملة في هذا المستوى بنفس الطرق مع فارق المعطيات الجزئية .

حوسبة الجمل

الجملة هي وحدة الكلام المفيدة القابلة لإفهام الطرف الآخر معطيات إدراكية، والجملة العربية: إما اسمية، أو فعلية، والجمل الاسمية، صيغ دالة على حالة ثابتة، بينما الجمل الفعلية، صيغ دالة على حالات متحركة. كانت حوسبة الجمل، من المجالات التي انصبت عليها جهود كثيرة في الحوسبة، فقامت بحوث الدلالة من منظور رقمي، والنحو والصرف الآليان، والإنتاج والتمييز الآلي للنصوص، وإنتاج وتمييز الكلام، والتعرف البصري على المحارف والمخطوطات، والترجمة الآلية، فكانت من أكثر المجالات ثراءً.

حوسبة النصوص

تعتبر النصوص عن أحداث متصلة، فتكون من العديد من المعطيات الإدراكية. وتندرج هنا فنون القصة، والشعر، والتعبير الإنساني المختلفة، ولا يوجد بُعدٌ إطار نظري أو تطبيقي للحوسبة في هذا المستوى؛ لأنه يتجاوز المستويات الآلية إلى المستويات الذكية العاقلة.

تطور الصرف المحوسب في الإنتاج والتعرف الآلي للمفردات والتراكيب العربية بشكل جيد، والتحدي الآن في التمثيل الرقمي للغة العربية، بناءً منظومة للتحليل الدلالي تربط بين المنصوص والمقصود (اللفظ و القصد). وقد شكل علماء اللغة العربية زيادة مبكرة فيما أصبح يعرف اليوم باللسانيات المحوسبة، فكانت ابتكارات الفراهيدي وسيبويه سابقه بقرون لوضع شكل علم التشكيل (المورفولوجيا)، وكانت ابتكارات الثعالبي سابقة للشبكات الدلالية، والأونتولوجي. عمل علماء في العصر الجديد على اثراء اللسانيات العربية المحوسبة، فنجد هذا

الفرع الآن في حالة تطور مستمر. بإيراد بعض الابتكارات القديمة، واستعراض بعض التطورات الحديثة نحفز، ونوثق للربط ما بين علوم العربية الراسخة، وما استجد في هذا الفرع الجديد، ونشكل مقارنة لإثراء الجديد من الكنز القديم. وقد تطور في الحوسبة علم يسمى تكنولوجيا اللغة، وهو علم، ومساق معرفي، وتجارب، ومعلومات تكنولوجية، من ضمنها، حوسبيات اللغة العربية، ويوجد مساق بدأ يتضح الآن يسمى (هندسة اللسانيات العربية).

قامت علي الفكر النحوي الخليي النظرية الخليلية الحديثة، والتي اتجهت إلى إعادة قراءة التراث اللغوي العربي الأصيل، والبحث عن خباياه، بغية التنبيه إلى الطفرة التلقائية المفاجئة، التي أحدثها الخليل، وسيبويه، وتلاميذهم، في تاريخ علوم اللسان البشري. مثل: عبد الرحمن الحاج صالح، زيادة بصياغته النظرية الخليلية الحديثة، مروجًا للنحو العربي العلمي، لا التعليمي، أو الفلسفي، والذي هو مجموع المثل والقواعد التي يمكن أن تفرع بها وعلمها جميع الإمكانيات التعبيرية الخاصة بالوضع العربي (عبد الرحمن الحاج صالح 2006). فهذا الجانب الديناميكي للغة تجهله اللسانيات البنيوية التقليدية؛ لأنها تركز كل اهتمامها على تشخيص الوحدات في ذاتها، وبالاعتماد على تقابل الصفات الذاتية التي تميزها عن غيرها.

تطور الإنتاج الآلي للمفردات، وذلك بتمكين التوليد الآلي من اشتقاق الأفعال، والأسماء المشتقة، والمصادر وتصريفها، انطلاقًا من جذورها الثلاثية، أو الرباعية (ريس وآخرون 2019 ب)، باستخدام قواعد النحو، والصرف، والمعجم المحوسب. وكمثال: لإنتاج كلمة من الجذر (فتح) تدل على آلة يستخدم وزن اسم الآلة (مفعال) فنقول: مِفْتاح. وأيضًا هناك في اسم الآلة دلالات معنوية مختلفة لكل وزن من الأوزان، وباختيار الدلالة المستهدفة مع الوزن المختار لذلك مع

الجذر يمكن إنتاج المفردة آلياً. الصرف الآلي للنصوص يعمل على المقارنة ما بين تركيبية موجودة، وتركيبات معيارية، فهذه المقارنة القصد منها، التعرف الآلي على التركيبية اللغوية. ونعني التراكيب اللغوية الكاملة في شكل جمل لأن الإنسان يفهم بالجمل، ويدرك بالجمل، فلا بد أولاً: أن يتعرف على الجملة، ويضعها في السياق الكامل للكلام، وبعد ذلك يحلل هذه الجملة لمفرداتها، ولدلالاتها المختلفة (دلالات الجملة ودلالات الكلمات).

كما يمثل التعرف الضوئي على المحارف Optical Character Recognition أحد المنجزات الهامة في هذا المجال، وبواسطته يتم التعرف على المخطوطات المصورة. إدماج هذا التطور في التعرف على الخطوط، والأصوات، مع التطور في إنتاج برمجيات حوسبة إنتاج الجمل، والكلمات، يمثل التحدي الآني في هذا المجال. كما يوجد تحدٍ علمي تكنولوجي أوسع في إدماج الإطار السابق في إطار أوسع يشمل وسائل التفاعل الحالية مع الحاسوب من لوحة مفاتيح وفارة وخلافه. تجسير الفجوة الرقمية في ما بين الإنسان والحاسوب، يجسر فجوة أخرى في ما بين الذين يتخاطبون بلغات مختلفة، فتصبح الحواسيب مترجمات وسيطة بين اللغات. وفي هذا السياق لابد من الإشارة للتطور الكبير الذي حدث في قضايا الترجمة الآلية في ما بين اللغات مع الاعتراف بأن المنتجات الحالية لا زالت في مراحلها المبكرة، وتساعد المترجمات الآلية المترجمين البشر الآن لزيادة إنتاجهم وتجويده بصورة كبيرة.

المدونات اللغوية Corpus

الجمع هو Corpora وتتكون من مجموعة متون، أو مجموعة من النصوص المكتوبة، أو المنطوقة، وتكون هذه المادة اللغوية كبيرة الحجم. للمدونات اللغوية أسماء عديدة منها: المتن اللغوي، الذخيرة اللغوية، مدونة نصية، لسانيات

المتون، لسانيات المدونة اللغوية. وتعتبر المدونات اللغوية الرقمية واحدة من أهم تطبيقات اللسانيات الحاسوبية المستخدمة في تطبيقات حاسوبية كثيرة منها: أنظمة استرجاع المعلومات، مثل: نظم استرجاع الوثائق، والرد الآلي، ونظم تصحيح القواعد، ونظم تصحيح الإملاء، والترجمة الآلية (ثنائية)، ومتعددة اللغة)، والصناعة المعجمية وغيرها. وتعمل تطبيقات المدونات في بناء المعاجم، البحث عن بعض خصائص اللغة كدراسة مترادفات والفروق بينها وبين تراكيبها، دراسة التطور الدلالي، دراسة المصطلحات والتعابير المسكوكة، ارتباط بعض المفردات ببعض التراكيب النحوية، دراسة بعض التراكيب النحوية وعلاقتها، ونوع النص، إيجاد العلاقات الدلالية، بناء الأنطولوجيا، وتعليم اللغة للناطقين بغيرها، دراسة اللغة المنطوقة، وكذلك التداولية التي -في الغالب- تعتمد على اللغة المنطوقة أكثر من المكتوبة، تصميم أدوات وبرمجيات حوسبة اللغة، تحليل الخطاب والدراسات اللسانية الاجتماعية، دراسة اللهجات، دراسة الأساليب في لسانيات النص. (عدني 2019).

لا بُدَّ من الإشارة هنا إلى أن من بين المجهودات الرائدة مجهودات المنظمة العربية للتربية، والثقافة، والعلوم، مشروع النهوض باللغة العربية، للتوجه نحو مجتمع المعرفة. ويعمل في البحث والتطوير في مجال المعالجة الآلية للغة العربية ولمّ شمل العديد من العلماء والباحثين من داخل، وخارج الوطن العربي. وتهيئة المناخ المناسب لهم لتبادل الآراء والمعارف فيما بينهم. وأثمرت هذه المجهودات عددًا من البرامج المجانية مفتوحة المصدر، مثل: نظام الاشتقاق والتصريف في اللغة العربية، وبرنامج الخليل الصرفي، ومحلّ آلي للغة العربية، ومشكّل آلي عربي مفتوح المصدر، والمساهمة في بناء الشبكات الدلالية للمفردات العربية، ومحرك البحث في المحتوى الرقمي العربي. كذلك لا بُدَّ من الإشارة للمدونة

العربية (المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية)، والمعجم التاريخي للغة العربية لاتحاد مجامع اللغة العربية، الذي يعمل عليه مجمع الشارقة مع بقية المجامع، والمعجم التاريخي للغة العربية بقطر، من ضمن المجهودات الرائدة الرائعة التي بدأت تحدث حراكًا، وتشكل أساسات قاعدية للأعمال المستقبلية في تطوير أدوات اللغة العربية لاستشراف المستقبل.

العالم الدلالي (Semantic Verse)

اللغة هي تعبير عن الفكر ولها وجهان: (الكلام و الكتابة)، وأتى التمثيل الرقمي ليُعبّر عن كلٍّ من الكلام، والكتابة، وذلك على مستويات عديدة أهمها: الحرف العربي، والمفردات، والجمل، ويتمثل التحدي في التمثيل الرقمي للغة العربية في بناء منظومة للتحليل الدلالي، تربط بين المنصوص والمقصود (اللفظ والقصد) (ريس 2019-ج). عرف السيد الشريف الجرجاني الدلالة بأنها: "الدلالة هي كون الشيء بحاله يلزم من العلم به العلم بشيء آخر، والشيء الأول هو الدال، والثاني هو المدلول، وكيفية دلالة اللفظ على المعنى باصطلاح علماء الأصول محصورة في عبارة النص، وإشارة النص، واقتضاء النص". وقسم فايز الداية الدلالة إلى: دلالة أساسية (أو معجمية)، ودلالة صرفية، ودلالة نحوية، ودلالة سياقية. والدلالة الصرفية تُعنى بدراسة التركيب الصرفي للكلمة، وإسقاط المعنى الصرفي للصيغة؛ ليضاف على المعنى المعجمي لمادة الكلمة؛ لاستنباط المعنى الكلي للكلمة. وقد عرّفها حيدر بأنها: "تلك الدلالة التي يعرب عنها مبنى الكلمة" وقد أشار الحازمي إلى أن "البنية الكلمة أهمية في تحديد معناها، فعن طريق البنية وصيغها المختلفة، تبرز المعاني وتُحدّد". وقال الخليل: "هي المعاني المستفادة من الأوزان، والصيغ المجردة" تطور حديثًا التحليل الدلالي Semantic Analysis ليعمل علي بناء القدرة على استيعاب وفهم النصوص

واستنباط قيم رقمية دالة عن المعاني والمضامين في إطار السياق. لو درسنا أعمال الثعالبي لوجدنا إرثاً ضخماً في هذا المضمار. تمثل الشبكات الدلالية علاقات دلالية بين المفاهيم، وغالباً ما تستخدم كطريقة لتمثيل المعرفة. تعتبر الشبكة مخططاً موجهاً، أو غير موجه مؤلفاً من عقد، التي تمثل المفاهيم بالإضافة إلى الخطوط التي تمثل العلاقات. التحليل الدلالي يعمل علي بناء الحقول الدلالية للكلمات المركزية التي تشكل أساساً بؤرياً بحيث تكون (شبكة معرفية) للكلمة الأساسية التي نبحث فيها، وذلك برصد الكلمات المرادفة، والمعاكسة، والقريبة للكلمة المركزية. حيث يعمل التحليل الدلالي علي بناء القدرة على استيعاب وفهم النصوص واستنباط قيم رقمية دالة عن المعاني والمضامين في إطار السياق. وتعرّف الويكيبيديا الشبكة الدلالية علي أنها شبكة تمثل علاقات دلالية بين المفاهيم، وغالباً ما تستخدم باعتبارها طريقة لتمثيل المعرفة. تعتبر الشبكة مخططاً موجهاً أو غير موجه مؤلفاً من عقد، التي تمثل المفاهيم بالإضافة إلى الخطوط .

أصبح مصطلح "العالم ذو الدلالات اللفظية" أو "العالم ذو المعاني" ثورة جديدة حيث تصبح المعلومات والبيانات قابلة للمعالجة منطقياً من قبل الذكاء الاصطناعي بحيث تتحول تلك المعلومات، والبيانات إلى شبكة بيانات ذات معنى، "أي أنه يمكن للبرامج الحاسوبية الخاصة أن تعرف ماذا تعني هذه البيانات". مع مجموعة من الطرائق والتقنيات المتبعة لجعل الآلات قادرة على فهم المعاني أو "الدلالات" للمعلومات على الميتافيرس، يصبح هذا العالم شبكة من البيانات التي يمكن معالجتها من قبل الآلات ويجعل العالم أكثر قابلية للفهم من قبل الآلات فيشكل بنية تحتية مناسبة للعملاء الأذكياء (Intelligent Agents) للقيام بعمليات معقدة لمستخدميهم، وهو التعريف الصريح عن المعلومات الموجودة

ودمج المعلومات بطريقة ذكية، وتوفير الوصول المعنوي الدلالي في الميتافيرس واستخراج المعرفة من النصوص.

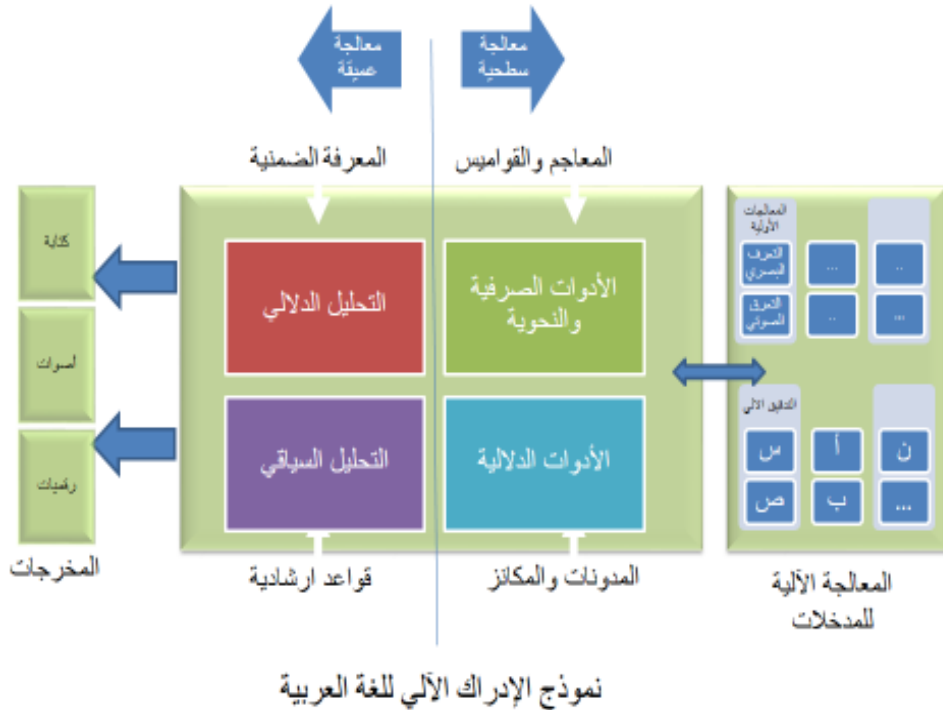
الإدراك الآلي للغة العربية

وقال سيد سليمان "اللغة هي الوجه المنطوق للفكر" (سليمان 2003). وفي العلاقة بين النص المنطوق والمكتوب يري أبو العباس القلقشندي صاحب صبح الأعشى: "أن الخط واللفظ يتقاسمان فضيلة البيان ويشتركان فيها من حيث إن الخط دال على الألفاظ والألفاظ دالة على الأوهام ولاشتراك الخط واللفظ في هذه الفضيلة؛ وقع التناسب بينهما في كثير من أحوالهما، وذلك أنهما يعبران عن المعاني، إلا أن اللفظ معنى متحرك، والخط معنى ساكن، وهو وإن كان ساكنًا فإنه يفعل فعل المتحرك بإيصاله كل ما تضمنه إلى الأفهام، وهو مستقر في حيزه قائم في مكانه، كما أن اللفظ فيه العذب الرشيق السائغ في الأسماع كذلك الخط فيه الرائق المستحسن"، وقال أيضًا "ولما اشترك اللفظ والخط في الفوائد العامة التي جعلت فيهما وقع الاشتراك أيضًا بين آلهما إذ آلة اللفظ اللسان، وآلة الخط القلم، وكل منهما يفعل فعلًا" (القلقشندي 821هـ). بدخول الحاسوب أضحت لدينا وسيلة جديدة للتعبير والتعرف على كل من الخط واللفظ، مع إدماجهما مع وسائط تعبيرية أخرى مثل الصورة.

من منظور لغوي تطبيقي نقدم مقاربة لغوية عربية لريادة هذه الفضاءات الجديدة، تعمل في فضاءات الواقع الافتراضي والمعزز، والذكاء الاصطناعي والويب 3 لتطوير قنوات ذكية أكثر تفاعلية باللغة العربية، حيث يكون بالإمكان التواصل الذكي بين الإنسان العربي والحاسوب، وتوجد فرصة للتفاعل بين الذكاء البشري والذكاء الاصطناعي. بتفكيك مفردات المسألة في الفصول السابقة تم تطوير النموذج اللاحق وبناء تصميم للحلول في المجال يبدأ بتطوير نموذج

الإدراك الآلي للغة العربية، كنواة صلبة تقبل البناء التجريبي حولها في مراحل؛ وصولاً للحلول المقبولة.

يهدف النموذج الموضح في الشكل أدناه لصناعة إطار ومنصة لتطوير برمجيات قادرة على إنتاج وفهم اللغة في صورتها الصوتية والمرئية (الكلام والكتابة)، في مجالات مثل: نظم الحوار، والاستفسارات عن قواعد البيانات، واسترجاع النصوص والتحكم بالنظم الخبيرة Expert system، والإنسان الآلي Robot، وتشكل المحللات الصرفية والأدوات الدلالية التطبيق الأهم لنظريات علاقات اللغات والآلات، وبالذات نظرية الآلة الذاتية (الأتوميتا)، حيث تقوم الآلة بحوسبة عمليات إدراك وإنتاج المفردات العربية (ريس 2020



الخاتمة والتوصيات

في هذه الورقة تم تقديم صورة عن توجهات البحوث الحالية والمستقبلية في مجال اللسانيات الحاسوبية العربية، وتم تقديم نموذج للإدراك الآلي للغة العربية لصناعة أطر وأدوات للغة العربية تواكب التطورات الكبيرة الحادثة في اللسانيات التطبيقية، وتحديات العوالم الافتراضية (المتافيرس)، وثورة الويب الثالث. تنبه الورقة لتخلف بحوث وحلول اللسانيات التطبيقية العربية خلف نظيراتها العالمية الأخرى، وضرورة إحداث تحول منهجي في مفاهيم وأطر هذا المجال لمواكبة الطفرة الهائلة التي تشهدها تطبيقات اللسانيات المحوسبة عالميًا. وقد لا يتأتى ذلك بدون تضافر القائمين علي بحوث اللسانيات العربية، وأهل بحوث الحاسوب والذكاء الاصطناعي، مع استجلاب ونقل المعرفة، والتكنولوجيا العالمية في هذا المجال. توصي الورقة بتطوير استراتيجية لغوية عربية جديدة استجابة للتحديات الجديدة، وتشكيل فريق عمل عربي للتنسيق ووضع السياسات لاستكمال تطوير المعارف اللغوية والبرمجيات اللازمة لبناء الطيف الكامل لحوسبة اللغة العربية. كما توصي بإنشاء مجموعات عمل مشتركة للعمل في قضايا المحتويات العربية في الفضاء الرقمي. كما تؤشر لأهمية العناية بتطوير محتويات عربية غنية متنوعة لمواكبة الفضاء الجديد، والاهتمام بربط المجتمع العلمي مع الصناعة واستقطاب التمويل، ودفع استخدام الأطر المحوسبة والمحتويات الرقمية في تعليم اللغة.

قائمة المراجع

أولاً: المراجع العربية

- 1- الحازمي، عليان، علم الدلالة عند العرب، مجلة جامعة أم القرى لعلوم الشريعة واللغة العربية وآدابها، ع27، 2004.
- 2- حيدر، فريد، "علم الدلالة - دراسة نظرية وتطبيقية"، مكتبة الآداب للطباعة والنشر والتوزيع، القاهرة، 2005.
- 3- خليل، حلمي، الكلمة دراسة لغوية معجمية، دار المعرفة الجامعية، الاسكندرية، ط2، 1996.
- 4- الداية، فايز، "علم الدلالة العربي - بين النظرية والتطبيق"، دار الفكر، دمشق، ط2، 1996.
- 5- الرصين، رصين صالح، "الإعلام والوسائط المتعددة عقبة في تطوير العربية أم فضاء للتمكين؟" الندوة الدولية - اللغة العربية وسؤال المستقبل 5-6 مارس 2015 وزارة الثقافة، قطر.
- 6- الرواشدة، زياد أحمد إبراهيم، أسامة عبد الوهاب ريس، دلالات مفردات القرآن وأبعاد النمذجة الرياضية لها، مجمع اللغة العربية الخرطوم، 2021.
- 7- ريس، أسامة عبد الوهاب، "اللغة العربية والحوسبة - مقدمة لهندسة اللسانيات العربية" مجلة مجمع اللغة العربية، السودان، 2007.
- 8- ريس، أسامة عبد الوهاب، اللغة العربية والتطور التكنولوجي، مجمع اللغة العربية الخرطوم، 2014.
- 9- ريس، أسامة عبد الوهاب، اللسانيات العربية المحوسبة وتطبيقاتها، الخرطوم، السودان، 2017.

- 10-ريس، أسامة عبد الوهاب، واقع وآفاق تطور تكنولوجيا اللغة العربية، مجمع اللغة العربية الخرطوم، 2017.
- 11-ريس، أسامة عبد الوهاب، "اللسانيات العربية المحوسبة ابتكارات قديمة ونماذج حديثة"، الخرطوم، السودان، 2017.
- 12-ريس، أسامة عبد الوهاب، "حوسبة الطيف اللغوي بالسودان، المجلة العربية للدراسات اللغوية، المنظمة العربية للتربية والثقافة والعلوم، ع 44، 2019.
- 13-ريس، أسامة عبد الوهاب، بابكر زين العابدين، محمد عدني حمد السيد، وزينب إدريس، الدلالات الصرفية لصيغ المصطلحات، ورشة عمل اللسانيات المحوسبة التطبيقية - نحو آفاق جديدة للغة العربية في الفضاء الرقمي عمان، الأردن، 2019.
- 14-ريس، أسامة عبد الوهاب، النظرية اللسانية الحديثة للغة العربية، ورشة عمل اللسانيات المحوسبة التطبيقية - نحو آفاق جديدة للغة العربية في الفضاء الرقمي، عمان، الأردن، 2019.
- 15-ريس، أسامة عبد الوهاب، آفاق الإدراك الآلي للغة العربية ومستقبل التفاعل بين الإنسان والحاسوب، المؤتمر العلمي الدولي الإلكتروني الأول، المجمع العلمي العراقي، بغداد، العراق، 2020.
- 16-ريس، أسامة عبد الوهاب، ابتكار النظرية الخيلية الحديثة عند عبدالرحمن الحاج صالح، ندوة الخليل، ندوة عبدالرحمن الحاج صالح مجمع اللغة العربية الخرطوم، 2021.
- 17-ريس، أسامة عبد الوهاب، اللغة العربية والمحتوى الرقمي وتحديات

المستقبل، ندوة مجامع اللغة العربية: بين الضرورة وتحديات المستقبل، اللجنة الوطنية القطرية للتربية والثقافة والعلوم والملتقى القطري للمؤلفين، 2021.

18- سليمان، سيد عبد الحميد، سيكولوجية اللغة والطفل، دار الفكر العربي، 2003.

19- الشريف، الجرجاني، "كتاب التعريفات، دار الكتب العلمية، بيروت، 1983م.
20- صالح، عبد الرحمن الحاج، أنماط الصياغة اللغوية الحاسوبية والنظرية الخليلية الحديثة، المؤتمر الخامس لمجمع اللغة - اللغة العربية في عصر المعلوماتية، دمشق، 2006.

21- عدني، محمد، المدونات اللغوية ونظم استرجاع المعلومات، ورشة عمل اللسانيات المحوسبة التطبيقية - نحو آفاق جديدة للغة العربية في الفضاء الرقمي، عمان، الأردن، 2019.

22- فائق، أحمد، مدخل عام لعلم النفس، مكتبة الانجلو المصرية، 2003.
23- القلقشندي، أحمد بن علي، صبح الأعشى في صناعة الإنشا، تحقيق: يوسف علي طويل، دار الفكر، دمشق، 1987.

24- الموسي، نهاد، "الفصحى والعامية تكامل أم تضاد؟"، الندوة الدولية - اللغة العربية وسؤال المستقبل 5-6 مارس 2015 وزارة الثقافة، قطر.

ثانيا: المراجع الأجنبية

- 1- A . Meduna, (2000) '*Automata and Language*,' Springer.
- 2- Cole R.A., J. Mariani, H. Uszkoreit, G. Varile, A. Zaenen, V. Zue, A. Zampolli (Eds.) (1997) '*Survey of the State of the Art in Human Language Technology*,' Cambridge University Press and Giardini.
- 3- OpenAI (2023) ChatGPT: *Optimizing Language Models for Dialogue*. Available at <https://openai.com/blog/chatgpt/> (Accessed 19 February 2023).
- 4- Thomas Sudkamp (1997), '*Language And Machine*,' Addison-Wesley.
- 5- Uszkoreit, H., V. Kordoni, V. Kubon, M. Rosner and S. Kirchmeier-Andersen. (2005). '*Language Technology from a European Perspective*'. In Chris Brew and Dragomir R. Radev (eds.), *Proceedings of the Second ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, pp. 43 - 48, University of Michigan - Ann Arbor, June 2005, Association for Computational Linguistics.
- 6- Uszkoreit, H., F. Xu, J. Steffen and I. Aslan (2006) '*The Pragmatic Combination of Different Cross-Lingual Resources for Multilingual Information Services*' In Proceedings of LREC, Genova, Italy, May, 2006.



تمثيل النصوص العربية وتصنيفها بناء علي نهج ثنائي متواليات الحروف الهجائية

**Representation and classification of Arabic
texts based on Alphabet**

د. فاطمة الغنام

دكتوراه في هندسة الحاسبات والنظم – جامعة بنها - مصر

الملخص

نظراً للتنامي المتزايد في أعداد الملفات عبر شبكة الإنترنت أصبحت عملية التصنيف الآلي للنصوص أحد الموضوعات البحثية المهمة في مجال معالجة اللغات الطبيعية. يستخدم التصنيف الآلي للنصوص في كثير من التطبيقات مثل إدارة المحتوى، البحث السياقي، التنقيب عن الآراء، تحليل مراجعة المنتجات، تصفية الرسائل غير المرغوب فيها، واستخراج المشاعر النصية. لتصنيف مجموعة من النصوص آلياً، يلزم أولاً تحويل النصوص إلى صورة رقمية مناسبة لخوارزميات التعلم الآلي. الطريقة الشائعة لاستخراج سمات النصوص وتمثيلها رقمياً هي حقيبة الكلمات. لكن تتمثل المشكلة الرئيسة لهذه الطريقة في الأبعاد العالية لمساحة متجه السمات، والتي تتناسب تناسباً طردياً مع أعداد الكلمات الموجودة بالمستندات، فضلاً عن الحاجة إلى أدوات وبرامج تعتمد على معالجة اللغة الطبيعية، والتي تمثل تحدياً كبيراً في حالة اللغة العربية. تشرح هذه الورقة نهجاً جديداً لاستخراج سمات النصوص وتمثيلها رقمياً يعتمد على الحروف الهجائية. هذا النهج يتبنى الفرضية المعروفة باسم رمزية الصوت التي تنص على الارتباط الظاهر بين متواليات صوتية معينة، ومعاني معينة في الكلام. النهج المتبع في هذه الورقة يقدم إسهامين رئيسين في مجال تصنيف النصوص. أولاً- أظهرنا إمكانية استخدام سمات ثابتة تستند إلى الحروف الهجائية دون الحاجة إلى المفردات الموجودة بالمستندات؛ وهذا يسهم بشكل كبير في تقليل الأبعاد العالية لمساحة متجه السمات. ثانياً- لا يتطلب هذا النهج استخدام أدوات وبرامج معقدة لمعالجة اللغة الطبيعية، ولا سيما اللغات عالية التصريف والاشتقاق كما في اللغة العربية. لقد أثبت العمل الحالي القدرة

على تصنيف مجموعات من المستندات النصية العربية بنجاح. حيث أظهر ما يقرب من 80٪ توفيرًا في مساحة متجه الملفات وتحسنًا في الأداء بنسبة 2٪ مقارنة بأفضل النتائج الأخرى المسجلة لتصنيف مجموعة بيانات الجزيرة الإخبارية.

الكلمات المفتاحية: معالجة اللغات الطبيعية، تصنيف النصوص، تمثيل النصوص رقميًا، الوثائق العربية، الحروف الهجائية، خوارزميات التعلم الآلي، رمزية الصوت.

Abstract

In text classification, texts must be converted into numerical representations that the learning algorithms can understand. The high dimensions of vector space and the requirement for language-specific tools are two major issues with the widely utilized bag of words technique. In the present study, a new alphabetic approach is used to construct feature terms and applied it to classify Arabic documents. The proposed approach has two main contributions to the text classification area. First, we have demonstrated the possibility of using constant feature terms that are based on the standard alphabet without the need for document's vocabularies; this definitely helps in reducing the dimensions of the vector space for large corpus. Second, it does not require natural language processing tools. The current work has proved the ability to classify collections of Arabic text documents successfully. It showed approximately 80% savings in vector space and 2% performance improvement compared to the best recorded results on Aljazeera News dataset.

Keywords: Natural language processing, Text classification, Text representation, Arabic documents, Alphabet, Machine learning algorithms, Phonemic symbolism.

المقدمة

يعد تصنيف النص أحد الموضوعات البحثية الصعبة؛ بسبب ضرورة تنظيم العدد المتزايد من الوثائق النصية الإلكترونية على الإنترنت وتصنيفها. تم تطبيق تصنيف النص بنجاح على مجموعة متنوعة من المجالات مثل: تصنيف الأخبار، تصفية البريد الإلكتروني العشوائي، التنقيب عن الآراء، استرجاع المعلومات، الفهرسة التلقائية، التلخيص، وغيرها. يمكن تعريف تصنيف النص على أنه مهمة تحديد انتماء نصوص لغة طبيعية إلى فئة واحدة أو أكثر من الفئات المحددة مسبقاً بناءً على محتواها (Sebastiani, 2002). مثال على ذلك تصنيف الملفات النصية الإخبارية في إحدى الفئات التالية: الأعمال، والترفيه، والسياسة، والرياضة، والتكنولوجيا.

أحدث الطرق المتبعة للتصنيف هي استخدام خوارزميات تعلم الآلة. تعلم الآلة يقوم على وجود بيانات وقرائنها (الفئات التي تنتهي إليها) الصحيحة عند وقت التعلم، بحيث تتمكن الآلة من التعلم والتعرف على البيانات والقرائن المختلفة والتمييز بينها، وتشكل هذه البيانات أمثلة حقيقية للتنبؤ، يمكن للنموذج التعلم منها للتنبؤ بقيم القرائن الجديدة. تصنيف النص باستخدام تعلم الآلة يتطلب أولاً- تحويل البيانات النصية وتمثيلها في صورة عددية لكي تكون مناسبة للاستخدام مع خوارزميات تعلم الآلة. حيث إن هذه الخوارزميات لا تتعامل مع الكلمات ولكن مع الأرقام فقط. العديد من الباحثين قاموا بتطبيق خوارزميات مختلفة للتعلم الآلي لدراسة فاعليتها على تصنيف النص، في حين أن القليل من الأعمال اهتمت باستكشاف الطرق الممكنة لتمثيل البيانات النصية في صورة عددية وتحليلها ودراسة تأثيرها على دقة أنظمة تصنيف النص.

لتمثيل البيانات النصية في صورة عددية تستخدم الغالبية العظمى من الأبحاث نموذج فضاء المتجه (متجه المصطلحات) (Salton et al., 1975) vector space model VSM والذي يستعمل طريقة حقيقية الكلمات (Salton et al., 1975) . Bag Of Words BOW في طريقة حقيقية الكلمات، يتم تحويل المستند النصي إلى متجه عددي يمثل وزن الكلمات. أولاً- يتم استخراج وتعيين جميع الكلمات التي وردت في مجموعة المستندات المراد تصنيفها. ثم يحسب متجه السمات لمستند ما بناءً على أعداد تكرارات هذه الكلمات في ذلك المستند. تؤدي المجموعة الكاملة من المتجهات لجميع المستندات قيد النظر إلى ما يسمى نموذج متجه السمات/المصطلحات. ويكون طول متجه المستند يساوي عدد الكلمات المختلفة الموجودة في جميع المستندات. تتمثل المشكلة الرئيسة في تلك الطريقة في الأبعاد العالية لمساحة متجهات السمات، والمصفوفة المتناثرة الضخمة، والتي بدورها تؤثر على أداء وتعقيد خوارزميات تعلم الآلة. بالنسبة لمجموعة كبيرة من الملفات النصية، يمكن أن يصل مساحة متجهات السمات إلى مئات الآلاف. لتجاوز هذه المشكلة، عادةً ما يتم تنفيذ طريقتين للمعالجة المسبقة من قبل معظم الباحثين لتقليص العدد الضخم من الكلمات. أولاً- عن طريق إزالة الكلمات التي تُستخدَم بشكل متكرر وشائعة بين جميع المستندات (ما يسمى بكلمات التوقف)، مثل الضمائر، وأسماء الإشارة، وحروف الجر، وعلامات العطف، وما إلى ذلك.

ثانياً- تحويل جميع الكلمات إلى أشكالها الأساسية (مثل الجذع أو الجذر) وإسقاط التكرارات. وهذا مما قد يساعد في تقليص أبعاد مساحة متجهات السمات ومتطلبات الذاكرة. رغم ذلك، فإن المعالجة المسبقة تمثل في حد ذاتها تحدياً إضافياً، خاصة في حالة اللغات عالية الصرف والاشتقاق، كما في اللغة

العربية. لذلك؛ فإن إيجاد طريقة لأداء تصنيف ناجح دون الحاجة إلى أدوات وبرامج تعتمد على اللغة، يعد إسهامًا كبيرًا في مجال تصنيف النصوص.

تنتمي اللغة العربية إلى عائلة اللغات السامية. وهي لغة القرآن في الكتاب المقدس للمسلمين. يستخدمها حوالي 250 مليون شخص لغة أولى. تتكون الهجائية العربية من 28 حرفًا، وعلى عكس اللغة الإنجليزية تُكْتَب من اليمين إلى اليسار. تتميز بوجود عدة أنواع من أحرف العلة القصيرة، التي يمكن استخدامها في توليد نطق مختلف للحرف. ويُعتَقَد أن اللغة العربية هي اللغة الأم التي انبثقت منها اللغات السامية الأخرى؛ لأنها من أغزر اللغات من حيث المادة اللغوية، حيث تحتوي اللغة العربية على ثروة لفظية ضخمة لا يعادلها أي معجم آخر 12.3 مليون كلمة، بينما تحتوي اللغة الإنجليزية 600 ألف كلمة فقط. اللغة العربية أيضًا تمتاز بنظام ثنائي الجنس، مذكر أو مؤنث وثلاثة أشكال للفعل: الماضي، والحاضر، والمستقبل. إحدى السمات الرئيسة - مثل اللغات السامية الأخرى - هي الجذر وأنماط الأوزان. ومفردات اللغة العربية غنية بالصيغ الاشتقاقية المختلفة، فكل كلمة في اللغة هي مزيج من مجموعة من الأحرف الأساسية (الجذر)، بالإضافة إلى نمط معين (وزن). إن غموض اللغة العربية عند معالجتها بالحاسوب، سواء بسبب طبيعتها الصرفية الغنية، أو نقص التشكيل في المنشورات العربية القياسية الحديثة، يجعل مرحلة المعالجة الآلية للغة أكثر تعقيدًا من حالة اللغة الإنجليزية. لذلك؛ البحث عن طريقة لأداء عملية التصنيف بنجاح دون الحاجة إلى أدوات وبرامج معالجة اللغة يمثل إسهامًا رئيسًا في مجال تصنيف النصوص.

هناك خلاف بين العلماء القدماء والمُحدِّثين حول جدلية الدلالة الصوتية للألفاظ. النظرة التقليدية للعلاقة القائمة بين صوت الكلمة ومعناها اعتباطية،

أي إن معنى الكلمة لا يتأثر بصوتها. تشير الفرضية البديلة، المعروفة باسم رمزية الصوت إلى الارتباط الواضح بين تسلسلات صوتية معينة، ومعاني معينة في الكلام (المبارك، 1968)، (مبروك، 2002)، (Nuckolls, 1999)، (Sapir, 1929). وقد عبرت الدراسة المبكرة لابن جني (المتوفي في 1001) عن هذه النظرية اللغوية الفطرية بقوله الشهير: "خلق الحروف على المعنى المقصود والغرض". اتبع عباس (1998) نهج رمزية الصوت، واتفق على أن "معنى الحرف العربي هو صدى صوته في الضمير أو الروح". وقد أثبتت دراسته لخصائص ومعاني الحروف العربية توافق 91-50٪ من الحروف العربية مع معانيها في 3523 جذر. الدراسة الحالية مستوحاة من مفهوم فرضية رمزية الصوت. لقد قمنا ببحث إمكانية توسيع فكرة رمزية الصوت إلى منطقة تصنيف النص من خلال اكتشاف توزيع تسلسل الحروف الهجائية؛ للتنبؤ بفئة المستند. في هذا الصدد، أجرينا تجربة أولية على أساس الحروف الهجائية. كانت دقة تصنيف تلك التجربة مشجعة. مما دفعنا إلى تطوير الفكرة والتوسع في استخدام الحروف الهجائية لاعتماد نهج ثنائي متواليات الحروف الهجائية.

في هذه الورقة، نقدم نهجًا جديدًا لتمثيل المستندات النصية، واستخدامها في عملية تصنيف النصوص. يعتمد إنشاء السمات في هذه الطريقة على جميع التباديل الممكنة لثنائي الحروف الهجائية، واستخدم عدد تكرار السمات لحساب قيم أوزان محتويات الوثيقة. تتمثل المساهمة الرئيسة لهذا العمل في استخدام سمات قياسية محددة مسبقًا ثابتة بدلاً من الاعتماد على الكلمات الموجودة بالمستندات لاستخراج السمات. وهذا يضمن تقليل الزيادة الهائلة في مساحة متجه المستندات مع زيادة حجم البيانات؛ مما يساهم بدوره في تحسين أداء خوارزميات التعلم. علاوة على ذلك، يتجنب النهج المقترح استخدام برامج

معالجة اللغة نظرًا لتعقيدها ووجود لبس في التمثيل خاصة للغات عالية الصرفية، بما في ذلك اللغة العربية. تم تطبيق العمل الحالي على تصنيف مجموعات النصوص العربية، بالإضافة إلى دراسة مستفيضة حول نتائج تصنيف النصوص العربية.

يتميز النهج المقترح بما يلي:

- يعتمد بناء متجه السمات على الحروف الهجائية.
- عدد السمات ثابت ولا يتأثر بعدد الكلمات في المستندات.
- التوفير الكبير في مساحة متجه السمات.
- لا تتأثر بيانات التدريب والاختبار باستخراج السمات أو بحساب متجه المستندات.
- سهل التطبيق ولا يتطلب أدوات وبرامج معالجة اللغة.

الفقرات التالية من البحث تشمل ما يلي: تقدم الفقرة 2 الأعمال السابقة. الفقرة 3 تشمل دراسة النهج المقترح لتصنيف النصوص. ثم الدراسات التجريبية، واستعراض النتائج في الفقرة 4، وتختتم الفقرة الأخيرة بملخص لهذه الورقة.

2- الأعمال السابقة

يعد تصنيف النصوص من أهم قضايا البحث في مجال التنقيب عن النصوص. تصنيف النصوص آليًا يمر بعدة مراحل: المرحلة الأولى: تتطلب استخراج سمات النصوص، وتمثيلها رقميًا لتناسب خوارزميات تعلم الآلة، حيث يعبر عن سمات المستندات بمتجهات رقمية. وتلعب هذه المرحلة دورًا محوريًا يؤثر بشكل مباشر على دقة التصنيف. ونستعرض فيما يلي نظرة عامة على تصنيف النصوص فيما يتعلق بالطرق المختلفة لاستخراج سمات النصوص وتمثيلها.

تركز معظم أبحاث تصنيف النص على طريقة حقيبة الكلمات BOW، حيث تناظر كل سمة كلمة واحدة موجودة في مجموعة نصوص التدريب. وبالتالي فإن كل الكلمات الموجودة في النصوص تسهم في تكوين متجه السمات. لتقليل أبعاد مساحة متجه السمات، استخدمت صورة الجذر، أو حذف السوابق واللاحق من الكلمات من قبل العديد من الباحثين (Syiam et al., 2006 & Zahran et al., 2017 & Bahassine et al., 2009)، من أجل دمج الاشتقاقات المختلفة للكلمة في سمة واحدة. مثال: (مؤتمر، المؤتمر، مؤتمرات، مؤتمراتهم) تدمج كلها في سمة واحدة (مؤتمر). يعد انتقاء السمات - أيضاً- أسلوباً شائعاً لتقليل المصطلحات غير ذات الصلة أو الصاخبة؛ بهدف تحسين أداء التعلم، وتوفير متطلبات الحساب (Guyon & Elisseeff, 2003). تم إجراء العديد من مناهج تقليل الأبعاد من خلال استخدام تقنيات انتقاء السمات، مثل اكتساب المعلومات والمعلومات المتبادلة ومربع كاي (Information Gain, Mutual Information, Chi-Square) وما إلى ذلك. تم ذكر تفاصيل وظائف الاختيار هذه في (Nigam, 2001). هناك طريقة أخرى تتمثل في تطبيق مُلخّص لاستخراج الجمل الهامة أولاً، واختيار أفضل الكلمات من المستندات الموجزة بدلاً من استخدام كل الكلمات. في العمل (Ker & Chen, 2000) طبقت هذه الطريقة على نصوص باللغة الانجليزية. وأشارت نتائجهم إلى أن التصنيف المستند إلى التلخيص يمكن أن يحقق أداءً مقبولاً وتوفيراً في وقت الحساب. قام (Al-Thwaib, 2014) بتطبيق طريقة التلخيص أولاً، على مجموعة بيانات عربية بغرض التصنيف، أظهرت نتائجها تحسناً في استخدام الكلمات المستخرجة من الملخص مقارنة باستخدام جميع الكلمات في تمثيل السمات.

من أجل تحسين دقة التصنيف هناك أعمال أخرى عملت على توسيع

التمثيل بمعلومات إضافية، مثل: متواليات الكلمات، والمعرفة الدلالية. يمكن استخدام عبارات متواليات الكلمات (n-grams) بشكل مستقل، أو بالاشتراك مع الكلمات المفردة (unigrams) لإنشاء السمات. بينما أظهرت النتائج التجريبية في (Bekkerman & Allan, 2004) أن استخدام سمات العبارات فقط يؤدي إلى انخفاض دقة التصنيف، مقارنة بخط الأساس في طريقة حقيبة الكلمات. استخدم (Fürnkranz, 1998) سمات مجموعات من الكلمات المفردة، وعبارات متواليات الكلمات. وأشارت نتائج تصنيفه إلى أن متواليات الكلمات ذات الطول 2 أو 3 عادةً ما تحسن أداء التصنيف، في حين أنَّ التسلسلات الأطول ليست مفيدة. على الرغم من أن عبارات متواليات الكلمات أكثر توصيفًا من الكلمات المفردة فقط، إلا أنَّها قليلة التوزيع والوجود، وهذا بدوره يفاقم مشكلة الأبعاد. نظرًا لأن عبارات متواليات الكلمات الثلاثية لديها متجه سمات أكبر بكثير من نظيرتها الثنائية، فقد ركزت معظم الأبحاث على الجمع بين الكلمات الأحادية، وعبارات متواليات الكلمات الثنائية مع استخدام تقنيات اختيار السمة المختلفة. يطبق العمل (Al-Shalabi & Obeidat, 2008) نفس طريقة التمثيل لتصنيف الوثائق العربية. حيث استخدمت السمات على مستوى الكلمات الأحادية والثنائية، وقارنوا نتائج التصنيف بنتائج طريقة حقيبة الكلمات التقليدية. أظهرت نتائجهم للتصنيف أن استخدام عبارات متواليات الكلمات ينتج دقة أفضل من استخدام الكلمات المفردة فقط في تمثيل السمة.

نهج آخر أكثر ثراءً من طريقة حقيبة الكلمات التقليدية، هو استخدام الأنطولوجيا لتمثيل المستندات. يحافظ نموذج الأنطولوجيا على معرفة المجال للمصطلح الموجود في المستند. ومع ذلك، لا يزال البناء الأنطولوجي التلقائي مهمة صعبة بسبب نقص قاعدة المعرفة المنظمة (Harish et al., 2010 & Zhang et

2015, al.) اقترح (Bloehdorn & Hotho, 2004) نهجاً يدمج المفاهيم من المعرفة الأساسية في تمثيل المستندات لتصنيف المستندات النصية. أظهرت تجاربهم أن دمج المفاهيم في تمثيل السمة يحسن نتائج التصنيف. عمل (Yousif et al., 2017) على تصنيف النصوص العربية باستخدام قاموس WordNet AWN كمصدر معجمي ودلالي. اقترحوا مخططاً للأوزان يعتمد على تواتر العلاقة في AWN ووثائق المجموعة. أظهرت نتائج التصنيف أن نهجهم المقترح تفوق على نهج حقيبة الكلمات.

بينما ركزت الطرق المذكورة أعلاه على الكلمات لتمثيل المستندات، استخدمت طرق أخرى نهجاً بديلاً حيث تعاملوا مع الحروف للتعامل مع المعالجة الصرفية. اقترح (Kanaris et al., 2007) التعامل على مستوى متواليات الحرف مع مصنفات نصية في إطار عمل تصفية مكافحة البريد العشوائي المستندة إلى المحتوى. أظهرت نتائجهم أن متواليات الحرف لحساب أوزان السمات أكثر موثوقية من الكلمات على الرغم من حقيقة أنها تزيد من مشكلة الأبعاد. تم اقتراح التمثيل المستند إلى مزيج من الحروف الثنائية، والثلاثية المستخرجة من الوثائق في (Berger et al., 2005). العمل (Santos et al., 2014) اقترح الشبكات العصبية العميقة. استخدم نهجهم الكلمات كأساس، حيث تشكل السمات على مستوى الحروف المستخرجة من الكلمة تمثيلاً موزعاً. اقترح (Khreisat, 2006) تصنيفاً للنص العربي باستخدام تردد الحروف الثلاثية. تم قياس مدى التقارب بين كل وثيقة وجميع المستندات الأخرى بناءً على ملف تعريف التردد للحروف الثلاثية باستخدام مقياسي (Dice & Manhattan). اقترح (Sawaf et al., 2001) مصنفًا قائمًا على الحروف للتعامل مع مشكلة البيانات المتفرقة بدلاً من تقليل السمات بواسطة القواعد الصرفية. حيث استخلصوا من النصوص تسلسلات

متصلة لنوعين من الوحدات: كلمات كاملة الشكل، أو الحروف الثلاثية. وأجروا تجاربهم لتصنيف النص على مجموعة عربية كبيرة ولم يستخدموا أي خطوات بوصفها مهمة معالجة مسبقة.

3- تصنيف النص باستخدام نهج ثنائي متواليات الحروف الهجائية

يبدأ النهج المقترح لتصنيف النص أولاً من خلال إنشاء مصطلحات السمات، والتي تعتمد فقط على الحروف الهجائية. واقتصرت المعالجة المسبقة على التوحيد الإملائي فقط. لكل مستند، يعتمد حساب قيم متجه سمات المستند بناءً على تكرارات سمات مصطلحات ثنائيات الحروف الهجائية في ذلك المستند. وبعد ذلك يتم تطبيق تقنية التحقق المتقاطع k-fold، واستخدام مصنف SVM الشهير لتصنيف المستندات. وسوف يتم عرض الخطوات التفصيلية للنهج الحالي في الفقرات التالية.

1-3 إنشاء مصطلحات السمات

اعتمد نهج تصنيف الوثيقة الحالي الحروف الهجائية لإنشاء مصطلحات السمات. ونظراً لأن الحروف هي البذرة الأساسية في مصطلحات السمات، فهي لذلك قياسية ولا تتأثر بطبيعة أو كلمات المحتوى. لإنشاء مصطلحات السمات، تُستخرج جميع الترتيبات الممكنة لحرفين هجائيين من حروف اللغة، على سبيل المثال: اب، ات، اث،....ين، يه، يو.

يسمى ترتيب عناصر مجموعة من الكائنات في متسلسلة التباديل (Permutation). ويكون P عدد العناصر الناتجة من الترتيبات المختلفة للكائنات r المأخوذة من الكائنات n مع السماح بالتكرار ما يلي [26]:

$${}^nP_r = n^r \quad (1)$$

حيث إن

- r هو طول المتسلسلة، ويساوي عدد الأحرف المختارة لبناء مصطلح ($r = 2$) لمصطلح ثنائي).

- n هو عدد الحروف الهجائية في اللغة (الهجائية العربية الأساسية تتكون من 28 حرفاً).

في العمل الحالي، عند إنشاء السمات، استبعدنا الحروف المتماثلة من الظهور في قائمة السمات (على سبيل المثال: "قق"، "تت") لتقليل مساحة متجه السمات. بينما أظهرت التجارب الأولية أن تجاهلها ليس له تأثير يذكر على أداء عملية التصنيف. بالإضافة إلى ذلك، لا يمكن اكتشاف الحرف المزدوج مباشرة في حالة تواجد الحرف مرتين متتاليتين في الكلمة، مع عدم وجود حرف علة بينهما. في هذه الحالة، يُكتب الحرف مرة واحدة فقط ويوضع فوقه حرف متحرك قصير "شدة" (أو مدة في حال كان الحرف ألفاً)، والذي لا يظهر غالباً في المستندات العربية القياسية الحديثة.

وبذلك، يكون P عدد العناصر الناتجة من الترتيبات المختلفة للكائنات r - بدون تكرار - المأخوذة من الكائنات n ما يلي (Permutations, 2022)

$${}^nP_r = \frac{n!}{(n-r)!} \quad (2)$$

$$r=2, n=28$$

وبالتالي، فإن العدد الإجمالي لترتيبات الحروف العربية الثنائية هو 756، وهو ما يمثل عدد السمات التي تم إنشاؤها في هذا العمل.

تجدر الإشارة إلى أنه بالإضافة إلى فوائد استخدام عدد محدود من السمات، فإن عملية إنشاء السمات نفسها في هذا النهج تختصر الكثير من

العمليات مثل تنظيف النص، وانتقاء ودمج السمات مقارنة بطريقة حقيقية الكلمات.

2-3 المعالجة المسبقة للنص

في اللغة العربية، بعض الحروف لها أشكال مختلفة بسبب الأصوات المتعددة التي تمثلها. على سبيل المثال، الحرف الأول في الهجائية العربية، والحرف الأكثر استخدامًا بينهم بسبب الأصوات المتعددة التي يمثلها، له أشكال مختلفة (ا، أ، آ، إ). (حرف الهاء) قد يكتب بأشكال مختلفة: التاء المربوطة والهاء (ة، ه) في النصوص المتداولة. في العمل الحالي، تقتصر خطوة المعالجة المسبقة فقط على التوحيد الإملائي لأشكال الحروف العربية. فقد تم توحيد الأشكال المختلفة للألف "ا"، أ، آ، إ إلى "ا"، وأشكالها من "ة، ه" إلى "ه". ومع ذلك، في تجاربنا الجانبية مؤخرًا، وجد أن التوحيد الإملائي للحروف العربية ليس له تأثير كبير على أداء مصنف النص. ويرجع هذا - في رأينا - إلى حقيقة أن هذه الحروف بشكل عام أقل بكثير من تلك الموجودة في مجموعة النصوص ككل. وبالتالي، يمكن تجاوز عملية التوحيد الإملائي في الأعمال التالية.

وتجدر الإشارة إلى أن العديد من خطوات المعالجة المسبقة التي شاع استخدامها في نموذج حقيبة الكلمات، بما في ذلك تنظيف الكلمات الوظيفية (مثل حروف الجر وغيرها، وتقليل الكلمات إلى جذعها أو جذرها) قد تم تجنبها عن عمد. فالهدف هو تقييم وفحص دقة النهج الحالي؛ لتصنيف المستندات باستخدام أبسط الخطوات. حيث إن المعالجة المسبقة تعتمد بشكل كبير على اللغة المستخدمة وهيكلها، وبالتالي تتطلب أدوات معالجة لغوية متعددة مختلفة لكل مرحلة حسب اللغة المستخدمة. ولهذا يعد تجاوز مثل هذه العمليات المعقدة مع الحفاظ على الأداء الجيد إسهامًا كبيرًا في مجال التصنيف.

3-3 حساب متجه المستند

الخطوة التالية هي تحويل كل مستند نصي إلى متجه رقمي يمكن استخدامه مدخلاً لنموذج التعلم الآلي. يتم حساب متجهات المستند بناءً على البيانات النصية ومصطلحات السمات، وهي تعبر عن وجود مصطلحات السمات داخل المستندات، مع الأخذ في الاعتبار أن كل مصطلح يعبر عن سمة واحدة. هناك العديد من الطرق لحساب الأوزان المستخدمة في عمليات التصنيف، نختار مصطلح التردد البسيط المعدل لطول الوثيقة والذي يسمى تردد المصطلح المعياري $ntf(t,d)$. تردد المصطلح $tf(t,d)$ ، هو عدد المرات التي يظهر فيها المصطلح t في مستند d . تردد المصطلح المعياري $ntf(t,d)$ هو عدد المرات التي يوجد فيها مصطلح في وثيقة معينة d ، يتم معيرتها من خلال العدد الإجمالي للمصطلحات الموجودة في تلك الوثيقة. تتراوح قيم تردد المصطلح المعياري 0 إلى 1.

$$ntf(t,d) = tf(t,d) \div \text{number of terms in } d \quad (3)$$

من الواضح أن طول متجه المستند يساوي العدد الإجمالي لمصطلحات السمات المنشأة.

4-3 تحديد السمات

ليست كل التركيبات المكونة من حرفين أبجديين شائعة في اللغة العربية. فهناك بعض التراكيب نادرة الحدوث ولا يؤثر عدم وجودها في مصطلحات السمات بشكل فعال على عملية التصنيف. ومن ثمّ، تقليل أبعاد المتجه وتجنب وجود سمات غير ذات صلة (تشويش)، كان من المستحسن تطبيق اختيار السمة. تضمن مرحلة اختيار السمة أن يتم اختيار تلك السمات التي تميل بشدة نحو فئة معينة في عملية التعلم. إحدى الطرق الإحصائية الأكثر شيوعاً لاختيار السمات المستخدمة في أدبيات التصنيف هي مربع كاي (χ^2 chi-square). وقد

أثبتت دقة عالية في تصنيف النصوص العربية والإنجليزية،

(Salton et al., 1975 & Anitha et al., 2013 & Al-Tahrawi et al., 2015).

χ^2 يقوم بحساب وجود استقلالية بين المصطلح t وفئة معينة i . لنفترض أن n هو العدد الإجمالي للمستندات في المجموعة، ويكون $p_i(t)$ هو الاحتمال الشرطي للفئة i للمستندات التي تحتوي على t ، ويكون P_i هو الجزء من المستندات ذات الفئة i ، ويكون $F(t)$ هو الجزء من المستندات التي تحتوي على مصطلح t . يتم تعريف χ^2 للمصطلح بين المصطلح t والفئة i على النحو التالي:

(Aggarwal et al., 2012)

$$\chi^2(t) = \frac{n \cdot F(t)^2 \cdot (p_i(t) - P_i)^2}{F(t) \cdot (1 - F(t)) \cdot P_i \cdot (1 - P_i)} \quad (4)$$

لذلك، استخدم χ^2 -statistic في العمل الحالي لتحديد السمات الأكثر تميزاً في تصنيف الوثائق العربية؛ تم تحديد نسب مئوية للسمات الأفضل لبناء المصنف واختبار نتائج قيم p المختلفة خلال التجارب في الفقرات 4.2، 4.4، 4.5.

5-3 عملية تعلم الآلة

بعد إنشاء متجهات المستند، يمكن تطبيق مرحلة اختيار المصنف المناسب. الهدف هو العثور على الخوارزمية التي تحقق الدقة القصوى مع تقليل وقت الحساب المطلوب للتدريب. هناك عدد من المصنفات الناجحة التي تم استخدامها في تصنيف النص. أثبت SVM أنه مناسب جداً لتصنيف النص لدقته العالية وقدرته المتأصلة في التعامل مع مساحات السمات الكبيرة (Anitha et al., 2000 & Aggarwal et al., 2012 & Cristianini et al., 2013). يتمثل المبدأ الرئيس لخوارزم SVM في تحديد الفواصل في مساحة البحث التي يمكنها

فصل الفئات المختلفة بشكل أفضل. تعتمد عملية الفصل على المسافة القصوى بين جانبي المستوى الفائق، وأقرب المتجهات في عينة بيانات التدريب. تتمثل إحدى سمات SVM في أنه نظرًا لأنه يحاول تحديد الاتجاه الأمثل للتمييز في مساحة السمة من خلال فحص المجموعة المناسبة من السمات، فهو مناسب جدًا للأبعاد العالية (Aggarwal et al., 2012).

يتطلب تدريب SVM حل مشكلة تحسين البرمجة التربيعية (QP) الكبيرة جدًا. يجد مصنف SVM مستويًا فائقًا مثاليًا يعمل على زيادة مسافة أي فئة إلى المستوى الفائق الفاصل، وفي نفس الوقت تقليل مخاطر سوء تصنيف بيانات التدريب. (SMO) هو خوارزمية لحل مشكلة التحسين التي تنشأ أثناء تدريب آلات المتجهات الداعمة. إنه يقسم هذه المشكلة إلى سلسلة من أصغر المشاكل الفرعية الممكنة، والتي يتم حلها بعد ذلك بشكل تحليلي. SMO؛ وهو إصدار WEKA (Witten et al., 2005) من خوارزمية SVM لبناء النموذج. في تجاربنا، تم اختبار مجموعة البيانات باستخدام تقنية التحقق من الصحة k-fold مع $k = 10$. في هذه التقنية، يتم تقسيم مجموعة البيانات الأصلية بشكل عشوائي إلى عينات فرعية k. يتم الاحتفاظ بعينة فرعية واحدة كبيانات التحقق من الصحة لاختبار النموذج، ويتم استخدام العينات الفرعية المتبقية $k - 1$ كبيانات تدريب. تتكرر العملية k مرة، مع استخدام كل عينة فرعية k مرة واحدة كبيانات اختبار. يمكن بعد ذلك حساب متوسط نتائج k لحساب التقدير النهائي.

4- البحوث والدراسات التجريبية

باستخدام النهج المقترح أُجريت تجارب لتصنيف مجموعات النصوص علي أربع مجموعات بيانات نصوص عربية، كما أُجريت دراسة مستفيضة حول السمات المستخدمة في التصنيف. في التجارب، تم استخدام الدقة (precision)، والاستدعاء

(recall)، ومقياس (F-measure) F كمقاييس أداء للتصنيف. في جميع التجارب التي تم إجراؤها، تم تدريب المصنف على مجموعة البيانات باستخدام تقنية التحقق من الصحة k-fold مع $k = 10$.

أجريت أربع تجارب منفصلة على مجموعات البيانات. تم استخدام كل من تطبيقات Weka (Witten et al., 2016)، Rapid Miner، (Rapid Miner Project،) (2013) أثناء التجارب. في ما يلي سوف نستعرض تفاصيل مجموعات البيانات المستخدمة ونتائج تقييم الأداء وتقليص السمات.

1-4 مجموعات البيانات

في هذا العمل، تم استخدام أربع مجموعات بيانات نصوص عربية مختلفة كما هو موضح في الجدولين 1، 2.

- **مجموعة بيانات الخليج الإخبارية: (Alkhaleej news dataset)**

مجموعة بيانات الخليج (Abbas, 2004) *Alkhaleej news*. تحتوي مجموعة البيانات على 5690 مستندًا تقابل ما يقرب من 3 ملايين كلمة. يتم تصنيف كل مستند بإحدى الفئات الأربع التالية {"الأخبار الدولية" و "الأخبار المحلية" و "الرياضة" و "الاقتصاد"}. اخترنا عشوائيًا مجموعة من 2000 وثيقة موزعة بالتساوي بين الفئات الأربعة.

- **بيانات الجزيرة الإخبارية 9 فئات: (Alj-News9)**

تحتوي مجموعة البيانات العربية Alj-News9 على 2700 مستندًا للمقالات الإخبارية باللغة العربية (Alj-News, 2004). صُنّف كل مستند بواحدة من الفئات التسع التالية {"الاقتصاد"، "الصحة"، "القانون"، "الأدب"، "السياسة"، "الدين"، "الرياضة"، "التكنولوجيا"، "الفن"} مع عدد متساوٍ من المستندات في كل فئة.

• بيانات الجزيرة الإخبارية 5 فئات: (Alj-News5)

Alj-News5 هي مجموعة بيانات أخرى مختلفة عن السابقة حيث تحتوي على 1500 مستنداً للمقالات الإخبارية (Alj-News, 2004). صُنف كل مستند بوحدة من الفئات الخمس التالية {"الفن"، "الاقتصاد"، "السياسة"، "العلوم"، "الرياضة"}. استخدمت بيانات Alj-News5 من قبل باحثين آخرين (Chantar et al., 2011 & Al-Tahrawi et al., 2015). في مجموعة التجارب التي أجريت سوف نقارن نتائج تصنيف بيانات Alj-News5 باستخدام النهج الحالي مع الأعمال البحثية الأخرى المذكورة.

• مجموعة بيانات بي بي سي العربية للأخبار: (BBC-Arabic News Dataset)

تحتوي مجموعة البيانات BBC العربية على 4763 وثيقة من بي بي سي عربي (Saad & Ashour, 2010)، تم جمعها من موقع بي بي سي عربي. صُنف كل مستند بوحدة من الفئات السبع التالية {"الشرق الأوسط"، "أخبار العالم"، "الأعمال"، "الرياضة"، "الصحف"، "العلوم"، "متفرقات"}.

جدول 1. قواعد البيانات العربية المستخدمة في التجارب

البيانات	عدد الفئات	عدد الوثائق
بيانات الخليج الاخبارية	4	2000
بيانات الجزيرة-9	9	2700
بيانات الجزيرة-5	5	1500
بي بي سي العربية للأخبار	7	4763

2-4 توزيع ثنائيات الحروف الهجائية في المفردات العربية

أجريت التجربة لدراسة توزيع متواليات ثنائيات الحروف الهجائية في المفردات العربية. في هذه التجربة، استخدم مربع كاي Chi-Square لقياس أهمية كل سمة. طبقت تجربة تصنيف الوثائق باستخدام مجموعات البيانات العربية الأربع بشكل منفصل.

عند دراسة نتائج أوزان مربع كاي على مصطلحات ثنائيات الحروف الهجائية، وجد أن هناك 125 مصطلحًا / سمة ذات أوزان صفرية متفق عليها في جميع مجموعات البيانات العربية الأربع. وهذا يمثل 16٪ من العدد الإجمالي للمصطلحات. حيث لوحظ أن هذه المصطلحات الثنائية لم تحدث تقريبًا في مفردات مجموعات البيانات. لذلك اعتبرنا أن هذه المصطلحات نادرة ويصعب حدوثها في المفردات العربية. وفقًا لذلك، أُسقطت قائمة المصطلحات النادرة بأمان من قائمة السمات، كما سيتم عرضها في التجربة التالية. مثال على ثنائي الحروف نادر الحدوث في المفردات العربية التي كشفت عنها هذه التجربة: (حخ، دذ، سش، شص، صض، ضظ، ذش).

3-4 التحقق من صحة نهج ثنائيات الحروف الهجائية

في هذه التجربة قمنا بتصنيف مجموعات بيانات عربية مختلفة بشكل منفصل باستخدام مصنف SVM-SMO وتمثيل متجه ثنائيات الحروف الهجائية. تهدف التجربة إلى معرفة ما إذا كان النهج الجديد قابلاً للتطبيق لتصنيف المستندات العربية. استُخدمت تقنية التحقق من الصحة k-fold مع k = 10 لتدريب المصنف. تم تطبيق التجربة باستخدام 631 سمة (العدد الكامل للخصائص باستثناء السمات النادرة ذات الوزن الصفري كما في الفقرة 4.2). يوضح الجدول 2 ملخص دقة التصنيف الإجمالية من حيث الدقة Precision،

والاستدعاء Recall، ومقياس F (F-measure) على مجموعات البيانات العربية المختلفة. في الجدول 2، يمكن ملاحظة أن نتائج مقياس F تقع في النطاق بين (0.949) و (0.874) في مجموعات البيانات العربية الأربع، وحصلت مجموعة بيانات Alj-News5 على أعلى مقياس F بينما مجموعة بيانات BBC-Arabic كانت الأدنى (علماً بأن مقياس F الأمثل هو الواحد الصحيح). هناك عدة عوامل متعلقة بمجموعة البيانات، والتي تؤثر على دقة التصنيف. يتضمن ذلك: نوع وطبيعة مجموعة البيانات، عدد المستندات والفئات، توزيع المستندات على طول الفئات الموجودة، وجود بيانات مشوشة، والتداخل بين الفئات.

جدول 2: دقة تصنيف المصنف SVM-SMO على مجموعات البيانات العربية المختلفة باستخدام النهج الحالي:

البيانات	Precision	Recall	F-measure
بيانات الخليج الاخبارية	0.905	0.905	0.905
بيانات الجزيرة-9	0.930	0.930	0.930
بيانات الجزيرة-5	0.949	0.949	0.949
بي بي سي العربية للأخبار	0.874	0.874	0.874

وسوف نعرض دراسة ومقارنة أكثر تفصيلاً مع أعمال آخرين في التجارب التالية.

4-4 تأثير تقليل مساحة السمات على دقة التصنيف

أُجريت التجارب بمساحات مختلفة للسمات لدراسة مدى تأثير اختيار عدد محدد من السمات على دقة التصنيف.

في البداية، تم ترجيح السمات المختارة بواسطة مربع كاي وترتيبها حسب

أوزانها بترتيب تصاعدي، ثم تحديد طول متجه السمة بنسبة 100٪ و 90٪ وهكذا حتى 10٪ من السمات الكاملة (حيث السمات الكاملة = 756 مصطلح ثنائي الحروف الهجائية)، وتدريب المصنف باستخدام تقنية التحقق من الصحة k-fold مع $k = 10$ على مجموعات البيانات العربية المختلفة بشكل منفصل.

يوضح الجدول 3 العرض التفصيلي لنتائج الدقة بمقياس F للنسب المئوية المختلفة لطول متجه السمات الأعلى مرتبة باستخدام محدد سمة مربع كاي. تشير أعمدة الجدول إلى النسبة المئوية لاختيار السمات، والعدد المقابل لهذه النسبة من السمات المحددة، ومقاييس F المسجلة لمجموعات البيانات العربية المختلفة.

جدول 3: النسب المئوية، وعدد السمات المقابلة، ونتائج التصنيف لمجموعات البيانات المختلفة.

بي بي سي العربية	الجزيرة-5	الجزيرة-9	الخليج	عدد السمات	% السمات المختارة
0.874	0.949	0.930	0.905	756	100%
0.874	0.949	0.930	0.905	680	90%
0.874	0.949	0.930	0.905	605	80%
0.874	0.949	0.928	0.904	529	70%
0.872	0.944	0.926	0.903	454	60%
0.871	0.946	0.925	0.902	378	50%
0.867	0.940	0.921	0.900	302	40%
0.861	0.930	0.916	0.894	226	30%
0.846	0.914	0.900	0.875	151	20%
0.799	0.890	0.859	0.841	75	10%

يوضح الجدول 4 قيم متوسط الانحراف عن خط الأساس بنسب مختلفة من السمات الأعلى مرتبة. تمت صياغة نتائج دقة التصنيف من حيث مقياس F في الرسمين .

جدول 4: متوسط الانحراف في الدقة عن خط الأساس باستخدام اختيارات مختلفة لأعداد السمات.

متوسط الانحراف	% نسبة اختيار السمات
0.000	90%
0.000	80%
0.001	70%
0.002	60%
0.003	50%
0.007	40%
0.014	30%
0.030	20%
0.067	10%

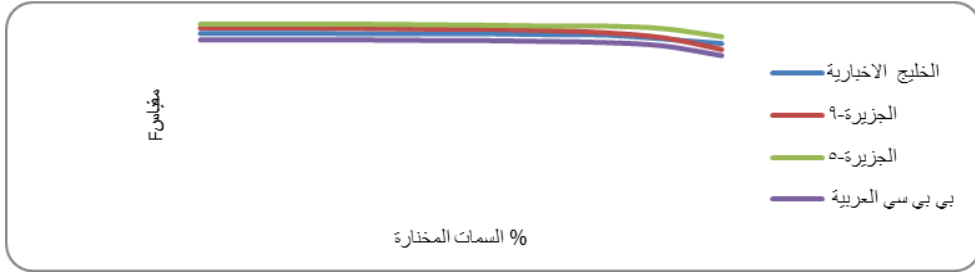
البيانين الموضحين في الشكلين 1 و 2. توضح الأرقام تأثير النسبة المئوية المختارة لعدد السمات باستخدام محدد السمة مربع كاي على الدقة التي تم تحقيقها عبر مجموعات البيانات العربية المختلفة.

يوضح الشكل 1 نتائج الدقة للنسب المئوية المختلفة للسمات الأعلى تصنيفًا، عبر مجموعات البيانات المختلفة.

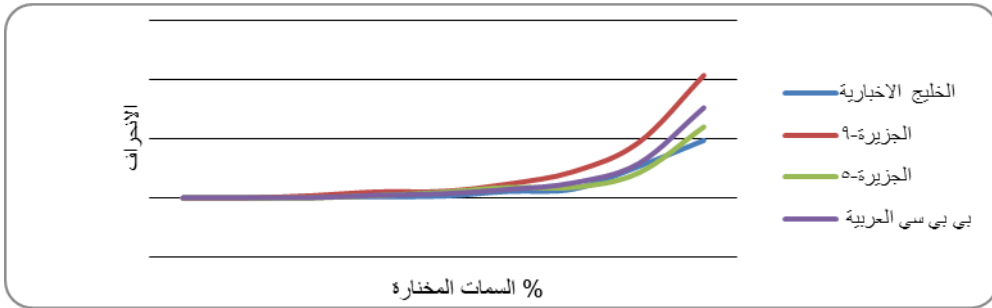
ويوضح الشكل 2 انحرافات الدقة عن خط الأساس عبر مجموعات البيانات للنسب المئوية المختلفة للسّمات. مع اعتبار خط الأساس هو الدقة الحاصلة عند استخدام العدد الكلي للسّمات. علمًا بأن الانحراف صفر يمثل أفضل النتائج.

يتضح من الجدول 3 أن جميع مجموعات البيانات حققت أفضل دقة عند استخدام كامل السّمات. ومع ذلك، كما يتضح من الجدول 4، أنه باستخدام لما يصل إلى 50٪ من السّمات الكاملة، لم يكن هناك تغيير يذكر في نتائج الدقة (أقل من 1٪) في جميع مجموعات البيانات. عند (40 - 20)٪ من السّمات الكاملة، يبدأ متوسط الدقة في الانخفاض تدريجيًا بنسبة (1-3)٪ مقارنة بالنتائج عند السّمات الكاملة. بينما لوحظ تدهور شديد في الدقة حوالي (7٪) عند استخدام 10٪ من السّمات الكاملة.

أظهرت النتائج أنه يمكن استخدام النهج المقترح بنفس الكفاءة الكاملة تقريبًا، من خلال استخدام 50٪ فقط من السّمات الكاملة، أي 378 مصطلحًا من ثنائي الحروف الهجائية العربية، بدلًا من 756. وعلى الرغم من وجود تدهور في كفاءة المصنف عند نسبة 10٪ مقارنة بحالة السّمات الكاملة، إلا أن المصنف لا يزال يعمل بنتائج معقولة. يشير هذا إلى وجود حوالي (10-50)٪ من المصطلحات الهجائية الثنائية والتي تلعب دورًا حاسمًا في تمثيل المستندات لعملية التصنيف. وفي التجربة القادمة سوف نقدم مزيدًا من التفاصيل والمقارنة مع الطرق الأخرى.



شكل 1. النسبة المئوية لاختيار السمات ودقة التصنيف المقابلة



شكل 2. الانحراف في الدقة عن خط الأساس عند اختيار نسب مختلفة لعدد السمات .

5-4 التحليل والمقارنة بالطرق الأخرى

سوف نسلط مزيداً من الضوء على نتائج التصنيف التفصيلية لبيانات الجزيرة-5 الإخبارية Alj- News5 ؛ من أجل تحليل النتائج من كتب، ومقارنة الدقة مع الطرق الأخرى. هناك صعوبات في إجراء مقارنة مباشرة مع الأعمال البحثية السابقة الأخرى في مجال تصنيف النصوص العربية؛ من بين هذه الصعوبات ما يلي: لا توجد مجموعة بيانات عربية معيارية، مجموعة البيانات المستخدمة غير متوفرة في بعض الحالات، يختار الباحثون بشكل عشوائي كمية

مختلفة من مستندات مجموعة البيانات الأصلية، وعدد مختلف من الحالات لكل فئة، وتنظيم مختلف للتدريب والاختبار، وعدم وضوح عدد السمات المختارة؛ لذلك تم اختيار بيانات الجزيرة-5 الإخبارية للمقارنة نظرًا لتوافر نتائج اختباراتهما من خلال أعمال أخرى.

النتائج التفصيلية لتطبيق نهج ثنائيات الحروف الهجائية مع مصنف SVM-SMO على بيانات الجزيرة-5 الإخبارية Alj-News5 تم عرضه في الجدول 5 الذي يوضح الدقة Precision، والاستدعاء Recall، ومقياس F لكل فئة. تم تدريب المصنف باستخدام تقنية التحقق من الصحة k-fold مع $k = 10$. كما هو مبين في الجدول 5، كان أعلى مقياس F في فئة "الرياضة" (0.995)، بينما كان أدنى قياس (0.901) في فئة "السياسة". وكان المتوسط لمقياس F لجميع الفئات (0.949). يوضح الجدول 6 مصفوفة الارتباك (Confusion Matrix) المقابلة، والتي تشير فيها القيم القطرية إلى الفئات التي تم التنبؤ بها بشكل صحيح، وتشير القيم غير القطرية إلى أخطاء، على سبيل المثال: استبدال 16 تصنيفًا من "الفن" بـ "السياسة".

جدول 5 نتائج التصنيف حسب الفئة لبيانات الجزيرة-5 الإخبارية

الفئة	Precision	Recall	F-Measure
فن	0.937	0.937	0.937
اقتصاد	0.946	0.930	0.938
سياسة	0.890	0.913	0.901
علوم	0.980	0.967	0.973
رياضة	0.993	0.997	0.995
المتوسط	0.949	0.949	0.949

جدول 6: مصفوفة الارتباك لبيانات الجزيرة-5 الإخبارية

A	B	C	D	E	Classified as
281	1	16	1	1	A= فن
2	279	16	3	0	B= اقتصاد
14	9	274	2	1	C= سياسة
2	6	2	290	0	D= علوم
1	0	0	0	299	E= رياضة

من أجل قياس مدى نجاح النهج المقترح، أجريت مقارنة مع عمليتين أخريين؛ النظام-1، اقترحه [36] ; بينما النظام-2، اقترحه (Al-Tahrawi et al., 2015)، وذلك لإجراء التصنيف علي نفس مجموعة بيانات الجزيرة-5 الإخبارية. نظراً للتأثير المهم لأبعاد مساحة السمات في عملية التصنيف، فقد تم أخذها في الاعتبار عند المقارنة. يعرض الجدول 7 أفضل نتائج لإجمالي مقياس F وعدد السمات المحددة للأنظمة الثلاثة في مجموعة بيانات الجزيرة-5 الإخبارية Alj-News 5. وعلى الرغم من صعوبة إجراء مقارنة دقيقة إلا أن تحليل النتائج أظهر ما يلي:

النظام-1 استخدم BPSO-KNN (Binary Particle Swarm Optimization- KNN) بوصفها طريقة لانتقاء السمات. وطبقت خطوات المعالجة المسبقة التي تشمل إزالة كلمات التوقف والكلمات النادرة، ولم يتم تطبيق الاشتقاق. وفي عملهم قاموا بتجربة خوارزميات مصنفات مختلفة، أفضل نتائجهم على مجموعة الجزيرة-5 الإخبارية العربية (Alj-News5) حصل عليها المصنف SVM حيث أظهرت نتائجهم متوسط مقياس F = 0.931 ومع ذلك، كانت قيمة الدقة

هذه على حساب متطلبات الذاكرة العالية نظرًا للعدد الكبير من السمات (2967)، التي تم استخدامها لبناء المصنف.

النتائج التي تم الحصول عليها من خلال النهج الحالي تفوقت على النظام 1 من حيث مقياس $F(0.949)$ ، وكذلك بالنسبة لعدد السمات المستخدمة (529) فقط. بالإضافة إلى ذلك، في النهج الحالي، لتحقيق نفس الدقة تقريبًا (0.930) التي حصل عليها النظام 1، يمكن استخدام 266 سمة فقط، كما هو موضح في الجدول 4.

النظام-2 في هذا العمل طبقت العديد من خطوات المعالجة المسبقة بما في ذلك الاشتقاق وإزالة كلمات التوقف. استخدم مصنف PN لتصنيف الوثائق. في عملهم، لتقليل عدد السمات بشكل أكبر، استخدم مربع كاي لاختيار السمة، وتم تحديد 1٪ فقط من سمات كل فئة لبناء المصنف. كانت نتائجهم على مجموعة الجزيرة-5 الإخبارية العربية (Alj-News5) متوسط مقياس $F = 0.893$ باستخدام 135 سمة. النهج الحالي له التفوق في متوسط مقياس $F(0.949)$ على الرغم من أن النظام 2 قد استخدم خوارزمية قوية مصنف PNs، إلا أنه يحتوي على قيود على الذاكرة. لذلك؛ كانت هناك حاجة لتقليل عدد السمات كما تم تنفيذها في عملهم. في هذا الصدد، تم تطبيق تجربة أخرى لاختبار نهج ثنائيات الحروف الهجائية الحالي باستخدام نفس عدد السمات التي تم استخدامها في النظام 2. أظهرت النتائج أن النهج الحالي باستخدام ثنائيات الحروف الهجائية حققت مقياس $F = 0.913$ باستخدام 135 سمة منتقاة، مقارنة بـ 0.893 التي حصل عليها النظام-2 بنفس عدد السمات.

جدول 7: أفضل نتائج مقياس F وعدد السمات المختارة للأنظمة الثلاثة في مجموعة بيانات الجزيرة-5 الإخبارية (Alj-News5)

العمل	عدد السمات	F-measure قياس
النظام-1	2967	0.931
النظام-2	135	0.893
النهج الحالي	529	0.949

بالإضافة إلى أن النهج المقترح قد أثبت نتائج عالية الدقة؛ هناك نوعان من المساهمات الأساسية في هذا المجال.

أولاً- السمات التي تم استخدامها في عملية التصنيف معيارية ومنفصلة عن محتويات المستندات؛ لذا فهي لا تزيد مع زيادة حجم البيانات، وهذه مسألة مهمة للحفاظ على متطلبات الذاكرة عند الحد الأدنى.

ثانياً- تم الحصول على النتائج دون الحاجة إلى أدوات وبرامج معقدة لمعالجة اللغات الطبيعية، لاسيما اللغات عالية التصريف والاشتقاق كما في اللغة العربية.

5- الخاتمة

في هذا العمل، صممنا نهجاً جديداً يعتمد على ثنائيات الحروف الهجائية؛ لبناء السمات وتم تطبيقها على تصنيف المستندات العربية. استخدمت كثافة تواجد مصطلحات ثنائيات الحروف الهجائية في النص لتمثيل محتوياته رقمياً. النهج الجديد لا يتطلب أدوات وبرامج معالجة اللغات الطبيعية. باستخدام مصنف SVM-SMO، أثبت النهج المقترح القدرة على تصنيف مجموعات

المستندات النصية العربية بنجاح. أيضًا، أظهرت النتائج على مجموعات البيانات العربية أنه يمكن استخدام النهج المقترح بنفس الكفاءة الكاملة تقريبًا من خلال استخدام 50٪ فقط من مصطلحات ثنائيات الحروف الهجائية. النهج الحالي له إسهامتان رئيستان.

أولاً- السمات قياسية ومنفصلة عن محتويات المستندات؛ هذا يساعد في تقليل الأبعاد العالية مع زيادة حجم البيانات.

ثانيًا- يمكن إجراء عملية التصنيف دون الحاجة إلى أدوات وبرامج معقدة لمعالجة اللغات الطبيعية، لاسيما اللغات عالية التصريف ولاشتقاق كما في اللغة العربية.

الملاحق

ملحق (أ) قائمة الأشكال

- شكل 1 النسبة المئوية لاختيار السمات ودقة التصنيف المقابلة. 22
- شكل 2 الانحراف في الدقة عن خط الأساس عند اختيار نسب مختلفة لعدد السمات. 22

ملحق (ب) قائمة الجداول

17	قواعد البيانات العربية المستخدمة في التجارب.	جدول 1
18	دقة تصنيف المصنف SVM-SMO على مجموعات البيانات العربية المختلفة باستخدام النهج الحالي.	جدول 2
20	النسب المئوية وعدد السمات المقابلة، ونتائج التصنيف لمجموعات البيانات المختلفة.	جدول 3
20	متوسط الانحراف في الدقة عن خط الأساس باستخدام اختيارات مختلفة لأعداد السمات.	جدول 4
23	نتائج التصنيف حسب الفئة لبيانات الجزيرة-5 الإخبارية.	جدول 5
23	مصفوفة الارتباك لبيانات الجزيرة-5 الإخبارية.	جدول 6
25	أفضل نتائج مقياس F وعدد السمات المختارة للأنظمة الثلاثة في مجموعة بيانات الجزيرة-5 الإخبارية.	جدول 7

قائمة المراجع

- 1- عباس، حسن، خصائص الحروف العربية ومعانيها، منشورات اتحاد الكتاب العرب، 1998.
- 2- المبارك، محمد، فقه اللغة وخصائص العربية، بيروت: دار الفكر للطباعة والنشر، ط3، 1968.
- 3- مبروك، مراد، من الصوت إلى النص، نحو نسق منهجي لدراسة النص الشعري، الاسكندرية: دار الوفاء لدنيا الطباعة والنشر، 2002.

References

- 1- Aggarwal, C. C., and Zhai, C. (2012), *"Mining text data"*, Springer Science & Business Media, 2012. Citado, pp.7-18.
- 2- Alj-News- Arabic Corpora. (2004), Retrieved September 07, 2016, from <https://filebox.vt.edu/users/dsaid/Alj-News.tar.gz>. Last access on January 2013.
- 3- Al-Shalabi, R.; and Obeidat, R. (2008), *"Improving KNN Arabic text classification with n-grams based document indexing"*, Proceedings of the Sixth International Conference on Informatics and Systems, Cairo, Egypt, pp.108-112.
- 4- Al-Tahrawi, M. M., and Al-Khatib, S. N. (2015), *"Arabic Text Classification using Polynomial Networks"*, Journal of King Saud University-Computer and Information Sciences, Vol. 27 No.4, pp.437-449.
- 5- Al-Thwaib, E. (2014), *"Text Summarization as Feature Selection for*

- Arabic Text Classification*”, World of Computer Science and Information Technology Journal (WCSIT), Vol. 4 No.7, pp.101-104.
- 6- Anitha, N., Anitha, B., and Pradeepa, S. (2013). *“Sentiment Classification Approaches”*, International Journal of Innovation Engineering and Technology, Vol. 3 No.1, pp.22-31.
- 7- Bahassine, S., Madani, A., and Kissi, M. (2017), *“Arabic Text Classification Using New Stemmer for Feature Selection and Decision Trees”*, Journal of Engineering Science and Technology, Vol. 12 No. 6, pp. 1475-1487.
- 8- Bekkerman, R., and Allan, J. (2004), *“Using Bigrams In Text Categorization”*, Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst.
- 9- Berger, H., Köhle, M., and Merkl, D. (2005), *“On the Impact of Document Representation on Classifier Per-formance in e-Mail Categorization”*, In ISTA pp. 19-30.
- 10- Bloehdorn, S., and Hotho, A. (2004), *“Boosting for Text Classification with Semantic Features”*, In International workshop on knowledge discovery on the web, Springer, Berlin, Heidelberg, pp. 149-166.
- 11- Chantar, H. K., and Corne, D. W. (2011), *“Feature Subset Selection for Arabic Document Categorization Using BPSO-KNN”*, In Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on, pp. 546-551. IEEE.
- 12- Cristianini, N., and Shawe-Taylor, J. (2000), *“An introduction to support vector machines and other kernel-based learning*

- methods*”, Cambridge university press.
- 13- Fürnkranz, J. (1998), “*A study using n-gram features for text categorization*”, Austrian School of Computer Science, Carnegie Mellon University, USA.
- 14- Guyon, I., and Elisseeff, A. (2003), “*An introduction to variable and feature selection*”, Journal of machine learning research, 3(Mar), pp. 1157-1182.
- 15- Harish, B. S., Guru, D. S., and Manjunath, S. (2010), “*Representation and classification of text documents: A brief review*”, IJCA, Special Issue on RTIPPR (2), pp. 110-119.
- 16- Joachims, T. (1998). “*Text categorization with support vector machines: Learning with many relevant features*”, In European conference on machine learning, Springer, Berlin, Heidelberg, pp. 137-142.
- 17- Kanaris, I., Kanaris, K., Houvardas, I., and Stamatatos, E. (2007), “*Words versus character n-grams for anti-spam filtering*”, International Journal on Artificial Intelligence Tools, Vol. 16(06), pp.1047-1067.
- 18- Ker, S. J., and Chen, J. N. (2000), “*A text categorization based on summarization technique*”, In Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics- Association for Computational Linguistics, Vol. 11, pp. 79-83.
- 19- Khreisat, L. (2006), “*Arabic text classification using N-gram*

- frequency statistics a comparative study*", In Proceedings of the 2006 International Conference on Data Mining, pp. 78–82.
- 20-Mourad Abbas - Arabic Corpora (2004), Retrieved September 04, 2018, from <https://sites.google.com/site/mouradabbas9/corpora>. Last access on January 2018.
- 21-Nigam, K. P. (2001), *"Using unlabeled data to improve text classification"*, PhD Thesis, School of Computer Science, Carnegie Mellon University, USA.
- 22-Nuckolls, J. B. (1999), *"The case for sound symbolism"*, Annual Review of Anthropology, Vol. 28, pp. 225–252.
- 23-Oraby, S., El-Sonbaty, Y., El-Nasr, M. A. (2013), *"Exploring the effects of word roots for Arabic sentiment analysis"*, In Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 471-479.
- 24-Permutations: Retrieved October 10, 2022, from <https://en.wikipedia.org/wiki/Permutation>
- 25-Rapid Miner Project RM (2013), *"The Rapid Miner Project for Machine Learning"*, Available: <http://rapid-i.com/> Last access on December 2017.
- 26-Saad, M. K., and Ashour, W. (2010, November), *"Osac: Open-source Arabic corpora"*, In 6th ArchEng Int. Symposiums, EECS, Vol. 10.
- 27-Salton G, Wong A, and Yang C (1975) , *"A vector space model for automatic indexing"*, Commun ACM , Vol. 18 No.11, pp. 613–620
- 28-Santos, C. D., and Zadrozny, B. (2014), *"Learning character-level*

- representations for part-of-speech tagging*”, In Proceedings of the 31st International Conference on Machine Learning ICML- Vol. 14 , pp. 1818-1826.
- 29- Sapir, E. (1929), “*A study in phonetic symbolism. Journal of Experimental Psychology*”, Vol. 12, pp. 239–255. doi:10.1037/h0070931.
- 30- Sawaf, H., Zaplo, J., and Ney, H. (2001), “*Statistical classification methods for Arabic news articles*”, Arabic Natural Language Processing Workshop, ACL’2001, pp. 127–132.
- 31- Sebastiani, F. (2002), “*Machine learning in automated text categorization*”, ACM computing surveys (CSUR), Vol. 34 No.1, pp. 1-47.
- 32- Syiam, M. M., Fayed, Z. T., and Habib, M. B. (2006), “*An intelligent system for Arabic text categorization*”, International Journal of Intelligent Computing and Information Sciences, 6(1), pp.1–19.
- 33- Witten, I. H., and Frank, E. (2005), “*Data Mining: Practical machine learning tools and techniques*”. Morgan Kaufmann.
- 34- Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016), “*Data Mining: Practical machine learning tools and techniques*”, Morgan Kaufmann.
- 35- Yousif, S. A., Samawi, V. W., and Elkabani, I. (2017), “*Arabic Text Classification: The Effect of the AWN Relations Weighting Scheme*”, In Proceedings of the World Congress on Engineering, Vol. 2.
- 36- Zahran, M. M., Kanaan, G., and Habib, M. B. (2009), “*Text feature selection using particle Swarm optimization algorithm*”, World

Applied Sciences Journal, 7(Special Issue of Computer , IT), pp. 69–74.

- 37-Zhang, S., Boukamp, F., and Teizer, J. (2015), *“Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA)”*, Automation in Construction, Vol. 52, pp.29-4



التحقق من النطق الصحيح باستخدام تقنية الحاسوب

**Computer-Aided Language (CALP)
Pronunciation**

دراسة تطبيقية على القرآن الكريم

إعداد:

أ.د. محسن عبد الرازق رشوان

كلية الهندسة - جامعة القاهرة - مصر

د. نعيم محمد عبد الغني

كلية الهندسة - جامعة القاهرة - مصر

ملخص البحث

ينطلق البحث من تعريف تقنية التَّحَقُّق من الأخطاء، ثم يبدأ في شرح كيفية بناء نموذج معياري للنطق الصحيح، ثم يرصد تقييم تطبيقي حفص، وسالم اللّذين اعتمدا على تقنية التحقق من النطق الصحيح باستخدام تقنية computer-aided language pronunciation (CALP)، وكان التقييم آلياً أولاً ثم بعدها بمقارنة المخرجات مع العقل البشري، ثم برصد أثر التقنية في التعليم بشكل واقعي عملي، وأخيراً من خلال تقييم التطبيقين مقارنة بالتطبيقات المشابهة المتاحة الآن على الهواتف المحمولة.

كلمات مفتاحية: النطق الصحيح - التحقق - التلاوة - أخطاء - الذكاء الاصطناعي - التعليم - التقنية.

Abstract

The research starts with the definition of error verification technology, then begins to explain how to build a standard model for correct pronunciation. The research work monitors the evaluation of Hafss and Salem applications, which relied on the technique of verifying correct pronunciation using Computer Aided Language Pronunciation technique (CALP), and the evaluation was automated first, then by comparing the outputs with the human, then by monitoring the impact of technology on education in a realistic and practical way. Finally, the two applications were compared to similar applications now available on mobile phones

Keywords: correct pronunciation - verification - recitation - errors - artificial intelligence - education - technology.

المقدمة

اللغة أصوات يعبر بها كل قوم عن أغراضهم⁽¹⁾، وتخضع هذه الأصوات للوصف من حيث المخارج والحركات التي يقوم بها جهاز النطق، من حيث الصفات، والظواهر الصوتية المصاحبة لهذا النطق⁽²⁾، وهذه الأصوات تحتاج إلى تعليم، ولا يكون ذلك إلا بتفاعل بين المعلم والمتعلم؛ وإدخال الحاسوب في ذلك إنما هو تحدٍ في أن يحاكي قدرات الإنسان على أن يستمع ويحلل ويتعلم اللغة.

وهذا البحث خلاصة جهود لفريق من الباحثين بالشركة الهندسية لتطوير النظم الرقمية (RDI) استمرت لأكثر من 25 عامًا في توظيف التقنيات لخدمة التعليم، وخاصة تعليم اللغة العربية والقرآن الكريم، وليست هذه الورقة البحثية وليدة تجربة استمرت وقتًا يسيرًا، بل هي خلاصة لعمل فريق من الباحثين المختصين في الهندسة مع نظرائهم اللغويين؛ حيث كانت الجهود حثيثة لبناء نموذج معياري في نطق أصوات القرآن الكريم، انطلاقًا من التنظير الذي تركه القدماء والمُحدِّثون، بداية من سيبويه (ت: 180 هـ)، وابن جني (392هـ)، وانتهاءً بعلمائنا المُحدِّثين العرب، والغرب مثل إبراهيم أنيس، وعبد الرحمن أيوب، وتمام حسان، وفيرث، وهنري فليش، وبرجشتراسر، وغيرهم من الرواد.

(1) انظر، ابن جني، أبو الفتح عثمان، الخصائص، تحقيق: محمد علي النجار، دار الكتب المصرية، د.ت، د.ط، 33/1

(2) عمارة، خليل أحمد، في التحليل اللغوي، مكتبة المنار، الأردن، 1987م، ص 15.

1-1 مفهوم التقنية

computer-aided language pronunciation (CALP)

التَّحَقُّق من النطق الصحيح من الوسائل التعليمية التي يحتاجها متعلم اللغات بشكل عام، ومتعلم القرآن الكريم بشكل خاص، وهذا البحث مخصص لدراسة أثر استخدام تقنية computer-aided language pronunciation (CALP)، وهي قريبة جداً من تقنية تعرف الصوت (Speech Recognition)، والفرق بينهما يكمن في أن التقنية الأولى تركز على التحقق من معيارية نطق المتحدث من حيث الصواب والخطأ لكلام نعرفه مسبقاً، أما التقنية الثانية فهي أننا لا نعرف ما الذي سيقوله المتكلم، والآلة هي التي ستعرفنا ما الذي يقوله.

وتقنية computer-aided language pronunciation (CALP) تعني النطق اللغوي بمساعدة الكمبيوتر، وهي ثمرة من تاريخ طويل في تطوير التقنيات الحاسوبية لتعليم اللغات بشكل عام، ويندرج تحته بطبيعة الحال تعلم النطق الصحيح⁽¹⁾.

إننا في هذه التقنية نضع البدائل المختلفة للنطق الصحيح منها، أو للنطق الخطأ، خذ مثلاً: لو أن أجنبياً يريد أن يتعلم العربية، وليس عنده حرف مثل "الطاء" ويريد أن ينطق كلمة طيب فسيقولها بالتاء "تيب"، فإذا كان النطق أقرب لـ"تيب" نعطي تغذية مرجعية بموضع الخطأ، ونطلب منه إعادة المحاولة، وربما نحيله إلى طريقة نطق الطاء تفصيلاً⁽²⁾.

(1) Abdou, S. M., Hamid, S. E., Rashwan, M., Samir, A., Abdel-Hamid, O., Shahin, M., & Nazih, W. (2006). Computer aided pronunciation learning system using speech recognition techniques. In: Proceedings of the ninth international conference on spoken language processing, pp. 849

(2) رشوان، محسن عبد الرازق، مدخل إلى حوسبة اللغة، ضمن كتاب مقدمة في حوسبة اللغة العربية، إصدار مركز الملك عبد الله لخدمة اللغة العربية، السعودية، 2019م، ص 44، 45

تقنية (التحقق من الأخطاء) وتعاملها مع اللغات

الهدف من التقنية كما ذكرنا هو تعرف أخطاء النطق في اللغات بشكل عام، ونحن قد استخدمناها في تعرف أخطاء النطق في القرآن الكريم أولاً، ثم اللغة العربية، وبدأنا بالقرآن الكريم؛ لأن القرآن الكريم أكثر ثراء في فونيماته، والناطقون به من العالم الإسلامي كثيرون؛ فتعليم نطق القرآن سيعطينا مادة غزيرة جداً يمكن أن نبني منها نظام تعرف الأخطاء خاصة إذا كانت اللغة العربية بها فونيمات وأصوات غير موجودة في لغات العالم، والشكل رقم (1) فيه دراسة مقارنة بين العربية وعشر لغات لرصد الصوامت التي تتفرد بها اللغة العربية عن باقي اللغات، وقد اخترنا الصوامت محلاً للدراسة؛ لأنها تمثل مجال الاختلاف، أما الحركات فهي مشتركة غالباً مع كل اللغات مع تفاصيل في كيفية إنتاج هذه الحركات ما بين التفخيم، والترقيق، والتقليل، والإمالة، وغيرها من صفات الحركات.

وإذا تأملنا الجدول وجدنا -مثلاً- أن حرف الباء موجود في كل اللغات إلا الصينية، وصوت الواو ليس موجوداً في التركية والأردية والروسية؛ أما صوت الثاء فليس موجوداً إلا في الإسبانية والإنجليزية، وصوت الذال موجود فقط في الإنجليزية، وصوت السين موجود في كل اللغات إلا اللغة التركية، وأما الصينية فتخلو من صوت الزاي، والياء غير موجودة في الأردية والصينية، والشين غير موجودة في الصينية والإسبانية، والجيم غير موجودة في الإسبانية والتركية والصينية والفرنسية والألمانية، وأما الغين فليست موجودة إلا في التركية والأردية، ومع أن صوت الغين قريب من الخاء فإن الخاء توجد في أربع لغات على عكس الغين توجد في لغتين فقط؛ فالخاء موجودة في الإسبانية والأردية والصينية والروسية والألمانية، وكذا الهمزة غير موجودة في الإندونيسية

والإسبانية والصينية والفرنسية والألمانية، أما الهاء فإنها لا توجد في الإسبانية والروسية والصينية والفرنسية.

وأما الأصوات المشتركة في كل اللغات فهي: الميم والنون واللام والكاف، وأما الأصوات التي تنفرد بها العربية دون غيرها فهي: أصوات الظاء والضاد والقاف والحاء والعين.

بيان الأصوات الصامتة

البيان أدناه يعرض صوامت اللغات محل الدراسة من واقع رموز الألفبائية الصوتية الدولية (IPA)، يشير (٠) إلى أن الصوت العربي غير موجود في تلك اللغة، هناك العديد من الأصوات في كل من هذه اللغات لا وجود لها في اللغة العربية وبالتالي غير مدرجة في هذا البيان.

طريقة النطق	التفخاري	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا / رعد	اللفظي	نصف حركة	استعصا / رعد	التفخاري	استعصا

هذا الجدول من تصميم الباحث من واقع رموز الألفبائية الدولية (IPA) للغات محل الدراسة.

شكل رقم (1) لبيان فونيمات اللغة العربية التي تتفق أو تختلف مع عشر لغات هي الأندونيسية والإسبانية والتركية والأردية والروسية والصينية والإنجليزية والفرنسية والألمانية والهوسا.

إذاً فهذه خصائص أشهر اللغات العشرة التي يتكلم بها العالم، وكما هو ثابت فإن اللغة العربية بها فونيمات أكثر من اللغات الأخرى، وأن لغة القرآن أكثر ثراء في الأصوات من حيث إضافة الغنن والمدود وغيرها من قواعد التجويد، بعد ذلك نريد أن نبحث في كيفية بناء نموذج لاكتشاف أخطاء النطق في القرآن وكيفية تصحيحه بشكل آلي.

2-1 كيفية بناء نموذج للتحقق من النطق الصحيح.

بناء نموذج رياضي للتحقق من النطق الصحيح يتم من خلال الخطوات التالية:

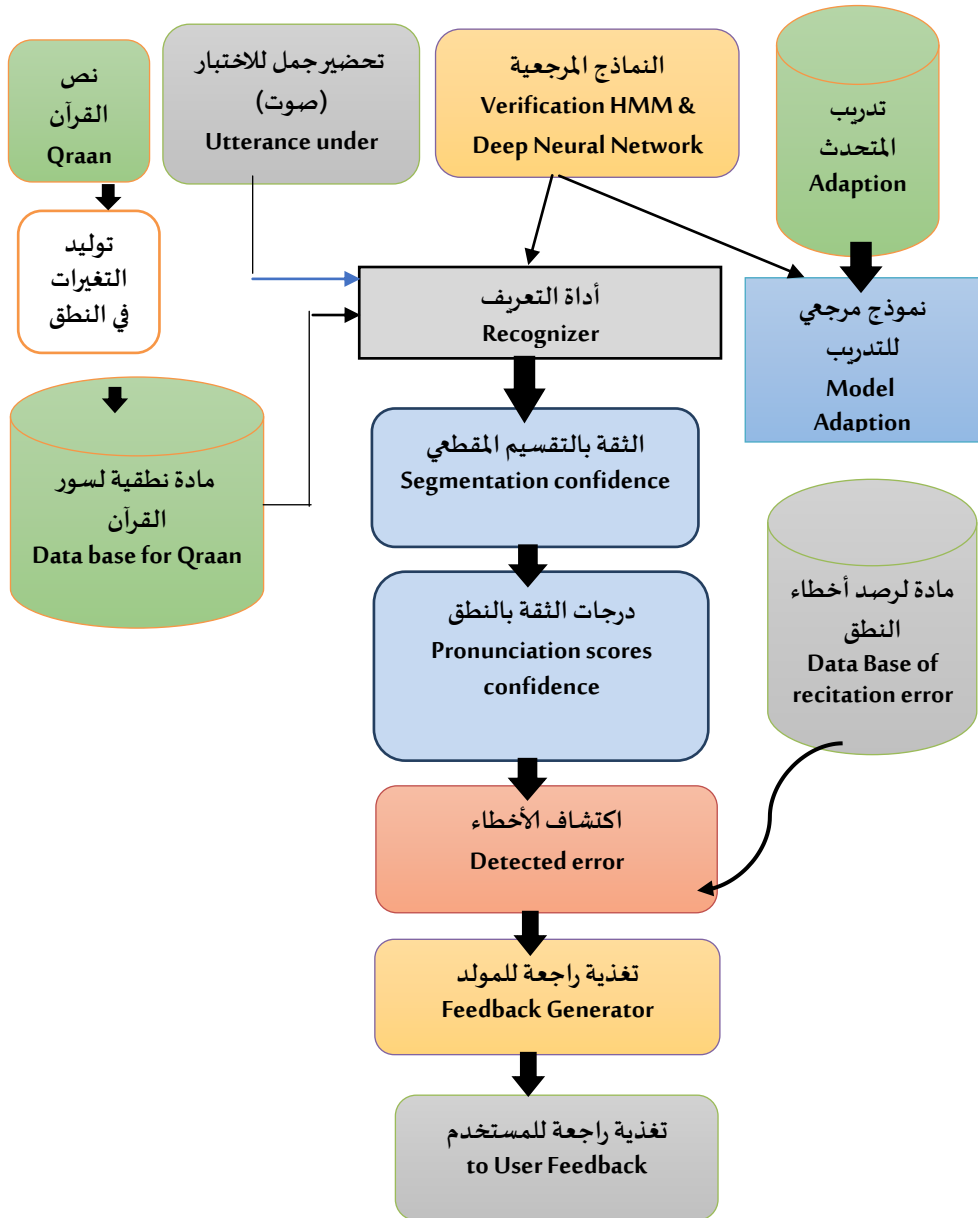
1- بناء نظام رياضي يوضح التتابع الصحيح لفونيمات الكلمة والقواعد الحاكمة لذلك، مثل عدم جواز التقاء الساكنين، وغيرها من القواعد المستنبطة من المدونات اللغوية، ويزود هذا النموذج بالنطق الصوتي الصحيح للقرآن الكريم.

2- بناء نموذج رياضي لمعيارية النطق الصحيح، من خلال تحضير جمل صوتية منطوقة بشكل سليم.

3- بناء نموذج رياضي للمقارنة بين الصواب والخطأ، من خلال بناء قاعدة بيانات للأخطاء.

4- اكتشاف الأخطاء وتحليلها على مستوى المولد، ثم على مستوى المستخدم النهائي.

5- توليد الردود على المتعلم، وبيان هذا النموذج في الشكل رقم (2):



شكل 2- نموذج رياضي لبناء تقنية اكتشاف الخطأ، ويظهر فيه معيارية الصواب وطريقة تحليل الأخطاء وتوليد الرسائل للمولد والمستخدم ش

ويظهر في النموذج خمسة محطات هي:

1- أداة التعريف التي تُغذّى بالنماذج المرجعية للتدريب وتحضير الجمل الصحيحة للاختبار، والنموذج المرجعي للتدريب يستقي مادته من تدريب المتحدثين والتسجيل لهم، وهو بدوره يغذي أيضاً أداة التعريف.

2- والمحطة الثانية تأتي في الثقة بالتقسيم المقطعي للكلام، ثم اختبار درجات الثقة بالنطق، ثم اكتشاف الأخطاء، ثم توليد التغذية الراجعة لمولد الأخطاء، ثم توليد تغذية راجعة للمستخدم؛ ومرحلة اكتشاف الأخطاء تُغذّى بقاعدة بيانات للأخطاء، وتفصيل ذلك في الخطوات التالية:

3-1 بناء نموذج معياري لتعليم الحاسوب.

بدأ العمل أولاً ببناء نموذج معياري للنطق الصحيح من خلال الاعتماد على كبار القراء في العالم الإسلامي، وخاصة الشيخ الحصري؛ حيث قمنا بما يأتي:

1- تحديد ساعات صوتية عشوائية من المصحف وتقطيعها إلى ملفات قصيرة.

2- مطابقة النص المكتوب مع المسموع.

3- الكتابة الصوتية للآيات القرآنية لتحديد الفونيمات.

4- التحليل الطيفي للفونيمات.

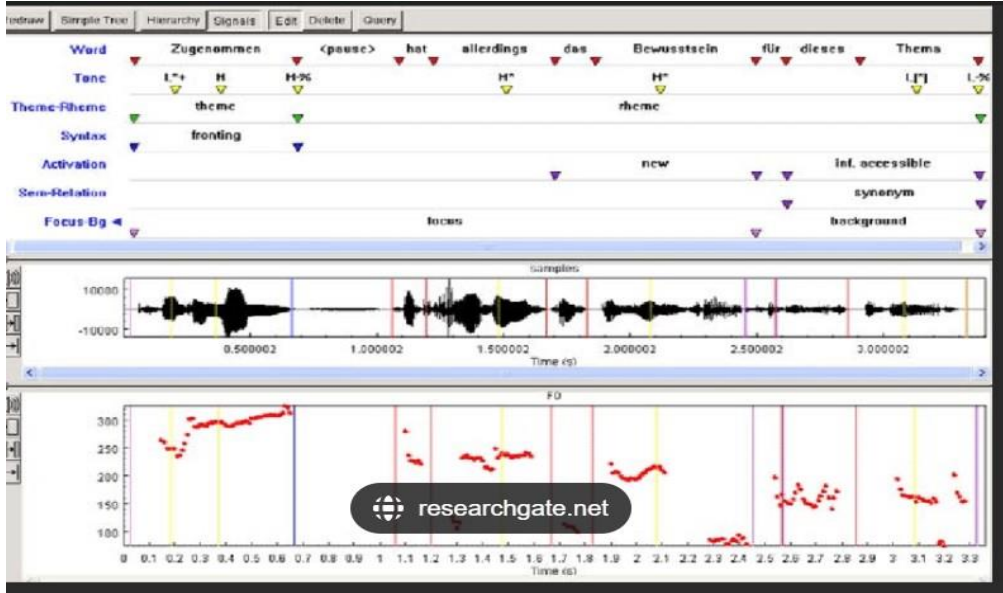
4-1 تمييز الفونيمات ومعرفة خصائصها.

في هذه المرحلة نحتاج إلى سماع كل فونيم على حدة، ورؤية شكله وخصائصه، ونستعين في هذه المرحلة بعدة برامج، منها (cool edit pro)، وبرنامج (EMU SPEECH)، وفي هذين البرنامجين يمكننا تمييز الفونيمات ومعرفة خصائصها بدقة شديدة؛ فنميز بين الصوت الاحتكاكي والانفجاري، إلخ، وصفات الفونيمات.

استخدام تقنيات التحليل الطيفي لمعرفة متوسط أطوال الفونيمات بشكل معياري، فمن المعروف أن الأداء يتحكم في زمن الفونيم، فمن الممكن أن ننطق الغنة بزمن طويل في القراءة البطيئة وتكون صحيحة، ومن الممكن أن نقرأها في القراءة السريعة التي تسمى بالحدُر في زمن أقل، وتكون معيارياً صحيحة أيضاً، وهناك معايير صوتية مثل النبر والتنغيم، نحتاج إلى رصدها لمعرفة أثرها في تحديد معيارية النطق الصحيح.

وفي هذا المقام نذكر رصداً لفونيمات القرآن الكريم لبناء قاعدة بيانات صوتية تفيد في تعليم الآلة أحكام تلاوة القرآن الكريم؛ فبناء على الإحصاء تبين - مثلاً- وورود صوت اللام في القرآن نحو 618465 مرة بنسبة 12 % تقريباً، بينما ورد صوت النون 338646 مرة بنسبة تقارب 7 % تقريباً، أما صوت الميم فورد 333858 مرة بنسبة تقارب 7 % أيضاً، بينما لم يرد صوت الضاد إلا 22640 مرة بنسبة 0.44 %، وصوت الغين 16693 مرة بنسبة 0.33 %، وصوت الظاء ورد 8959 مرة بنسبة 0.18 %

من هذه الإحصاءات يمكن أن نبني نموذجاً يتعلم من خلاله الحاسوب النطق الصحيح أولاً، ثم يقارن بين هذه المعيارية وما يستمع إليه من نطق، وعلى أساسها يتم تحديد الخطأ تماماً؛ حيث نحدد أطوال الفونيمات بدقة كما يظهر في الشكل رقم (3) الذي يشرح كيفية تحديد أطوال الفونيمات بالتحليل الطيفي.



شكل 3: لصورة التحليل الطيفي من برنامج EMU

5-1 تعليم الحاسوب اكتشاف الأخطاء.

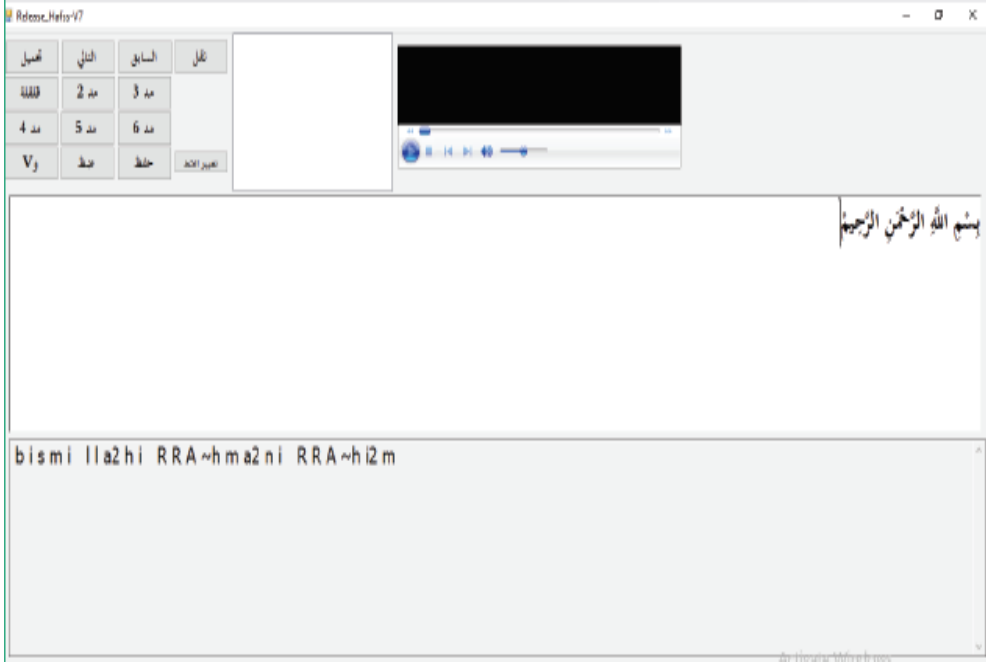
هذه المرحلة تعد أصعب من المرحلة السابقة؛ فتحديد الأصوات الصحيحة في القرآن أمر سهل؛ لأن المعيارية فيها واضحة ومحددة، مع ما يكتنفها من قضايا كالتفاوت في سرعة القراءة ونغمة القارئ، لكن الأخطاء ليست محصورة، وهي متنوعة لعوامل كثيرة، منها العيوب النطقية، واختلاف اللغات واللهجات، وتنوع المستوى المعرفي والثقافي، ولكي تصل نسبة الدقة في تعرف الأخطاء إلى أكثر من ذلك ينبغي أن تكون لديك عينة ضخمة من المتحدثين الذين تتنوع أعمارهم، وتختلف ثقافتهم وتتعدد لغاتهم، وهذا ما قمنا به بالفعل، والصفحات التالية ترصد خلاصة ما قمنا به في هذا الجانب:

أولاً: قمنا بحصر الأخطاء التي تكلم عنها علماء التجويد، حصراً نظرياً، فمثلاً

من الأخطاء التي يقع فيها القراء أن الكاف مرققة دائماً، وأكثر الأخطاء تقع فيها بسبب التفخيم الذي تتفاوت درجاته حتى يصل إلى أن تبدل القاف كافاً، وهكذا في كل الحروف، والغُن والمدود وغيرها.

ثانياً: قمنا بتسجيل جُمَل قرآنية وحروف لنحو 1200 متحدثاً من الذكور والإناث من مختلف الأعمار والثقافات؛ لتعليم الآلة كيفية اكتشاف الأخطاء.

ثالثاً: بعد التسجيل تخصص باحثون في أن يسمعون المقاطع الصوتية، ويكتبوها كما سمعوها، فمثلاً: إذا قال المتحدث: "الحمت لله لب العالمين" فإننا نكتبها كما هي ولا نكتبها "الحمد لله رب العالمين"، وهذا جعلنا نزيد في رموز الكتابة الصوتية بما يتوافق مع أنواع الأخطاء التي تزيد باستمرار، فمثلاً: كلمة (رب) تتنوع فيها الأخطاء إلى: (يب: إبدال الراء ياء)، وإلى: (لب: إبدال الراء لاما مرققة أو مفخمة)، وعندما نكتبها فإننا نراعي ما حدث تماماً، ونصفه حتى نستطيع أن نُعَلِّم الآلة هذا الخطأ؛ والأداة التي يستخدمها الباحث صُمِّمَت لهذا الغرض، فيها رمز للمد والقلقلة وإبدال الواو (V)، كما يظهر في الشكل التالي الذي يصف مرحلة توسيم الأخطاء (Annotation)، وهي المدونة التي سيتعلم عليها نظام التحقق من الأخطاء.



شكل: 4 أداة برمجية لكتابة المسموع فقط وتوسيمه.

المبحث الثاني: التحقق من قدرة الحاسوب على اكتشاف الأخطاء.

في هذه المرحلة نخبر قدرة التقنية على التحقق من الأخطاء بوسيلتين:

الوسيلة الأولى: التحقق الذاتي: ونقصد بها اختبار التقنية ذاتيًا على اكتشاف أخطاء التلاوة من خلال اختيار عينة عشوائية، ومقارنة أدائها بمعيارية الصواب، وهذه المرحلة تتم بشكل آلي.

الوسيلة الثانية: المقارنة البشرية: ونقصد بها إجراء تجارب معملية مع شيوخ مُجَازِينَ متقنين لتقييم الطلاب في أحكام التلاوة مقارنة بتقنية تعرف الأخطاء.

تفاصيل المرحلة الأولى:

في هذه المرحلة يحدث تقييم للنظام لنعرف المشكلات التي على الباحث أن يجد لها حلاً، مثل: عدم اكتشاف التفخيم والترقيق، أو تبديل فونيم مكان آخر، أو إضافة فونيم ليس موجوداً، أو حذف فونيم موجود... إلخ، هذه هي الأخطاء التي سنذكر لها أمثلة، مع صورة من المصفوفة التي رصدت هذه الأخطاء، وفي هذا الإطار نذكر تقنية حفص التي أنتج منها مؤخراً تطبيق سالم الذي يعلم سورة الفاتحة وحروف الهجاء بشكل تفاعلي، وفيه استخدمنا هذه التقنية بواسطة بناء ثلاثة نماذج، نموذج التحقق من الأخطاء عن طريق:

أ- الحرف وحركته.

ب- الكلمة.

ج- الجملة.

أ- على مستوى الحرف.

وهنا يجدر بنا أن نذكر النتائج التي توصل إليها تدريب الآلة على تقنية تعرف الأخطاء، وسنبداً في الجدول (1) بذكر عدد المتحدثين وما سجلوه من ملفات صوتية، وإجمالي عدد الساعات المسجلة للمتحدثين من جنسيات مختلفة.

جدول (1) يبين المادة التي سجلت لرصد الأخطاء التي يقع فيها القراء.

عدد المتحدثين	عدد الملفات الصوتية	عدد الساعات
3284	522691	348

وهذه المادة دُرِّبَ النظام فيها على أكثر من نصف مليون ملف صوتي، ما بين الحروف والكلمات والجملة، ومقارنة ذلك بالنموذج الصحيح، وعند اختبار

كفاءة النظام اخترنا خمس ساعات بشكل عشوائي، وظهرت النتائج على هذه المصفوفة؛ حيث تشير إلى أن هناك حروفاً وكلمات وجمالاً نطقت خطأ، ولم يستطع النظام أن يكتشفها على النحو التالي:

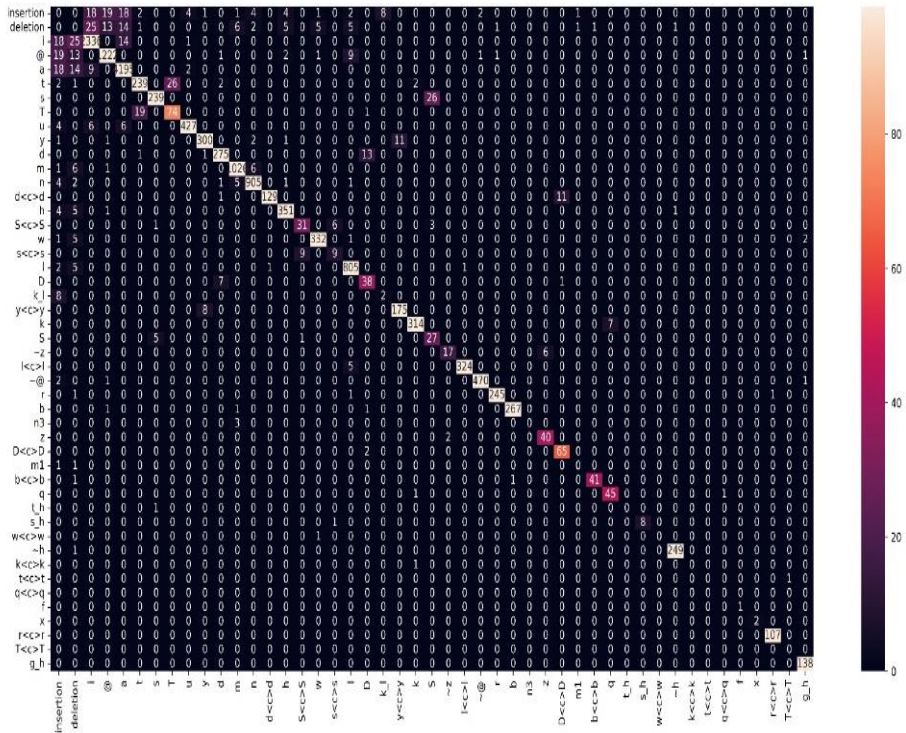
أولاً: أمثلة على مستوى الحروف.

نقصد بالحروف، الحرف وحركته، وهما فونيمان، فالوحدة هنا ليست بالفونيم وإنما بالفونيمين. أ - ب... إلخ (a - ba-....ect)، وهنا في خمس ساعات، أي على نحو عشرين متحدثاً كانت هذه الأخطاء التي لم يستطع الحاسوب اكتشافها، أما الباقي فقد استطاع تمييزها، كما يظهر من الجدول (2) ثم المصفوفة التي تليه.

الجدول (2): وفيه رصد لأخطاء النظام في عدم قدرته على اكتشاف الأخطاء

الخطأ	معدل تكرار عدم اكتشافه
الخلط بين الدال والضاد	22
الخلط بين الثاء والسين	12
الخلط بين الذال والزاي	18
الخلط بين الشين والسين	1
الخلط بين التاء والطاء	10
الخلط بين القاف والكاف	18

ويمكن استقرار المزيد عبر هذه المصفوفة



المصفوفة (1) وفيها اختبار كفاءة النظام في اكتشاف الأخطاء
على مستوى الحروف.

على مستوى الكلمة.

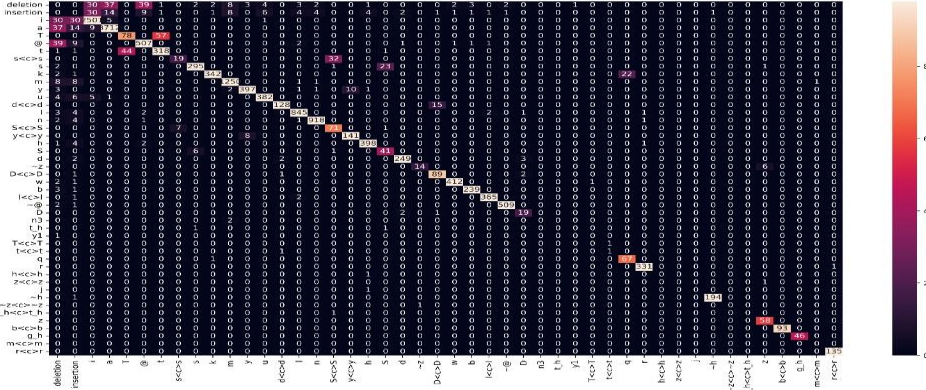
اكتشاف التقنية الخطأ على مستوى الكلمة يمكن رصده في الجدول رقم (3).

الجدول رقم (3) ويظهر فيه عدم قدرة النظام على اكتشاف بعض الأخطاء على مستوى الكلمة.

عدد مرات عدم اكتشافه	موضع الخطأ في الكلمة
19	الخلط بين التاء والطاء
31	يصل الخلط بين السين المشددة والصاد المشددة
26	الخلط بين الصاد والسين
11	الياء المشددة والياء غير المشددة
17	الذال والزاي
14	الخلط بين الكسرة والفتحة

أمثلة من المصفوفة التي ترصد اختبار التقنية في مدى قدرتها على اكتشاف الأخطاء على مستوى الكلمة، ويلاحظ أيضاً أن التقنية تكتشف الفونيم الخطأ على مستوى الكلمة، وهذا أيضاً ما يتم في اكتشاف الأخطاء على مستوى الجملة، والمزيد يمكن اكتشافه في المصفوفة رقم (2).

ويمكن رصد المزيد من عدم قدرة التقنية على اكتشاف الأخطاء عبر هذه المصفوفة أيضاً:



المصفوفة رقم (3) تظهر فيما عدم قدرة النظام على اكتشاف بعض الفونيمات على مستوى الجملة.

ويمكن إجمال النتائج التي خرجنا بها من النماذج الثلاثة من اختبار قدرة التقنية على اكتشاف الأخطاء فيما يأتي:

- 1- إن تخصيص نموذج واحد للتحقق من اكتشاف الأخطاء يقلل من دقة التطبيق؛ ولذا ينبغي النظر عند تعليم الأصوات إلى النطق على مستوى الحرف والكلمة والجملة، وبناء خوارزمية لكل مستوى؛ لأن هذا يجعل النتائج أكثر دقة.
- 2- كلما كبرت الوحدة فإن الأمر يكون أكثر صعوبة على التقنية، وهذا هو التحدي الذي يقوم به الباحثون لتقليل الأخطاء.
- 3- يطول الزمن عند اكتشاف الخطأ عندما تكبر الوحدة، بمعنى أن زمن

اكتشاف الخطأ في ملف الصوت الذي به حرف، يكون أقصر من زمن اكتشاف الخطأ في الملف الذي به كلمة أو الذي به جملة. هذه النتائج تساعد كثيرًا في تحسين النظام ليكون مقاربًا في التقييم من العنصر البشري، بل أحيانًا يكون متفوقًا عليه، وهذا ما سنؤكدده من خلال التجربة في المبحث التالي: الوسيلة الثانية: المقارنة بين الحاسوب والعقل البشري في اكتشاف الأخطاء. في هذه الوسيلة قمنا بعدة تجارب من مناطق مختلفة: في مصر، والكويت، والسعودية، على النحو التالي:

اختبارات دقة التقنية.

أولاً: أجريت هذه التجربة في مصر حيث اخترنا 9 شيوخ مُجَازين، وحددنا لهم في التقنية مجموعة من الكلمات والجمل، وطلبنا منهم تسجيلها مع تعمد الخطأ من واقع خبرتهم، وتوفر لدينا في هذه التجربة أخطاء كثيرة، وكانت النتيجة وصول دقة برنامج اكتشاف الأخطاء إلى 84 %.

ثانيًا: اخترنا (300) جملة، وطلبنا من شخص عادي معرفته بأحكام التجويد متوسطة، أن يقرأ هذه الجمل، وطلبنا من (4) شيوخ تحكيم قراءة الجمل الثلاثمائة، ثم أدخلنا هذه الجمل على التقنية فكانت النتائج أن متوسط اتفاق التقنية مع كل شيخ على حدة ~ 80 %، ونسبة مخالفة التقنية لإجماع الشيوخ حوالي 5%. وهذه تكاد تتطابق مع توافق تحكيم الشيوخ فيما بينهم.

ثالثاً: اختبار التقنية في قياس مقدرة التعلم.

أجريت هذه التجربة في مصر، حيث حددنا 9 دروس مختلفة مثل: القلقة، وإدغام النون الساكنة والتنوين، والإقلاب... إلخ، وقمنا بترك المتعلمين يتدربون عن طريق التقنية حفص، وكان متوسط التدريب على التقنية نحو ساعة وربع لكل درس، وكانت النتيجة المذهلة ارتفاع مستوى المتعلم من 39 % إلى 74 %.

رابعاً: اجراء التجربة:

- أجريت هذه التجربة في السعودية، على 42 طالباً بإحدى مدارس الرياض، 21 طالباً في العينة الضابطة، ومثلهم في العينة التجريبية، من خلال:
- اختيار دروس عشوائية من تقنية حفص، هي: الإدغام، والإظهار في النون الساكنة.
 - تحديد عينة ضابطة تتم بشكل تقليدي.
 - تحديد عينة تجريبية تتعلم من خلال التقنية.
 - اخترنا الطلاب على أساس التكافؤ في المرحلة والسنة الدراسية، فهم في الصف الخامس الابتدائي.
 - كانت النتيجة: حصلت العينة الضابطة في الاختبار على 33 %، بينما حصلت العينة التجريبية التي تعلمت من خلال التقنية على 38 %.
 - تعميم التقنية على العينة الضابطة والعينة التجريبية، وبعد أسبوعين أجري الاختبار فكانت النتيجة حصول العينة الضابطة على متوسط درجات 55 %، أما عينة القياس فبلغ 77 %.
 - بلغت نسبة التحسن في استخدام الوسائل التقليدية في التعليم حوالي 65 %، بينما بلغت نسبة التحسن عن طريق التقنية 104 %.
- خامساً: أجريت هذه التجربة في الكويت، على حلقين من حلقات تحفيظ القرآن الكريم بوزارة الأوقاف الكويتية، هما: مركزا الشاطبي، والوازن؛ حيث قمنا بتحديد العينة الضابطة التي ستتعلم بشكل تقليدي والعينة التجريبية التي ستتعلم بتقنية حفص، وكانت هذه هي الإجراءات.

1- إجراء اختبار تحديد المستوى يشمل: التحريري والاستماع والشفوي ودرجاتهم 20 و30 و50 على الترتيب.

2- تحديد أحكام القلقلة وحكم إدغام الميم الساكنة.

ويلاحظ في هذه العينة أن نسبة التحسن أعلى بكثير من العينة الضابطة التي كانت تعتمد على الطريقة التقليدية. وخلاصة هذه النتائج كما يأتي:

1- كانت النتيجة: حصلت العينة الضابطة في الاختبار على 50 %، بينما حصلت العينة التجريبية التي تعلمت من خلال التقنية على 45 %.

2- تعميم التقنية على العينة الضابطة والعينة التجريبية، وبعد أسبوعين أجري الاختبار فكانت النتيجة حصول العينة الضابطة على متوسط درجات 70 %، أما عينة القياس فبلغ 81 %.

3- بلغت نسبة التحسن في استخدام الوسائل التقليدية في التعليم حوالي 40 % بينما بلغت نسبة التحسن عن طريق التقنية 80 % .

المبحث الثالث: تقييم تقنية حفص وسالم بين التقنيات المشابهة.

بعد التقييم الذاتي لتقنية حفص، والتقييم من خلال المقارنة مع العنصر البشري؛ فإننا سنُجري تقييمًا آخر لمعرفة تقدم تقنية حفص وسط التقنيات المشابهة، وقد بحثنا في التطبيقات المتاحة التي تعلم القرآن من خلال العنصر التفاعلي، واخترنا ثلاثة تطبيقات سنرمز لها بتطبيق 1 وتطبيق 2 وتطبيق 3.

أولاً: تطبيق (1)

تطبيق (1) من خلاله يستطيع المستخدم تسميع آيات القرآن الكريم، ونعتمد على ما جاء في التطبيق من وصف تعريفي له؛ حيث إن التطبيق في نشرته التعريفية يذكر هذا النص: "التطبيق الأول والوحيد الذي يختص بتعليم تلاوة القرآن الكريم باستخدام تقنية التعرف على الصوت، كما يتميز بعدة خصائص فريدة؛ فهو يدعم أي لغة حتى الآن؛ بالإضافة إلى تسهيل عملية الحفظ، وإحدى أهم مزاياه هي خاصية البحث في القرآن الكريم".

ووفقاً لما جاء في التطبيق فإن نظام تسميع مختص بما يأتي:

1- المساعدة في تذكر الكلمات والجمل.

2- الخطأ في الكلمات ووضعها في غير موضعها.

3- يستخدم تقنيات تعرف الصوت.

وبناء على ما سبق فإنه يتيح للمستخدم ثلاث خواص، هي:

أولاً: تسميع: خاصية تلاوة المستخدم للآية أمام الجهاز وتصحيحها من قبل التطبيق.

ثانياً: الحفظ: خاصية ترديد الآيات أمام الكمبيوتر مع إخفاء بعض الكلمات أثناء القراءة للمساعدة على الحفظ.

ثالثاً: المصحف: خاصية القراءة من المصحف، مع تفسير الآيات ومعاني كلمات القرآن.

رابعاً: تسميع بالذكاء الاصطناعي: خاصية تلاوة المستخدم للصفحة التي يحفظها أمام الكمبيوتر، واختيار الحفظ من قبل التطبيق بتقنية سماع صوته وإظهار الأخطاء.

خامساً: تسميع عبر الحوسبة السحابية للإنترنت ذات السرعة العالية: وهي مثل الخاصية السابقة لكن مع الاتصال بالإنترنت.

سادساً: تسميع بالنقر على الشاشة، حيث تظهر الكلمة بالنقر على الشاشة.

سابعاً: تسميع عرض آلي: حيث تظهر الكلمة على الشاشة بشكل آلي.

ثامناً: التطبيق به نحو 20 لغة هي العربية والإنجليزية والفرنسية والألمانية والهندية، وغيرها.

تاسعاً: التطبيق به إمكانية لاختيار قارئ من 10 قراء.

عاشرًا: فيه إمكانية للتنبيه عند الخطأ، وتكرار الآية حسب رغبة المستخدم، وإظهار التفسير وإخفاؤه، والانتقال التلقائي للآية.

نقد التطبيق.

الوصف السابق للتطبيق تحقق إلى حد كبير، فبعد تجربته على 100 جملة تغطي أحكام التجويد من خلال ثلاثة باحثين تعمدوا الأخطاء في أحكام التجويد لتقييم البرنامج تبين أنه:

- 1- لا يقوم بأي تصحيح لأحكام التلاوة.
- 2- لا يكتشف أي خطأ في نطق الحروف.
- 3- يُلون كلمات على أنها تنطق خطأ رغم نطق الباحث لها بشكل سليم.
- 4- عدم اكتشاف الخطأ أصلاً رغم نطقه والمبالغة في إظهار الخطأ.
- 5- لا يعطي رسالة تعليمية عن نوعية هذا الخطأ إلا فيما يختص فقط بالتسميع؛ حيث يلون الكلمة التي يخطئ فيها المستخدم باللون الأحمر.

والخلاصة: أنه بهذه الطريقة لا يعتمد تقنية التحقق من الأخطاء؛ بل يستخدم تقنية تعرف الصوت وتحويله إلى نص.

مثال تطبيقي لتقييم التقنية من الباحثين في الجدول رقم (5)

الجدول رقم (5) يشرح تقييمات الباحثين للتطبيق (3) في قدرته على اكتشاف الأخطاء.

الباحث	اكتشف الخطأ	لم يكتشف الخطأ
1	43 مرة	67 مرة
2	14	86
3	22	88

وينبغي ملاحظة أن عبارة اكتشاف الخطأ جاءت مصحوبة بملاحظات هي:

1- اكتشاف الخطأ مع تحديد كلمات، قال البرنامج إنها خاطئة رغم نطقها بشكل صحيح.

2- اكتشاف أخطاء بكثرة رغم نطقها بشكل صحيح.

3- اكتشاف الخطأ مع ثقل في البرنامج وطول المدة في تحليل الجملة.

أما عبارة لم يكتشف الأخطاء فجاءت أيضاً مصحوبة بملاحظات هي:

1- عدم اكتشاف الخطأ المقصود واكتشاف أخطاء أخرى.

2- أظهر الآية كلها بشكل خاطئ.

3- أظهر الآية كلها صواباً رغم نطقها بطريقة خاطئة.

4- عدم اكتشاف الخطأ والانتقال إلى الصفحة التالية.

ثانياً: تطبيق (2)

تطبيق (2) يعتمد على تقنية تعرف الصوت؛ فيظلل الكلمة أو الجملة التي يقرأها المستخدم فقط، ولا ينميه إلى الصواب والخطأ؛ ومن ثم فكل المواضع التي تم اختبار الترتيل فيها من الباحثين الثلاثة لم يستطع البرنامج أن يكتشف فيها أي أخطاء.

ثالثاً: تطبيق (3)

تطبيق (3) يستخدم تقنيات الذكاء الاصطناعي في تعرف صوت القارئ، ويدعم التطبيق (52) لغة، بالإضافة إلى تسهيل عملية الحفظ، وإحدى أهم مزاياه خاصة البحث بالصوت؛ مما يسهل على الأشخاص ذوي الإعاقة من استخدامه، ويعمل على تطويره شركة اقرأ للتقنية، وقد حصل على المركز الأول في جائزة الكويت الدولية .

خواص التطبيق.

- 1- التردد أمام القارئ المفضل للمستخدم والتصويب من قبل البرنامج.
- 2- التلاوة المرسله وتصويبها من التطبيق.
- 3- إظهار الكلمات المتشابهة لفظاً.
- 4- الحفظ حيث يردد الطالب الآيات مرات، ويقوم التطبيق بإخفاء ما يريد الطالب تسميعه وتصويب ذلك.
- 5- وجود المصحف مع تفسير الآيات.

وصف عمل التطبيق بالذكاء الاصطناعي واكتشاف أخطاء التلاوة.

- 1- يبدأ التطبيق باختيار الموضع الذي يقرأه المستخدم.
- 2- بعد الفراغ من القراءة تلون الأحكام التي أخطأ فيها القارئ.

3- ثم بالوقوف مطوّلًا على الكلمة يظهر له إن كان قرأها صوابًا أم خطأً.

4- وعند الخطأ يبين له موضع الخطأ لكن لا يشرحه له.

اختبار التطبيق.

تطبيق (3) هو الوحيد بين التطبيقات الثلاثة محل الاختبار الذي يخرج أخطاء التلاوة، ويعتمد على تقنية (CALP)، ولذلك سنفصل القول فيه بناءً على أحكام التلاوة التي حددنا أمثلة لها ومواقع أخطاء محددة يخطئ فيها ثلاثة باحثين بثلاث طرائق مختلفة.

وقد رصد الباحثون ملاحظاتهم على تطبيق المعلم على النحو التالي:

- 1- لم يكتشف التطبيق الخطأ المقصود بالاختبار واكتشف أخطاءً أخرى.
- 2- ظهور الخطأ في نصف الحكم فقط، مع أن نطق الحكم كله خطأً.
- 3- ظهور الخطأ صوابًا والصواب خطأً.
- 4- عدم ظهور الأخطاء أصلاً مع كثرتها.
- 5- عدم ظهور الأخطاء مع كثرتها.
- 6- ظهور الخطأ مع تلوين الكلمات قبل نطقها.
- 7- اكتشاف الخطأ ووصفه بدقة.
- 8- اكتشاف الخطأ ولم يصفه.
- 9- اكتشاف الخطأ ووصفه بشكل خاطئ.
- 10- نطق الباحث الخطأ مرتين فيكتشفه التطبيق مرة ولا يكتشفه مرة أخرى.

تحليل بيانات نتائج اختبار الباحثين لتطبيق 3

وبالنظر إلى هذه الملاحظات جميعها، فإنه يمكن دمجها في ملاحظتين فقط

هما:

هل اكتشف الخطأ المراد اختباره ؛ أم لم يكتشف الخطأ ؟ .

وهل اتفق الباحثون في هذا أم اختلفوا ؟ .

وإن كانوا قد اختلفوا فما نسبة هذا الاختلاف ؟

والجواب عن ذلك نلخصه في الجدول رقم (6)، وسنرمز لاكتشاف الخطأ بالرمز موجب (+) وبعدم اكتشاف الخطأ بالرمز (-)، وتكتب العلامتان بترتيب الأمثلة كما في قاعدة البيانات الأساسية.

الجدول رقم (6) وفيه نرمز لاكتشاف الخطأ بالرمز موجب (+) وبعدم اكتشاف الخطأ بالرمز (-)، وتكتب العلامتان بترتيب الأمثلة كما في قاعدة البيانات الأساسية).

م	الحكم	عدد الأمثلة	الباحث الأول	الباحث الثاني	الباحث الثالث
1	الإظهار الحلقي	2	--	--	--
2	الإدغام بغنة (ن)	2	(+)	-	-
3	الإدغام بغنة (م)	2	++	-	+ _
4	الإدغام بغنة (ي)	2	++	+	- +
5	الإدغام بغنة (و)	2	++	-	- +
6	الإدغام بغير غنة (ل)	2	- +	--	- +
7	الإدغام بغير غنة (ر)	2	-	-	-
8	الإقلاب	2	++	-	- +
9	إخفاء حقيقي مع (ص)	2	++	-	- +
10	إخفاء حقيقي (ذ)	2	++	-	- +

م	الحكم	عدد الأمثلة	الباحث الأول	الباحث الثاني	الباحث الثالث
11	إخفاء (ث)	2	- +	-	--
12	إخفاء (ك)	2	++	-	-
13	إخفاء (ج)	2	++	-	++
14	إخفاء (ش)	2	+	+	- +
15	إخفاء (ق)	2	+	- +	- +
16	إخفاء (د)	2	++	-	- +
17	إخفاء (ط)	2	++	-	- +
18	إخفاء (س)	2	++	- +	- +
19	إخفاء (ز)	2	- +	-	- +
20	إخفاء (ف)	2	++	-	-
21	إخفاء (ت)	2	++	-	+
22	إخفاء (ض)	2	++	-	+
23	إخفاء (ظ)	2	++	-	-
24	الميم الساكنة-إظهار شفوي	2	++	-	++
25	الميم الساكنة-إخفاء شفوي	2	- +	-	- +
26	إدغام المتماثلين	2	+	- +	- +
27	اللام القمرية	2	- +	+	- +
28	اللام الشمسية	2	++	-	-
29	المتقاربان ل-ر	1	+	-	+
30	المتقاربان ك-ق	1	+		-

م	الحكم	عدد الأمثلة	الباحث الأول	الباحث الثاني	الباحث الثالث
31	متجانسان صغير ت - د	1	+		-
32	متجانسان ت - ط	1	+	-	--
33	متجانسان ث - ذ	1	+		+
34	متجانسان د ت	1	+		+
35	متجانسان ذ - ظ	1	+	+	+
36	متجانسان ب - م	1	+	-	+
37	متجانسان ط - ت	1	+	-	+
38	المد الأصلي	2	-	- +	- +
39	المد المتصل	2	- +	--	+ -
40	المد المنفصل	2	++	+	-
41	المد العارض للسكون	2	++	-	+
42	المد اللازم	1	+	+	+
43	القلقلة - ق	1	+	+	+
44	القلقلة ط	1	-	-	+
45	القلقة - ب	1	+	-	+
46	القلقلة - ج	1	+	+	+
47	القلقة - د	1	-	-	-
48	التفخيم - خ	1	-	-	+
49	التفخيم - ص	1	+	-	-
50	التفخيم - ض	1	-	+	-

م	الحكم	عدد الأمثلة	الباحث الأول	الباحث الثاني	الباحث الثالث
51	التفخيم - غ	1	+	-	+
52	التفخيم - ط	1	+	-	-
53	التفخيم - ق	1		-	-
54	التفخيم - ظ	1		-	-
55	أخطاء التشكيل	4	++++	++++	++++
56	تغيير الحرف المتقارب	10	-----	-----	-----

استخلاص النتائج لتطبيق (3) للباحثين الثلاثة.

1- تمت المقارنة على 56 خطأً، تمثل كل أحكام التجويد، ولكل خطأً مثال أو مثالان، أو أربعة أو عشرة أمثلة حسب تمثيله في القرآن الكريم بمجموع 100 مثال.

2- اتفق الباحثون في 10 أحكام فقط من مجموع المادة، وهي باللون الأخضر

3- اتفق باحثان في 23 حكماً من مجموع المادة، وهي باللون البني الفاتح.

4- اختلف الباحثون في تحكيم الأمثلة في 23 موضعاً.

5- أكثر الأحكام أو (الأخطاء المطلوب اكتشافها) التي اتفق فيها الباحثون هي: في أخطاء التشكيل، وتغيير الحروف المتقاربة.

وخلاصة ما توصلنا إليه في هذه المقارنة ما يأتي:

1- تطبيق ترتيب لا يهتم بأخطاء التلاوة مطلقاً واستخدامه لتقنيات الذكاء الاصطناعي مقتصر على البحث بالكلمة أو الجملة فقط.

- 2- تطبيق تسميع لا يهتم بأخطاء التلاوة، ويقتصر على بيان مواضع الخطأ على مستوى المفردة أو الجملة، ولا يكتشف حتى أخطاء التشكيل.
- 3- تطبيق المعلم يكتشف أخطاء التلاوة لكنه لا يصنفها ويبين نوع الخطأ، وكثيراً ما يصف الخطأ صواباً والصواب خطأً كما تبين في الرصد السابق.

خاتمة البحث

بعد استعراض هذه التجارب فإننا خلصنا إلى النتائج التالية:

- 1- التحقق من النطق الصحيح باستخدام تقنية computer-aided language pronunciation (CALP) تقنية فعّالة في تعليم النطق الصحيح بشكل عام، وتطبيقها على القرآن في اللغة العربية، يعد مستوى متقدماً، مقارنة بتعليم النطق الصحيح في اللغة العادية.
- 2- تعليم الحاسوب معيارية النطق الصحيح أسهل بكثير من تعليمه الأخطاء.
- 3- استخدام نموذج واحد في بناء نموذج لاكتشاف الأخطاء ليس مناسباً، بل كل تطبيق ينبغي أن يصمم لوحده نماذج متنوعة حتى يخرج بالدقة المناسبة؛ فتطبيق سالم الذي صدر هذا العام كان التعليم فيه على مستوى الحرف والكلمة والجملة من خلال التركيز على سورة الفاتحة وكانت الدقة عالية بسبب تصميم النماذج التي عرضناها في البحث.
- 4- مقارنة التطبيقات المشابهة أمر مهم للاستفادة من جهود الآخرين، وللبناء على الجديد الذي يتوصلون إليه.
- 5- إن هذا البحث خلاصة تجربة تزيد على 25 عامًا، وفيه سجلت آلاف الساعات كما أظهر في البحث، وهذه عملية مكلفة جداً، وتحتاج إلى وقت وجهد كبيرين، ويوصي البحث بأن تتبنى الجهات والمؤسسات مثل هذه المشروعات التي تعد أداة مهمة من أدوات التعليم والتعلم معاً.

مصطلحات البحث

التحقق من النطق الصحيح computer-aided language pronunciation (CALP): النطق اللغوي بمساعدة الكمبيوتر، وترتكز على التحقق من معيارية نطق المتحدث من حيث الصواب والخطأ لكلام نعرفه مسبقًا.

تقنية تعرف الصوت (Speech Recognition): تقنية تستطيع التعرف على صوت المتكلم وتحدد دلالة الكلام.

تعليم الآلة Machine Learning: وسيلة لتدريب الحاسوب على محاكاة العقل البشري ليؤدي وظائف الإنسان.

قائمة المراجع

أولاً: المراجع العربية:

- 1- ابن جني، أبو الفتح عثمان، الخصائص، تحقيق: محمد علي النجار، دار الكتب المصرية، د.ت، د.ط.
- 2- رشوان، محسن عبد الرازق، مدخل إلى حوسبة اللغة، ضمن كتاب مقدمة في حوسبة اللغة العربية، إصدار مركز الملك عبد الله لخدمة اللغة العربية، السعودية، 2019م.
- 3- عمايرة، خليل أحمد، في التحليل اللغوي، مكتبة المنار، الأردن، 1987م.

ثانياً: المراجع الأجنبية:

- 1- Abdou, S. M., Hamid, S. E., Rashwan, M., Samir, A., Abdel-Hamid, O., Shahin, M., & Nazih, W. (2006). *Computer aided pronunciation learning system using speech recognition techniques. In: Proceedings of the ninth international conference on spoken language processing.*
- 2- Juang, B. -H., & Furui, S. (2000). *Automatic recognition and understanding of spoken language - A first step toward natural human-machine communication. Proceedings of the IEEE.*



إشكالات المعالجة الآلية للغة العربية وآفاق تطويرها بنك المشجرات العربية أنموذجاً

**Problems of Automated Processing of the
Arabic Language and the Prospects
for its Development,
the Arabic Treebank as a Model**

د. سندس كرونتا

المعهد العالي للغات بتونس - جامعة قرطاج

الملخص

نهتم في هذه الورقة بتقديم بنك المشجرات العربية في جامعة بنسلفانيا Upenn Arabic Treebank تقديمًا نقديًا والنظر في إمكانيات تطويره. ويندرج هذا البحث في إطار سعينا إلى النقل المعرفي لمنهجية بنك المشجرات بجامعة بنسلفانيا؛ من أجل استثمارها في تطوير بنك المشجرات العربية وتجاوز نقائصه. والهدف الأسمى من ذلك تحقيق وسم نحوي آلي للغة العربية يراعي خصوصياتها اللسانية. ومن أبرز الإشكاليات التي نطرحها في هذا البحث: كيف يمكن تطبيق منهجية بنك المشجرات Treebank Methodology في الوسم النحوي الآلي للغة العربية مع مراعاة خصوصياتها اللسانية، كما قدّمت في نظرية النحو العربي من جهة، والكليات اللغوية التي اعتمدت في منهجية التحليل الآلي في جامعة بنسلفانيا من جهة ثانية؟ وإلى أي مدى تمكّنت منهجية بنك المشجرات العربية في صيغته الحالية من تحقيق هذه المعادلة واجتناب الاختزالية التي قد تتعارض مع الوصف اللساني الصحيح؟ وما هي سبل تطوير هذه الطريقة في المعالجة الآلية للغة العربية وإعادتها إلى موطنها الأصلي في مراكز البحث والجامعات العربية المهتمة باللسانيات الحاسوبية؟

يتكوّن البحث من ثلاثة أقسام: نمهد في القسم الأول بتقديم مشروع بنك المشجرات العربية تقديمًا موجزًا من حيث نشأته وأساسه النظرية والمنهجية. ثم نهتم في القسم الثاني بعرض بعض الإشكالات التي واجهها فريق بنك المشجرات العربية في مرحلته الأولى بسبب عدم الإلمام الدقيق بخصائص النحو العربي. ونختم عملنا بالإشارة إلى بعض الهنات التي مازالت قائمة في الصيغة الحالية لبنك المشجرات العربية، والتي يمكن تجاوزها بتطوير هذه المنهجية في مراكز

بحث وجامعات عربيّة مختصّة في المعالجة الآلية للغات.

يمثّل هذا البحث محاولة لتوطين منهجية بنك المشجرات بجامعة بنسلفانيا في البحث اللساني العربي في مجال المعالجة الآلية للغة العربيّة، وتطوير سبل رقميتها وتوظيف المدوّنات النصيّة العربيّة المحشّاة annotated في الترجمة الآليّة من اللغة العربيّة وإليها، واستخراج المعلومات وتطوير البرمجيات والتطبيقات المعتمدة في تعليم اللغة.

كلمات مفتاحيّة: بنك المشجرات العربيّة - مبادئ النحو العربي - منوال العمل والربط - تحشية - المحلّل الصرفي - المحلّل التركيبي .

Abstract

In this paper, we are aiming at making a critical presentation of the Penn Arabic Treebank and researching possibilities for its development. This research falls within the framework of our endeavor to transfer knowledge by investing the methodology of the Treebank at the University of Pennsylvania in developing the Arabic Treebank and overcoming its shortcomings to achieve an automated grammatical analysis of the Arabic language that considers its linguistic peculiarities. Among the most important problems that we will address are: How can the Treebank Methodology be applied, taking into consideration the peculiarities of the Arabic language as presented in the theory of Arabic grammar on the one hand, and the linguistic schools that were adopted in the automated analysis methodology at the University of Pennsylvania on the other hand? To what extent was the methodology of the Arab Treebank in its current form able to achieve this equation and avoid the reductionism that may conflict with the correct linguistic description of Arabic syntactic structures? What are the ways to develop the methodology of Arabic Treebank and return it to its original homeland in Arab research centers and universities interested in computational linguistics?

The research consists of three sections: In the first section, we briefly introduce the Arab Treebank project in terms of its creation circumstances and its theoretical and methodological backgrounds. Then, in the second section, we present some of the problems faced by the Arabic Treebank team in its first stage due to a lack of accurate

knowledge of the characteristics of Arabic grammar. We conclude our work by referring to some of the challenges that still exist in the current version of the Arabic Treebank, which can be overcome, and discussing the possibility of developing this methodology in Arab research centers and universities specialized in the field of Natural Language Processing.

This research represents an attempt to settle the Arab Treebank in Arabic linguistic research in the field of automatic processing of the Arabic language, developing ways to digitize it, and employing annotated Arabic text blogs in the field of machine translation, information extraction, software development and applications used in language education.

Keywords: Arabic Treebank - principles of Arabic grammar - Government and Binding model - annotation - morphological analyzer - syntactic analyzer.

المقدمة

لم يكن الاهتمام بالمعالجة الآلية للغة العربية في الجامعات والمراكز البحثية الأوروبية والأمريكية في الكثير من الأحيان نابعا من شواغل علمية صرفة. وإنما كان مرتبطا بتحقيق أهداف اقتصادية مثل: تطوير البرمجيات والتطبيقات الموجهة إلى العالم العربي، أو أهداف سياسية وعسكرية وأحيانا مخبرية. ولكن هذا لا يمنعنا من الاعتراف بأن هذه المراكز البحثية المتقدمة من حيث إمكانياتها التقنية والعلمية، قد أنجزت خطوات هامة في مسار المعالجة الآلية للغة العربية، وتحشية مدونات نصية ضخمة باللسان العربي (شفوية ومكتوبة) لتدريب الحاسوب على الترجمة الآلية من العربية وإليها، وتحليل النصوص والخطابات العربية تحليلا صرفيا وإعرابيا ودلاليا. ونعتقد أن مثل هذه المشاريع التي قد توقفت بعضها بسبب توقف التمويل من الجهات السياسية، أو التجارية التي كانت تمويلها، نظرا إلى توقف الحاجة الظرفية التي بعثت من أجلها، ينبغي أن تواصل وتطور في جامعات ومراكز بحث عربية، تكون غايتها الأساسية تنمية اللغة العربية واستعمالاتها الرقمية، وتطوير برامج الترجمة الآلية واستخراج المعلومات آليا من النصوص العربية information retrieval.

ويعد بنك المشجرات العربية بجامعة بنسلفانيا أحد أهم هذه المشاريع البحثية، التي اهتمت بتطوير آلة بناء مشجرات treeeditor لتحليل النصوص العربية المعاصرة، تحليلا نحويا من الناحية الصرفية والإعرابية. وقد أنجز هذا المشروع في مجمع المدونات اللسانية (LDC) Linguistic Data Consortium بجامعة بنسلفانيا بالولايات المتحدة الأمريكية (2001-2014). وكانت الباحثة قد شاركت في إنجاز هذا المشروع، وإعداد الدليلين التوجيهيين في التحليل الصرفي،

وفي التحليل الإعرابي بإشراف محمّد المعموري وأن بيز⁽¹⁾. ولعلّ أبرز الإشكالات التي واجهت هذا العمل على أهميته وضخامة المدوّنة النصيّة التي اشتغل عليها، هي عدم اعتماده منذ البداية على مبادئ النحو العربي، واستنساخه في الدليل التوجيهي الأوّل⁽²⁾ القواعد والأحكام النحويّة التي استعملت في بنك المشجرات الإنجليزيّة، مع إدخال بعض التعديلات الطفيفة. وهذا ما جعل الفريق يواجه جملة من الإشكالات في المعالجة الآلية للغة العربيّة، التي ترجع في أغلبها إلى عدم مراعاة خصوصيات اللغة العربيّة ومبادئها النحويّة المختلفة كليّاً عن مبادئ النحو الإنجليزي نظراً إلى تباعد اللسانين في تصنيف الألسنة البشريّة.

ستقوم الباحثة في قسم أوّل من هذا البحث بتقديم مشروع بنك المشجرات العربيّة، تقديمًا موجزًا من حيث نشأته، وأسسها النظرية والمنهجية. ثمّ تهتمّ في القسم الثاني بعرض بعض الإشكالات التي واجهها فريق بنك المشجرات العربيّة في مرحلته الأولى؛ بسبب عدم الإلمام الدقيق بخصائص النحو العربي، مقارنة بالأنحاء الأخرى، وخاصّة النحو الإنجليزي. وتختتم الباحثة عملها بذكر بعض الهنات التي مازالت حاضرة في الصيغة الحاليّة لبنك المشجرات العربيّة، والتي يمكن تجاوزها بإيجاد ضرب من الموازنة بين القواعد الضمنية في النظرية

-
- (1) Maamouri, M., Bies, A, Krouna, S., Gaddeche, B. and Bouziri. B., "Arabic Treebank Morphological and Syntactic Annotation Guidelines", 2008, in <http://projects.ldc.upenn.edu/ArabicTreebank/>, Linguistic Data Consortium, University of Pennsylvania.
- (2) Maamouri, M & Bies, A. « Developping an Arabic Treebank: Methods, Guidelines, Procedures and Tools", In Proceedings of Workshop on Computational Approaches to Arabic Script Based Language, p.p. 2-9, Geneva, Switzerland, COLING, 2004. In LDC Papers in <https://www.ldc.upenn.edu/language-resources/papers/ldc-papers#a2001>.

النحويّة العربيّة، ومقتضيات المبادئ اللسانية التوليدية التي تنطلق منها منهجيّة بنك المشجّرات في جامعة بنسلفانيا. وتأمل الباحثة أن تسهم مثل هذه المقترحات في تطوير هذه المنهجية في مراكز بحث وجامعات عربيّة مختصة في مجال اللسانيات الحاسوبية. وذلك لغاية تكوين منهجية قادرة على الوسم النحوي الآلي لمدوّنات نصيّة ضخمة شفويّة ومكتوبة، قديمة ومعاصرة، من أنماط خطاب مختلفة كالخطاب اليومي والخطاب الأدبي والخطاب الديني.

1- تقديم موجز لبنك المشجرات العربيّة

قدّمت الباحثة في مقال سابق مع مجموعة من الباحثين من جامعة القصيم، الوسم النحوي الآلي في منهجية بنك المشجرات العربيّة في جامعة بنسلفانيا، من حيث ظروف نشأته، والأسس النظرية والمنهجية التي قام عليها. وسنعود في هذه الفقرة إلى بعض ما جاء في هذا التقديم تمهيداً لعرضنا أهم إشكالات المعالجة الآلية للغة العربيّة في هذا المحلّ الآلي⁽¹⁾.

ارتبط الوسم النحوي للغات، بطرق التمثيل البصري للعلاقات النحويّة، الذي ظهر مع التوزيعيين، ثمّ تطوّر استعماله في المدرسة التوليدية. ويقوم على تحليل الجمل والمركّبات إلى مكّوناتها المباشرة في مستويات مختلفة. وقد أسهم تطوّر الذكاء الاصطناعي والمعالجة الآلية للغات الطبيعيّة، في تطوّر الوسم النحوي الآلي. فأنشئت مدوّنات نصيّة ضخمة مصحوبة بتحشية annotation، أي توصيف نحوي وصرفي ودلالي يُمكن "من تقسيم النصوص الخام المعتمدة إلى

(1) المجدوب، عز الدين وآخرون، "منهج الوسم النحوي الآلي للعربية في منهجية بنك المشجرات النحوية في جامعة بنسلفانيا"، مجلّة اللسانيات العربيّة، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربيّة، ع 9، ذو القعدة 1440 هـ/ يوليو 2019 م.

كلمات وتوزيعها إلى أقسام وضبط الوظائف النحوية التي تشغلها ووضع حدود بين الجمل. وهي خطوات أساسية للوصول إلى معاني النصوص ومضامينها⁽¹⁾. وقد أنشئ بنك المشجرات العربية باعتباره أحد هذه المدونات النصية المحشاة في مجمع المدونات اللسانية⁽²⁾ Linguistic Data Consortium منذ سنة 2001. واعتمد منهجية بنك المشجرات كما قدّمها ميتش ماركوس Mitch Marcus، الذي وضع أول محلّل حاسوبي للظواهر الصرفية، والتركيبية باعتماد هذه المنهجية. وقد طبّقت في البداية في تحليل اللغة الإنجليزية (1989-1996)، ثمّ تواصل المشروع، فشمّل السنة أخرى منها العربية والصينية والتشكيكية والكورية. إذ يفترض ماركوس أنّ منواله الحاسوبي صالح للتحليل الصرفي والتركيبى لعامة الألسنة على اختلاف خصائصها التركيبية والصرفية⁽³⁾ ولئن صرّح منظّرو هذه المنهجية باستنادهم إلى مبادئ النحو الكلي كما افترضها التوليديون في منوال العمل والربط (1981)⁽⁴⁾، فإنّهم قد انطلقوا في الواقع من

(1) المجدوب، "منهج الوسم النحوي الآلي"، ص 6.

(2) تأسس مجمع المدونات اللسانية LDC سنة 1992 بمقتضى منحة من وكالة البحث الخاصة بوزارة الدفاع الأمريكية (وكالة مشاريع البحوث المتقدمة في مجال الدفاع Defense advanced research projects agency (DARPA)). وكان المجمع مركزاً بحثياً تابعة لجامعة بنسلفانيا. من أهم أهدافه تيسير تبادل المدونات اللغوية المحوسبة وتوزيعها بين مؤسسات البحث على الصعيد العالمي وتذليل الصعوبات الفنية والقانونية التي تحول دون ذلك. وقد طوّر في البداية تحشية مدونات باللغة الإنجليزية ثم انتقل المجمع إلى معالجة بمدونات من لغات مختلفة أهمّها العربية والصينية والكورية والتشكيكية، وأعطى الأولوية للمنطوق في المعالجة فانتقل في مرحلة ثانية إلى دراسة بعض اللهجات العربية ووسمها صرفياً وتركيبياً ودلالياً.

(3) Marcus, Mitchell & al., "Building a large annotated corpus in English: The Penn Treebank", Computational Linguistics, Vol. 19, N°2, October 1993 (accessed 12/09/2022 in https://repository.upenn.edu/cis_reports/237)

(4) Chomsky, N., Lectures on Government and Binding: The Pisa Lectures, Walter de Gruyter, 1993, p.371.

تحليل اللسان الإنجليزي، ثم عمّموا طريقة التحليل المعتمدة في هذا اللسان في معالجتهم لبقية الألسنة الأخرى، مع إدخال بعض التعديلات الطفيفة التي لا تراعي بالضرورة الأنحاء الخاصة لهذه الألسنة. فجعلت العربية خاصّة في المرحلة الأولى من العمل، كما ذكرت سميّة المكي: "نقاد لتدعن حوسبياً إلى ما يسمح به المحلل الحاسوبي دون اعتبار الكليات اللسانية المسيّرة لنظام اللغة الطبيعية الحوسبي، إضافة إلى ذلك فقد أضحت تطوّع أساساً إلى الخصائص النحويّة المتحقّقة في الإنجليزية"⁽¹⁾. وهذا ما وُلد في رأينا إشكالات عديدة في معالجة اللغة العربيّة آلياً وتحليلها صرفياً وتركيبياً كما ستبين الباحثة في القسمين المواليين من هذا البحث.

1-1 خلفيات بنك المشجرات العربية النظرية ومبادئه المنهجية:

تستند منهجية بنك المشجرات في جامعة بنسلفانيا، كما قدّمنا إلى المنوال الحاسوبي الذي وضعه ميتش ماركوس، والذي يعتمد بدوره بمبادئ منوال العمل والربط أساساً نظرياً. ومن أهمّ مبادئ هذا المنوال التي وظّفها بنك المشجرات:

1- حوسبة النقل: تتمثل في ظهور مكّونات من الجملة في البنية السطحية، في مواضع غير مواضعها الأصليّة في البنية العميقة. فيُمثّل النقل بالإشارة إلى ارتفاع العنصر المنقول إلى موضعه الجديد، مع ترك أثر $trace$ (*T*) في موضعه الأصلي من البنية متقارن إحصائياً معه. واعتمد بنك المشجرات حوسبة النقل للربط بين العنصر المنقول وموضعه الأصلي، والتأشير لعلاقة التقارن الإحصائي بينهما كما هو مبين في المشجر أدناه.

(1) المكي، سميّة، "التشجير الآلي للجملة العربيّة بين الكليات والمقاييس (بنك المشجرات العربي نموذجاً)"، مجلّة أنساق، مج.2، ع.2، قسم العربيّة – كلية الآداب والعلوم بجامعة قرطاج، يونيو 2018، ص.185.

Al+jaliyd+u::the+ice+[def.nom.])الجليد(S (NP-TPC-1

::sayoTar+a::dominate _[verb]سيطر (VP

(NP-SBJ-1 *T*)

::EalaY::on/aboveعلى (PP-CLR

::Al+Turuq+i::the+roads)الطرق (NP

(1)الجليد سيطر على الطرق

2- حوسبة المزج التي تقوم على التوليف بين عنصرين لإنتاج عنصر ثالث (مثال

م.اس = {أداة تعريف، اسم}). وتتسم هذه العملية بخاصية التكرارية، التي

تميّز الألسنة البشرية. فاعتبرها تشومسكي "آلية ضرورية لتحقيق الخاصية

التكرارية"⁽²⁾. واعتمد بنك المشجّرات هذه الآلية في بناء المشجّر من أعلى إلى

أسفل على النحو التالي:

(م.اس (م.اس مقابلة

(م.اس كرة

(م.اس ال+قدم)

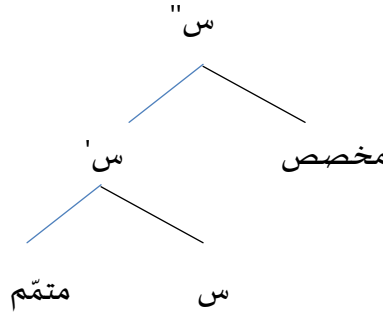
((م.وص. ال+نهائية))

(1) هذا المثال مأخوذ من الدليل التوجيهي التركيبي

Maamouri, M., Bies, A, Krouna, S., Gaddeche, B. and Bouziri. B., Arabic Treebank Morphological and Syntactic Annotation Guidelines, 2008.

(2) Chomsky, N., "On the Nature, Use, and Acquisition of Language". In: Ritchie, W.C. and Bhatia, T.K., Eds., Handbook of Child Language Acquisition, Academic Press, San Diego, CA, 1999, 33-54.

3- توظيف نظرية- س' (X-Bar Theory) في تمثيل الأبنية الإعرابية. وتمكن هذه النظرية من تمثيل موحد متكرر للعلاقات الإعرابية في المركبات وفق المخطط السلبي التالي:



وتتسم هذه البنية بخاصية التكرارية أيضا مما يُمكن من اختزال كبير في حوسبة التحليل التركيبي للجملة. ولئن وظّف بنك المشجرات العربية هذه البنية في تمثيل الجمل والمركبات، فإنّه ألغى الإسقاط الأوسط في المركّب الفعلي، وجعل الفعل والفاعل والمفعول، أي الرأس والمتّم والمخصّص، واقعة في نفس المستوى من التمثيل لتجنّب التعقيد في التشجير. وهذا ما يتعارض مع خاصية الهرمية والتكرارية اللتين تميّزان الأبنية الإعرابية في الألسنة البشرية⁽¹⁾.

وإلى جانب استناد المنهجية لما يعتبر كليّات لغوية عابرة للألسن، سعى فريق بنك المشجرات العربية إلى الاستئناس ببعض قواعد النحو العربي، في حدود ما توفرّ لديهم في بداية المشروع من معارف حول النظرية النحوية العربية. فاختاروا اعتماد ترتيب (فعل فاعل مفعول) باعتباره الترتيب الأساسي في الجملة العربية. فيكون توليد الفاعل في المشجر داخل المركّب الفعلي، ويرد مباشرة بعد الفعل العامل فيه. واعتبروا أنّ الفاعل إذا كان غير متحقّق معجميًا، أي ضميرًا مستترًا،

(1) المكي، سمية، التشجير الآلي للجملة العربية، 186-187

يمثل بمقولة فارغة معجميًا توافق ما يعبر عنه في منوال العمل والربط بـ "ضم صغير" (pro)⁽¹⁾.

ولكنّ الباحثة لاحظت عند التحاقها بالفريق تغليب الرافد اللساني التوليدي، الذي اعتمده بنك المشجّرات الإنجليزّة على الرافد النحوي العربي، في معالجة اللغة العربيّة. ولعلّ هذا كان من ضمن الأسباب التي تفسّر الصعوبات في المعالجة الصرفيّة والإعرابيّة التي واجهها الفريق بعد خمس سنوات من انطلاق المشروع، كما ستبيّن الباحثة في القسم الثاني من هذا المقال.

2-1 مراحل التحليل النحوي وبرمجياته

عُيّن بنك المشجّرات العربيّة في مرحلة أولى بتحليل مدوّنات نصيّة ضخمة من الصحافة المكتوبة، (الوكالة الفرنسيّة للأنباء AFP، جريدة الحياة وجريدة النهار اللبنانية)⁽²⁾ باعتبارها تمثّل نماذج من اللغة العربيّة المعياريّة الحديثة Modern Standard Arabic (MSA)⁽³⁾؛ ثمّ اتسع مجال تحشية المدوّنات العربيّة، فشمّل الصّحافة المرئيّة، والمسموعة، والتحاوّر الشفوي، باعتماد العربيّة المعياريّة، واللهجات العربيّة المحليّة، مثل: اللهجة الشاميّة، والعراقيّة، والمصريّة، والتونسيّة⁽⁴⁾.

(1) في مقابل ضم كبير PRO وهو الفاعل المتصل بالأفعال غير المصرفة infinitive form في بعض الألسنة
(2) يبلغ حجم هذه المدوّنات المكتوبة المحشّاة أكثر من نصف مليون كلمة (542.543): مدونة وكالة الأنباء الفرنسيّة (128810 كلمة) وجريدة الحياة (125698 كلمة) وجريدة النهار (293035 كلمة). انظر المجدوب وآخرون، الوسم النحوي الآلي، ص.15.

(3) Maamouri, M & Bies, A. « Developping an Arabic Treebank: Methods, Guidelines, Procedures and Tools », In Proceedings of Workshop on Computational Approaches to Arabic Script Based Language, p.p. 2-9, Geneva, Switzerland, COLING, 2004.

(4) Maamouri, M., "Developping and using a pilote dialectical arabic treebank", In ==

وتكون المعالجة الآلية لهذه النصوص في مرحلتين:

أ. مرحلة التحليل الصرفي

اعتمد فيها محلّل باكوولتر Backwalter Analyser، الذي أنشأه اللساني المختصّ في القواميس الإلكترونية، تيم باكوولتر Tim Buckwalter (BAMA). وقد أدخلت عليه العديد من التعديلات التي اقتضتها معالجة المدوّنات النصيّة في مشروع بنك المشجّرات العربيّة. فأصبح في صيغته المعدّلة ملكًا لمجمع المدوّنات اللسانية في جامعة بنسلفانيا (SAMA 2009). ويقدمّ هذا المحلّل كل الاحتمالات الممكنة لتحليل الكلمة صرفيًّا، ويضمّ كل تحليل صرفيّ جملة من المعطيات الصرفية والمعجميّة (انظر الصورة أسفل الجدول):

المعطيات الصرفية والمعجمية	مثال
قطعة كلاميّة مكتوبة باعتماد كتابة ألفبائيّة خاصّة بالمحلّل (Buckwalter Transliteration) مع تحديد الحركات المناسبة للسياق وللوضع الإعرابي للكلمة.	fymA
اللّمة أو المدخل المعجمي للكلمة	mA-1
الصّرافم المكوّنة للقطعة الكلاميّة (السوابق، الجذع، اللواحق)	fy + mA
قسم الكلم المناسب لكلّ جزء من القطعة الكلاميّة	fy/ PREP + mA/ REL-PRON
المقابل الإنجليزي الحرفي gloss	in, by + what, which

==

Proceedings of the Fifth International Conference On Language Resources and Evaluation (LREC'60), Genoa, Italy, 2006.

إشكالات المعالجة الآلية للغة العربية وآفاق تطويرها

أستاذ عوض كذلك لأسمع رأيك بدايةً فيما يتعلق بهذا اللقاء

File

Transliteration: flymā

Comment:

Annotation File: Y:\rachida\ATB5\ATB5-partc-pass1\xml\submitted-batch2\ALAM_VITHEVENT_ARB_20070206_205801.qtr.xml

Paragraph Number: 76 out of total 457 (word 7)

Selected Analysis: [flymā] [mā_1] fly/PREP+mā/REL_PRON : in, by + what/which

Formatted Selection: fly+mā PREP+REL_PRON in, by + what/which

flymā	SUB_CONJ	While, whilst, as, during
flymā	NOUN_PROP	FEMA (Federal Emergency Management Agency)
fly+mā	PREP+REL_PRON	in, by + what/which
fly+mā	PREP+INTERROG_PRON	in, by + what/which
fly+mā	PREP+SUB_CONJ	in, by + that/if/unless/whether/as long as/as soon as

X-Solution	Set Comment	No match (-)	Passive form (-)	-u	-N	-hi	NOUN -> ADJ	Number (-)	Punctuation (-)	
Left	Right	MISSING_BEFORE	MISSING_AFTER	Hamza problem (-)	-a	-F	WCE	ADJ -> NOUN	Clitic	Grammar Problem
Prev. Paragraph	Goto Paragraph	Next Paragraph	VOC_VAR	Gloss problem	-i	-K	NCE	NOUN -> Adverb	Dialectal	Remove Selection
Play Audio	Stop Audio	ChangeVowel	Partial	TRANSERR	Typo	Del CASE	DISFL	REG ACC	NoSpeech	FOREIGN

يقدم محلّل ساما، كما هو مبين في هذا المثال، كل الاحتمالات الممكنة لتقطيع الكلمة "فيما" إلى مورفيمات، وتحديد حركات كل مورفيم، وقسم الكلم الخاص به. ويقدم أيضًا ما قد يقابله في اللغة الإنجليزية. وتكون مهمة المحللين البشريين اختيار الوصف المناسب لسياق الكلمة الوارد في أعلى الشاشة. وهو في هذا المثال حرف جرّ + اسم موصول (PREP + REL_PRON). وتمثّل مخرجات التحليل الصرفي دخلًا للمحلّل التركيبي يمكن من وصف كلّ وحدة إعرابية بحسب طبيعتها الصرفية المعجميّة وموضعها في الجملة أو المركّب.

بد التحليل الإعرابي

اعتمد بنك المشجّرات في جامعة بنسلفانيا المحلّل التركيبي الآلي، الذي وضعه دان بيكال Dan Bickel's parser وقد واصل بنك المشجّرات العربيّة استعمال هذه الأداة في بناء المشجّرات لتمثيل الأبنية الإعرابية المكوّنة للوحدات الجمليّة. إن المدوّنة المحشّاة بالتحليل الصرفي للألفاظ تعالج بواسطة هذا المحلل الآلي الذي يُمكن من:

- بيان حدود الجمل المستقلة إعرابياً.
- تحديد المكونات المباشرة للجملة وللمركبات باعتماد وسم كل عقدة (node label) بحسب رأسها المعجمي (مركب اسمي، مركب حرفي، مركب وصفي...).
- تسمية الوظائف الإعرابية والدلالية لكل مكون؛ لتمييزه عن بقية المكونات التي في نفس المستوى (معمولات الفعل مثل الفاعل والمفعول به/ مخصّصاته الظرفية مثل الزمان، المكان، الهيئة، السبب...).
- التمييز بين علاقات الإتمام complementation، وتمثل بالتفرّع في المشجّر (motherhood)، وعلاقات الإلحاق أو التخصيص Adjunction التي تكون عناصرها في نفس المستوى تحت تحكّم عقدة أمّ واحدة (sisterhood).
- تحديد المقولات الفارغة الناتجة عن النقل (ICH، RNR، Trace) أو عن الضمير الصغير pro-drop (الضمير المستتر في العربية)، والذي يرمز إليه برمز *، والتأشير على التقارن الإحالي بين العنصر المنقول وموضع نقله في الجملة (indexation).

ويكون على المحلّلين البشريين تمثيل الجملة في شكل مشجّر تركيبّي، يعتمد آلية المزج من أعلى إلى أسفل، بحسب العلاقات الإعرابية بين مكونات الجملة، ووفق صيغ مختزلة من نظرية س' (كما بينا في الفقرة السابقة)⁽¹⁾. تعرض نفس المدونة في مرحلة أولى على محلّلين بشريّين مختلفين. ثمّ تعرض في مرحلة ثانية مخرجات التحليل الأوّل على محلّلين آخرين؛ لإصلاح ما قد يكون تسرّب في التحليل الأوّل من أخطاء. وفي مرحلة أخيرة، تعالج المدونة المحشاة تركيبياً

(1) لمزيد من التفاصيل انظر المكي، سمّة، التشجير الآلي للجملة العربية، 186-188؛ والمجدوب وآخرون، منهج الوسم النحوي الآلي.

بواسطة أداة مراقبة الجودة QC tool التي تستخرج آلياً الأخطاء المحتملة ويقوم أحد المحللين المدربين بإقرار هذه الأخطاء، أو تجاهلها إن لم تكن تمثل خطأ فعلياً في التحليل.

وكانت الغاية من هذه المراحل المتعددة تقديم تحليل تركيبى متجانس ومنسجم مع مبادئ التحليل، في منهجية بنك المشجرات العربية⁽¹⁾ حتى تكون ناجعة في تدريب المحلل التركيبى الآلى والاستغناء تدريجياً على التحليل البشرى.

وعلى الرغم من هذا الحرص على المراجعة والتثبت، واجه بنك المشجرات العربية جملة من الصعوبات منذ انطلاقه لأسباب عديدة؛ لعل أهمها ثراء اللغة العربية الصرفي والإعرابي، مقارنة بالألسنة الحديثة الأخرى مثل الإنجليزية. وتعمقت إشكالات التحليل الآلى للمدونات المحشاة بعد خمس سنوات من بداية المشروع (2001-2006)، مما كان يهدد استمراريته. وذلك بسبب تدني نسبة التوافق بين المحللين inter-annotators agreement التي كانت ضعيفة مقارنة ببنكي المشجرات الإنجليزية والصينية. ولكن ممّولى المشروع، ربّما نظراً إلى أهمية اللغة العربية الاستراتيجية بالنسبة إليهم، منحوا الفريق فرصة لمدة سنة لتحسين نسبة نجاح المحلل الآلى في بنك المشجرات العربية (f-scores). فعملنا في هذه السنة التي التحقت فيها بالفريق على مراجعة مبادئ التحليل الصرفي والإعرابي، التي كانت معتمدة بما يتماشى مع خصوصيات اللغة العربية ومقتضيات النحو العربي.

(1) Maamouri, M., Bies, A, Krouna, S. Gaddeche, B. and Bouziri. B., Arabic Treebank Morphological and Syntactic Annotation Guidelines, 2008, in <http://projects ldc.upenn.edu/ArabicTreebank/>, Linguistic Data Consortium, University of Pennsylvania.

2- إشكاليات المعالجة الآلية للغة العربية التي واجهها بنك المشجرات العربية:

1-2 صعوبات مرحلة التأسيس 2001-2006

واجه بنك المشجرات العربية، كما ذكرنا آنفاً، جملة من الإشكالات المتعلقة بالمعالجة الآلية للغة العربية، منذ بداية المشروع سنة 2001، حين حاول الفريق تطبيق المنهج الذي اعتمد في بنك المشجرات الإنكليزية، كما هو تقريباً، في التحليل الصرفي والتركيب للغة العربية. ويرجع ذلك إلى الاختلاف الكبير بين اللسان العربي، واللسان الإنكليزي في العديد من الوجوه الصرفية، والإعرابية. وقد أجمل محمد معموري، وأن ييز، هذه الصعوبات فيما يلي:

- عدم استعمال اللغة العربية المعيارية في التواصل اليومي التلقائي في المجتمعات العربية. وهذا ما يجعل حذق هذه اللغة، ومعرفة نحوها، وصرفها، معرفة دقيقة مقصورين على نخبة من المتكلمين بها. فكانت مهمة إيجاد محللين بشريين يمكنهم القيام بالتحشية الصرفية، والنحوية، على نحو تلقائي ومنسجم فيما بينهم أمراً في غاية الصعوبة.

- غياب الحركات وبعض الرموز المطبعية المهمة، مثل: التضعيف، وهمزة القطع في الكتابة العربية المعاصرة، المعتمدة في الصحافة المكتوبة التي اعتمدت مدونة للتحليل. وهذا ما يعيق قرائة النص العربي لدى غير المتمكنين من العربية المعيارية⁽¹⁾.

وما زاد من حدة هذه الصعوبات في المعالجة الآلية للغة العربية، نشأة هذا المشروع في غير موطنه الأصلي. وهو الجامعات العربية التي تزخر بالكفاءات العلمية في اللسانيات والنحو العربي، وفي علوم الحاسوب، والتي يمكنها تجاوز

(1) Maamouri, M & Bies, A, "Developing an Arabic Treebank", 1-2

الكثير من هذه المشاكل دون الحاجة إلى إسقاط المنوال الإنجليزي في تحليل الأبنية العربية. وهو الحل الأسهل الذي لجأ إليه الفريق، وكان سبباً فيما لوحظ من تدني نسبة التوافق بين المحللين البشريين، التي أثرت سلباً في نسبة نجاح المحلل الآلي، أو ما يسمى بـ f scores. فكانت هذه النسبة في بنك المشجرات العربية أقلّ بتسع نقاط من نظيره الصيني، وبأربع عشرة نقطة من بنك المشجرات الإنجليزية⁽¹⁾. وهو ما دفع مموّلي المشروع إلى حثّ فريق بنك المشجرات العربية على القيام بجملة من المراجعات لتحسين نسب دقة المحلل الآلي ورفع نسبة التوافق بين المحللين البشريين.

2-2 التعديلات التي أدخلت على المحلل الصرفي BAMA

لئن كان القاموس الإلكتروني لتيم باكولتر ثريًا، من جهة احتوائه على عدد هام من السوابق والجذوع واللواحق المعتمدة في اللسان العربي⁽²⁾، فإنّه كان في بداية استعماله في هذا المشروع يفتقر إلى الدقة في مستوى الوصف الصرفي، وتحديد أقسام الكلم. إذ اقتصر في البداية على أقسام الكلم الثلاثة: (الاسم، والفعل، والحرف)، ثمّ أضيفت إليه عند استعماله في بنك المشجرات العربية

(1) Maamouri, M., Bies, A., Kulick, S., "Enhancing the Arabic Treebank: A Collaborative Effort toward New Annotation Guidelines", In Conference Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco,

<https://dblp.uni-trier.de/db/conf/lrec/lrec2008.html#MaamouriBK08>, [accessed 16/09/2022].

(2) يتضمّن معجم باكولتر في صيغته 2.0 لسنة 2004 أي قبل مراجعته وتطويره 299 سابقة Prefix و 618 لاحقة suffix و 82158 جذعا (تمثّل 38600 لمة). راجع

Buckwalter, Tim, Buckwalter Morphological Analyser Version 2.0, in <https://catalog.ldc.upenn.edu/docs/LDC2004L02/> [accessed 15/09/2022]

سنة 2001 بعض التقسيمات الأخرى، التي استعملت في بنك المشجرات الإنجليزية من قبيل الظرف ADV، واسم الاستفهام INTEROG، والموصولات الحرفية SUB_CONJ، وغيرها⁽¹⁾.

ولكن هذه التفرعات لم تف بما يحتاجه التحليل الصرفي في اللغة العربية من تدقيقات دلالية، فضلاً عن الوقوع في بعض الأخطاء في تصنيف الألفاظ؛ بسبب التأثير بمقابلها الإنجليزي. ولعل من أبرز هذه الأخطاء التي وقع إصلاحها في النسخة المعدلة من المحلل الصرفي لتيم باكولتر.

- ضبط قائمة الحروف العاملة في الأسماء PREP المتكوّنة من حروف الجرّ، وحروف القسم، وحرف الاستثناء. وإخراج الظروف، وبعض الأسماء لازمة الإضافة، من قبيل: تحت، وقبل، وخارج، وأمام، وغير، وسوى، ومثل، ... من هذه القائمة التي كانت تصنّف ضمنها استناداً إلى تصنيف ما يقابلها في الإنجليزية.

- معالجة ظاهرة الاشتراك اللفظي في الكثير من الألفاظ العربية، التي يكون فيها لنفس الجذع لمات lemmas متعددة، مثل: اللفظ "ما" التي قد تكون حرفاً، أو اسماً، أو مشبهة بفعل ناقص، وتصنّف ما الحرفية حسب سياقها التركيبي، إلى: ما الموصولة، وما النافية، وما الزائدة، وتنقسم ما الاسمية إلى: اسم استفهام، واسم موصول، واسم شرط، واسم تعجب، واسم ميم. وتعامل ما المشبهة بليس، أو "ما الحجازية" معاملة الفعل الناقص، في مثل قوله تعالى: (مَا هَذَا بَشَرًا إِنْ هَذَا إِلَّا مَلَكٌ كَرِيمٌ) [يوسف: 31]. والأمثلة كثيرة على هذه الألفاظ، التي راجعنا تصنيفها في محلّل باكولتر. واقترحنا تفرع

(1) انظر

Buckwalter, "Tim. Buckwalter Arabic Morphological Analyzer Version 1.0 LDC2002L49". Web Download. Philadelphia: Linguistic Data Consortium, 2002.

اللمات بما يتناسب مع دورها الإعرابي، كما وصف في النحو العربي. وقد ساعدت هذه التصنيفات المحللين البشريين في مستوى التحليل الصرفي، على اختيار الحل المناسب في وصفهم للقطعة الكلامية في سياقها من الجملة. وهذا التدقيق الصرفي مكن من جهة ثانية من حل بعض المشكلات في مستوى التحليل الإعرابي؛ إذ يسهل اختيار قسم الكلم الدقيق عملية تحديد البنية الإعرابية، التي يقتضها اللفظ في الجملة أو المركب.

- تصنيف إن وأخواتها ضمن الحروف المشبهة بالأفعال PSEUDO_VERB، بعد أن كانت تلحق بالموصولات الحرفية SUB_CONJ على الرغم من أنها تتصدر جملا مستقلة وتعبّر عن الفعل الإنشائي للمتكلم.

لقد اكتفينا بذكر بعض الأمثلة من المراجعات، والتدقيقات التي أدخلناها على محلل باكولتر، عند التحاقنا بفريق بنك المشجرات العربية؛ لأنّ المجال ههنا لا يتسع لإدراجها كلّها⁽¹⁾. وقد ساهمت هذه التعديلات في ترفيع نسبة نجاح المحلل الصرفي SAMA⁽²⁾ بسبب التحسن الملحوظ في نسبة الاتفاق بين المحللين البشريين، الذين وجدوا فيها ما يقربهم من مبادئ الصرف العربي المشتركة بينهم.

(1) يمكن الرجوع إلى هذه المراجعات بالنظر في

Maamouri & al., "Arabic Treebank Morphological and Syntactic Annotation Guidelines", 2008.

Maamouri, M., Bies, A., Kulick, S., "Enhancing the Arabic Treebank".

(2) بعد إدخال هذه التعديلات اقتنى مجمع المدونات اللسانية محلل باكولتر BAMA وسوّى المحلل الصرفي للعربية المعيارية (SAMA) Standard Arabic Morphological Analyser.

2-3 المراجعات التي قدمت في التحليل التركيبي

من البديهي أن التعديلات التي أدخلناها على المحلل الصرفي ستُلقي بظلالها على التحليل التركيبي؛ لأنّ تغيير قسم الكلم (مثال تحوّل: فوق من حرف جرّ إلى اسم ظرف) سيؤدّي حتمًا إلى تغيير البنية الإعرابية التي ترد فيها تلك الكلمة (فوق الشجرة: مركّب بالجرّ سابقًا يحلّل مركّبًا إضافيًا بعد التعديل). والأمثلة كثيرة على هذه التعديلات التي أمكن القيام بها آليًا. ومن التغييرات المهمة في مستوى التحليل الصرفي، التي أثرت في المشجّرات التركيبية، وسم إنّ وأخواتها بالحروف المشبهة بالأفعال، مما جعلها تعامل في مستوى التركيب معاملة الفعل. فتكون رأس مركّب فعلي كما هو مبين في المثال التالي:

(ج م. ف إنّ)

(مبتدأ (الطقس))

(خبر (جميل))

فتكون بنيتها شبيهة ببنية الجملة الاسمية التي يتصدّرها الناسخ الفعلي. وإن كنّا نوافق سمية المكّي في أنّ هذا التحليل يبقى مشكلاً من الناحية اللسانية التوليديّة⁽¹⁾، فإنّه يظلّ الوصف الذي تتيحه منهجيّة بنك المشجّرات، ولا يتنافى مع التصوّر النحوي التقليدي لهذه الحروف عند النحاة العرب، الذين وسموها بـ"الحروف المشبهة بالأفعال". وهذا لا يمنع إمكانية التفكير في تمثيل أفضل لهذه البنية المشكلة في العربية، يستوعب في الآن نفسه التصوّر اللساني الحديث ومقتضيات الحوسبة الآليّة.

(1) اعتبرت سمية المكّي "إنّ" تعبر عن القوّة الإنشائية وتكون رأس طور، وتسقط في موقع المصدر لا في موقع الفعل، راجع المكّي، التشجير الآلي للجملة العربية، ص. 191.

وقد قدّمنا إلى جانب هذه التعديلات المرتبطة بمراجعة التحليل الصرفي في محلّ ساما، جملة من المراجعات الأخرى تتعلّق بطبيعة الأبنية الإعرابيّة في اللغة العربيّة، التي لم تراعى في بداية المشروع. ومن أهمّها نذكر مراجعة بنية الإضافة اللفظية التي يكون رأسها صفة. فقد وجدت الباحثة في الصيغة الأولى لبنك المشجّرات العربيّة، تحليلاً مزدوجاً لبنية الإضافة اللفظية بحسب الطبيعة المعجميّة لرأسها. فإذا كان الرأس مشتقاً مصدرًا مثل مقابلة الوزير، فإنّ المركّب يعدُّ مركّباً اسمياً إضافياً، ويمثّل بعلاقة الإتمام complementation

(م. اس (مقابلة)

(م. اس. (الوزير))

أمّا إذا كان الرأس صفة، فإنّه لا يعتبر رأس المركّب بل ملحقاً بالاسم الذي يرد بعده، حتّى وإن لم يكن هناك مطابقة في الجنس، والعدد، والإعراب، والتعريف والتنكير، بينهما التي تقتضيها علاقة النعتية في العربيّة.

وكان هذا التحليل الخاطئ للبنية الإعرابيّة، ناتجاً عن تأثر المحلّلين ببنية المركّب النعتي في اللسان الإنجليزي، التي يكون فيها النعت سابقاً للمنعوت. وهي بنية كما هو معروف لا تتحقّق في اللسان العربي، في المركّب النعتي. وكان سبب عدم قبول بنية الإضافة اللفظية في حالة المركّب الوصفي؛ هو رفضهم أن يكون الاسم متممًا للصفة؛ لأنّ مثل هذه البنية لا نجدها في اللغات الهندوأوروبيّة. وقد أثبت البحث في أصول بنية الإضافة اللفظيّة؛ أنّها من خصائص اللغات الساميّة، إذ توجد في العربيّة، وفي العبريّة القديمة. وقد أحدث قرار اعتماد بنية الإضافة في تحليل هذه المركّبات الوصفيّة تغييراً كبيراً في بنك المشجّرات العربيّة نظراً إلى كثرة استعمال الإضافة اللفظية من هذا الصنف في النصوص العربيّة. وحقّق تحسّناً في نسبة التوافق بين المحلّلين البشريين، الذي كانوا متردّدين بين بنية

النت التي فرضها عليهم الدليل التوجيهي الأول وبنية الإضافة التي يملها عليهم
حدهم اللغوي ونحوهم الضمني.

ومن التعديلات المهمة أيضًا التي تمكنت الباحثة من إقناع فريق بنك المشجرات
العربية باعتمادها في التحليل التركيبي، تغيير طريقة تمثيل المركبات شبه
الإسنادية، أي المركبات الاسمية التي يتصدرها مشتق يعمل عمل الفعل فيقتضي
معمولاته (فاعل، مفعول به، مفعول فيه ...). فأصبحت هذه المشتقات تعامل
معاملة الفعل في المشجر، وتكون رأسًا لمركب فعلي ضمن مركب جملي S-NOM
أو S-ADV أو صلة لمركب موصولي يكون فيه اسم الموصول محذوفًا SBAR-
NOM كما هو مبين في المشجرات التالية⁽¹⁾.

اختلاس (S-NOM (VP

(NP-SBJ) مسؤولين

(NP-OB) جزءا

(PP) من

(NP (NP هذا)

(NP) المبلغ))))))

اختلاس مسؤولين جزءًا من هذا المبلغ.

(S (VP - خلقت

(NP-SBJ-1 طائرة)

(1) هذه الأمثلة من الدليل التوجيهي التركيبي لبنك المشجرات العربية

Maamouri, et al., "Arabic Treebank Morphological and Syntactic Annotation Guidelines", 2008.

(NP-LOC) فَوْقَ

(NP) مَنَاطِقِ

(NP) الجَنُوبِ)))

(S-ADV (VP) مُتَّجِهَةً

(NP-SBJ-1 *)

(NP-DIR-CLR) شَمَالاً))))

حلقت طائرة فوق مناطق الجنوب متجهة شمالاً.

نشير في ختام هذه الفقرة إلى أنّ هذه التعديلات التي أدخلت على بنك المشجّرات العربيّة انعكست إيجابياً على التحليل الآلي للمدوّنة. إذ لوحظ تحسّن واضح في نسبة اتفاق المحلّلين البشريين سواء في مستوى التحليل الصرفي أو التحليل التركيبي. وهو ما رَفَع من قياس f من % 84.98 إلى % 94,30. وهي نسبة فاقت توقّعات الفريق الذي كان هدفه تحقيق نسبة نجاح تساوي نسبة بنك المشجّرات الصينيّة الذي وصل إلى حدود % 93,8⁽¹⁾.

ترجع هذه النتائج الإيجابية، إلى أنّ هذه المراجعات التي عرفها بنك المشجّرات العربيّة، كانت متماشية مع مبادئ النحو العربي، ومع خصوصيات اللغة العربيّة. فقد تفتّن أعضاء الفريق عند مراجعتهم لمواضع عدم الاتفاق بين المحلّلين البشريين، إلى أنّ السبب يعود إلى التزام بعضهم بتعليمات الدليل التوجيهي المستندة إلى النحو الإنجليزي، واستناد الآخرين إلى معرفتهم الضمنيّة أو المكتسبة للنحو العربي.⁽²⁾

(1) Maamouri, M., Bies, A., Kulick, S., Enhancing the Arabic Treebank, p.5

(2) Ibid., p.6

وقام الفريق بعد هذه المراجعة بإعادة تحليل أهم المدونات المحشاة في السنوات السابقة صرفيًا، وتركيبًا، بالاستناد إلى الدليلين التوجيهيين الجديدين، اللذين شاركنا في صياغتهما، وفي تدريب المحللين البشريين على استعمالهما في التحشية الصرفية والتركييبية. ثم انتقل اهتمام مموّلي المشروع إلى تحشية المدونات الشفوية، في بعض اللهجات العربية، مثل: اللهجة الشامية، والمصرية، والعراقية، والتونسية، والمغربية. وقد شاركت الباحثة في جزء من هذه الأعمال، مع الفريق الذي كوّنته في تونس (2008-2014). واعتمدت تقريبًا نفس منهجية التحليل التي توخّاها الفريق في تحشية المدونات النصية المكتوبة باللغة العربية المعيارية، مع إدخال بعض التعديلات المتعلقة بشكل خاص بطريقة الكتابة،⁽¹⁾ وبالتحليل الصرفي⁽²⁾. وتعتبر المعالجة الآلية للهجات العربية من البرامج البحثية المهمة، والواعدة في مجال اللسانيات الحاسوبية، لما فتحة من آفاق، وما تمثّله من تحديات هامة أمام عمليات تطوير رقمنة اللغة العربية، في تواصلنا اليومي على الشبكة العنكبوتية.

(1) اعتمدنا مواضع الكتابة اللهجية التي وضعها نزار حبش. انظر

Habash, N., Diab, M., Rambow, O., "Conventional Orthography for Dialectical Arabic", in Proceedings of the 8 th International Conference on Language Resources and Evaluation (LREC), European Language Resource Association, Turkey, 2012.

(2) للاطلاع على مزيد التفاصيل راجع

Maamouri, M., Krouna, S., Tabassi, D., Habash, N. & al., "Egyptian Arabic Morphological Annotation Guidelines", 2012,
https://www.researchgate.net/publication/331315455_Egyptian_Arabic_Morphological_Annotation_Guidelines [accessed in 17/09/2022].

3- قراءة نقدية في الصيغة الحالية لبنك المشجرات العربية

حقّق بنك المشجّرات العربيّة في صيغته الحاليّة، كما قدّمنا في الفقرات السابقة، نسب نجاح عالية في درجة اتفاق المحلّلين البشريين، والقدرة على الوسم النحوي الآلي للنصوص المكتوبة، والشفويّة، سواء في اللغة العربيّة المعيارية، أو في بعض اللهجات العربيّة. وهذا ما جعله مُعْتَمَدًا في مراكز بحث مختلفة حول العالم، ولدى باحثين كُثُر في مجال المعالجة الآلية للغة العربيّة. ولكنّه لم يخل من هنات على المستوى اللساني، في وصفه لأبنية اللغة العربيّة، وتفسير كيفية اشتغالها. وقد رصدت سميّة المكي⁽¹⁾ بعض الإخلالات، التي وقع فيها بنك المشجّرات العربيّة، في التحليل التركيبي خاصّة. ونذكر من هذه الهنات: عدم مراعاته لقيود النقل في النظرية التوليدية، وعدم استيعاب الخاصية التكرارية التي تميّز اللغة البشرية؛ مما أدّى إلى تمثيل الرؤوس، والمخصّصات، والمتّمات في نفس المستوى. وهذا يتعارض مع نظرية س' التي تعتمد الإسقاط الثنائي في مستويات تمثيلية مختلفة. وقد أرجعت الباحثة جلّ هذه الإخلالات في الوصف اللساني، إلى طبيعة "النظام الحاسوبي الآلي غير القادر على استيعاب النظرية اللسانية، وتشغيلها على نحو يعطي الأولويّة للتحليل اللساني الصحيح، والتفاعل الجادّ والمثمر مع المختصّين في اللسانيّات"⁽²⁾. وأكّدت أنّ ما عمّق هذا الإشكال هو "سعي بنك المشجّرات العربي، إلى تحقيق الاقتصاد الحوسبي، والتناسق مع بنك المشجّرات الإنجليزي، على حساب خصائص تشكّل البنية الإعرابية في اللغة عمومًا، وفي العربيّة على وجه

(1) المكي، سميّة، التشجير الآلي للجملة العربيّة، 194

(2) المكي، سميّة، التشجير الآلي للجملة العربيّة، 193

الخصوص⁽¹⁾. ولعلنا نوافق سميّة المكي، في أنّ المعالجة الآلية في بنك المشجّرات العربية، وغيره من أشكال التحليل الآلي للغات الطبيعية، تواجه تحدّيًا أساسيًا، وهو كيفية استيعاب النظام الحوسبي الصناعي للحوسبة النحوية الطبيعية. ولا يمكن مواجهة هذا التفاعل إلاّ باعتماد منهج التضافر المعرفي الفعلي، بين اللسانيات وعلم الحاسوب، والتفاعل الجاد بين المختصّين في هذين المجالين.

وقد كان هذا التفاعل قائمًا إلى حدّ ما في فريق بنك المشجّرات العربية، الذي كان يضمّ لسانيين وحاسوبيين. ولكنّ الهنات التي لاحظتها المكي في مقالها، وما سنستعرضه في بقية هذا البحث من إشكالات، يرجع إلى عدم معرفة دقيقة بالنحو العربي من جهة، وعدم مواكبة التطوّرات التي حصلت في النظريات اللسانية الحديثة من جهة ثانية. وتجزم الباحثة أنّ هذا المشروع لو كان قد أنجز في مركز بحث عربي، لما وقع في هذه الأخطاء في وصف الأبنية العربية وتمثيلها.

وستعرض الباحثة في هذه الفقرة، أمثلة من هذه الأخطاء لم تتعرّض إليها سميّة المكي في مقالها، والتي ترى أنّها تحتاج إلى مراجعة وتصحيح؛ لدعم الكفاية الوصفية لبنك المشجّرات العربية.

1-3 معالجة الجملة الاسميّة

اختار بنك المشجّرات العربية مبدأ أنّ الجملة في العربية تستند إلى ترتيب ف فامف (SVO). وهو اختيار يتماشى مع ما تتضمنه مبادئ النحو العربي، باعتبار أنّ الفعل هو عماد الجملة في العربية. وهو المولّد لبقية المواضع من خلال حركة العمل الإعرابي من العامل في اتجاه المعمول. ولكنّ تقسيم الجمل في بنك المشجّرات العربية ورد مختلفًا عن التصنيف المعتمد في النحو العربي. إذ اعتبروا

(1) م.ن، ص.ن.

أنَّ الجملة الاسميّة تقتصر على ما يسمّى "الجملة التعادليّة" (equational sentence) من قبيل:

السماء صافيّة

زيد في البيت

أي الجمل الاسميّة التي لا يتصدّر الخبر فيها فعل. أمّا الجمل التي يخبر فيها عن المبتدأ بالفعل، فقد اعتبرت جملاً فعليّة وقع فيها تحويل نقل الاسم إلى موقع المصدر، أو ما يسمّى بالموضعة Topicalisation. ويعدّ هذا النقل إجباريًا مهما كان موضع الاسم في الجملة التي يتصدّرها الفعل.

زيد₁ مرض₁ Ø

زيد₁ مرض أبوه₁

زيد₁ وجدت المجوهرات في بيته₁

فكلّ هذه الجمل تمثّل في المشجّر جملاً فعليّة تستند إلى بنية (ف فا مف)، وتقدّم الاسم فيها ليس إلا تحويلاً طارئاً على البنية العميقة يظهر في البنية السطحيّة.

هذا التحليل كما ذكرت سميّة المكيّ يخرق قيود النقل وخاصّة قيد التحتيّة⁽¹⁾ ولا يراعي المجالات التي يمتنع فيها نقل العنصر خارج حدودها، أو ما يسمّى بالجزيرة لدى روس Ross⁽²⁾. وهو يطرح من جهة نظريّة العمل في النحو العربي أيضاً جملة من التساؤلات. إذ إنّ عامل الرفع في "زيد" في الأمثلة السابقة

(1) انظر

Chomsky, N., Barriers, Cambridge MA, MIT Press, 1986a.

(2) Ross, J.R., Constraints on Variables in Syntax, thesis (Ph. D) at MIT, September 1967, in https://www.researchgate.net/publication/34708215_Constraints_on_variables_in_syntax_microform, [accessed in 19/19/2022].

هو حسب النحاة العرب الابتداء، وهو عامل معنوي يستدلّ عليه بغياب العوامل اللفظية⁽¹⁾. أمّا إذا اعتبرنا فرضية النقل والموضوعة، فإنّ "زيد" يكون في (أ) فاعلاً للفعل مرض (وهذا ما يتعارض مع مبدأ عدم عمل الفعل في فاعله إذا تقدّم عليه)، وفي (ب)، و(ج) يكون في موضع مضاف إليه مجرور (على الرغم من امتناع تقدّم المضاف إليه على المضاف في العربية). والملاحظ أنّ "زيد" يكون مرفوعاً في كلّ الحالات المذكورة.

فهذا الاختيار في التمثيل لا يراعي الحالات الإعرابية للاسم، ويخرق مبادئ حركة العمل الإعرابي، التي تقتضي أنّ الفاعل هو تمام الفعل والمضاف إليه هو تمام المضاف. فكلّ منهما بمثابة الكلمة الواحدة، على حدّ تعبير النحاة القدامى. وهذا ما يمنع تقدّم الثاني على الأوّل، إلّا إذا أخرج عن مجال عمله وصار تحت تحكّم عامل آخر، وهو الابتداء في حالتنا هذه. فالابتداء هو عامل الرفع في الأمثلة السابقة وغيرها من الجمل التي يتصدّرها اسم مرفوع، ويكون خبرها مركّباً إسنادياً.

وقد اقترحنا على الفريق تحليل هذه الجمل، باعتبارها جملاً اسمية يكون الخبر فيها جملياً S-PRD. ولكنّ الإشكال كان مصطلحياً بالأساس؛ لأنّ مصطلح المحمول predicate المنطقي المعتمد في الأنحاء الغربية، يختلف عن مصطلح الخبر ذي الأصول النحوية الدلالية الذي استعمله النحاة العرب. فالمحمول هو جزء من القضية، ولا يمكن أن يكون قضية، أمّا مصطلح الخبر، فهو كلّ ما يخبر به مهما كان شكله النحوي. ويمكن الإخبار في العربية بالجمل.

(1) انظر تعريف الابتداء باعتباره عاملاً معنوياً في الجرجاني، عبد القاهر، العوامل المنة النحوية في أصول علم العربية، شرح خالد بن عبد الله الأزهرى الجرجاوي، تج. محمّد عبد السلام شاهين، القاهرة، دار الكتب العلمية.

فكان تمسك الفريق بهذا التمثيل الذي تعتبره الباحثة غير مناسب للجملة العربية، لا يستند إلى أسباب لسانية حقيقية. وإنما يرجع إلى مبررات مصطلحية تتعلق بترجمة مصطلح الخبر بالمحمول predicate، التي قد لا تنقل كل خصائص مصطلح الخبر المفهومية. ويرجع الاختلاف بين المصطلحين إلى استناد الأنحاء الغربية إلى المنطق الأرسطي خلافاً للنحو العربي، الذي كانت بداياته سابقة لاطلاع العرب على منطق أرسطو عبر الترجمة السريانية. ونعتقد أنّ تعريب هذه المصطلحات يرفع هذا الحرج المصطلحي، ويمكننا من إدخال هذا التعديل على بنك المشجرات العربية في صيغته العربية دون أي إشكال.

2-3 اعتبار كان وليس فعلين مساعدين:

لاحظنا في بنك المشجرات العربية أنّ تمثيل الجملة التي تتصدرها كان، أو ليس، يكون بثلاثة طرق مختلفة بحسب الشكل النحوي للخبر:

كان زيدٌ مريضاً

(ج.م.ف. كان)

(م.اس.- مبتدأ زيدٌ)

((م.اس.- خبر مريضاً))

كان زيدٌ يضحكُ

(ج.م.ف. كان)

(م.اس.- 1 زيدٌ)

(م.ف. يضحكُ)

((((م.اس.- فاعل - 1*T*)))

(3) كانَ يضحكُ

(ج.م.ف.كان)

(م.ف.يضحك)

(م.اس.*(((

وهذا التعدد في تمثيل نفس البنية في الحقيقة لا موجب له، وإنما هو تعقيد يتعارض مع الاختزالية في التمثيل، التي تقتضيها منهجية التحليل الآلي في بنك المشجرات العربية. ويرجع اعتماد هذا الاختيار إلى سببين:

- ما ذكرناه من رفض مفهوم الخبر المركب الإسنادي S-PRD من الناحية النظرية في منهجية بنك المشجرات في الجمل من قبيل (2).

- التأثر بالأنحاء الغربية في اعتبار، كان، وليس، تتصرف إعرابياً مثل مقابلها الإنجليزي to be الذي يكون فعلاً مساعداً، يدخل على صيغة الفعل مباشرة، ويكون معه الصيغ التصريفية المركبة (present progressive، present perfect، past perfect، الخ).

وهذا يتعارض مع مبدأ اقتضاء الفعل في العربية للفاعل بعده في بنية (ف.فا.مف). ولئن كان الناسخ الفعلي "كان" لا يقتضي فاعلاً باعتباره فعلاً ناقصاً يدخل على الجملة الاسمية ويدلّ على الزمان، فإنه يعمل الرفع في اسمه، والنصب في خبره، ولا يدخل مباشرة على الفعل كما هو ممثّل في الجملة (3).

فقد كان يمكن أن تمثّل هذه الجمل دون تأويل نقل لا موجب له في مثل (2) و(3) على النحو التالي:

(ج.م.ف.كان)

(م.اس.- مبتدأ-1 زيد)

(مركب إسنادي- خبر (م.ف.يضحك)

(((((م.اس-فاعل-1)*)))

(ج.م.ف. كان

(م.اس.- مبتدأ-1*)

(م.إسنادي- خبر (م.ف. يضحك

(م.اس.- فاعل-1*)

ونشير ههنا إلى أن استعمال نفس المصطلح Subject للدلالة على المرفوعات بوظائفها المختلفة (الفاعل، المبتدأ، نائب الفاعل، اسم الناسخ الفعلي) يؤدي إلى خلط مصطلحي يمكن تجاوزه، بتعريب هذا المصطلح باعتماد المصطلحات المناسبة لكل وظيفة من هذه الوظائف التي تختلف في البنية الإعرابية التي تقتضيها. وهذا يمكننا من تمثيل هذه الجمل تمثيلاً متناسقاً داخلياً ومنسجماً مع النظرية النحوية العربية ومنهجية بنك المشجرات في مبادئها العامة.

نتبين حينئذ أن اعتبار 'كان' ومقابلها 'ليس' فعلين مساعدين، كان من الأخطاء التي لم يقع إصلاحها في بنك المشجرات العربية، على الرغم من أن بقية أخوات كان (أصبح، مازال، صار، ظل...) كانت تحلل على نحو يتوافق إلى حد ما مع كيفية تمثيلها في النحو العربي. فتكون متحركة في اسمها وخبرها، سواء أورد الخبر مفردة، أو مركباً اسمياً، أو حرفياً، أم ورد مركباً إسنادياً. وفي هذه الحالة يعتبر الخبر جملة متممة S complement. والسؤال المطروح هو: لماذا أُفردت كان وليس عن بقية عناصر مجموعتهما التي تتصرف مثلها تركيبياً؟ ولعلّ الجواب هو التأثير ببنك المشجرات الإنجليزي في تمثيله للفعل to be الذي يدخل على الفعل مباشرة ويكون فعلاً مساعداً (auxiliary verb)، ويكون مختلفاً عن تمثيل أفعال الحالة مثل remain, stay, become (التي تقارب في دلالتها أخوات كان، أصبح، وظلّ، وبات...) والتي تحلل بوصفها أفعالاً تامة.

3-3 عدم التناسق في معالجة المركبات شبه الإسنادية⁽¹⁾

إحتاج فريق بنك المشجرات العربية عند مراجعته لمبادئ التحليل سنة 2006، إلى إعادة النظر في بعض المركبات التي يتصدّرها مشتق يكون عاملاً عمل الفعل على الرغم من صيغته الاسمية. فأحدثت صفة صرفية في المحلل الصرفي ساما خاصّة بهذا النوع من الأسماء المشتقة وهي DV (deverbal). وأصبحت هذه الأسماء المشتقة توسم في المشجر باعتبارها رأس مركّب فعلي VP. ولكن الإشكال في التحليل الآلي لهذه المركبات، يكمن في تدريب المحللين البشريين على التمييز بين السياقات الفعلية والسياقات الاسمية، التي يمكن أن يرد فيها المشتق. فقد يكون المشتق في بعض السياقات التركيبية غير عامل، فيؤوّل بالاسم أو الصفة، ويكون رأس مركّب اسمي أو مركّب وصفي، أو يؤوّل تأويلاً فعلياً فيكون رأس مركّب فعلي.

مقابلة الوزير لن تفيدك : تأويل اسمي

مقابلة الوزير اليوم لن تفيدك : تأويل فعلي

فوضعت جملة من الشروط التي تحدّد السياقات التي يكون فيها هذا التأويل أو ذاك. وهي شروط شكلية لا تستجيب في بعض الأحيان لطبيعة المشتقات العاملة المتأرجحة بين الفعلية والاسمية. فاعتبار دخول الألف واللام على المشتق شرطاً مانعاً من أن يكون عاملاً عمل الفعل لا يفسّر التكافؤ الإعرابي والدلالي الذي نجده بين هذين القولين:

(1) يحيل هذا المصطلح المعتمد في نحو المركبات في الجامعة التونسية إلى المركبات الاسمية التي يتصدّرها مشتق (مصدر، اسم فاعل، صفة مشبهة الخ). يكون عاملاً عمل الفعل ويقضي معمولاته بعده.

أحبّ استمتاعك بالموسيقى في هدوء

أحبّ الاستمتاع بالموسيقى في هدوء

ففي الحالتين نعتبر أن المصدر "استمتع" مشتق عامل عمل فعله استمتع سواء أكان مضافاً أم معرفاً بالألف واللام. أمّا في بنك المشجّرات العربيّة، فإنّ المركّب الذي يتصدّره المشتق يحلّل في الجملة الأولى باعتباره جميلة مصدرية متمّمة للفعل S-NOM، فيما يعتبر في الجملة الثانية مركّباً اسمياً مفعولاً به. وهذا ما يخلق عدم انسجام في تحليل المركّبات شبه الإسناديّة التي لم يراع فيها محافظة المشتق العامل، على خصائص الاسم التصريفية، مثل: التعريف بالألف واللام، والتنوين، ودخول حروف الجرّ عليه، على الرغم من كونه يعمل عمل الفعل ويقتضي معمولاته.

ومن مظاهر عدم الانسجام أيضاً في تمثيل هذا الصنف من المركّبات، اقتصارهم في المشتقات العاملة على المصدر، واسم الفاعل، واسم المفعول. أمّا بقيّة المشتقات مثل الصفة المشبهة، وصيغة المبالغة، فهي لا يمكن اعتبارها مشتقات عاملة، حتّى وإن توقّرت فيها شروط العمل، من قبيل ورود اسم مرفوع، أو منصوب بعدها، مثل: "فلانٌ طاهرةٌ سريره، قليلٌ كلامه، قريبٌ من ربّه". فمثل هذه المركّبات لا تعتبر في بنك المشجّرات العربيّة، مركّبات شبه إسناديّة لأنّ المشتق فيها صفة، وليس اسم فاعل، أو اسم مفعول. وهذا الاختيار يعود أيضاً إلى تأثرهم بالنحو الإنجليزي، الذي لا تكون فيه الصفات Adjectives عاملة عمل الفعل. في حين أنّ العربيّة باعتبارها لغة اشتقاقية، يمكن للصفات المشتقة فيها سواء أكانت دالة على الحدوث والتجدّد، مثل: اسم الفاعل، واسم المفعول، أم دالة على الثبات والدوام، مثل الصفة المشبهة أن تكون عاملة عمل الفعل.

لقد أشرنا في هذه الفقرة إلى جملة من الهنات والأخطاء، التي مازالت قائمة في

بنك المشجّرات العربيّة في صيغته الحاليّة، على الرغم من المراجعات الهامّة التي أدخلناها عليه منذ سنة 2006. وتجدر الإشارة إلى وجود هنات أخرى في تمثيل الأبنية الإعرابيّة، لم نذكرها مكتفين بذكر أهمّ هذه الأخطاء من ناحية أثرها في التحليل اللساني الصحيح للجملة العربيّة. وقد لاحظنا أنّ أغلبها يعود إلى عدم المعرفة الدقيقة بالنظريّة النحويّة العربيّة القديمة وقراءاتها الحديثة، والتأثر الواضح ببنك المشجّرات الإنجليزي والنحو الإنجليزي بشكل عام. ولعلّ هذا الأمر كان مقصودًا باعتبار أنّ الغاية من منهجيّة بنك المشجّرات هو جعلها طريقة في التحليل الآلي عابرة للألسن. وذلك بالتركيز على الكليّات اللغويّة، والتقليل من أهميّة المقاييس، والسمات الخصوصية للألسنة المختلفة. ولكنّ هذا التمشّي ينعكس بالضرورة على سلامة الوصف اللساني للغة العربيّة، وخاصّة إذا وقع إسقاط خصائص نحو لسان آخر، هو اللسان الإنجليزي في هذا السياق في تحليل الأبنية العربيّة.

الخاتمة

خلاصة القول أنّ مشروع بنك المشجّرات العربيّة، على الرغم مما ذكرناه من نقائص، يعدّ من المشاريع المهمّة في الوسم النحوي الآلي للغة العربيّة. وقد اعتمدت المدوّنات النصيّة الضخمة التي وقّرها بنك المشجّرات العربيّة في بحوث ومشاريع لسانية حاسوبية في مختلف أنحاء العالم. وهذا يدلّ على الحاجة الملحة لمثل هذه المشاريع العلميّة، رغم حدود كفايتها الوصفية، والتفسيرية، في معالجة اللغة العربيّة آلياً وتحليلها صرفياً وإعرابياً ودلالياً. وقد استعملت مدوّنات بنك المشجّرات لإنشاء البنك القضوي Propbank، الذي يعنى بالمدى التركيبي الدلالي للأفعال. وظهرت بالتوازي مع بنك المشجّرات العربيّة بجامعة بنسلفانيا مشاريع بحثية أخرى، تعتمد منهجية التحليل المشجّري الآلي في معالجة الأبنية الإعرابية العربيّة. ومن أبرز هذه المشاريع التي استفادت من منهجية جامعة بنسلفانيا (وإن كانت تستند نظرياً إلى نحو التعلّق dependency grammar، لا إلى نحو المركّبات phrase structure grammar)، نذكر بنك مشجّرات التعلّق العربية ببراق Prague Arabic Dependency Treebank (PADT)⁽¹⁾ وبنك المشجّرات العربيّة بجامعة كولومبيا CATiB⁽²⁾. وقد يكون أحدث

(1) انظر

Jan Hajič, Otakar Smrž, Petr Zemánek, Jan Šnaidauf, and Emanuel Beška. "Prague Arabic Dependency Treebank: Development in Data and Tools". In Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools, pages 110-117, Cairo, Egypt, September 2004.

(2) انظر

Nizar Habash and Ryan M. Roth, "CATiB: The Columbia Arabic Treebank". In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 221–224. Association for Computational Linguistics.

هذه المشاريع بنك المشجرات الذي أنشأه مخبر المقاربات الحاسوبية في تنميط اللغة (CAMEL) Computational Approaches in Modeling Language بجامعة نيويورك بأبو ظبي. وقد سعى هذا المخبر بإشراف د. نزار حبش، الذي شارك في معالجة اللهجات في بنك المشجرات العربية في جامعة بنسلفانيا إلى بناء بنك مشجرات عربية يقوم على علاقات التعلق الكلية⁽¹⁾. ولا شك أن كل هذه المشاريع قد أسهمت في تطور المعالجة الآلية للغة العربية، وتحسين برمجيات الترجمة الآلية من اللغة العربية إليها.

ولكن هذا الوسم النحوي الآلي، على الرغم من نجاعته الحاسوبية، يحتاج كما بينت الباحثة في هذا المقال، إلى مراجعات من وجهة النظر النحوية، واللسانية العربية، من شأنها أن تحسن قدرة هذه المحللات الصرفية، والتركيبية الآلية، على وصف أبنية اللغة العربية، وتفسير اشتغال المعنى، وكيفية انبثائه في الخطاب العربي سواء أكان شفويًا أم مكتوبًا، معياريًا مشتركًا أم لهجيًا محليًا. والسبيل المثلى لتحقيق ذلك تكمن في توطين مثل هذه المشاريع البحثية في الجامعات، ومراكز البحث العربية، والاستفادة من الكفاءات اللسانية العربية في إنشاء بنك مشجرات عربية، يكون ذا قدرة وصفية وتفسيرية عالية تراعي خصوصيات اللغة العربية والمبادئ الكلية اللسانية في الآن نفسه.

وقد حاولنا من خلال قراءتنا النقدية لبنك المشجرات العربية في جامعة بنسلفانيا، الإشارة إلى السبل الممكنة لتطوير هذه المنهجية. وهو أمر يحتاج إلى

(1) لمزيد التفاصيل

Taji, D., Habash, N., Zemman, D., "Universal Dependencies for Arabic", in Proceeding of the 3rd Arabic Natural Language Processing Workshop (WANLP), Valencia, Spain, April, 3, 2017, p.p. 166-176

تضافر جهود اللسانيين، وعلماء الحاسوب العرب، وتفاعلهم تفاعلاً مثمرًا دون تغليب وجهة نظر الحاسوبي، على وجهة نظر اللساني، أو العكس. فاللسانيات الحاسوبية تعدُّ من الحقول المعرفية البينية، التي تقوم على التضافر المعرفي الحقيقي بين علمي الحاسوب واللسانيات.

قائمة المراجع

أولاً: المراجع العربية

- 1- الجرجاني، عبد القاهر، العوامل المنة النحوية في أصول علم العربية، شرح خالد بن عبد الله الأزهرى الجرجاوي، تح. محمد عبد السلام شاهين، القاهرة، دار الكتب العلمية، د.ت.
- 2- المجذوب، عز الدين وآخرون، "منهج الوسم النحوي الآلي للعربية في منهجية بنك المشجرات النحوية في جامعة بنسلفانيا"، مجلة اللسانيات العربية، مركز الملك عبدالله بن عبد العزيز الدولي لخدمة اللغة العربية، ع 9، ذو القعدة 1440 هـ/ يوليو 2019 م.
- 3- المكي، سميرة، "التشجير الآلي للجملة العربية بين الكلّيات والمقاييس (بنك المشجرات العربي نموذجاً)"، مجلة أنساق، مج.2، ع.2، قسم العربية - كلية الآداب والعلوم بجامعة قرطاج، يونيو 2018.

ثانياً: المراجع الأجنبية

- 1- Buckwalter, "Tim. 'Buckwalter Arabic Morphological Analyzer Version 1.0 LDC2002L49,' (Philadelphia, Linguistic Data Consortium, 2002). Available at <https://catalog.ldc.upenn.edu/LDC2004L02> Web Download (Accessed: 9 September 2022).
- 2- Buckwalter, Tim, *Buckwalter Morphological Analyser Version 2.0*, (Pensylvania, LDC, 2002) Availabe at <https://catalog.ldc.upenn.edu/docs/LDC2004L02/> (Accessed at 15 Septembre 2022)

- 3- Chomsky, N., *Barriers*, (Cambridge MA, MIT Press, 1986a).
- 4- Chomsky, N. 'On the Nature, Use, and Acquisition of Language'. In: W.C. Ritchie & T.K., Bhatia, (eds.) *Handbook of Child Language Acquisition* , (San Diego, CA, Academic Press, 1999), 33-54.
- 5- Habash,N., Diab, M., Rambow, O., 'Conventional Orthography for *Dialectical Arabic*', Proceedings of the 8 th International Conference on Language Resources and Evaluation (LREC), (Turkey, European Language Resource Association, 2012), Available at: http://www.lrec-conf.org/proceedings/lrec2012/pdf/579_Paper.pdf (Accessed: 10 September 2022).
- 6- Nizar Habash and Ryan M. Roth, "CATiB: The Columbia Arabic Treebank". In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, (Association for Computational Linguistics, 2009).
- 7- Jan Hajič, Otakar Smrž, Petr Zemánek, Jan Šnidauf, and Emanuel Beška. "Prague Arabic Dependency Treebank: Development in Data and Tools". In Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools, (Cairo, Egypt, September 2004).
- 8- Maamouri, M., 'Developing and using a pilote dialectical arabic treebank', Proceedings of the Fifth International Conference On Language Resources and Evaluation (LREC'60), (Genoa, Italy, 2006), Available at: https://www.researchgate.net/publication/271451888_Developing_an_Arabic_treebank (Accessed: 10 September 2022).
- 9- Maamouri, M & Bies, A., 'Developing an Arabic Treebank:

Methods, Guidelines, Procedures and Tools, In Proceedings of Workshop on Computational Approaches to Arabic Script Based Language, (Geneva, Switzerland, COLING, 2004) . p.p. 2-9, Available at:

<https://aclanthology.org/W04-1602> (Accessed at 10 September 2022).

- 10- Maamouri, M., Bies, A, Krouna, S., Gaddeche, B. and Bouziri. B., ‘*Arabic Treebank Morphological and Syntactic Annotation Guidelines*’ (Pensylvania, Linguistic Data Consortium, 2008). Available at <http://projects.ldc.upenn.edu/ArabicTreebank/>. (Accessed: 17 September, 2022).

- 11- Maamouri, M., Bies, A., Kulick, S. ‘*Enhancing the Arabic Treebank: A Collaborative Effort toward New Annotation Guidelines*’, Conference Proceedings of the International Conference on Language Resources and Evaluation, LREC, 26 May - 1 June 2008, (Marrakech, Morocco, 2008) Available at: <https://dblp.uni-trier.de/db/conf/lrec/lrec2008.html#MaamouriBK08>, (Accessed 16 September 2022).

- 12- Maamouri, M., Krouna, S., Tabassi, D., Habash, N. & al. ‘*Egyptian Arabic Morphological Annotation Guidelines*’, Available, (Pensylvania, Linguistic Data Consrtium, 2012) at: https://www.researchgate.net/publication/331315455_Egyptian_Arabic_Morphological_Annotation_Guidelines (Accessed: 17 September 2022].

- 13- Marcus, Mitchell & al ‘*Building a large annotated corpus in English:*

- The Penn Treebank*. Computational Linguistics, Vol. 19, N°2, October. (Pensylvania, 1993) Available at:
https://repository.upenn.edu/cis_reports/237 (Accessed: 12 September 2022).
- 14- Ross, J.R., Constraints on Variables in Syntax, thesis (Ph. D) at MIT, September 1967, in
https://www.researchgate.net/publication/34708215_Constraints_on_variables_in_syntax_microform, [accessed in 19/19/2022].
- 15-Taji, D., Habash, N., Zemman, D., *“Universal Dependencies for Arabic”*, in Proceeding of the 3rd Arabic Natural Language Processing Workshop (WANLP), (Valencia, Spain, April,3, 2017)



استكشاف أداء نموذج التميمي المحسن للتوسيم النحوي الآلي على سور من القرآن الكريم

**Exploring the Performance of the Enhanced
Tamimi's Model of the POS Tagging in a Sample
of the Holy Qur'an**

إعداد
د. أفراح التميمي
مدير قسم المدونات اللغوية
بمجمع الملك سلمان العالمي للغة العربية

الملخص

نموذج التمييزي المحسن للتوسيم النحوي، نموذج مطور من نموذج أولي يستعمل خوارزمية الحقول العشوائية المشروطة Conditional Random Fields (CRF)، إحدى خوارزميات تعلم الآلة الموجهة، المستعملة لبناء نماذج احتمالية تمييزية، تتنبأ بوسوم كلمات متتالية مع أخذ السياق بعين الاعتبار. وقد درب النموذج على مدونة من 60 ألف كلمة تقريبًا، موسمة يدويًا بأقسام الكلام. ويستعمل هذا الموسم النحوي مجموعات وسوم تنطلق من نظرية تمام حسان في تقسيم الكلام العربي. وتأتي مجموعة الوسوم الأساسية فيه على سبعة أقسام هي: الأسماء - الأفعال - الصفات - الضمائر - الظروف - الأدوات - الخوالف، بالإضافة إلى التقسيمات غير اللغوية، التي يستلزمها التحليل النصي الحاسوبي، وهي: علامات الترقيم - الاختصارات - الكلمات الأجنبية - الرموز. وتعتمد هذه الورقة على المنهج التجريبي لاستكشاف أداء نموذج التمييزي المحسن للتوسيم النحوي، بمجموعة الوسوم الأساسية على سور من القرآن الكريم؛ سعيًا لتوسيم القرآن الكريم به كاملاً. وأظهرت النتائج بعد توسيم تلك السور بهذا النموذج، ثم مراجعتها يدويًا، بأن الأفعال شكلت ما نسبته 18% تقريبًا من المدونة العينة، وأن النموذج نجح في توسيم 1846 فعلًا من 2081 فعلًا، محققًا نسبة من الصحة بلغت 89%. إن هذه النسبة قد تجعل من أهمية تحسين أداء وسم الأفعال حسب خصوصية القرآن الكريم، أولوية في مشاريع توسيم القرآن الكريم مستقبلاً، وفي تحسين أداء النموذج في جميع مجموعات وسومه.

الكلمات الدلالية: التوسيم النحوي؛ النموذج المحسن؛ مجموعة وسوم؛

تدريب؛ اختبار؛ خوارزمية؛ مدونة

Abstract

The enhanced Tamimi's model for grammatical tagging is an improved model from an initial model that uses Conditional Random Fields (CRF). It is one of the supervised machine learning algorithms used to build discrimination probability models that predict a sequence of words taking context into account. This model has been trained on a corpus that consists of approximately 60,000 tokens manually annotated with parts of speech. It uses tag-sets based on Tammam Hassan's theory of parts of speech. The basic tag-set is seven parts of speech: nouns, verbs, adjectives, pronouns, adverbs, particles, and exclamations, and non-linguistic categorization required by automated textual analysis: punctuation, abbreviations, foreign words, and symbols. This paper depends on the experimental method to explore the performance of the enhanced automated Tamimi's model in a sample of the Holy Quran; to tag the Holy Qur'an in its entirety with all tag-sets of the model. After tagging the sample corpus with this model and correcting manually, the results show that verbs constituted approximately 18 % and the model correctly tagged 1846 verbs from 2081, achieving a percentage of accuracy it stood at 89%. This percentage may make the importance of improving verb tags according to the specificity of the holy Qur'an a priority in future Qur'an tagging projects, and in improving model performance.

Keywords: POS tagging; enhanced model; tag set; training set; test set; algorithm; corpus

المقدمة

تعرف عملية التوسيم النحوي بالعملية التي تسند فيها الرموز، التي تشير إلى نوع الكلمة (فعل، واسم،... إلخ) لكل وحدة معجمية في النص⁽¹⁾. وتعرف هذه العملية أيضًا بالتوسيم النحوي Grammatical، أو التوسيم التركيبي الصرفي morpho-syntactic أو التوسيم بأقسام الكلام POS Tagging. وهي مختلفة عن التوسيم بالتحليل التركيبي Parsing الذي يستعمل التشجير والأقواس⁽²⁾.

ويعد التوسيم النحوي، أو التوسيم بأقسام الكلام الآلي، من أهم المهام في معالجة اللغات الطبيعية وأولها. وتكمن أهميته في أن دقته تؤثر كثيرًا على دقة المهام التالية له في عمليات المعالجة الحاسوبية لاعتمادها عليه. وقد بدأت أولى المحاولات في بناء أنظمة التوسيم النحوي للعربية منذ أوائل الألفية الثانية، رغم وجود محاولات سابقة غير مسجلة لشركتي صخر وزيروكس، وغيرهما من الشركات التجارية. إذ بنيت للعربية العديد من النماذج النحوية التي انطلق بعضها من نظرية لغوية عربية، وبعضها الآخر قدم مجموعات توسيمية مقترحة، أو اعتمد على مجموعة وسوم باكولتر المنطلقة من اللغة الإنجليزية⁽³⁾.

وتكمن أهمية هذه الورقة في أن دقة معظم هذه الموسومات النحوية كانت تتراوح ما بين 90% و98% في الغالب، وقيمتها لا يمكن أن تقاس بالدقة التقنية

(1) McEnery, T. et al. (2006) Corpus-based Language Studies: An advanced resource book. USA: Routledge. p. 34.

(2) McEnery, T. and Wilson, A. (2011) Corpus linguistics: An introduction. Edinburgh: Edinburgh University Press. p.46

(3) حبش، نزار، مقدمة في المعالجة الطبيعية للغة العربية، (ترجمة: هند الخليفة)، جامعة الملك سعود، الرياض، 2014، ص. 144.

التي تصل إليها وحسب؛ إذ لا بد من أن تقيم تقييماً يتضمن النظر في مجموعة الوسوم نفسها، بل وفي محتوى هذه الوسوم ونتائج النموذج نفسه على نصوص خارجة عن النصوص التي دربت عليه. ومن ثمَّ ليست الدقة التي يرصدها مطورو الموسومات، أو النماذج بعد اختبار أنظمتهم، هي المعيار الوحيد لاختيار الموسم المناسب للنص المراد توسيمه. فثمة معايير أخرى تحدد اختياراتنا للأدوات عند توسيم النصوص الهدف، ومنها ما أشار ليتش Leech حين ذكر أن قبولنا للتوسيم باعتباره وسيلة مفيدة وثرية بالمعلومات، ينبغي أن يتوقف على خبرات من أضافوا التوسيم للمدونة، أو مدى فائدة المقترح التوسيمي الذي تبناه أو اعتمده⁽¹⁾. كما أن ماننق Manning وجد أن الدقة تنقص بشكل ملحوظ إذا اختلفت طبيعة البيانات، أو اختلف أسلوب الكاتب فيما بين ما درب عليه الموسم وبين البيانات المراد توسيمها⁽²⁾.

ومن بين أنظمة التوسيم النحوية التي يمكن الوصول إليها، والاستفادة منها، نظام التميمي المحسن للتوسيم النحوي، الذي اختارته الباحثة؛ لانطلاقه من نظرية عربية حديثة في مجموعات وسومه، ولأنه قد درب على مدونة موسمة يدوياً، تنوعت أوعية نصوصها، وأزمنتها، وأمكنها، وضمت نزراً يسيراً من أي القرآن الكريم.

وإن كان هذا النظام قد حقق نسبة صحة بلغت 93,72 % في جميع أقسام الكلام، و92% في الأفعال على مدونة عامة قد تنوعت نصوصها وامتدت أزمنتها

(1) Garside, R. et al. (2013) Corpus annotation: Linguistic information from computer text corpora. USA: Routledge. p.6

(2) Manning, C. (2011) "International Conference on Intelligent Text Processing and Computational Linguistics," in Computational Linguistics and Intelligent Text Processing. Berlin, Heidelberg: Springer, pp. 171–189.

وأمكنها، أهدف في هذه الورقة لاستكشاف أدائه في سبع سور من القرآن الكريم فقط، وهذه المرة في الأفعال تحديداً؛ سعيًا لتوسيم مدونة القرآن الكريم به كاملة. وقد اتخذت فيه من التجريب منهجًا لاستكشاف أدائه، بمجموعة وسومه الأساسية فقط، متخذة من قائمة الأفعال في هذه السور معيارًا ذهبيًا gold standard لقياس الصحة accuracy، مجيبة عن السؤال الآتي: ما مستوى أداء نموذج التميمي المحسن للتوسيم النحوي الآلي على سور من القرآن الكريم؟

1- الأبحاث السابقة

على الرغم من أهمية التوسيم النحوي للنصوص الفصحى من عربية التراث، والنصوص الفصحى من العربية المعاصرة، إلا أن غالب الموسومات النحوية بنيت بالاعتماد على نصوص معاصرة، في مجال الصحافة، وتقدم موسماتها للعربية الفصحى المعاصرة، أو العربية الكلاسيكية. ويلاحظ في تلك الموسومات النحوية الآلية أنها تعطي نتائج متقاربة في الصحة (بين 90% إلى 98%)، رغم اختلاف منهجية التوسيم المتبعة، واختلاف خوارزمياتها⁽¹⁾، ذلك أن كل موسم منها اختبر على نفس نوع المدونات التي درب عليها. غير أن معظم الموسومات النحوية العربية بنيت بتدريها على نصوص من العربية الفصحى في مجال الصحافة، فتظهر معظمها نتائج سيئة عند تطبيقها على القرآن، أو على نصوص من التراث. وقد قارنت دراسة الربيعه⁽²⁾ بين نظامي: مدى، والخليل، المدرّبتين على نصوص عربية

(1) التميمي، أفراح، نظام آلي للتقطيع والتوسيم النحوي العربي، دار كنوز المعرفة، الأردن، 2020، ص. 121.

(2) Alrabiah, M. et al. (2014) "KSUCCA: A key to exploring Arabic historical linguistics," International Journal of Computational Linguistics (IJCL), 5(2), pp. 27–36.

فصيحة في مجال الصحافة؛ من أجل توسيم مدونة الذخيرة KSUCCA - ذات النصوص القديمة - بأحدهما، فأسفرت النتائج عن انخفاض في دقتهما بلغ من 10% إلى 15%، وذلك بالتجريب على خمس عينات من أوعية مختلفة من ذات المدونة. وركزت دراسة العصيمي وأتويل⁽¹⁾ على تقييم المحللات الصرفية، والموسمات النحوية، المتاحة للأغراض البحثية، والمصممة للغة العربية الفصحى الحديثة، أو القديمة، وقارنت بينهما. ولاحظ الباحثان انخفاضاً في الدقة والتغطية عند تطبيقها على عربية التراث. وقد ارتأيا أن يكون للغة التراث محللات صرفية، وموسمات نحوية خاصة، أو أن يحسن المتوفر منها حالياً والمدرّب على النصوص الصحفية فقط؛ ليدعم بعربية التراث. كما اقترحا حلاً آخر، وهو أن تدمج تلك الأنظمة وفق المنهج التجميعي ensemble method؛ لتزيد من كفاءة الأداء. أما عماد محمد⁽²⁾ فقد قام في دراسته بتقطيع مدونة دينية صغيرة الحجم، وتوسيمها بأقسام الكلام بطريقة شبه آلية. وهي مدونة تتكون من الكتيبات الآتية: "الأحاديث النووية"، و"متن أبي شجاع"، و"المنقذ من الضلال". وقد تضمنت 9000 كلمة تقريباً، درب نموذجها عليها، بالاعتماد على نفس مجموعة وسوم البنك الشجري العربي. وأظهرت النتائج أن أدواته في توسيم النصوص العربية التراثية، تفوقت على الأداة التي دربت على البنك الشجري العربي، على الرغم من أن البنك الشجري كان أكبر بإحدى وعشرين مرة. وقد استكشف أبو زينة، وعبد

(1) Alosaimy, A. and Atwell, E. (2017) "Tagging classical Arabic text using available morphological analysers and part of speech taggers," Journal for Language Technology and Computational Linguistics, 32(1), pp. 1–26.

(2) Mohamed, E. (2018) "Morphological segmentation and part-of-speech tagging for the Arabic heritage," ACM Transactions on Asian and Low-Resource Language Information Processing, 17(3), pp. 1–13

الباسط⁽¹⁾ أداء موسم ستانفورد لأقسام الكلام على عربية التراث، وهو الموسم المدرب على نصوص من العربية الفصيحة في مجال الصحافة. وفي هذه الدراسة قيّم الباحثان أداء ستانفورد في القرآن الكريم على أفعال الأمر فقط. وضمت مجموعة الاختبار 741 فعل أمر ظهرت في 1848 موضعاً في القرآن الكريم. وعلى الرغم من الدقة التي رُصدت عند بناء نظام ستانفورد التي بلغت 96.26% لكل الوسوم، حقق الموسم في أدائه على أفعال الأمر في القرآن دقة منخفضة جداً بلغت 7.28%. وخلصت الدراسة إلى أن انخفاض الأداء يعزز من الحاجة إلى المزيد من الأبحاث للكشف عن سبب عدم دقة الموسومات النحوية في النصوص التراثية، التي قد تكون بسبب عدم التمييز في التدريب بين العربية المعاصرة وعربية التراث. وبينت التمييزي⁽²⁾ واقع ثلاثة موسومات نحوية آلية، طبقت على مدونات عربية مختلفة لم تدرب عليها، وذلك من خلال موقع سكيث إنجن المضافة عليه. وقد استكشفت أدائها من خلال أول خمسين كلمة فقط من الكلمات الأكثر تكراراً في كل وسم لكل مجموعة. وكان أحد هذه الموسومات، موسم مدى MADA الذي طبق على مدونة الذخيرة KSUCCA⁽³⁾. وبعد التحقق اليدوي من صحة الخمسين كلمة الأولى الأكثر تكراراً في كل وسم فيه، والنظر في سياقات الكلمات الغامضة، وجدت التمييزي أن 988 كلمة، من 1460 كلمة، وُسّمت بوسم خاطئ، و472 كلمة وُسّمت بالوسم الصحيح، فجاءت نسبة الصحة 32%.

(1) AbuZeina, D. and Abdalbaset, T.M. (2019) "Exploring the performance of tagging for the classical and the modern standard arabic," Advances in Fuzzy Systems, 2019, pp. 1–10.

(2) التمييزي، نظام آلي للتقطيع والتوسيم النحوي العربي، صص. 120-166.

(3) Alrabiah, M. (2015) Building A Distributional Semantic Model for Traditional Arabic and Investigating its Novel Applications to The Holy Qur'an. dissertation.

أما هذه الورقة، فتستكشف سعيًا لتوسيم القرآن الكريم كاملاً أداء نظام التيمي المحسن للتوسيم النحوي، وبمجموعة وسومه الأساسية المنطلقة من نظرية تمام حسان في تقسيم الكلم⁽¹⁾. وهو نموذج درب على 60 ألف كلمة من نصوص من العربية الفصحى التراثية والمعاصرة. والنظر في هذا النموذج، قد يجعل من أهمية تحسين وسوم الأفعال حسب خصوصية القرآن الكريم، أولوية في مشاريع توسيم القرآن الكريم مستقبلاً.

2- الأدوات والبيانات

سأستعمل في هذه الورقة الأدوات والبيانات الآتية:

1-2 موسم التيمي النحوي المحسن

يستعمل موسم التيمي النحوي المحسن⁽²⁾ مجموعة وسوم⁽³⁾ تنطلق من نظرية تمام حسان في تقسيم الكلام العربي⁽⁴⁾، وهو تقسيم منطقي قابل للتمثيل الهيكلي، ويمكن من خلاله استقراء جوانب الكلمات، واستيعابها بحيث لا تفلت أيُّ منها من أيِّ قسم من أقسام الكلام، فضلاً عن إمكانية تكييفه مع تعليمات المجموعة الاستشارية الخبرة لمعايير هندسة اللغة EAGLES، الساعية لتوحيد معايير شكل ومحتوى التوسيم النحوي⁽⁵⁾. وتأتي مجموعة التيمي الأساسية في

(1) التيمي، أفرح، "نحو تحسين أداء نموذج التيمي للوسم النحوي الآلي للغة العربية"، المجلة الدولية للسانيات العربية، المغرب، ع 7، مج 1، 2021، ص. 61-79

(2) التيمي، "نحو تحسين أداء نموذج التيمي للوسم النحوي الآلي للغة العربية"، ص. 61-79

(3) التيمي، نظام آلي للتقطيع والتوسيم النحوي العربي، ص. 183.

(4) حسان، تمام، اللغة العربية، معناها ومبناها، دار الثقافة، المغرب، 1994، صص. 90-120.

(5) McEnery, T. and Wilson, A. (2011) Corpus linguistics: An introduction. p.38

المستوى الأول على سبعة أقسام، مع إمكانية توسيعها، وهي: الأسماء - الأفعال - الصفات - الضمائر - الظروف - الأدوات - الخوالب" بالإضافة إلى التقسيمات غير اللغوية التي يستلزمها التحليل النصي الحاسوبي، وهي: علامات الترقيم- الاختصارات- الكلمات الأجنبية- الرموز. ويظهر الجدول (1) الأقسام الكلامية الأساسية مع وسمها، بالإضافة إلى الأقسام غير اللغوية ووسومها.

جدول (1) الأقسام الكلامية الأساسية مع وسمها

القسم الكلامي الأساسي	وسمه
الاسم	N
الفعل	V
الصفة	A
الضمير	P
الأداة	RP
الخالفة	I
الظروف	D
علامات الترقيم	PUNC
الرموز	SYMB
ألفاظ أجنبية	FOREIGN
اختصارات	ABBREV

والنموذج المستعمل هو نموذج مطور من نموذج أولي درب على مدونة من 10 آلاف كلمة. ويستعمل خوارزمية الحقول العشوائية المشروطة Conditional Random Fields (CRF). وهي خوارزمية من خوارزميات تعلم الآلة الموجه

المستعملة؛ لبناء نماذج احتمالية تمييزية تتنبأ بوسوم كلمات متتالية، مع أخذ السياق بعين الاعتبار⁽¹⁾. وفي النموذج المحسن دربت الخوارزمية على مدونة من 40 ألف كلمة (مدونة التدريب)، التي تشكل 80% من المدونة الكاملة ذات الـ 60 ألف كلمة والموسومة نحويًا من خلال منهجية إعادة تدريب النموذج model retraining. وقد كانت نصوصها امتدادًا للإطار الذي اتخذ في النموذج الأولي، فضمت نصوصًا عربية من الفصحى، على امتداد عصورها وأزمنتها وأمكنها وموضوعاتها⁽²⁾. جرب النموذج المحسن على 20% من كامل المدونة (مدونة الاختبار)، وحقق درجة صحة بلغت 93,72%. وكان ذلك باستعمال مكتبة sklearn_crfsuite في البايتون، وبالإعدادات الافتراضية للخوارزمية، بالإضافة إلى استخلاص بعض الخصائص المتعلقة بالتوسيم، بعد تجهيز بيانات التدريب واختبار من خلال دالة (features). وقد كانت الخصائص مطبقة على كل كلمة داخل جملة، وهي: الكلمة نفسها، والكلمة السابقة، واللاحقة لها، والكلمة نفسها، وحروفها الثلاثة الأولى والأخيرة، وما إذا كانت الكلمة رقمًا أم لا، وما إذا كانت الكلمة في أول جملة أو نهايتها.

(1) Lafferty, J. D.; McCallum, A. & Pereira, F. C. N. (2001), Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in 'Proceedings of the Eighteenth International Conference on Machine Learning', Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 282-289.

(2) مدينة الملك عبد العزيز للعلوم والتقنية، المدونة اللغوية العربية، نسبة توزيع النصوص على الأوعية، متاح على الرابط (شاهد يوم 2022-09-09):

https://corpus.kacst.edu.sa/medium_chart.jsp

2-2 مدونة عينة خام لسور من القرآن الكريم

وتمثل هذه المدونة العينة سبع سور من القرآن الكريم بالرسم الإملائي، تقع في أربع أجزاء منه، وهن على التوالي: إبراهيم، والحجر، والنحل، والإسراء، والكهف، ومريم، وطه، وتضم 9572 كلمة. ويشير الجدول (2) إلى عدد الكلمات قبل المعالجة في كل سورة، مع مجموعها فيهن.

جدول (2) عدد الكلمات قبل المعالجة في كل سورة

رقم السورة	السورة	عدد الكلمات قبل المعالجة
14	إبراهيم	882
15	الحجر	756
16	النحل	1972
17	الإسراء	1669
18	الكهف	1693
19	مريم	1069
20	طه	1486
المجموع		9527

2-3 قائمة بالأفعال الواردة في السور السبع:

أُستعملت قائمة الأفعال بوصفها معيارًا ذهبيًا gold standard، وجمعت من السور السبع: إبراهيم، والحجر، والنحل، والإسراء، والكهف، ومريم، وطه، بالاستعانة بموقع مدونة القرآن الكريم، إذ يضع قاموسًا لكل الأفعال في القرآن الكريم، ويسردها مرتبة وفق الأجزاء في سياقاتها العبارية، ملونة باللون الأحمر،

ومربوطة برقم الجزء والآية⁽¹⁾. وقد ساعد هذا الأمر في استخلاصها، وفصل كل المباني التقسيمية عنها، وتجهيزها للاعتماد عليها في العمل بعد حذف كل ما لم يكن فعلاً حسب النظرية المنطلق منها في بناء النموذج المحسن، فحذفت من الأفعال كان وأخواتها، وكاد وأخواتها. وقد مثلت الأفعال ما يقارب 18% (2081 كلمة) من كامل المدونة. ويشير الجدول (3) إلى عدد الأفعال الفعلية في كل سورة، بجانب مجموعها في كامل المدونة العينة.

جدول (3) عدد الأفعال الفعلية في كل سورة

رقم السورة	السورة	عدد الأفعال الفعلية
14	إبراهيم	189
15	الحجر	141
16	النحل	401
17	الإسراء	362
18	الكهف	394
19	مريم	206
20	طه	388
المجموع		2081

(1) Dukes, Kais, The Quranic Arabic Corpus, available at (seen on 2022-09-09): <http://corpus.quran.com/morphologicalsearch.jsp>

4-2 مقياس الصحة accuracy

في تعلم الآلة، يتطلب قياس أداء أنظمة التوسيم النحوي وجود نسختين من مجموعة البيانات: بيانات اختبار خام، وبيانات اختبار موسومة نحويًا بدقة؛ تتخذ معيارًا ذهبيًا gold standard. فتتم مقارنة نتائج التوسيم النحوي الآلي لبيانات الاختبار بالتوسيم اليدوي لها. ولتعذر توفير المعيار الذهبي لكامل كلمات عينة الدراسة، الذي يمكننا من الكشف عن بقية مقاييس الأداء المتبعة في قياس أنظمة التوسيم النحوي؛ فإن مقياس الأداء في هذه الورقة هو مقياس الصحة، الذي يعرف بأنه: نسبة التنبؤات الصحيحة في بيانات الاختبار (قائمة الأفعال)، وحسب بسهولة بقسمة عدد التنبؤات الصحيحة على العدد المجمل للتنبؤات⁽¹⁾.

3- التطبيق

عولجت المدونة، العينة الخام المكونة من سور من القرآن كاملة، وهيئت تمهيدًا لتوسيمها آليًا بالموسم النحوي بمجموعة وسوم التمييز الأساسية للكشف عن أداء النموذج المحسن في الأفعال. وقد مرت بالمراحل الآتية:

1-3 التفريق tokenization

وفيهما استعنتُ بأداة المشذب العربي⁽²⁾ لحذف الرموز وعلامات التشكيل، مع

(1) Paroubek, P. (2007) "Evaluating Part-of-Speech Tagging and Parsing," in L. Dybkjær, H. Hemsén, and W. Minker (eds) Evaluation of Text and Speech Systems. Dordrecht: Springer, pp. 99–124.

(2) مدينة الملك عبد العزيز للعلوم والتقنية، المدونة اللغوية العربية، أداة لمعالجة المدونات العربية، متاح على الرابط (شاهد يوم 2022-09-09):

<https://sourceforge.net/projects/ghawwasv4/files/?source=navbar>

الاحتفاظ بالأعداد الرقمية؛ لخلو النص القرآني من علامات الترقيم التي تساعد في الفصل بين الجمل، واعتمادي على الأرقام في الفصل بينها. فأصبحت مثلاً الآية الكريمة من سورة الكهف: وَرَبَطْنَا عَلَى قُلُوبِهِمْ إِذْ قَامُوا فَقَالُوا رَبُّنَا رَبُّ السَّمَوَاتِ وَالْأَرْضِ لَنْ نَدْعُو مِنْ دُونِهِ إِلَهًا لَقَدْ قُلْنَا إِذَا شَطَطًا (14) بعد التفريق بهذه الصورة:

"وربطنا على قلوبهم إذ قاموا فقالوا ربنا رب السماوات والأرض لن ندعو من دونه إلها لقد قلنا إذا شططا 14"

2-3 التقطيع segmentation

في هذه المرحلة استعملت المقطع⁽¹⁾ المتوافق مع نموذج التميمي المحسن، الذي يفصل متغيرات التقطيع التي تقتضيها عملية التوسيم النحوي به⁽²⁾، وهي كما تظهر في الجدول (4). ومن ثَمَّ فإن الآية المذكورة سابقاً بعد التقطيع تكون بعد التقطيع بهذه الحالة:

وربطنا على قلوبهم إذ قاموا ف قالوا ربنا رب السماوات و الأرض لن ندعو من دونه إلها ل قد قلنا إذا شططا 14.

وكما نلاحظ في الآية الكريمة بعد التقطيع، فصلت المباني التقسيمية التي بني على أساسها الموسم النحوي، كالواو والفاء واللام وذلك بوصفها أدوات. وقد راجعت أدائه على النص القرآني، وصححت ما أخطأ فيه يدوياً، وظهرت الحاجة أيضاً بعد التقطيع إلى عملية التسوية الهجائية في بعض الكلمات، التي قد أخطأ فيها المقطع بسبب طبيعة الكلمات العربية، التي يتغير شكلها الكتابي بعد فصلها.

(1) التميمي، نظام آلي للتقطيع والتوسيم النحوي العربي، صص. 238-243

(2) التميمي، "نحو تحسين أداء نموذج التميمي للوسم النحوي الآلي للغة العربية"، صص. 61-79

فكلمة مثل: مَمَّنْ، أصبحت بعد التقطيع (م من)، وكان لا بد من إعادة (م) لشكلها الطبيعي قبل إدغامها في كلمة (من) لتكون كلمتين منفصلتين ومستقلتين، على النحو التالي: (مِنْ مَنْ)، وكلمات أخرى أيضاً، كالكلمات المعرفة بأل التي تتصل بها الأداة (ل)، نحو: للمؤمنين، وحذفت ألفها.

جدول (4) متغيرات التقطيع

المتغير	مثاله
همزة الاستفهام والتسوية (أ)	أيرضيك ذلك؟ - "وسواء عليهم أأنذرتهم أم لم تنذرهم"
حرف الجر والقسم (ب)	ببيتك - بالله
تاء القسم (ت)	تالله
حرف الاستقبال (س)	سيزور
عن المدغمة نونها في ميم من وما (ع)	عمّن - عمّا - عم
فاء الشرطية والسببية والاستئنافية والعاطفة (ف)	إذا مررت فسلم عليهم - ولا يؤذن لهم فيعتذرون - فتعالى الله عما يشركون - دخل محمد فخالد
حرف الجر (ك)	كالقمر
حرف الجر ولام الأمر ولام التأكيد (ل)	للمنزل - لينته عنه - إن في ذلك لعبرة
من المدغمة نونها في ميم من وما (م)	ممن - مما
هاء التنبيه (ها - ه)	أيها - هأنت
حرف العطف والمعية والحال والقسم (و)	كبر وعظم - سرت والليل - مضيت وأنا سعيد - والله
من الجارة المدغمة ميمها في ميم الاستفهام (م)	مم

المتغير	مثاله
في الجارة المتصلة بمن (في)	فيمن
حروف الجر المتصلة بما الاستفهامية المحذوفة الألف (ب-في-ل- على-إلى- حتى)	بم - فيم-لم-علام-إلام-حتام
إذ الظرفية المتصلة ببعض الكلمات	حينئذ-عندئذ

3-3 التوسيم tagging

بعد تفريق المدونة، وتقطيعها أصبحت الكلمات الفعلية tokens، في مدونة القرآن العينة 11600 كلمة. وفي هذه المرحلة وسمت المدونة العينة آلياً بالنموذج المحسن باستعمال وسومه الأساسية. وهكذا ظهرت نتيجة التوسيم كما يأتي في الآية الكريمة السابقة الذكر:

و_ربطنا V_على RP_قلوبهم N_إذ D_قاموا V_ف RP_قالوا V_ربنا N_رب N_السموات N_و RP_الأرض N_لن RP_ندعو V_من RP_دونه N_إلها N_ل RP_قد RP_قلنا V_إذا D_شططا N_14 DIGIT

ثم حددت جميع الأفعال فيها يدوياً في كل سورة على حدة، حسب قائمة الأفعال المعالجة تفريقاً، وتقطيعاً، وحذفاً لما لا يدخل في الأفعال، وفق نظرية تمام حسان في تقسيم الكلم. ويظهر الفعل مرمّزاً وفق مجموعة وسوم التمييزي بالرمز V، وقد وقعت أزمنة الأفعال في مجموعة التمييزي الأساسي تحت هذا الوسم. بعد ذلك حصرت جميع الكلمات التي وسمت فعلاً، وقد تبين أن هناك ما وسم فعلاً وهو ليس كذلك، فاستبعد من النظر. ويظهر الجدول (5) عدد الكلمات الفعلية في المدونة العينة في كل سورة بعد تفريقها وتقطيعها، بإزاء عدد الأفعال الفعلي في كل سورة، ومجموع ما وسم فعلاً سواء كان فعلاً أم لم يكن.

جدول (5) بيانات المدونة العينة بعد التفريق والتقطيع والتوسيم

رقم السورة	السورة	عدد الكلمات الفعلية tokens	عدد الأفعال الفعلي في السورة	مجموع ما وسم فعلا
14	إبراهيم	1091	189	175
15	الحجر	922	141	128
16	النحل	2423	401	374
17	الإسراء	2037	362	320
18	الكهف	2031	394	362
19	مريم	1287	206	203
20	طه	1809	388	352
	المجموع	11600	2081	1914

4- النتائج التجريبية

لقد بينت دراسة الربيعة انخفاض أداء المحللين: مدى، والخليل، في النصوص التراثية⁽¹⁾، وكشف بحث العصيمي، وأتويل، عن انخفاض أداء أكثر من موسم على العربية التراثية⁽²⁾، وتوصل عماد محمد إلى أن أدواته المبنية على مدونة تراثية صغيرة الحجم، كانت أجود أداء على النصوص التراثية من الأداة

(1) Alrabiah, M. et al. (2014) "KSUCCA: A key to exploring Arabic historical linguistics," International Journal of Computational Linguistics (IJCL), 5(2), pp. 27–36.

(2) Alosaimy, A. and Atwell, E. (2017) "Tagging classical Arabic text using available morphological analysers and part of speech taggers," Journal for Language Technology and Computational Linguistics, 32(1), pp. 1–26.

المدرّبة على نصوص صحفية تمثل أضعاف حجم مدونته التراثية⁽¹⁾، ويُنَّ أبو زينة، وعبد الباسط، تدهور أداء ستانفورد في توسيم أفعال الأمر في القرآن الكريم⁽²⁾، وكشفت دراسة التميمي، عن ضعف موسم مدى، في توسيمه لمدونة الربيعية التراثية⁽³⁾. وحيث إنَّي أسعى لتوسيم القرآن الكريم بهذا النموذج المحسن، تبين لي بعد النظر في قائمة الأفعال الواردة في المدونة العينة، التي تضم سبع سور من القرآن الكريم، وبعد مقارنة هذه الأفعال بما نتج من عملية التوسيم الآلي للأفعال بالنموذج المحسن، تبين أن 1846 من مجموع الأفعال وسم فعلاً، و235 وسمت خطأً بوسوم أخرى، من وسوم أقسام الكلام، تنوعت بين الاسم، والصفة، والأداة. وقد أخذ الاسم فيها الحيز الأكبر؛ حيث وسم به خطأً 214 فعلاً، كما وسم 12 فعلاً بوسم الصفة، ووسمت 9 أفعال بوسم الأدوات. فحقق النموذج درجة صحة بلغت 88.70% وهي نسبة أقل إذا ما قارناها بأداء النموذج نفسه في الأفعال على نصوص متنوعة من العربية الفصحى، حيث بلغت دقته في تعيين الأفعال على نفس نوع المدونة التي درب عليها 92%.

ويبين الجدول (6) المعلومات المتصلة بالأفعال الواردة في المدونة بعد التوسيم، وأعداد الوسوم الصحيحة في الأفعال في كل سورة بجانب الوسوم الخاطئة فيها، ونسب الصحة أيضاً التي حسبت بتقسيم عدد النتائج الصحيحة

(1) Mohamed, E. (2018) "Morphological segmentation and part-of-speech tagging for the Arabic heritage," ACM Transactions on Asian and Low-Resource Language Information Processing, 17(3), pp. 1–13.

(2) AbuZeina, D. and Abdalbasat, T.M. (2019) "Exploring the performance of tagging for the classical and the modern standard arabic," Advances in Fuzzy Systems, 2019, pp. 1–10.

(3) التميمي، أفرح، نظام آلي للتقطيع والتوسيم النحوي العربي، 2020

على مجموع الأفعال وضربها في 100. ويلاحظ تقارب أداء النموذج في السور جميعاً حيث تراوحت نسبة الصحة بين 84 % و 88% وذلك قد يكون إشارة إلى ثبات الأداء بين سور القرآن الكريم، أو أن ذلك يعود لكون العينة من السور المكية.

لقد تضمنت المدونة العينة الموسومة جميع أقسام الكلام، لكنها خلت من علامات الترقيم، التي تخلو منها نصوص القرآن بطبيعتها، وهذا يعني صفريتها في الاحتمالات الانتقالية عند تطبيق النموذج، الأمر الذي سيؤثر بلا شك على أداء النموذج، خصوصاً أن أوزان علامات الترقيم مع الأفعال في الاحتمالات الانتقالية transition probabilities⁽¹⁾ من الاحتمالات الانتقالية الأكبر وزناً في النموذج المحسن. وفضلاً عن ذلك، فقد درب النموذج المحسن على نصوص من جمل مقطعة، وتلكم بمثابة المعلومات الإضافية التي تعين على التنبؤ بأقسام الكلام في النموذج. غير أنه في مرحلة تقطيع المدونة العينة في الدراسة لم تكن الجمل قد فصلت عن بعضها، وإنما كان التقطيع للكلمات ذات المباني التقسيمية فقط بمتغيرات التقطيع المشار إليها سابقاً، وقد يكون لذلك أثر على أداء النموذج في نصوص القرآن الكريم.

(1) يقصد بالاحتمالات الانتقالية: "احتمالية أن يلي وسوما معينة من وسوم أقسام الكلام وسوما أخرى في نص متصل. وعادة ما يحدد ذلك بالتحليل الإحصائي للمدونات الموسومة. فعلى سبيل المثال، إذا كانت هناك 50 ألف كلمة فعلية في مدونة ما موسومة باعتبارها اسم، و30 ألف كلمة منها يلها فعل، فإن الاحتمال الانتقالي لاسم ← فعل هو 60 %"، انظر: التميمي، "نحو تحسين أداء نموذج التميمي للوسم النحوي الآلي للغة العربية"، صص. 61-79.

جدول (6) الإحصاءات المتعلقة بالأفعال في المدونة العينة بعد التوسيم

نسبة الصحة	الوسوم الصحيحة في الأفعال	الوسوم الخاطئة في الأفعال	مجموع ما وسم فعلا	عدد الأفعال	السورة	رقم السورة
84.84	168	21	175	189	إبراهيم	14
84.39	119	22	128	141	الحجر	15
93.26	374	27	374	401	النحل	16
88.12	319	43	320	362	الإسراء	17
87.56	345	49	362	394	الكهف	18
87.86	181	25	203	206	مريم	19
87.62	340	48	352	388	طه	20
88.70	1846	235	1914	2081	المجموع الكلي	

لقد كانت معظم الأفعال الموسمة خطأً في المدونة العينة واردة مرة واحدة حيث بلغت 162 فعلا، وثمة فعل واحد تكرر 5 مرات، وآخر تكرر 4 مرات، وفعلا تكرر ثلاث مرات، و8 أفعال تكررت مرتين. وفي هذا السياق، وبالعودة للمدونة التي درب عليها نموذج التميمي المحسن نستطيع أن نصنف الأخطاء التي وقع النموذج فيها وفق الجدول (7)، وفيما يأتي بيان ذلك.

جدول (7) أنواع الأخطاء التي وقع فيها النموذج المحسن في بعض سور القرآن الكريم

نوع الخطأ	عدد الأخطاء	نسبة الخطأ
أفعال غير معروفة	193	82.12
الفجوة المعجمية	22	9.4
الغموض اللغوي	16	6.8
أخطاء في مدونة التدريب	4	1.7
المجموع	235	%100

إن إحدى مشكلات الطرائق الإحصائية في معالجة اللغة الطبيعية، هي مشكلة البيانات المتناثرة التي تتأتى من صفيرية، أو انخفاض تكرارات الكلمات، أو الخصائص، ومن ثم يحدث التمثيل غير الدقيق لها. فالكلمات التي لم ترد في مدونة تدريب النموذج، ستعطى احتمالية صفيرية. وسواء زاد حجم مدونة التدريب أو قل، ستشكل الكلمات غير المعروفة جزءاً كبيراً من البيانات الجديدة. فقد أفاد ايسين Essen وستاينبيس Steinbiss⁽¹⁾ أن 12% من مجموعة الاختبار التي تشكل 25% من مدونة مليونية، كانت عبارة عن كلمات غير معروفة لم تحدث في مدونة التدريب.

لقد جاءت معظم الأخطاء أفعالاً غير معروفة unknown word، وقد شكلت ما نسبته 82% من مجموع الأخطاء. وأعني بالأفعال غير المعروفة، الأفعال التي لم ترد في مدونة التدريب أبداً. وكونها لم ترد في مدونة التدريب، سيعتمد النموذج

(1) Essen, U. and Steinbiss, V. (1992) "Cooccurrence smoothing for Stochastic language modeling," [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 161–164. Available at: <https://doi.org/10.1109/icassp.1992.225947>.

على سياقات خصائصها فقط، والسياقات غالبًا مهمة. إن هذه الأفعال في معظمها أفعال يختص بها القرآن الكريم أولاً، كما أن صغر حجم المدونة وليس نوعية النصوص المدرب عليها النموذج المحسن، وندرة هذه الأفعال على مستوى النص القرآني نفسه، وليس على مستوى العربية وحسب، نحو: أوتين، أحتكن... قد تكون من أسباب عدم ترجيح وجود بعض هذه الأفعال فيها؛ لأن النموذج المحسن قد درب على نزر يسير من آيات القرآن الكريم.

وثمة أخطاء أخرى أصنفها تحت ما يصطلح عليه بالفجوة المعجمية lexical gap⁽¹⁾. ويقصد بالفجوة المعجمية هنا، أن ترد الكلمة مرارًا في مدونة التدريب موسومة بوسم معين ترد به مرارًا وتكرارًا، ثم يرد نفس شكل الكلمة في نص آخر لكنه يكون فعلًا فيسمه خطأً بالوسم الذي تدرب عليه. أو أن ترد الكلمة التي لا تكون إلا فعلًا في مدونة التدريب، وتتكرر في سياقات معينة وسمت فيها صحيحةً، ولكنها لم ترد بالسياق نفسه في القرآن الكريم فيقع الخطأ. وقد حدث هذا النوع من الأخطاء في 9% تقريبًا من مجموع الأخطاء. لقد وردت مرارًا كلمة (نقص) في مدونة التدريب باعتبارها اسمًا، فوسمها النموذج في النص القرآني بهذا الوسم الوحيد لها في مدونة التدريب، نحو:

ك RP_ ذلك RP_ نقص V_ عليك RP_

فوسم النموذج (نقص) باعتبارها اسمًا؛ لأنها وردت في كل سياقاتها بمدونة التدريب مصدرًا للفعل نَقَصَ لا قَصَّ. كما أخطأ النموذج في توسيم كلمة (نريد) ووسمها بوصفها اسمًا مع ورودها في مدونة التدريب في السياق الآتي:

(1) Manning, C. (2011) "International Conference on Intelligent Text Processing and Computational Linguistics," in Computational Linguistics and Intelligent Text Processing. Berlin, Heidelberg: Springer, pp. 171–189.

نريد_V أن RP نعتمد_V على RP أنفسنا_N

إلا أنها وردت في القرآن الكريم بسياق مختلف، فوسمها خطأً بوسم الاسم:

له RP فيها RP ما P نشاء_V ل RP من P نريد_N

ذلك أن النموذج يميز بين السياقات، ولم تكن الكلمة (نريد) وقعت في نفس السياق التي وردت فيه في السورة الكريمة.

ولقد كان للغموض اللغوي، الذي سببه تشابه صيغ بعض الأفعال في سور القرآن المبحوثة بكلمات من أقسام الكلام الأخرى - خصوصاً مع غياب التشكيل - أثر واضح في أخطاء توسيم أفعال من نحو: (آت) الشبيه بصيغة الفاعل: آتٍ، إذ قد وسمت في عينة الدراسة باعتبارها صفة؛ لأن شبيهاتها في البنية في مدونة التدريب للنموذج المحسن وردت صفة مرة واسماً مرة أخرى، نحو: آسى وآل. وهذا النوع من الأخطاء قد شكل من أخطاء النموذج ما نسبته 6% تقريباً، وقد يفيد فيها ربط النموذج بالتشكيل وحسابات التصاحب correlation.

وثمة أخطاء قليلة لم تتجاوز نسبة 2% جاءت من البيانات التدريبية نفسها. وهذا يعني أن النموذج دُرّب على بيانات وسم فيها الفعل خطأً، على نحو ما ورد في مدونة التدريب للنموذج المحسن:

قد RP طال_N مقامهم_N، PUNC

إذ ينبغي أن توسم (طال) بوصفها فعلاً، والنتيجة هو أن النموذج وسمها بوسم خاطئ في الآية الكريمة: أ RP ف RP طال_V عليكم RP العهد_N رغم تشابه السياقين؛ لأنه درب على ذات الخطأ. ورغم أن هذا النوع من الأخطاء المتعلق بأخطاء مدونة التدريب والموسمين البشر شكل أقل أنواع الأخطاء في تقييم النموذج إلا أن حله من أسهل الحلول، إذ يكون بالرجوع لمدونة التدريب وتصحيح الأخطاء فيها، ومن ثم إعادة تدريب النموذج.

الخاتمة

لقد استكشفت هذه الورقة أداء موسم التميمي النحوي المحسن الذي درب على نصوص من العربية ممتدة عبر جميع أزمنتها وأمكنها وموضوعاتها. وكشفت النتائج التجريبية أهمية التمييز بين البيانات اللغوية عند تدريب النماذج، فالنماذج التي تدرب لتوسيم القرآن الكريم لا بُدَّ أن تختلف عن النماذج التي تدرب لتوسيم مستويات لغوية أخرى من العربية. وإن كان النموذج المحسن متنوع البيانات عند التدريب إلا أن نصوص القرآن فيه لا تمثل إلا ما يمثله القرآن الكريم من نسبة بين نصوص المحتوى العربي المتاح، وهي نسبة قليلة جدًا ولا تكاد تذكر. ويمكن أن يركز العمل مستقبلاً في توسيم القرآن الكريم بهذا النموذج على النظر في نفس العينة لوسوم أقسام الكلام الأخرى التي يتضمنها النموذج، كالأسماء، والصفات، والأدوات، والضمائر، والظروف، والخوالب. كما يمكن أن يستكشف أداء النموذج في وسوم الأفعال في المستوى الفرعي من مجموعة الوسوم في النظام نفسه، والمستوى الموسع منها أيضاً. فقد تكشف التجربة والبصيرة اللغوية أن انخفاض الأداء له أسباب أخرى تتعلق بمجموعة الوسوم نفسها وخصوصيتها بالنص القرآني.

قائمة المراجع

المراجع العربية

- 1- التميمي، أفراح، نظام آلي للتقطيع والتوسيم النحوي العربي، دار كنوز المعرفة، الأردن، 2020.
- 2- التميمي، أفراح، "نحو تحسين أداء نموذج التميمي للوسم النحوي الآلي للغة العربية"، المجلة الدولية للسانيات العربية، المغرب، ع 7، مج 1، 2021، ص. 61-79.
- 3- حبش، نزار، مقدمة في المعالجة الطبيعية للغة العربية، (ترجمة: هند الخليفة)، جامعة الملك سعود، الرياض، 2014.
- 4- حسان، تمام، اللغة العربية، معناها ومبناها، دار الثقافة، المغرب، 1994.
- 5- مدينة الملك عبد العزيز للعلوم والتقنية، المدونة اللغوية العربية، نسبة توزيع النصوص على الأوعية، متاح على الرابط (شاهد يوم 2022-09-09):
https://corpus.kacst.edu.sa/medium_chart.jsp
- 6- مدينة الملك عبد العزيز للعلوم والتقنية، المدونة اللغوية العربية، أداة لمعالجة المدونات العربية، متاح على الرابط (شاهد يوم 2022-09-09):
<https://sourceforge.net/projects/ghawwasv4/files/?source=navbar>

المراجع الأجنبية

- 1- AbuZeina, D. and Abdalbasat, T.M. (2019) "Exploring the performance of tagging for the classical and the modern standard arabic," *Advances in Fuzzy Systems*, 2019, pp. 1–10. Available at: <https://doi.org/10.1155/2019/6254649>.

- 2- Alosaimy, A. and Atwell, E. (2017) *"Tagging classical Arabic text using available morphological analysers and part of speech taggers," Journal for Language Technology and Computational Linguistics*, 32(1), pp. 1–26. Available at: <https://doi.org/10.21248/jlcl.32.2017.212>.
- 3- Alrabiah, M. (2015) *Building A Distributional Semantic Model for Traditional Arabic and Investigating its Novel Applications to The Holy Qur'an*. dissertation.
- 4- Alrabiah, M., Al-Salman, A. and Atwell, E. (2014) *"KSUCCA: A key to exploring Arabic historical linguistics," International Journal of Computational Linguistics (IJCL)*, 5(2), pp. 27–36.
- 5- Dukes, Kais, The Quranic Arabic Corpus, available at (seen on-09 2022-09): <http://corpus.quran.com/morphologicalsearch.jsp>
- 6- Essen, U. and Steinbiss, V. (1992) *"Cooccurrence smoothing for Stochastic language modeling," [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 161–164. Available at: <https://doi.org/10.1109/icassp.1992.225947>.
- 7- Garside, R., Leech, G. and McEnery, T. (2013) *Corpus annotation: Linguistic information from computer text corpora*. USA: Routledge.
- 8- Lafferty, J. D.; McCallum, A. & Pereira, F. C. N. (2001), *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, in 'Proceedings of the Eighteenth International Conference on Machine Learning', Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 282-289.

- 9- Manning, C. (2011) *"International Conference on Intelligent Text Processing and Computational Linguistics,"* in *Computational Linguistics and Intelligent Text Processing*. Berlin, Heidelberg: Springer, pp. 171–189.
- 10- McEnery, T. and Wilson, A. (2011) *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- 11- McEnery, T., Tono, Y. and Xiao, R. (2006) *Corpus-based Language Studies: An advanced resource book*. USA: Routledge.
- 12- Mohamed, E. (2018) *"Morphological segmentation and part-of-speech tagging for the Arabic heritage,"* *ACM Transactions on Asian and Low-Resource Language Information Processing*, 17(3), pp. 1–13. Available at: <https://doi.org/10.1145/3178459>.
- 13- Paroubek, P. (2007) *"Evaluating Part-of-Speech Tagging and Parsing,"* in L. Dybkjær, H. Hemsén, and W. Minker (eds) *Evaluation of Text and Speech Systems*. Dordrecht: Springer, pp. 99–124



التصنيف الموضوعي للنص العربي باستخدام التصنيف غير المتوازن: مواضيع القرآن الكريم نموذجاً

**Themes' Classification of Arabic Text using
Imbalanced Classification: An Example of
Quranic Topics**

د. أكرم محمد زكي

كلية تقنية المعلومات والاتصالات،
الجامعة الإسلامية العالمية بماليزيا، ماليزيا

د. بسام عركوك

كلية العلوم التطبيقية والإنسانية - المهرة، جامعة حضرموت، اليمن

الملخص

يعد التصنيف الموضوعي من المواضيع الحيوية في الوقت الحاضر؛ بسبب كثرة المواد المنتشرة في الفضاء الإلكتروني. حيث يهدف التصنيف الموضوعي إلى وضع عنوان مناسب للنص بناءً على معايير، ومحددات يتم استخلاصها من النص. وقد وصل التصنيف الموضوعي إلى مراحل متقدمة في اللغة الانجليزية، ولا يزال يحتاج إلى المزيد من البحث، والتطوير في اللغة العربية. وهناك الكثير من التقنيات والخوارزميات التي تساعد في هذا الشأن، ومنها تقنيات التصنيف غير المتوازن، وقد تم تطبيق تقنيات التصنيف غير المتوازن على نطاق واسع في مجال التنقيب عن البيانات. حيث يتم استخدامها لتصنيف الفئات غير المتوازنة، والتي لا تتساوى في عدد عيناتها. وتكمن مشكلة الفئات غير المتوازنة في أن أداء التصنيف يميل إلى الفئة التي تحوي مزيداً من العينات، بينما تحصل الفئة ذات العينات القليلة على أداء ضعيف. يمكن أن تحدث هذه المشكلة في التصنيف القرآني؛ بسبب الاختلاف في عدد الآيات للمواضيع القرآنية المطلوب تصنيفها. صنف العديد من الدراسات البحثية المواضيع القرآنية، والتي اعتمدت على التصنيف التقليدي. وفي المقابل، هناك دراسات قليلة صنفت الموضوعات القرآنية على أساس تقنيات التصنيف غير المتوازن. لذلك، تهدف هذه الدراسة إلى التصنيف الموضوعي للنص العربي (متمثلاً بآيات من القرآن الكريم) باستخدام بعض تقنيات التصنيف غير المتوازن، ومقارنة أدائها مع طريقة التصنيف التقليدي. ومن ثمَّ تحديد أفضل طريقة من بين هذه التقنيات في تصنيف آيات القرآن الكريم. استعرضت هذه الدراسة أيضاً الأدبيات التي ستناقش الطرق السابقة، التي تعتمد على قواعد البيانات القرآنية، والمصنفة حسب أساليب مختلفة. ويليهِ سيتم تقديم منهجية البحث

لتوضيح خطوات ومراحل عملية التصنيف. وأخيراً، سيتم الكشف عن النتائج والتجارب لاستخلاص بعض النتائج وإظهار أداء تصنيف المصحف الشريف باستخدام التصنيف غير المتوازن أو بدونه. تعمل هذه الدراسة على تطبيق طرق التصنيف غير المتوازن؛ لتصنيف المواضيع القرآنية غير المتساوية في عدد آياتها. في هذه الدراسة استخدمنا التقنية الأشهر في التصنيف غير المتوازن، والتي هي: (SMOTE). حيث تم استخدام العديد من المقاييس لتقييم النتائج التجريبية، والتي أظهرت تحسن في النتائج في أداء التصنيف القرآني عند تطبيق تقنية SMOTE.

الكلمات المفتاحية: التصنيف الموضوعي، التصنيف غير المتوازن، مواضيع القرآن الكريم، تقنيات إعادة التوازن، تصنيف النصوص.

Abstract

Imbalanced classification techniques have been widely applied in the field of data mining. It is used to classify imbalanced classes that do not have the same number of samples. The problem with imbalanced classes is that the classification performance tends to the majority class that are with more samples, while the class with fewer samples will perform poorly. This problem can occur in the Quranic classification due to the difference in the number of verses of Quranic topics to be classified. Many research studies have classified Qur'anic topics, based on the traditional classification. However, there are few studies that classify Qur'anic topics based on these imbalanced classification techniques. Therefore, this paper aims to apply imbalanced classification methods to classify Qur'anic topics that are unequal in the number of their verses. In this research, we used the most famous technique in the imbalanced classification area, which is: (SMOTE). Several metrics were used in this research to evaluate the experimental results. These metrics are Sensitivity, Specificity, G-mean, MCC. The results showed an improvement in the performance of Quranic classification when applying the SMOTE technique.

Keywords: Theme Classification, Imbalanced Classification, Quranic Topics Re-sampling techniques, Text Classification.

المقدمة

القرآن الكريم هو أحد الكتب السماوية، التي أنزلها الله سبحانه وتعالى على رسله. ويعتبر القرآن الكريم الذي أنزله الله على خاتم النبيين، محمد صلى الله عليه وسلم، المرجع الأساس لما يقدر بنحو 1.6 مليار مسلم حول العالم. وقد اعتنى المسلمون بالقرآن الكريم، وخدموه بمختلف الوسائل قديماً وحديثاً، ومن هذه الخدمات، ما نجده اليوم من مئات البرامج والتطبيقات التقنية، التي تخدم القرآن الكريم. ومنها على سبيل المثال أبحاث الذكاء الاصطناعي، التي أجريت على القرآن الكريم، مثل تصنيف النص التلقائي. حيث يمكن اعتبار التصنيف القرآني التلقائي أحد الأبحاث الأساسية في عالم الذكاء الاصطناعي. ويمكن تعريف التصنيف الموضوعي التلقائي للقرآن الكريم، هو عبارة عن تعيين موضوع أو أكثر بشكل تلقائي للنص القرآني وفقاً لمحتواه.

صنفت بعض الدراسات النص القرآني، على أنه إما توجيهي أو قصصي، وللنص ارتباط عميق ضمنى ببعض المفاهيم ذات الصلة. وتتداخل المواضيع الجوهرية للقرآن الكريم ومعانيه، من سورة إلى سورة، ومن آية إلى آية⁽¹⁾. لذلك، فإن اكتشاف الروابط الضمنية سيحتاج إلى وقت ودراسة أكثر تعمقاً لاكتشاف الموضوعات المخفية في القرآن الكريم. ولهذه الأبحاث أهمية بالغة، من خلال استخراج المعرفة الموحدة في القرآن الكريم؛ لمساعدة المسلمين على فهم النص القرآني جيداً⁽²⁾.

(1) Siddiqui, M. A., S. M. Faraz, and S. A. Sattar. "Discovering the thematic structure of the Quran using probabilistic topic model," in 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. 2013. IEEE, doi: 10.1109/nooric.2013.55.

(2) Salloum, S. A, et al., "A survey of Arabic text mining," in Intelligent Natural Language
==

يعمل تصنيف النصوص غير المتوازنة، من خلال تعيين فئة واحدة أو أكثر للمستند وفقًا لمحتوى النص. ويعد التصنيف غير المتوازن من الأساليب شائعة الاستخدام، عندما تكون مجموعة البيانات غير متساوية من حيث الحجم. ويطلق على مجموعة البيانات غير المتوازنة، عندما تكون فئة واحدة أو أكثر من فئاتها تحوي عينات أكثر بشكل ملحوظ من الفئات الأخرى. حيث يُطلق على الفئة التي تحتوي على مزيد من العينات بفئة الأغلبية، بينما يُطلق على الفئة التي تحوي عينات قليلة بفئة الأقلية⁽¹⁾⁽²⁾⁽³⁾، وتعتبر مجموعة البيانات غير متوازنة إذا كان معدل التوازن بين الفئات يتراوح بين 0.01 - 29.1٪.

اهتم الباحثون بالتصنيف غير المتوازن لمجموعات البيانات، نظرًا لأهميتها في العديد من الممارسات الواقعية، مثل: تصنيف النصوص، وتحليل البيانات البيولوجية، واكتشاف الاحتيال، وتصنيف الصور⁽⁴⁾. حيث كثيرًا ما يفشل التصنيف التقليدي عند تطبيقه في تصنيف البيانات غير المتوازنة؛ لأن دقة

==

Processing: Trends and Applications, Springer. pp. 417-431, 2018, doi: 10.1007/978-3-319-67056-0_20.

- (1) Ali, H., et al., "A review on data pre-processing methods for class imbalance problem," International Journal of Engineering & Technology, vol. 8, no. 3, pp. 390-397, 2019, doi: 10.14419/ijet.v8i3.29508.
- (2) Haixiang, G., et al., "Learning from class-imbalanced data: Review of methods and applications," Expert Systems with Applications, vol. 73, pp. 220-239, 2017, doi: 10.1016/j.eswa.2016.12.035.
- (3) Nayal, A., H. Jomaa, and M. Awad, "KerMinSVM for imbalanced datasets with a case study on arabic comics classification," Engineering Applications of Artificial Intelligence, vol. 59, pp. 159-169, 2017, doi: 10.1016/j.engappai.2017.01.001.
- (4) Khalilia, M, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," BMC medical informatics and decision making, vol. 11, no. 1, pp. 51, 2011, doi: 10.1186/1472-6947-11-51.

التصنيف تميل لفئة الأغلبية، بينما فئة الأقلية تحصل على دقة منخفضة⁽¹⁾. وللمعالجة هذه المشكلة استخدم الباحثون تقنية إعادة توازن العينات، والتي تعد من أكثر التقنيات شيوعاً لحل هذه المشكلة. والتي تعتمد على توليد نماذج جديدة للطبقة الأقلية أو حذف عينات من الطبقة الأغلبية؛ لإعادة توازن مجموعات البيانات غير المتوازنة. حيث يتم تطبيق تقنية إعادة التوازن قبل عملية التصنيف.

على الرغم من ذلك تمّ تطوير مئات الخوارزميات بلغات مختلفة في العقد الماضي⁽²⁾؛ لتصنيف مجموعات البيانات غير المتوازنة. إلا أن الأبحاث التي تخص اللغة العربية تعتبر قليلة جداً⁽³⁾⁽⁴⁾⁽⁵⁾ لا سيما الأبحاث التي تخص القرآن الكريم. ذلك على الرغم من وجود العديد من الدراسات التي صنفت آيات القرآن الكريم باللغة العربية، وغير العربية بناءً على تقنيات مختلفة. لكن هذه الدراسات

-
- (1) Alibeigi, M, S. Hashemi, and A. Hamzeh, "DBFS: An effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets," Data & Knowledge Engineering. vol. 81, pp. 67-103, 2012, doi: 10.1016/j.datak.2012.08.001.
 - (2) Haixiang, G., et al., "Learning from class-imbalanced data: Review of methods and applications," Expert Systems with Applications, vol. 73, pp. 220-239, 2017, doi: 10.1016/j.eswa.2016.12.035.
 - (3) Nayal, A., H. Jomaa, and M. Awad, "KerMinSVM for imbalanced datasets with a case study on arabic comics classification," Engineering Applications of Artificial Intelligence, vol. 59, pp. 159-169, 2017, doi: 10.1016/j.engappai.2017.01.001.
 - (4) Al-Azani, S and E.-S. M. El-Alfy. "Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text," in ANT/SEIT, 2017, doi: 10.1016/j.procs.2017.05.365.
 - (5) Mountassir, A, H. Benbrahim, and I. Berrada, "Some methods to address the problem of unbalanced sentiment classification in an arabic context," in 2012 Colloquium in Information Science and Technology, 2012, IEEE, doi: 10.1109/CIST.2012.6388061.

لم تهتم بتقنيات التصنيف غير المتوازن. علمًا أن العديد من الموضوعات القرآنية يمكن أن ينطبق عليها مفهوم مجموعات البيانات غير متوازنة؛ بسبب الاختلاف في عدد الآيات في الموضوعات المختلفة. كما لوحظ على سبيل المثال في الفهرس الموضوعي المستخدم في أحد البحوث⁽¹⁾، فعلى سبيل المثال فإن موضوع التوحيد يحتوي على 940 آية، بينما موضوع الشرك يحوي 94 آية فقط. إذن، يُعتبر موضوعات التوحيد والشرك فئتين غير متوازنتين إذا صُنِفَتا معا. لذلك فإن أداء التصنيف سوف يتجه لعينات موضوع التوحيد، بينما عينات موضوع الشرك ستحصل على أداء تصنيف ضعيف.

(1) محمد، زكي خضر، "المعجم المفهرس للتراكيب المتشابهة لفظًا في القرآن الكريم، دار عمار، الأردن"، 2002.

مصطلحات البحث

- التصنيف الموضوعي: هو التصنيف الموضوعي التلقائي للنصوص، والذي هو عبارة عن تعيين موضوع أو أكثر بشكل تلقائي للنص المكتوب وفقاً لمحتواه.
- التصنيف القرآني: هو التصنيف الموضوعي التلقائي للقرآن الكريم، والذي هو عبارة عن تعيين موضوع أو أكثر بشكل تلقائي للنص القرآني وفقاً لمحتواه.
- التصنيف غير المتوازن: يُعتبر حالة خاصة من عملية التصنيف، والذي يهتم بتصنيف الفئات غير المتساوية في عدد عيناتها.
- التصنيف التقليدي: هو التصنيف الذي يقوم بتصنيف الفئات دون مراعاة عدم التساوي في عدد عيناتها.
- SMOTE: أشهر تقنيات التصنيف غير المتوازن، والتي يهتم بإعادة توازن عينات الفئات الأقلية حسب جيرانها.

أهداف الدراسة

تهدف هذه الدراسة إلى تصنيف الموضوعات القرآنية، باستخدام بعض تقنيات التصنيف غير المتوازن، ومقارنة أداها مع طريقة التصنيف التقليدي، ومن ثمَّ تحديد أفضل طريقة من بين هذه التقنيات في تصنيف آيات القرآن الكريم. استعرضت هذه الدراسة أيضاً (في القسم التالي)، الأدبيات التي ستناقش الطرق السابقة التي تعتمد على قواعد البيانات القرآنية، والمصنفة حسب أساليب مختلفة. ويليهِ سيتم تقديم منهجية البحث لتوضيح خطوات ومراحل عملية التصنيف. وأخيراً سيتم الكشف عن النتائج والتجارب لاستخلاص بعض النتائج وإظهار أداء تصنيف المصحف الشريف باستخدام التصنيف غير المتوازن أو بدونه.

مراجعة الأدبيات

هناك العديد من الدراسات التي صنفت الآيات القرآنية لمجموعات اللغة العربية، وغير العربية لتحديد الموضوع المقابل لها. حيث كانت المحاولة الأولى لتصنيف المواضيع القرآنية العربية عن طريق محمد الكعبي ومن معه⁽¹⁾؛ حيث صمم المؤلفون في هذا البحث نموذجًا إحصائيًا لتصنيف الآيات، والذي يعمل بشكل تلقائي لسورتي الفاتحة وياسين لمواضيع محددة سلفًا. يقوم المصنف المقترح بتصنيف الآيات في كل سورة، بناءً على حساب درجة لكل آية مقابل كل فئة كمرحلة أولية، ثم بعد ذلك، يتم تخصيص الآيات للفئات ذات الأعلى درجة. تم اعتبار هذه الدراسة محدودة وبدائية في التصنيف الموضوعي للآيات القرآنية⁽²⁾.

ناقش نصورو تصنيف السور القرآنية، اعتبارًا للمراحل الأولية لرسالة النبي محمد صلى الله عليه وسلم، باستخدام خوارزميات التعلم الآلي⁽³⁾، حيث تم تصنيف السور القرآنية وفقًا لأماكن الوحي، باستخدام مصنفات Naïve Bayes و Support Vector. علاوة على ذلك، فقد تم تصنيف السور القرآنية إلى مكية ومدنية، كما أنه تم تجميع السور المكية والمدنية في مجموعات تعتمد على مكان الوحي. خلصت هذه الدراسة إلى أن السور المدنية حصلت على أداء أعلى من السور المكية، في كل من SVM و Naïve Bayes.

(1) Al-Kabi, M. N, et al., "Statistical classifier of the holy Quran verses (Fatiha and Yaseen chapters)," Journal of Applied Sciences. vol. 5, no. 3 pp. 580-583, 2005, doi: 10.3923/jas.2005.580.583.

(2) Al-Kabi, M. N., et al., "A topical classification of Quranic Arabic text," in Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. 2013.

(3) Nassourou, M, "Using Machine Learning Algorithms for Categorizing Quranic Chapters by Major Phases of Prophet Mohammad's Messengership," International Journal of Information and Communication Technology Research: Citeseer, vol. 2, 2012.

في دراسة أخرى، صنف المؤلفون آيات القرآن الكريم وفقاً لموضوعاتها، حيث طبق الباحثون خوارزميات التصنيف الأكثر شهرةً وهي: KNN، و Decision trees، و (NB)، (SVM) لثلاثة موضوعات مختارة. فقد قام الباحثون بتنفيذ مرحلة المعالجة الأولية على مجموعة البيانات القرآنية، واستخدموا ستة مقاييس أداء لتقييم النتائج. ولكن، تم استخدام 1227 آية فقط من إجمالي 6236 آية من القرآن الكريم، لمرحلة التدريب وتقييم الموضوعات المختارة. هذه الموضوعات هي: الجاهلون بالدين، ووحدانية الله، وجزاء المرتدين. هذه الموضوعات تعتبر موضوعات غير متوازنة (غير متساوية في عدد آياتها)، وفقاً لعدد آيات هذه المواضيع. فقد كان هناك 1218 آية لوحداية الله، وثلاث آيات فقط لموضوع الجاهلين بالدين، واحتوى موضوع جزاء المرتدين على ستة آيات فقط. صنف المؤلفون هذه الموضوعات غير المتوازنة باستخدام التصنيف التقليدي⁽¹⁾.

أكور وآخرون⁽²⁾ هدفوا إلى تقييم التشابه في الوثائق والنص في اللغة العربية، على أساس القرآن الكريم، فقد تم استخدام الآيات القرآنية بوصفها استعلامات للبحث عن أوجه التشابه، وتقييمها باستخدام طريقتهم المقترحة (MQVC).

تتكون مجموعة البيانات المستخدمة من أكثر من 2000 آية من القرآن، وقد طبقوا طريقة MQVC بشكل عشوائي لـ 70 سورة من أصل 114 سورة من القرآن الكريم. تم إجراء التجارب باستخدام تقنية الـ N-gram، بالإضافة إلى خوارزمية

(1) Al-Kabi, M. N., et al., "A topical classification of Quranic Arabic text," in Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. 2013.

(2) Akour, M, I. M. Alsmadi, and I. Alazzam, "MQVC: Measuring quranic verses similarity and sura classification using N-gram," WSEAS Transactions on Computers, vol. 13, pp. 485-491, 2014.

التعلم الآلي (مصنف LibSVM) لتصنيف السور المختارة من القرآن إلى سور مكية ومدنية.

بينما طبق حسن وآخرون، خوارزمية KNN لتصنيف آيات التفسير للقرآن الكريم، إلى فئات محددة سلفاً لإجراء التصنيف، حيث تم تقسيم 1000 آيات من القرآن الكريم، إلى مجموعتين: مجموعة التدريب، ويتكون من 800 آية، ومجموعة الاختبار، تتكون من 200 آية. لعملية التصنيف، تم اختيار سبعة مواضيع من نصوص التفسير: الميراث، والزواج، واحترام الآباء، والصلاة، والزكاة، والجهاد، والحلال، بعد ترجمتها من اللغة العربية إلى اللغة الملاوية. ومن ثمّ تم اختبار مجموعة البيانات بناءً على مقاييس التقييم (Recall and Precision). أقل قيمة Recall لفئة "الميراث" كانت 0.74، في حين أن فئة "الزواج" لديها أعلى قيمة Recall والتي تبلغ 0.9⁽¹⁾.

في بحث آخر⁽²⁾؛ والذي هدف إلى تطوير كفاءة استرجاع المعلومات من القرآن الكريم، بالاعتماد على نظام الإجابة على الأسئلة (QAS)، من خلال تصنيف الآيات القرآنية، باستخدام تقنية الشبكة العصبية (NN). استخدمت هذه الدراسة أكثر الترجمات الإنجليزية شعبية، لمصحف عبد الله يوسف علي باعتبارها مجموعة بيانات. تناولت QAS بعض المشاكل من خلال توسيع السؤال، وذلك باستخدام WordNet، والاستفادة من جمع المصطلحات

(1) Hassan, G. S., S. K. Mohammad, and F. M. Alwan, "Categorization of 'Holy Quran-Tafseer' using K-nearest neighbor algorithm," Int. J. Comput. Appl, vol. 129, no. 12, pp. 1-6, 2015.

(2) Hamed, S. K and M. J. Ab Aziz, "A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification," J. Comput. Sci., vol. 12, no. 3, pp. 169-177, 2016, doi: 10.3844/jcssp.2016.169.177.

الإسلامية؛ لتجنب الاختلافات في الأسئلة والترجمات. صنفت هذه الدراسة سورة البقرة إلى موضوعين: الصوم، والحج. ومن ثَمَّ، فإن QAS يسترجع الآيات التي لها صلة بالسؤال، بناءً على تقنية N-gram، ثم يقوم بترتيب الآيات المسترجعة بناءً على أعلى درجة تشابه لإرضاء رغبة المستخدم.

في دراسة حديثة، قدم المؤلفون منهجية اختيار الميزات لتسمية الآيات القرآنية تلقائياً؛ لربط الآيات بأركان الإسلام الخمسة: "الشهادة"، و"الصلوات"، و"الزكاة"، و"الصوم"، و"الحج". طريقة اختيار الميزة المقترحة استخدمت عدداً المصطلحات على أساس المجموعة؛ لتمثيل الميزات المستخرجة من ثلاث ترجمات قرآنية مختلفة. اعتمدت الطريقة على خوارزميات اختيار الميزات لتقليل مساحة الميزات، [200 آية مختارة عشوائياً على أساس مؤشر اليدوي، والتي تقتصر على أركان الإسلام الأوليين (الشهادتان، والصلاة). أخيراً، تم تنفيذ المصنفات KNN و SVM و NB بشكل مستقل على خوارزميات اختيار الميزات لتحديد عضوية الصنف لكل آية، وقياس النتائج من حيث المنطقة الواقعة تحت منحنى خصائص تشغيل المستقبل AUC⁽¹⁾.

رستم وماليم⁽²⁾ اقترحوا طريقة لتصنيف الأصناف المختارة، من خلال تحديد العلاقة المتبادلة بين مصادر القرآن، والحديث، باستخدام تقنيات تصنيف النصوص. تم تنفيذ العديد من التجارب باستخدام مجموعة بيانات القرآن

(1) Adeleke, A. O., et al., "Comparative analysis of text classification algorithms for automated labelling of Quranic verses," Int. J. Adv. Sci. Eng. Inf. Technol, vol. 7, no. 4, p. 1419, 2017, doi: 10.18517/ijaseit.7.4.2198.

(2) Rostam, N. A. P and N. H. A. H. Malim, "Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting," Journal of King Saud University-Computer and Information Sciences, 2019, doi: 10.1016/j.jksuci.2019.03.007.

والحديث. تم اختيار ثلاثة مواضيع: الحج، والصلاة، والزكاة، والتي تمت مقارنتها باستخدام ثلاثة مصنفات: SVM، و Naïve Bayes و KNN، عبر تردد الوثيقة العكسي (TF-IDF) بالإضافة إلى ترجيح المصطلح. علاوة على ذلك، تم تجربة طرق التجميع لمعرفة ما إذا كانت مناسبة لتصنيف مجموعات البيانات المترابطة كمصادر للقرآن والحديث.

قام الباحث (علوم الدين) وآخرون⁽¹⁾، ببناء نموذج لتصنيف نصوص قرآنية متعددة العناوين والمترجمة إلى الإنجليزية باستخدام مصنف KNN. في هذه الدراسة، تم دمج مصنف KNN مع الطرق TF-IDF و Weighted TF-IDF لتحسين تصنيف النصوص القرآنية. وفقًا للنتائج التي تم الحصول عليها في هذه الدراسة، حققت طريقة TF-IDF الموزونة نتائج أفضل مقارنةً بـ TF-IDF. في بحث آخر⁽²⁾ قام المؤلفون بدمج تقنيتي اختيار الميزات لتحسين أداء التصنيف، وفي نفس الوقت لتقليل وقت التشغيل الحسابي. الخوارزميات المدمجة عبارة عن (CFS) chi-square and wrapper correlation-based. أولاً، تم تنفيذ CHI لتقليل أبعاد مساحة الميزات، وتم تنفيذ CFS بعد ذلك لتحديد الميزات الأساسية من الميزات المخفضة. تم تطبيق هذه التقنية المقترحة لتصنيف الآيات القرآنية باستخدام المصنفات القياسية SVM و NB و J48. لقد حقق التصنيف القرآني نتائج أفضل من طرق الاختيار الأساسية عند تطبيق المنهج المقترح.

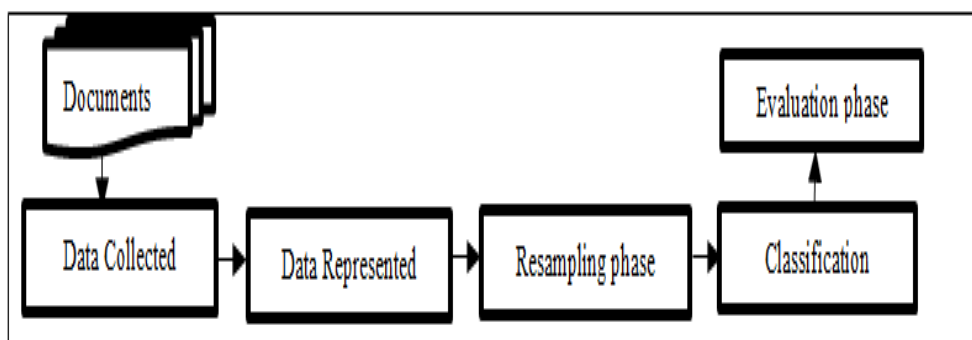
(1) Ulumudin, G, A. Adiwijaya, and M. Mubarak. "A multilabel classification on topics of qur'anic verses in English translation using K-Nearest Neighbor method with Weighted TF-IDF," in Journal of Physics: Conference Series. 2019. IOP Publishing, doi: 10.1088/1742-6596/1192/1/012026.

(2) Adeleke, A., et al., "A two-step feature selection method for quranic text classification," Indones. J. Electr. Eng. Comput. Sci., vol. 16, no. 2, pp. 730-736, 2019, doi: 10.11591/ijeecs.v16.i2.pp730-736.

اقترح المتولي والصيات⁽¹⁾ نموذجاً للآيات القرآنية متعددة المواضيع، والتي استخدمت لتصنيف السور الستة الأولى في القرآن الكريم. استخدمت هذه الدراسة المصنفات SVM و Naïve Bayes و J48 و KNN لتصنيف النص القرآني بشكل تلقائي. وفي النهاية، تم بناء نموذج تصنيف متعدد الطبقات لأسماء السور من هذه الآيات القرآنية باستخدام GaussianNB، و المصنف الدعم الموجه (SVC).

منهجية البحث

هذا القسم يشرح منهجية البحث بالتفاصيل، والتي تضم العديد من الخطوات قبل مهمة تصنيف آيات القرآن الكريم، والتي تم عمل مخطط لها في الشكل 1.



الشكل 1: المخطط الانسيابي لهذا البحث

وسوف يتم شرح كل خطوة من خطوات منهجية البحث في الأقسام الفرعية التالية:

(1) Elmitwally, N. S and A. Alsayat, "The Multi-Class Classification for the First Six Surats of the Holy Quran," International Journal of Advanced Computer Science and Applications, vol. 11, no. 1, pp. 327-332, 2020, doi: 10.14569/IJACSA.2020.0110141.

أ- جمع مجموعة البيانات

تصنف هذه الدراسة آيات قرآنية لثمانية مواضيع قرآنية، وفقاً لمعجم لمواضيع القرآن الكريم، والتي جمعت من قِبَل علماء مسلمين. هذه الموضوعات عبارة عن مجموعة بيانات غير متوازنة، والتي لا يمكن تصنيفها بالطرق التقليدية بشكل صحيح، يقدم الجدول 1 البيانات التي تم جمعها مع عدد العينات لكل مجموعة بيانات ثنائية، والنسب غير المتوازنة (IR) بين هذه الموضوعات، ورقم الميزات.

الجدول 1: يقدم البيانات التي تم جمعها مع عدد العينات لكل مجموعة بيانات ثنائية، IR وعدد الميزات.

	المواضيع القرآنية المختارة	من فئة الأغلبية	من فئة الأقلية	IR	من الميزات
1	التوحيد_الشرك	940	94	10	1119
2	الصلاة_الزكاة	186	37	5.03	645
3	العمل_العلوم	537	76	7.07	947
4	النبوة_الدين	322	86	3.74	743
5	القصص القرآنية - العلاقات السياسية	301	32	9.41	726
6	الجهاد_الصوم	153	29	5.28	624
7	إسلام_الإيمان	1291	993	1.30	1477
8	تنظيم العلاقات المالية - الدعوة إلى الله	211	140	1.51	760

شرح البيانات

تم استخدام الجذور القرآنية للموضوعات القرآنية المستخرجة، عن طريق البروفيسور محمد زكي خضر، حيث قام المؤلف بالاستعانة ببعض المصادر اللغوية؛ لاستخراج الجذور القرآنية: مثل لسان العرب، والمعجم المفهرس لألفاظ القرآن الكريم. لذلك، تم إلغاء خطوة المعالجة المسبقة؛ للحصول على الجذور القرآنية، وتم تجاهلها في هذا البحث. بعد ذلك، تم تمثيل الموضوعات التي جُمعت باستخدام تقنية مصطلح تردد الوثيقة العكسية (TF-IDF) ؛ لأجل تمثيل مصفوفة التصنيف. وهذه المصفوفة تتم بالخطوات التالية: أولاً، تم تنسيق جميع الجذور القرآنية في مساحة متجهة. بعد ذلك، تم حساب TF-IDF لجميع الكلمات القرآنية لكل آية، وتم وضع قيمتها تحت ميزتها في مساحة المتجهة. بعد ذلك، تم تمثيل المصفوفة من خلال مجموعة من جميع هذه المساحات لجميع الآيات. أخيراً، يتم تمييز هذه المتجهات في المصفوفة بناءً على المعجم المستخدم.

علاوة على ذلك، تتم إزالة جميع الآيات المشتركة لكلا الصنفين من المصفوفة. على سبيل المثال، توجد 24 آية في موضوعي: التوحيد، والشرك في نفس الوقت، لذلك تم حذف هذه الآيات من الصنفين: التوحيد، والشرك. حيث يتم حذف الآيات المشتركة من مجموعات البيانات لأجل الحصول على بيانات نقية؛ لمنع الارتباك أثناء مرحلة التصنيف، وتجنب التداخل بين مجموعات بيانات التدريب والاختبار .

اختيار الميزات

في هذه المرحلة، يتم تقليل ميزات البيانات المنسقة في المصفوفة، واختيارها بواسطة خوارزمية Chi-square تم استخدام خوارزمية Chi-square لتحديد الميزات المهمة التي تساعد في تحسين أداء التصنيف. يتم اختيار الميزات بواسطة

خوارزمية Chi-square، والتي تم استخدامها على نطاق واسع لتصنيف النص العربي⁽¹⁾⁽²⁾. بعد تحديد الميزات، يتم حذف العينات المكررة مرة أخرى من البيانات الممثلة. في هذا الحذف، تتم إزالة العينات المكررة للحصول على بيانات نقية، وأيضاً لتقييم الطرق المختلفة بشكل أكثر دقة. بمعنى آخر، عند وجود نفس العينة في كل من مجموعات بيانات التدريب والاختبار، فإن هذه العينة تتمتع بدقة 100٪. لذلك، يتم أخذ هذه المشكلة في الاعتبار للحصول على نتائج أكثر دقة وتقييم الأساليب المختلفة بدقة.

تقسيم مجموعات البيانات

يتم تقسيم مجموعات البيانات إلى 70٪ كبيانات تدريب و 30٪ كبيانات اختبار لبناء نماذج التدريب واختبار المصنفات. يأخذ هذا التقسيم في الاعتبار النسب غير المتوازنة لمجموعة البيانات القرآنية، مما يعني أن 70٪ من كل فصل يتم اختياره لبيانات التدريب، ونفس السياسة لبيانات الاختبار. يعتمد هذا التقسيم على إحدى الدراسات السابقة⁽³⁾، والتي خلصت إلى أن دقة التصنيف تتحسن عندما تكون مجموعة بيانات الاختبار أصغر من مجموعة بيانات التدريب. يوضح الجدول 2 مجموعات البيانات النهائية بعد تطبيق الخطوات السابقة مع عدد

-
- (1) Khorsheed, M. S. and A. O. Al-Thubaity, "Comparative evaluation of text classification techniques using a large diverse Arabic dataset," Language resources and evaluation, vol. 47, no. 2, pp. 513-538, 2013, doi: 10.1007/s10579-013-9221-8.
 - (2) Alabbas, W., H. M. Al-Khateeb, and A. Mansour. "Arabic text classification methods: Systematic literature review of primary studies," in 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), IEEE. 2016, doi: 10.1109/CIST.2016.7805072.
 - (3) Khorsheed, M. S. and A. O. Al-Thubaity, "Comparative evaluation of text classification techniques using a large diverse Arabic dataset," Language resources and evaluation, vol. 47, no. 2, pp. 513-538, 2013, doi: 10.1007/s10579-013-9221-8.

العينات لكل موضوع في مرحلة التدريب، والنسب غير المتوازنة (IR)، وعدد ميزات كل مجموعة بيانات.

الجدول 2: يوضح كيفية مجموعات البيانات النهائية
بعد تطبيق جميع خطوات السابقة

من الميزات	IR	من فئة الأقلية	من فئة الأغلبية	المواضيع القرآنية المختارة	
53	3.25	57	185	التوحيد_الشرك	1
18	2	16	32	الصلاة_الزكاة	2
98	4.47	51	228	العمل_العلوم	3
36	1.86	51	95	النبوة_الدين	4
86	5.5	22	121	القصص القرآنية - العلاقات السياسية	5
34	1.35	20	27	الجهاد_الصوم	6
133	1.47	529	778	الإسلام_الإيمان	7
60	1.34	67	90	تنظيم العلاقات المالية - الدعوة إلى الله	8

مرحلة إعادة توازن العينات

في هذه الدراسة سوف يتم استخدام التقنية الأكثر شيوعاً في إعادة توازن مجموعة البيانات غير المتوازنة، هذه التقنية هي،

SMOTE (Synthetic Minority Over-sampling Technique)

تم اقتراح SMOTE في إحدى الدراسات⁽¹⁾، وذلك عند الإفراط في أخذ عينة من

(1) Liu, X.-Y, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 2, pp. 539-550, 2009, doi: 10.1109/TSMCB.2008.2007853.

فئة الأقلية. SMOTE هي الطريقة الأكثر شيوعًا في عملية الإفراط في أخذ العينات لمجموعات البيانات غير المتوازنة⁽¹⁾⁽²⁾. في طريقة SMOTE، تُستخدم العينات الأصلية لفئة الأقلية، كنقطة انطلاق لزيادة فئة الأقلية عن طريق توليد عينات اصطناعية؛ لإعادة توازن العينات بين الفئات. إنها تفوق عينات فئة الأقلية بناءً على عدد العينات المجاورة. بعد ذلك، تمت إضافة العينات الجديدة من فئة الأقلية الاصطناعية، عن طريق استيفاء أزواج من أقرب الجيران بشكل عشوائي في فئة الأقلية. تم تطبيق طريقة SMOTE في هذه الدراسة للإفراط في أخذ عينات من موضوع الشرك. في هذه الدراسة أيضًا، تراوح عدد العينات المجاورة "K" من 1 إلى 10 لاختيار أفضل قيمة لطريقة SMOTE لاحقًا. لذلك، يتم تنفيذ طريقة SMOTE 10 مرات للحصول على أفضل النتائج لكل مجموعة بيانات.

مرحلة التصنيف

يستخدم هذا العمل أداة WEKA لتحليل المعرفة (WEKA 3.6) لتحميل مجموعات البيانات القرآنية المجمعة، وتنفيذ المصنفات المطلوبة. تم استخدام 9 مصنفات مختلفة وهي: Naive Bayes, KNN, J48, Random Forest Random Tree, SMO, Voted perceptron, LibSVM, and Simple CART. والتي ركبت مع تقنية SMOTE لتصنيف المواضيع القرآنية المختارة في هذا البحث. ولكن، قبل عملية التصنيف، يجب إعادة أخذ عينات مجموعات البيانات التدريبية من خلال طرق إعادة أخذ العينات. تم تطبيق عدة مصنفات في هذا البحث، لانتقاء

(1) Chawla, N. V, et al., "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research. vol. 16, pp. 321-357, 2002, doi: 10.1613/jair.953.

(2) Tang, Y, et al., "SVMs modeling for highly imbalanced classification," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 1, pp. 281-288, 2009, doi: 10.1109/TSMCB.2008.2002909.

افضلهم للتصنيف القرآني للمواضيع غير المتساوية في عدد آياتها.

مرحلة التقييم

تشتمل هذه الخطوة على مرحلة تقييم واختبار تقنيات التصنيف غير المتوازن، والتي تم انتقاؤها في هذه الدراسة، والمنفذة مع المصنف NB. يتم تقييم اختبار التجارب باستخدام النماذج المدربة أثناء عملية تصنيف الآيات القرآنية غير المتساوية في عدد آياتها، للكشف عن أداء المصنفات. طالما أن مجموعة البيانات غير متوازنة، فإن مقياس الدقة لا يكفي لاختبار أداء التصنيف غير المتوازن؛ لأن التنبؤ يميل إلى الصنف الأغلبية⁽¹⁾⁽²⁾⁽³⁾ لذلك، في هذه المرحلة، يتم استخدام مقاييس مختلفة هنا لتقييم أداء التصنيف غير المتوازن، والمخدمة في التصنيف غير المتوازن⁽⁴⁾ هذه المقاييس ومعادلتها هي كالتالي:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \dots \dots \dots (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \dots \dots \dots (2)$$

-
- (1) Al-Azani, S and E.-S. M. El-Alfy. "Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text," in ANT/SEIT, 2017, doi: 10.1016/j.procs.2017.05.365.
 - (2) Patel, H and G. S. Thakur, "Classification of imbalanced data using a modified fuzzy-neighbor weighted approach," International Journal of Intelligent Engineering and Systems, vol. 10, no. 1, pp. 56-64, 2017, doi: 10.22266/ijies2017.0228.07.
 - (3) Blagus, R. and L. Lusa, SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics, 2013. 14(1): p. 106.
 - (4) Haixiang, G., et al., "Learning from class-imbalanced data: Review of methods and applications," Expert Systems with Applications, vol. 73, pp. 220-239, 2017, doi: 10.1016/j.eswa.2016.12.035.

$$G - Mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \dots \dots \dots (3)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \dots \dots (4)$$

النتائج التجريبية والتحليل

في هذا القسم، يتم عرض النتائج التجريبية التي تم الحصول عليها من خلال هذا البحث. أولاً، يتم تقديم متوسط نتائج المصنفات دون استخدام طريقة التصنيف غير المتوازن، والتي هذه نتائج التصنيف التقليدي. يتم عرض التصنيف التقليدي، لإثبات ضعف التصنيف التقليدي عند استخدامه لتصنيف الموضوعات القرآنية غير المتساوية في عدد آياتها . ثانياً، يتم عرض متوسط النتائج المتحصل عليها باستخدام تقنيات التصنيف غير المتوازن، والموضحة في الخطوة الخامس من خطوات منهجية البحث لهذا العمل. يوضح الجدول 3 متوسط النتائج لجميع مجموعات البيانات القرآنية، لجميع التقنيات المستخدمة، والتي تم تنفيذها مع المصنف Naïve Bayes .

1- النتائج التجريبية التي تم الحصول عليها في هذا البحث

في هذا القسم سيتم مناقشة النتائج التي تم الحصول عليها باستخدام المصنفات المختارة. الجدول رقم 3 يقدم هذه النتائج لجميع تقنيات التصنيف غير المتوازن، والتي تم انتقاؤها وتقديمها في القسم 3.5. في الجدول يعرض متوسط النتائج لجميع مجموعة البيانات الثمانية للمواضيع القرآنية المختارة.

الجدول 3: النتائج التجريبية لجميع التقنيات المستخدمة
وفقاً لمقاييس التقييم المختارة:

التقنية المستخدمة	Sensitivity	Specificity	G-mean	MCC
NB	0.83	0.80	0.81	0.62
SMOTE_NB	0.88	0.70	0.78	0.59
KNN	0.86	0.55	0.68	0.45
SMOTE_KNN	0.84	0.74	0.79	0.58
J48	0.87	0.61	0.72	0.48
SMOTE_J48	0.84	0.72	0.77	0.54
Random Forest	0.88	0.64	0.75	0.55
SMOTE_RF	0.88	0.68	0.77	0.58
SMO	0.91	0.66	0.77	0.61
SMOTE_SMO	0.85	0.83	0.83	0.67
Voted Perceptron	0.85	0.64	0.73	0.50
SMOTE_VP	0.84	0.87	0.85	0.68
LibSVM	0.93	0.63	0.76	0.62
SMOTE_LibSVM	0.87	0.80	0.83	0.66
Random Trees	0.83	0.53	0.65	0.39
SMOTE_RT	0.86	0.61	0.72	0.49
Simple CART	0.89	0.63	0.75	0.54
SMOTE_SC	0.87	0.70	0.78	0.56

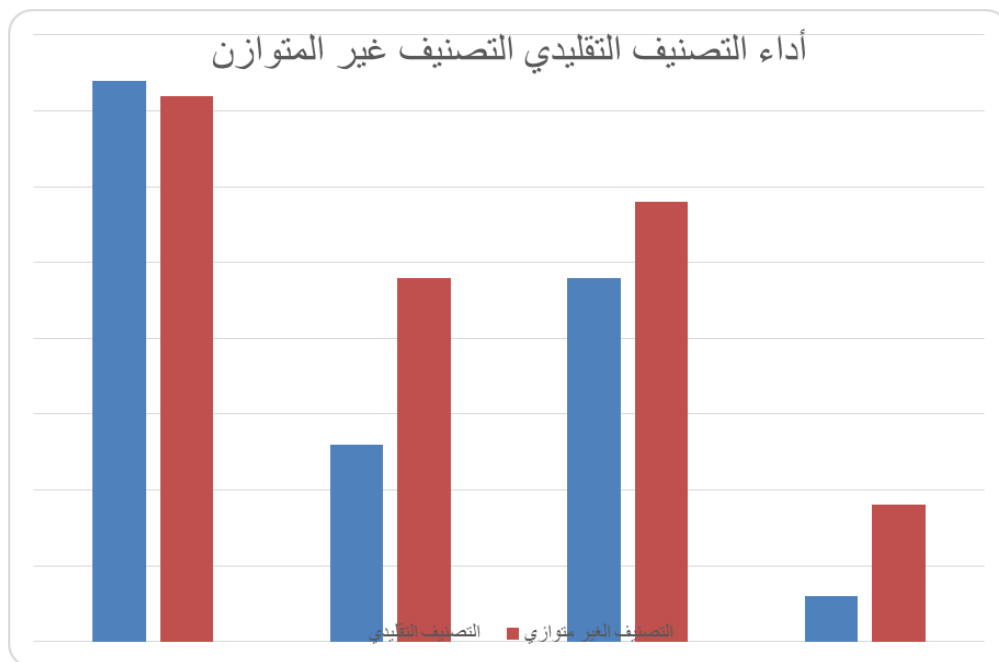
من خلال الجدول أعلاه، نلاحظ أن أداء جميع المصنفات تحسنت عندما تم تنفيذها مع SMOTE ما عدا الخوارزمية NB، والتي يعني أن التصنيف غير المتوازن حسنت من أداء التصنيف التقليدي. كما إنني ألاحظ من خلال النتائج، أن المصنف Voted Perceptron مع SMOTE حصلوا على أفضل الأداء من بين بقية التقنيات. كما ننوه أن الخوارزمية NB هي الخوارزمية الوحيدة التي لم يتحسن أداؤها عند تركيبها مع SMOTE.

تحليل النتائج بشكل عام

في هذا القسم سيتم مناقشة النتائج التجريبية بشكل عام، وذلك من خلال إيجاد متوسط أداء المصنفات التقليدية، والتي نُفذت بدون استخدام الطريقة SMOTE. كما أوجدنا أيضاً متوسط أداء التصنيف غير المتوازن ليتم مقارنتها مع التصنيف التقليدي.

الجدول 4: متوسط أداء التصنيف غير المتوازن وأداء التصنيف التقليدي

التقنية المستخدمة	Sensitivity	Specificity	G-mean	MCC
T_Classification	0.87	0.63	0.74	0.53
IM_Classification	0.86	0.74	0.79	0.59



الشكل 2: مقارنة بين أداء التصنيف التقليدي والتصنيف غير المتوازن

من خلال الجدول رقم 4 والشكل رقم 2، نلاحظ أداء التصنيف غير المتوازن تفوقت على أداء التصنيف التقليدي بشكل كبير في المقاييس Specificity, MCC. كما أنها أيضاً تفوقت على التصنيف التقليدي في المقياس G-mean بشكل ملحوظ. وحصلت على نتائج قريبة جداً للتصنيف التقليدي في المقياس sensitivity على الرغم من أن التصنيف التقليدي يميل أداؤها إلى الفئات الأغلبية.

من خلال نتائج البحث أعلاه، يمكننا تلخيصها بالنقاط التالية:

1- التصنيف غير المتوازن حسنت أداء التصنيف القرآني مقارنة بالتصنيف التقليدي.

2- التصنيف التقليدي أداؤه ضعيف للتصنيف القرآني؛ بسبب ميول التصنيف

للمواضيع القرآنية التي لها آيات كثيرة.

3- جميع خوارزميات التصنيف التقليدي المطبقة في هذا البحث تحسّن أدائها عندما نُفذت مع خوارزمية SMOTE ما عدا المصنف NB التي ضَعُفَ أدائها مع التصنيف غير المتوازن.

الخاتمة

يعد التصنيف الموضوعي أحد أهم الجوانب المستخدمة على نطاق واسع في الوقت الحاضر. صنفت العديد من الدراسات النص القرآني باستخدام طرق مختلفة. ومع ذلك، لا توجد دراسة لتصنيف النص القرآني على أساس التصنيف غير المتوازن. يهدف هذا العمل إلى تصنيف الموضوعات القرآنية غير المتساوية في عدد آياتها باستخدام طرق إعادة توازن العينات؛ لإعادة توازن مجموعة البيانات القرآنية غير المتوازنة قبل مرحلة التصنيف. كما تم في هذه الدراسة مقارنة أداء المصنفات قبل وبعد استخدام طريقة إعادة أخذ العينات. أظهرت النتائج أن أداء تصنيف مواضيع القرآن الكريم قد تحسن بشكل ملحوظ من خلال تطبيق التصنيف غير المتوازن للعديد من المقاييس المستخدمة في تقييم أداء التصنيف غير المتوازن. العمل المستقبلي لهذه الدراسة هو تطبيق تقنيات التصنيف غير المتوازن في تصميم مصنف عام، والذي يشمل جميع مواضيع القرآن الكريم. كما يمكن تطبيق تقنيات التصنيف غير المتوازن لتصنيف الأحاديث النبوية لتحسين التصنيف النبوي.

شكر وتقدير: تم دعم هذه الدراسة من قبل الجامعة الإسلامية العالمية في ماليزيا في إطار مشروع بحثي تحت عنوان (FRGS/1/2018/ICT04/UIAM/02/3)

قائمة المراجع

- 1- محمد، زكي خضر، المعجم المفهرس للتراكيب المتشابهة لفظاً في القرآن الكريم، دار عمار، الأردن، 2002.
- 2- Adeleke, A. O., et al., "Comparative analysis of text classification algorithms for automated labelling of Quranic verses," Int. J. Adv. Sci. Eng. Inf. Technol, vol. 7, no. 4, p. 1419, 2017, doi: 10.18517/ijaseit.7.4.2198.
- 3- Adeleke, A., et al., "A two-step feature selection method for quranic text classification," Indones. J. Electr. Eng. Comput. Sci., vol. 16, no. 2, pp. 730-736, 2019, doi: 10.11591/ijeecs.v16.i2.pp730-736.
- 4- Akour, M, I. M. Alsmadi, and I. Alazzam, "MQVC: Measuring quranic verses similarity and sura classification using N-gram," WSEAS Transactions on Computers, vol. 13, pp. 485-491, 2014.
- 5- Alabbas, W., H. M. Al-Khateeb, and A. Mansour. "Arabic text classification methods: Systematic literature review of primary studies," in 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), IEEE. 2016, doi: 10.1109/CIST.2016.7805072.
- 6- Al-Azani, S and E.-S. M. El-Alfy. "Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text," in ANT/SEIT, 2017, doi: 10.1016/j.procs.2017.05.365.
- 7- Ali, H., et al., "A review on data pre-processing methods for class imbalance problem," International Journal of Engineering &

- Technology, vol. 8, no. 3, pp. 390-397, 2019, doi: 10.14419/ijet.v8i3.29508.
- 8- Alibeigi, M, S. Hashemi, and A. Hamzeh, "DBFS: An effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets," Data & Knowledge Engineering. vol. 81, pp. 67-103, 2012, doi: 10.1016/j.datak.2012.08.001.
- 9- Al-Kabi, M. N, et al., "Statistical classifier of the holy Quran verses (Fatiha and Yaseen chapters)," Journal of Applied Sciences. vol. 5, no. 3 pp. 580-583, 2005, doi: 10.3923/jas.2005.580.583.
- 10- Al-Kabi, M. N., et al., "A topical classification of Quranic Arabic text," in Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. 2013.
- 11- Al-Khair, A. A and M. A. Kabbani, "Quran teacher intonation," The Tunisian Company for Distribution, 2003. 2003
- علي أبو الخير، محمد عربي القباني، مصحف معلم التجويد، الشركة التونسية للتوزيع.
- 12- Blagus, R. and L. Lusa, SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics, 2013. 14(1): p. 106.
- 13- Chawla, N. V, et al., "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research. vol. 16, pp. 321-357, 2002, doi: 10.1613/jair.953.
- 14- Elmitwally, N. S and A. Alsayat, "The Multi-Class Classification for the First Six Surats of the Holy Quran," International Journal of

- Advanced Computer Science and Applications, vol. 11, no. 1, pp. 327-332, 2020, doi: 10.14569/IJACSA.2020.0110141.
- 15- Haixiang, G., et al., "Learning from class-imbalanced data: Review of methods and applications," Expert Systems with Applications, vol. 73, pp. 220-239, 2017, doi: 10.1016/j.eswa.2016.12.035.
- 16- Hamed, S. K and M. J. Ab Aziz, "A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification," J. Comput. Sci., vol. 12, no. 3, pp. 169-177, 2016, doi: 10.3844/jcssp.2016.169.177.
- 17- Hassan, G. S., S. K. Mohammad, and F. M. Alwan, "Categorization of 'Holy Quran-Tafseer' using K-nearest neighbor algorithm," Int. J. Comput. Appl, vol. 129, no. 12, pp. 1-6, 2015.
- 18- Khalilia, M, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," BMC medical informatics and decision making. vol. 11, no. 1, pp. 51, 2011, doi: 10.1186/1472-6947-11-51.
- 19- Khorsheed, M. S. and A. O. Al-Thubaity, "Comparative evaluation of text classification techniques using a large diverse Arabic dataset," Language resources and evaluation, vol. 47, no. 2, pp. 513-538, 2013, doi: 10.1007/s10579-013-9221-8.
- 20- Liu, X.-Y, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 2, pp. 539-550, 2009, doi: 10.1109/TSMCB.2008.2007853.
- 21- Mountassir, A, H. Benbrahim, and I. Berrada, "Some methods to

- address the problem of unbalanced sentiment classification in an arabic context," in 2012 Colloquium in Information Science and Technology, 2012, IEEE, doi: 10.1109/CIST.2012.6388061.
- 22-Nassourou, M, "Using Machine Learning Algorithms for Categorizing Quranic Chaptersby Major Phases of Prophet Mohammad's Messengership," International Journal of Information and Communication Technology Research: Citeseer, vol. 2, 2012.
- 23-Nayal, A., H. Jomaa, and M. Awad, "KerMinSVM for imbalanced datasets with a case study on arabic comics classification," Engineering Applications of Artificial Intelligence, vol. 59, pp. 159-169, 2017, doi: 10.1016/j.engappai.2017.01.001.
- 24- Patel, H and G. S. Thakur, "Classification of imbalanced data using a modified fuzzy-neighbor weighted approach," International Journal of Intelligent Engineering and Systems, vol. 10, no. 1, pp. 56-64, 2017, doi: 10.22266/ijies2017.0228.07.
- 25-Rostam, N. A. P and N. H. A. H. Malim, "Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting," Journal of King Saud University-Computer and Information Sciences, 2019, doi: 10.1016/j.jksuci.2019.03.007.
- 26- Salloum, S. A, et al., "A survey of Arabic text mining," in Intelligent Natural Language Processing: Trends and Applications, Springer. pp. 417-431, 2018, doi: 10.1007/978-3-319-67056-0_20.
- 27-Siddiqui, M. A., S. M. Faraz, and S. A. Sattar. "Discovering the thematic structure of the Quran using probabilistic topic model," in

- 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences. 2013. IEEE, doi: 10.1109/nooric.2013.55.
- 28-Skryjomski, P. and B. Krawczyk. Influence of minority class instance types on SMOTE imbalanced data oversampling. in first international workshop on learning with imbalanced domains: theory and applications. 2017.
- 29-Tang, Y, et al., "SVMs modeling for highly imbalanced classification," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 1, pp. 281-288, 2009, doi: 10.1109/TSMCB.2008.2002909.
- 30-Ulumudin, G, A. Adiwijaya, and M. Mubarak. "A multilabel classification on topics of qur'anic verses in English translation using K-Nearest Neighbor method with Weighted TF-IDF," in Journal of Physics: Conference Series. 2019. IOP Publishing, doi: 10.1088/1742-6596/1192/1/012026.



مزالق الترجمة الآلية

دراسة تطبيقية على نصوص مختارة

The Pitfalls of Machine Translation: Applied
Study on Selected Texts

إعداد

الأستاذ الدكتور صديق جوهر

أكاديمي وناقد

خبير الترجمة في الأرشيف والمكتبة الوطنية / وزارة شؤون الرئاسة
(أبوظبي)

ملخص البحث

في ظل سياق العولمة أصبحت الاتصالات بين البلدان والثقافات متكررة بشكل متزايد، مما اقتضى استخدام بعض التقنيات الإلكترونية للمساعدة في ترجمة اللغات المختلفة، من أجل التواصل الإنساني، ومنها الترجمة الآلية، أو الترجمة بمساعدة الحاسوب. تعتمد دراسة الترجمة الآلية على ثلاثة تخصصات هي: اللغويات، والرياضيات، وعلوم الحاسوب. والترجمة الآلية هي فرع من فروع علم اللغة الحاسوبي، الذي يستخدم الآليات المحوسبة لترجمة السياق أو المحادثة من لغة بشرية إلى أخرى. لقد تم طرح مفهوم الترجمة الآلية لأول مرة بواسطة الباحث وارين ويفر في عام 1947، بعد عام واحد فقط من تطوير أول جهاز حاسوب رقمي إلكتروني، وحاسوبي متكامل. منذ ذلك الحين اعتبرت الترجمة الآلية واحدة من أصعب المهام في مجال معالجة اللغة الطبيعية. وعلى الرغم من أن الترجمة الآلية، لاتزال تواجه عوائق متنوعة، لا يمكن إنكار أن الترجمة بمساعدة الحاسوب قد حققت تقدماً هائلاً، خاصة بعد الاستعانة بتطبيقات الذكاء الاصطناعي، وتطوير شركة جوجل، وغيرها من الشركات، لأنظمة الترجمة العصبية في القرن الحادي والعشرين. في هذا السياق تستطلع هذه الورقة البحثية من خلال استخدام نظريات الترجمة الآلية المعاصرة بوصفها إطاراً منهجياً، لتطور أنظمة الترجمة الآلية منذ أربعينيات القرن المنصرم، حتى السنوات الأخيرة، مع تسليط الضوء على الترجمة الآلية، القائمة على البيانات، والترجمة الآلية القائمة على الأمثلة، والترجمة الآلية الإحصائية، والترجمة الآلية العصبية، من أجل رصد مزايا وعيوب الترجمة الآلية من خلال التطبيق على نصوص متنوعة ومختارة. خلصت الدراسة إلى تأكيد وجود قصور لا مندوحة عنه في برامج الترجمة الآلية،

خاصة عند الترجمة من اللغة العربية إلى اللغات الأوروبية. وتؤكد الدراسة عدم قدرة أنظمة الترجمة الآلية على التعامل بكفاءة مع العديد من النصوص، خاصة الأدبية والدينية والتراثية والتاريخية. ورداً على الادعاءات المتكررة عن القدرة الخارقة لآليات الذكاء الاصطناعي، التي أدخلت على أنظمة الترجمة الآلية، نؤكد على صعوبة اختراع مترجم فوري آلي في الوقت الراهن، واستحالة الاستغناء عن المترجمين البشريين، حتى لو استطاع العلماء إنشاء أنظمة ترجمة آلية عصبية متطورة، تطابق الدماغ البشري في المعرفة العامة، والقدرة على التفكير؛ لأن العلماء حتى الآن لم يستطيعوا فك شفرات الدماغ البشري نفسه، فكيف يمكنهم اختراع آلة تحاكي الدماغ البشري الذي لم يكتشفوا سوى بعض تعقيداته وخباياه؟

الكلمات المفتاحية: الترجمة الآلية- الذكاء الاصطناعي- الترجمة القائمة على القواعد النحوية- الترجمة العصبية- مزالق الترجمة الآلية.

Abstract

In the context of globalization, contacts between countries and cultures have become increasingly frequent, which necessitates the use of some electronic technologies to help translate different languages for human communication, including the use of machine translation or computer-assisted translation. The study of machine translation is based on three disciplines: linguistics, mathematics, and computer science. Machine translation is a branch of computational linguistics that uses computerized mechanisms to translate context or conversation from one human language to another. The concept of machine translation was first introduced by researcher Warren Weaver in 1947, just one year after the development of the first integrated electronic and digital computer. Since then, machine translation has been considered one of the most difficult tasks in the field of natural language processing. Although machine translation still faces various obstacles, it is undeniable that computer-assisted translation has made tremendous progress, especially after the use of artificial intelligence applications and the development of neural translation systems by Google and other companies in the twenty-first century. In this context, this paper explores, using contemporary machine translation theories as a methodological framework, the development of machine translation systems from the 1940s until recently, with a focus on data-based machine translation, example-based machine translation, statistical machine translation, and neural machine translation, in order to monitor the advantages and disadvantages of machine

translation by applying MT dynamics to various selected texts. The study concludes by emphasizing the existence of an inevitable shortcoming in machine translation programs, especially when translating from Arabic into European languages. The study confirms the inability of machine translation systems to deal efficiently with many texts, especially literary, religious, cultural, and historical texts. In response to the repeated claims about the extra-ordinary potentialities of artificial intelligence mechanisms that have been introduced to machine translation systems, the paper stresses the difficulty of inventing an automated interpreter at the present time and the impossibility of dispensing with human translators even if scientists create advanced neural machine translation systems that match the human brain in general knowledge and ability to think. Since scientists have not yet been able to decipher the human brain itself, how can they invent a machine that simulates the human brain, which they have only discovered some of its intricacies and mysteries?

Keywords: machine translation, artificial intelligence, rules-based translation, neural translation, machine translation scandals.

المقدمة

نبذة عن الترجمة الآلية:

يتجاهل بعض المتخصصين الترجمة الآلية، أو الترجمة بمساعدة الحاسوب⁽¹⁾؛ بينما يرى آخرون أنها فرصة لتطوير صناعة الترجمة وتحسينها. لقد أدت الترجمة الآلية إلى تغيير طريقة تعامل البشر مع اللغات على نطاق واسع. ولذلك يسعى المترجمون المحترفون، وأساتذة الجامعات، وشركات الترجمة للتكيف مع هذه التغييرات الجذرية⁽²⁾ التكلفة، والسرعة، هما السببان الرئيسيان اللذان يؤديان إلى الاستخدام الواسع لأنظمة الترجمة الآلية. يوفر توسع المعلومات عبر الإنترنت مصدرًا ضخمًا للمعرفة لجميع البشر، في جميع أنحاء العالم؛ للوصول إلى مصادرها وقراءتها. يمكن بعد ذلك كتابة المعلومات بلغة محددة، وترجمتها إلى جميع اللغات والثقافات الأخرى. توفر العديد من أنظمة استرجاع المعلومات ومحركات البحث، مثل جوجل تفسيرها الخاص للمعلومات والبيانات المنشورة على صفحات الويب، عند ترجمة نص من لغة إلى أخرى. ينبغي بعد ذلك تقييم صحة أو دقة هذا القاموس/ المعجم، خاصة عندما يتعلق الأمر بترجمة النصوص الحساسة، مثل الكتب المقدسة أو النصوص الدينية.

من ناحية أخرى تشكل تقنيات الترجمة مجالًا جديدًا ومهمًا للدراسة متعددة التخصصات، يقع في منتصف الطريق بين علوم الحاسوب والترجمة⁽³⁾.

(1) Machine Translation- (MT) -Computer Assisted Translation- (CAT)

(2) David Kauchak. Contributions to research on machine translation. Unpublished PhD Dissertation. California: University of California in San Diego, 2006. P.25.

(3) Michael Denkowski. Machine Translation for Human Translators. Unpublished PhD Dissertation. Pittsburgh, PA: Carnegie Mellon University, 2015. P.19.

وفي ظل سياق العولمة أصبحت الاتصالات بين البلدان والثقافات متكررة بشكل متزايد، مما اقتضى استخدام بعض التقنيات الإلكترونية للمساعدة في ترجمة اللغات المختلفة، من أجل التواصل، ومنها الترجمة الآلية، أو الترجمة بمساعدة الحاسوب. والترجمة الآلية، هي فرع من فروع علم اللغة الحاسوبي⁽¹⁾ الذي يحلل استخدام الأدوات المحوسبة لترجمة السياق، أو المحادثة من لغة بشرية إلى أخرى. بشكل عام تعتمد دراسة الترجمة الآلية على ثلاثة تخصصات، هي: اللغويات، والرياضيات، وعلوم الحاسوب. ولذلك تتأثر الترجمة الآلية بهذه التخصصات الثلاثة والجهود المشتركة بينهم.

الترجمة الآلية / الترجمة بمساعدة الحاسوب:

يستخدم هذا الفرع من الترجمة برامج الحاسوب للمساعدة في الترجمة، التي تسهل على المترجمين، خاصة الهواة، والمبتدئين، وغير المؤهلين، إنجاز الترجمة بشكل أفضل، بسهولة وفاعلية في بعض الأحيان، خاصة عند ترجمة المصطلحات العلمية، والتقنية، والتجارية، لذلك تستعين شركات الترجمة التجارية بأنظمة الترجمة الآلية توفيراً للنفقات. "لقد ثبت أن الترجمة الآلية جديرة بالاهتمام، حيث إنها تساعد على توفير الوقت، والمكاسب الإنتاجية، خاصة في مجالات الترجمة الفنية والإدارية"⁽²⁾. والجدير بالذكر أن الترجمة الآلية أو الترجمة بمساعدة الحاسوب، هي مجال تخصصي ناشئ حديثاً مقارنة بالمجالات الأخرى. لقد ظلت الدراسات في هذا المجال متعثرة نسبياً لوقت طويل، ولكنها تسارعت بوتيرة مرتفعة في الآونة الأخيرة.

(1) Computational Linguistics.

(2) Austermühl, Frank. 2001. Electronic tools for translators. Manchester: St. Jerome Publishing. P.64.

بشكل عام تشتمل الترجمة الآلية، أو الترجمة بمساعدة الحاسوب، على عدة أجزاء رئيسية، تتضمن ذاكرة الترجمة، وإدارة المصطلحات، وبرامج التوافق والمواءمة بين اللغة المصدر واللغة الهدف. ومنذ نشأتها شرعت الترجمة الآلية في مسار متعرج وطويل من التطور، والذي يمكن تقسيمه إلى المراحل الأربع التالية: فترة التدشين والانتشار (1947-1964)، ثم فترة الكساد (1965-1975)، يليها فترة التعافي (1976-1989)، وأخيرًا الفترة الجديدة من (1990 إلى الوقت الحاضر)، وهي فترة الازدهار، حيث تزايد الاهتمام بالترجمة الآلية بشكل ليس له مثيل من قبل. وبسبب السرعة والأداء العالي لأنظمة الترجمة الآلية في المرحلة الأخيرة، أحدثت الترجمة بمساعدة الحاسوب ثورة في أسواق الترجمة، ضد الترجمة اليدوية التقليدية/البشرية، وأصبح كل من يمتلك جهاز حاسوب، يدعي أنه مترجم ببساطة؛ لأنه يستطيع استخدام آليات الترجمة على الحاسوب، خاصة جوجل ترانزليت. ونظرًا للإعجاب بمزاياها المذهلة، فقد جذبت الترجمة بمساعدة الحاسوب والآلة الانتباه من جميع أنحاء العالم، مما أدى - بالتالي - إلى تحفيز تطورها. وفي سياق متصل لا تزال الترجمة الآلية، أو الترجمة بمساعدة الحاسوب، تواجه عوائق متنوعة أثناء التعامل معها. ومع ذلك لا يمكن إنكار أن الترجمة الآلية قد حققت تقدمًا هائلًا، ولديها مستقبل مشرق من حيث التقنيات والتطبيقات.

تاريخ مختصر للترجمة الآلية:

الترجمة الآلية باختصار، هي دراسة كيفية استخدام أجهزة الحاسوب للترجمة من لغة إلى أخرى. لقد تم طرح مفهوم الترجمة الآلية لأول مرة، بواسطة الباحث وارين ويفر في عام 1947، بعد عام واحد فقط من تطوير أول جهاز حاسوبي متكامل. منذ ذلك الحين اعتُبرت الترجمة الآلية واحدة من أصعب المهام

في مجال معالجة اللغة الطبيعية⁽¹⁾. من حيث المنهجية، تنقسم أنظمة الترجمة الآلية بشكل أساسي إلى فئتين: الأنظمة المستندة إلى القواعد النحوية⁽²⁾ والأنظمة القائمة على (مكانز البيانات)⁽³⁾، أو مجموعة البيانات اللغوية التي تم جمعها من خلال الباحثين.

منذ الوقت الذي تم فيه اقتراح فكرة الترجمة الآلية لأول مرة حتى التسعينيات، كانت الأنظمة القائمة على القواعد هي السائدة. تستخدم أساليب الترجمة الآلية المستندة إلى القواعد⁽⁴⁾ قواميس، ومعاجم ثنائية اللغة، وقواعد مكتوبة يدويًا لترجمة نصوص لغة المصدر إلى نصوص لغة الهدف. لاحقًا اتضح أن طرائق كتابة القواعد النحوية يدويًا، تتطلب عمالة مكثفة وتكاليف زائدة. علاوة على ذلك يصعب الحفاظ على بيانات القواعد، ويصعب تعديلها ونقلها من مجال إلى آخر ومن لغة إلى أخرى. وبالتالي بات من الصعب تطوير الأنظمة المستندة إلى القواعد والمستخدم في الترجمة ذات النطاقات المفتوحة، وسياقات الترجمة متعددة اللغات والثقافات.

تاريخيًا، وفي بداية المشوار كانت أنظمة الترجمة الآلية مصممة بشكل أساسي لخدمة التطبيقات العسكرية الأمريكية، إبان الحرب الباردة مع الاتحاد السوفيتي السابق. في عام 1954 أكملت جامعة جورج تاون الأمريكية تجربة الترجمة الآلية، من الروسية إلى الإنجليزية؛ لأغراض استخباراتية، وذلك لأول مرة باستخدام حاسوب من طراز (آي بي إم-701). وقد تمت التجربة بالتعاون مع

(1) Natural Language Processing

(2) Rule-based methods.

(3) Corpus-based methods.

(4) Rule-based machine translation

شركة تصنيع الحاسوب المعروفة الآن باسم (آي بي إم)⁽¹⁾. ويدل ذلك على أن حلم الترجمة الآلية قد أصبح حقيقياً منذ فترة طويلة.

ظلت الترجمة الآلية موضوعاً جدلياً ساخناً لأكثر من عقد، بعد النجاح الذي حققته جامعة جورج تاون في عام 1954. لكن هذه الطفرة انتهت فجأة بسبب تقرير اللجنة الاستشارية⁽²⁾ للمعالجة التلقائية للغات في عام 1966، الذي ألقى بظلال الشك على نجاعة الترجمة الآلية واستمراريتها. بعد التقرير الذي كان متشككاً جداً في الترجمة الآلية، وأدى إلى خفض كبير في تمويل أبحاث الترجمة الآلية، أصبح من الصعب للغاية آنذاك العمل على مشروعات الترجمة الآلية. ومن الجدير بالذكر أنه في عام 1962 خلال فترة الانتشار، كان اسم الجمعية العلمية المهيمنة هو (جمعية الترجمة الآلية واللغويات الحاسوبية)⁽³⁾، والتي أصبحت فيما بعد (جمعية اللغويات الحاسوبية)⁽⁴⁾، بعدما تم إسقاط عبارة (الترجمة الآلية) من اسمها في عام 1968، أثناء فترة الكساد في أعقاب تقرير اللجنة الاستشارية لمعالجة اللغات.

خلال المراحل الباكرة من الترجمة الآلية، كانت المقاربات قائمة على القواعد. لاحقاً اعتمدت معظم طرق البحث على البيانات، ومع ذلك لا تزال أنظمة الترجمة التجارية فقط تستخدم مناهج قائمة على القواعد (ص10)⁽⁵⁾. فيما بعد واصل باحثو الترجمة الآلية، محاولة تحسين جودة الترجمة. في عام 1965 عقد باحثو

(1) International Business Machines Corporation

(2) Language Processing Advisory Committee

(3) Association for Machine Translation and Computational Linguistics

(4) Association for Computational Linguistics.

(5) O'Hagan, Minako and David Ashworth. 2002. Translation-Mediated Communication in a Digital World. Clevedon: Multilingual Matters. p.41.

البرمجة اللغوية العصبية، أول مؤتمر دولي حول اللغويات الحاسوبية، والذي ركز على التحليل القائم على القواعد والإعراب والترجمة. وابتداءً من سبعينيات القرن الماضي، أصبحت الأنظمة القائمة على القواعد أكثر نضجًا. وفي عام 1978 أطلقت (سيستران)، وهي إحدى أولى شركات الترجمة الآلية، نظام ترجمة تجاري ناجح قائم على القواعد. ومن الجدير بالذكر أن شركة جوجل، قد استخدمت خدمة (سيستران) حتى عام 2007.

مع توافر البيانات ثنائية اللغة، أصبحت الأنظمة القائمة على البيانات هي السائدة بعد العقد الأول من القرن الحادي والعشرين، وانتشرت الترجمة الآلية القائمة على البيانات. ثمة ثلاثة أنظمة تعتمد عليها الترجمة الآلية القائمة على البيانات⁽¹⁾ على النحو التالي: الترجمة الآلية القائمة على الأمثلة⁽²⁾ والترجمة الآلية الإحصائية، والترجمة الآلية العصبية⁽³⁾. "من مشكلات أنظمة الترجمة الآلية الإحصائية،⁽⁴⁾ أنها تفترض أن الجمل في النص مستقلة، وتتجاهل خاصية الترابط الموجودة في كل الخطاب النصي؛ مما يؤدي إلى ترجمات مغلوطة"⁽⁵⁾.

لاحقًا، قامت شركة (جوجل) بتطوير نظام الترجمة الآلية العصبية، وتم تقديمه في نوفمبر 2016. وفي الحقيقة تستخدم هذه المنظومة شبكة عصبية اصطناعية لزيادة الطلاقة والدقة في الترجمة. وبعبارة أخرى، تجدر الإشارة إلى أن الترجمة الآلية العصبية، هي طريقة للترجمة الآلية، تستخدم شبكة عصبية

(1) Corpus-based machine translation.

(2) Example-based machine translation.

(3) Neural machine translation.

(4) Statistical machine translation.

(5) Allen, Jeff. 2003. "Post-editing". Harold Somers, ed. Computers and Translation. A Handbook for Translators. Amsterdam: John Benjamins. P.83.

صناعية؛ لتوقع احتمالية توالد الكلمات بوتيرة متتالية. والترجمة الآلية العصبية العميقة، هي امتداد للترجمة الآلية العصبية، وكلاهما يستخدم شبكة عصبية صناعية كبيرة، والفرق بينهما أن الترجمة الآلية العصبية العميقة، تعالج طبقات متعددة للشبكات العصبية بدلاً من طبقة واحدة فقط.

في منتصف الثمانينيات، تم اقتراح استخدام الترجمة الآلية القائمة على الأمثلة؛ لترجمة بعض نصوص لغة المصدر، عن طريق استرجاع أزواج الجمل المتشابهة من البيانات المتاحة ثنائية اللغة⁽¹⁾. وكانت نتائج الترجمة بهذه الطريقة ذات جودة عالية، فقط عندما تم استرجاع أزواج الجمل المتشابهة في لغتين مستهدفتين. ومع ذلك فقد فشلت هذه الطريقة بشكل عام؛ لأن المادة، والبيانات ثنائية اللغة المتوفرة، لا يمكنها تغطية جميع الظواهر اللغوية للأزواج اللغوية المستهدفة⁽²⁾.

في عام 1990 اقترحت فكرة الترجمة بمساعدة الحاسوب، حيث تتعلم الآلات تلقائياً معرفة الترجمة، من خلال التعرف على كمية كبيرة من البيانات، بدلاً من الاعتماد على الخبراء البشريين في كتابة القواعد، وجمع البيانات، ثم تغذية الآلات بها، كما حدث مع النموذج السابق. وقد أُطلق على هذا النظام اسم، الترجمة الآلية الإحصائية، حيث يتم إنشاء الترجمات على أساس النماذج الإحصائية، التي تستمد معلوماتها من تحليل النصوص ثنائية اللغة. يتناقض النهج الإحصائي مع الأساليب القائمة على القواعد في الترجمة الآلية وكذلك مع الترجمة الآلية القائمة على الأمثلة؛ لأنهما من صنع البشر.

(1) Zetzsche, Jost. 2005. The Translator's Tool Box. A Computer Primer for Translators.

Winchester Bay: International Writers' Group. P.39.

(2) Somers, Harold ed. 2003. Computers and Translation. A Handbook for Translators:

John Benjamins. Translation Studies. P.67.

الترجمة الآلية الإحصائية:

تمت صياغة فكرة الترجمة الآلية الإحصائية بشكل أكثر وضوحًا، بعد تطوير خمسة نماذج مختلفة من هذا النظام من الترجمة الآلية في عام 1993. لم يتم اعتماد أساليب الترجمة الآلية الإحصائية على نطاق واسع في تلك الآونة؛ بسبب تعقيداتها وهيمنة الأنظمة الأخرى على التطبيقات التجارية في الأسواق خلال الثمانينيات والتسعينيات⁽¹⁾. وفي أعقاب ظهور الأنظمة الإحصائية في الترجمة الآلية، بدأ مؤتمر آخر للغة البرمجة اللغوية العصبية أعماله - مؤتمر الطرق التجريبية في معالجة اللغة الطبيعية في عام 1996 - بهدف الجمع بين الأساليب التجريبية، من مجموعة من التخصصات المختلفة، مع الأخذ في الاعتبار الطرائق المستندة إلى مجموعة من اللسانيات، والمعلومات النظرية من فروع الهندسة ذات الصلة. وفي عام 1999، عقد الباحثون ورشة عمل صيفية في جامعة جونز هوبكنز الأمريكية، حيث قاموا بإعادة إنتاج خمسة نماذج بالتعاون مع شركة (آي بي إم)، وأصدروا مجموعة أدوات تخص الترجمة الآلية الإحصائية، تسمى (مصر)⁽²⁾ ولاحقًا تم إصدار مجموعة أدوات جديدة تسمى (جيزة) و(جيزة بلس- بلس)⁽³⁾.

في عام 2003 تم اقتراح طرق وأنظمة للترجمة الآلية قائمة على تخزين العبارات، مما أدى إلى زيادة تحسين جودة الترجمة. واستنادًا إلى أساليب الترجمة الآلية الإحصائية القائمة على العبارات، تم إصدار أنظمة مفتوحة المصدر مثل

(1) Dunne, Keiran J. ed. 2006. Perspectives on Localization. Amsterdam: John Benjamins Publishing. P.128.

(2) Egypt.

(3) GIZA- GIZA.++

نظام (فرعون)⁽¹⁾، ونسخته المطورة (موسى)⁽²⁾، مما يعزز بشكل كبير تطوير أنظمة الترجمة الآلية الإحصائية. بعد ذلك تم اعتماد أساليب الترجمة الآلية الإحصائية على نطاق واسع؛ بسبب توافر مجموعات من الأدوات المتاحة. وفي عام 2006 أطلقت جوجل خدمة الترجمة عبر الإنترنت، على أساس أنظمة الترجمة الآلية الإحصائية، القائمة على العبارات. كما أطلقت شركات أخرى مثل: مايكروسوفت، وبايدو⁽³⁾، خدمات ترجمة مماثلة في السنوات التالية. وتجدر الإشارة إلى أنه من الصعب على نموذج، أو نظام، أو طريقة واحدة، التعامل مع متطلبات الترجمة المختلفة، وفق اللغات المتعددة، وبالتالي تستخدم الأنظمة غالبًا طرقًا هجينة، تدمج نماذج الترجمة الآلية المختلفة من أجل تحسين أداء الترجمة.

بعد نجاح نماذج الترجمة الآلية الإحصائية، اقترح العديد من الباحثين نماذج جديدة؛ لزيادة تحسين أداء أساليب هذا النظام، بما في ذلك نماذج الترجمة الآلية الإحصائية، التي تم فيها تقديم المعلومات المورفولوجية (علم الصرف)، ونماذج الترجمة الآلية الإحصائية الهرمية⁽⁴⁾، ونماذج الترجمة الآلية الإحصائية، القائمة على بناء الجملة مع التحليل التحويلي الرأسي لجوانب من لغة المصدر / أو قطاعات من اللغة المستهدفة. وعلى الرغم من أن استخدام نماذج الترجمة الآلية الإحصائية، أدى إلى تحسين جودة الترجمة بشكل كبير، فإن هذه الأساليب تستخدم نماذج لوغاريتمية خطية؛ لدمج عدة مكونات

(1) Pharaoh.

(2) Moses.

(3) Microsoft – Baidu.

(4) Gambier, Yves and Henrik Gottlieb. eds. 2001. (Multi) Media Translation. Amsterdam: John Benjamins. P.96.

مصممة يدويًا⁽¹⁾: نموذج خاص بالترجمة، ونموذج خاص باللغة، ونموذج إعادة الترتيب في المكونات اللغوية. وقد نجمت مشاكل تقنية خطيرة أثناء إعادة ترتيب الأزواج اللغوية البعيدة، التي تنتمي إلى عائلات لغوية مختلفة الأصل.

مع التقدم الحثيث في تكنولوجيا التعلم العميق، في الكلام، والرؤية، والمجالات الأخرى، بدأ الباحثون في التخلي التدريجي عن الأنظمة الآلية، في مقابل تطبيق تقنية التعلم العميق على أنظمة الترجمة الآلية. والجدير بالذكر، أن النظام الآلي يتعلق بقدرة أجهزة الحاسوب على التفكير، والتصرف بتدخل بشري أقل، أما التعلم العميق، فيُقصد منه تعلم الحاسوب التفكير باستخدام هياكل اصطناعية، على غرار الدماغ البشري. لذلك يمكن لأنظمة التعلم العميق تحليل الصور، ومقاطع الفيديو والبيانات غير المهيكلة، بطرق لا يمكن لأنظمة التعلم الآلي القيام بها. في سياق متصل، اقترحت نماذج ترجمة شاملة للشبكات العصبية الصناعية في 2014، واستُخدم رسميًا مصطلح "الترجمة الآلية العصبية". تقوم الفكرة الأساسية للترجمة العصبية، على تقسيم اللغة المصدر وتوزيعها على شكل نماذج دلالية كثيفة، ثم يتم توليد الترجمة المطلوبة باستخدام آليات مناسبة.

بالتزامن مع هذه التطورات، اقترح علماء من الصين إطارًا للترجمة متعدد اللغات، يعتمد على الترجمة الآلية العصبية، والذي يُعتبر بمثابة ورقة بحثية متقدمة للترجمة متعددة اللغات، نشرت في عام 2015. كذلك نشرت شركة بايدو أول نظام للترجمة الآلية العصبية واسع النطاق في العالم في عام 2016، وأطلقت جوجل أيضًا نظام للترجمة الآلية العصبية، والذي تبعته شركات أخرى

(1) Krings, Hans P. 2001. Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes. Kent, Ohio: The Kent State University Press. P.78.

بإصدار أنظمة الترجمة الآلية العصبية الخاصة بها. وبالتالي فقد استغرق نشر نظام الترجمة الآلية العصبية عبر الإنترنت، حوالي عام واحد فقط منذ اقتراحه لأول مرة في عام 2014، بينما استغرق وضع أنظمة الترجمة الإحصائية في الخدمة عبر الإنترنت حوالي 16 عامًا.

بعد ذلك تم اقتراح نموذج ترجمة ترميزي التفافي (تلافيفي)، ينتقل من تسلسل إلى تسلسل، مع استخدام نماذج تحويلية، مما أدى مرة أخرى إلى تحسين جودة الترجمة بشكل كبير. وقد أثار هذا التحسن الكبير نقاشًا واسع النطاق، حول ما إذا كانت الترجمة الآلية جيدة مثل الترجمة البشرية أم لا. اجتذب النجاح الكبير الذي حققته الترجمة الآلية الإحصائية العديد من الباحثين، الذين طوروا طرقًا مختلفة، مثل النماذج غير الانحدارية، ونماذج الترجمة الآلية الإحصائية، غير الخاضعة للإشراف البشري، ونماذج أخرى تهدف إلى تحسين جودة وكفاءة الترجمة متعددة اللغات.

الترجمة الآلية: رفاهية واشكاليات

مزايا الترجمة الآلية:

بلا شك، فإن ترجمة الآلة، هي أسرع أنواع الترجمة المتاحة، ويمكن أن تكون فورية في معظم الحالات؛ لأنها ترجمة حرفية -كلمة بكلمة- ولا تأخذ في الاعتبار أي عوامل أخرى، ولذلك لا تستغرق وقتًا طويلاً. عند مقارنة الترجمة الآلية بالترجمة البشرية الاحترافية، سوف نجد فرقًا شاسعًا فيما يخص الوقت والمجهود المبذولين، وكذلك جودة الترجمة. تعتبر الترجمة الآلية مثالية عندما ترغب في ترجمة قدر كبير من المحتوى، يتطلب ترجمة سريعة ولا تحتاج إلى أي شيء أكثر من معنى عام للنص، بدون دقة أو انضباط لغوي أو دلالي.

تمثل الطبيعة المجانية للترجمة الآلية، أكبر عامل جذب للشركات ذات

الميزانيات المحدودة. هناك عدد لا يحصى من برامج الترجمة الآلية المجانية المتاحة على الإنترنت، ومن خلال البحث يمكنك العثور على العديد من المواقع المجانية، ولكن جوجل ترانزليت هو الأكثر شعبية. وبينما لا تزال الترجمة الآلية بعيدة عن الكمال، فإنها تتحسن باستمرار مع التقدم في التكنولوجيا. وأحد هذه التطورات يتعلق بالترجمة الآلية العصبية، حيث يتم استخدام شبكة عصبية صناعية كبيرة لتعظيم أداء الترجمة الآلية.

عيوب الترجمة الآلية:

إذا كان المحتوى الذي تريد ترجمته يحتاج إلى دقة بنسبة كبيرة، وإذا كان المحتوى سيشاهده عملاؤك، فلا يجب أن ترجمه آلياً. دقة الترجمة الآلية ضعيفة فيما يتعلق ببناء الجملة، واستخدام الكلمات، والمعاني الصحيحة، والدلالات، والتراكيب اللغوية، والأبنية الصرفية، خاصة في حالة الترجمة من العربية إلى اللغات الأجنبية. غالباً ما تبدو الترجمة الآلية أشبه بالنص المصدر أكثر من كونها نصاً جديداً موجهاً لجمهور مختلف، علاوة على صعوبة ترجمة المختصرات، والمصطلحات الفنية آلياً. نظراً لطبيعتها الحرفية لا تمتلك الآلات القدرة على إضافة سياق، أو نبرة صوت، أو إحساس إلى الترجمات. الكلمات ليس لها معنى بالنسبة للآلة، ولا يمكنها التمييز بين المهم وغير المهم. والنتيجة النهائية كارثية، حيث تبدو الترجمة مثل كتلة نصية آلية لها معنى، ولكن تبدو وكأنها كتبت من خلال شخص لا يفهم الموضوع، ولا يدرك دلالات النص.

قد يؤدي استخدام الترجمات الآلية في ترجمة كل النصوص إلى تشويه سمعة المترجم، أو فصله من عمله، أو القضاء على مستقبله الوظيفي. لذلك لا ننصح بترجمة التقارير الحساسة، أو المستندات القانونية، أو كتيبات التعليمات آلياً؛ لأنها يجب أن تكون دقيقة بنسبة 100٪. يمكن أن تكلفك الأخطاء في هذا

السياق مبالغ ضخمة من المال، أو تسبب ضررًا دائمًا لسمعة شركتك، وسمعتك بوصفك مترجمًا. كما ذكر أعلاه، فإن الترجمات الآلية تتحسن دائمًا مع تحسن التكنولوجيا، لكن هذه ليست تحسينات شاملة، وقد لا تكون الترجمة الآلية المناسبة للغة ما، تصلح بالنسبة إلى لغة أخرى.

تُظهر البحوث والدراسات جودة الترجمة بالنسبة للغة واحدة تم إجراء البحوث عليها، ولكنها لا تشير إلى مدى جودة نفس نظام الترجمة فيما يتعلق باللغات الأخرى. لذلك غالبًا ما تكون الترجمات الآلية غير دقيقة؛ لأنها تحصل على الكلمات من القاموس/المعجم، وتلتزم بالقيود الشرطية التي وضعها من صمم برامجه. ومع ذلك فقد رحب المترجمون بالترجمة الآلية؛ لأنها تعمل على توفير الوقت، وتقليل التكاليف. ولكن لا يمكن للمترجمين، خاصة الهواة الذين يستخدمون أنظمة الترجمة الآلية، إنتاج ترجمات عالية الجودة، تحقق الغرض من الترجمة، وتتوافق مع جميع المواصفات الفنية، وتلبي احتياجات المستخدم النهائي.

نماذج من الترجمة الآلية والبشرية من العربية إلى الإنجليزية

1- لا توجد روضة من الروضات المبنية على القبور.

A- There is no kindergarten built on graves (MT).

B- There is no meadow/garden built on graves (HT).

2- ويوجد بينها وبين (القاهرة) جزيرة فيها مساكن حسان وعلالي مشرفة.

A- And there is an island between it and (Cairo) where the residences of Hassan and Alali are honorable (MT).

B- And there is an island between it and (Cairo) beautiful houses with overlooking attics (HT).

3- تعرض في الأسواق الذخائر النفيسة من الجواهر والياقوت والأحجار الكريمة.

A- Precious munitions are displayed in the markets, gems, rubies and stones (MT).

B- Precious goods are displayed in the markets, gems, rubies and stones (HT).

4- مالكا لنافية اللغتين اليونانية والعربية.

A- Possessing the cornerstone of the Greek and Arabic languages (MT).

B- Mastering both Greek and Arabic languages (HT)

5- بعدما وضع الآباء المؤسسون لبناتها الأولى.

A- After the founding fathers had laid its first daughters (MT).

B- After the founding fathers had laid its foundations (HT).

6- أسماء الأعلام.

A- Names of flags (MT).

B- Proper names (HT).

7- دوام الحال من المحال.

A- Time course of shops (MT).

B- Nothing lasts forever (HT).

8- الاعتراف بالجميل

A- Recognition of the beauty (MT).

B- Recognition of gratitude (HT).

9- الفريق صالح عبد الرحمن قائد قوات البحرية.

- A- The team Saleh Abdul-Rahman, Commander of the Naval Forces (MT).
B- Lieutenant general Saleh Abdul-Rahman, Commander of the Naval Forces (HT).

10- شهدت منطقة الخليج العديد من التحولات

- A- The Gulf region has witnessed many transformations (MT).
B- The Gulf region went through many transformations (HT).

11- أهل مكة أدرى بشعابها.

- A- The people of Mecca know more about its reefs (MT).
B- The people of Mecca know more about its secret roads (HT).

12- أصدر الوزير قرارًا بتعيين المحافظ.

- A- The Minister issued a decision to appoint the Governor (MT).
B- The Minister issued a decree to appoint the Governor (HT).

13- تحكمها سنة التطور.

- A- Governed by the year of development (MT).
B- Governed by rule/norm/criterion/pattern (HT).

14- لقد استطاعت العمارة الإسلامية أن تنتقل من بيوت الشعر في البوادي إلى الأكواخ في القرى.

- A- Islamic architecture has been able to move from the houses of poetry in the deserts to the huts in the villages (MT).
B- Islamic architecture has been able to move from tents in the deserts to the huts in the villages (HT).

15- لقد افاض في وصف الجارية.

- A- He elaborated on the description of the current. (MT).
B- He elaborated on the description of the female slave (HT).

16- الهاربون من المدارس.

A- School fugitives (MT).

B- Students who drop off schools (HT).

17- العتبات النصية.

A- Text thresholds (MT).

B- Paratexts (HT).

18- الدراسات الإعلامية.

A- Media Studies (MT).

B- Onomastics (HT). The study of the history and origin of proper names.

- يلاحظ في الأمثلة السابقة، أن أخطاء المترجم الآلي بشكل عام ترجع إلى عدم قدرة الآلة على فهم السياق، ولذلك تقع الأخطاء بسبب الترجمة الحرفية للعبارات والكلمات، دون مراعاة لموقعها خارج المعجم اللغوي الذي يغذي المترجم الآلي بالنص المترجم. ويزداد الأمر سوءاً عند ترجمة اللهجات المحلية، أو اللغات المحكية في الأقطار العربية المختلفة، مثل: الإمارات، ومصر، والسودان، على سبيل المثال كما يتضح مما يلي:

1- قصة العقوب في التراث الإماراتي.

A- The story of the shank in the Emirati heritage (MT).

B- The story of the sand dune in the Emirati heritage (HT).

2- لو سمحت طرش لي كل الأغراض.

A- Please throw all the things for me (MT).

B- Please send all the things for me (HT).

- في اللهجة المحلية الإماراتية (العرقوب)، هو (الكثيب الرملي)، وكذلك فعل (طرش)، يعني (أرسل)، ويتضح عدم قدرة الآلة على ترجمة اللغة المحكية الإماراتية بشكل صحيح، وكذلك الحال بالنسبة للعامية المصرية كما يلي:
- 1- قال له البائع (لو سمحت يا باشا: توتو على كبُوتو).

A- The salesman said to him (Pasha, please: Tutu on Kabuto)-MT.

B- The salesman told him (Pasha, please pay all the money now). HT.

- 2- ما تقوله هو (هبل في الجبل).

A- What you're saying is (Hubble in the mountain)-MT.

B- What you're saying is (nonsense)- HT.

- 3- يا نهار أسود الرجل ودانا في إداهية.

(إبراهيم أصلان-رواية مالك الحزين)

A- O black day, the man and Dana are in a shrewd (MT).

B- "Damn it! he's really got me in deep shit" (HT).

- 4- وبعدين بقه في الليلة اللي مش فايتة دي

(إبراهيم أصلان-رواية مالك الحزين)

A- And then stay on the night that I didn't miss this "(MT).

B- Okay, okay. Do you have to ruin our whole evening? (HT).

- توضح الأمثلة أعلاه فشل المترجم الآلي في نقل المعنى المقصود في العبارات المقتبسة من اللهجة العامية المصرية، ونفس الشيء ينطبق على اللهجة السودانية كما يلي:

- 1- (حاج عبد الصمد . كدي قول بسم الله واقعد نجيب لك فنجان جبنة).

(الطيب صالح- رواية عرس الزين).

A- (Hajj Abd al-Samad. Please say in the name of God and sit down, and we will bring you a cup of cheese).MT.

B- 'Hajj Abdul Samad, just put your trust in God and sit down and have a cup of coffee with us. HT.

2- (يا زول أنا مو فاضي لك، ولا فاضي لي خبرياتك. باقي أنا عارفك مستهبل داير تطنتش علي قروشي).

(الطيب صالح- رواية عرس الزين).

A- (--Oh Zul, I am not empty for you, nor am I missing your experiences. The rest, I know you, is a start to the day, you are harassing my Qurashi)-MT.

B- 'I've not got the time, neither for you nor for your bit of news. I know well enough that you're trying to fool me and talk me out of my money.' HT.

- في ذات السياق يستحيل على المترجم الآلي ترجمة النصوص الدينية ترجمة صحيحة خاصة معاني آيات القرآن الكريم كما يلي:

1- والسماء والطارق (سورة الطارق).

A-And Al-Samaa and Tariq (MT).

B- By the sky and the night comer/the Pulsar (HT).

2- هل في ذلك قسم لذي حجر. (سورة الفجر).

A-Is there a section that has a stone (MT).

B-There surely is an oath for men of perception(HT).

3- والنازعات غرقًا. (سورة النازعات).

A- And disputes by drowning (MT).

B- By the (angels) who tear out (the souls of the wicked) with violence (HT).

4- والعاديات ضبحًا. (سورة العاديات).

A- And the usual at dawn (MT).

B- By the steeds that run with panting (HT).

نماذج من الترجمة الآلية والبشرية من الإنجليزية إلى العربية

1- The coast of the sea was very barren.

أ- كان ساحل البحر عقيمًا جدًا (ترجمة آلية).

ب- كان ساحل البحر وعراً (ترجمة بشرية).

2- The Annual West Beverly underachievement awards party honors the very best of the worst.

أ- يقام حفل جوائز الإنجاز التحتاني في غرب بيفرلي على شرف أفضل السيئين (ترجمة آلية).

ب- يقام حفل غرب بيفرلي السنوي لجوائز التقصير في الأداء/ الأداء المتدني على شرف أفضل السيئين. (ترجمة بشرية).

3- Emails, minutes and other correspondence

أ- رسائل البريد الإلكتروني ودقائق ومراسلات أخرى (ترجمة آلية).

ب- رسائل البريد الإلكتروني ومحاضر الاجتماعات ومراسلات أخرى (ترجمة بشرية).

4- Deposit books

أ- كتب الإيداع (ترجمة آلية).

ب- سندات الإيداع (ترجمة بشرية).

5- specific instructions about the listing, boxing and transfer of the records to their custody

أ- تعليمات محددة حول الإدراج والملاكمة ونقل السجلات إلى عهدهم (ترجمة آلية).

ب- تعليمات محددة عن وضع السجلات في قوائم وفي صناديق ونقلهم إلى عهدهم. (ترجمة بشرية).

6- Bank fees and charges .

أ- الرسوم المصرفية والشحنات. (ترجمة آلية).

ب- الرسوم البنكية والغرامات (ترجمة بشرية).

- يتضح من الأمثلة أعلاه عدم قدرة المترجم الآلي على إنتاج ترجمة دقيقة بسبب إخفاق الآلة في فهم سياق ودلالات الكلمات والعبارات. من ناحية أخرى فقد فشل المترجم الآلي في ترجمة المختصرات والأوائليات⁽¹⁾ كما يلي:

1- An OHC works with employers and workers to create a healthy and safe work environment by detecting Occupational Health and Safety (OHS) hazards and developing practical approaches to eliminate or control those hazards.

أ- تعمل OHC مع أصحاب العمل والعاملين؛ لخلق بيئة عمل صحية وأمنة من خلال الكشف عن مخاطر (OHS)، وتطوير نهج عملية للقضاء على هذه المخاطر، أو السيطرة عليها. (ترجمة آلية).

ب- تعمل (لجان الصحة والسلامة المهنية)، مع أصحاب العمل والعاملين؛ لخلق بيئة عمل صحية وأمنة، من خلال الكشف عن مخاطر الصحة والسلامة المهنية، وتطوير نهج عملية للقضاء على هذه المخاطر أو السيطرة عليها. (ترجمة بشرية).

(1) هي الأحرف الأولى من بعض العبارات التي جرت العادة على اختصارها.

2- The PAC examines the value for money of Government projects, programs and service delivery.

أ- يفحص PAC القيمة مقابل المال للمشاريع والبرامج الحكومية وتقديم الخدمات. (ترجمة آلية).

ب- تفحص (لجنة الحسابات العامة) القيمة مقابل المال للمشاريع والبرامج الحكومية وتقديم الخدمات. (ترجمة آلية).

3- BOT:

- Back On Topic

أ- العودة للموضوع-ترجمة آلية.

- BOT: Build-Operate-Transfer

ب- بناء - تشغيل - نقل (ترجمة بشرية).

4-PPP:

- Paycheck Protection Program أ

أ- برنامج حماية شيكات الرواتب (ترجمة آلية).

- Public-private partnership

ب- الشراكة بين القطاعين العام والخاص (ترجمة بشرية).

- يجد المترجم الآلي صعوبة في ترجمة المختصرات، وغالبًا تتم الترجمة بشكل مغلوط، كما هو وارد في المثال أعلاه. بالإضافة لذلك يستحيل على المترجم الآلي ترجمة بعض العبارات ذات التركيب اللغوي الغامض من الناحية النصية، أو الدلالية، أو النحوية كما هو موضح أدناه:

1- Public anger over passport red tape.

أ- الغضب العام على الشريط الأحمر جواز السفر (ترجمة آلية).

- ب- الغضب العام بسبب روتين استخراج جواز السفر (ترجمة بشرية).
- 2- Roster of staff to the company garage office vehicles
- أ- قائمة موظفين مرآب سيارات مكتب الشركة (ترجمة آلية).
- ب- قائمة الموظفين المرسله إلى مكتب المركبات ذي الصلة بمرآب سيارات الشركة. (ترجمة بشرية).
- 3- You have to make concession and breathe a huge relief.
- أ- عليك أن تقدم تنازلاً وتتنفس راحة كبيرة (ترجمة آلية).
- ب- عليك أن تقدم تنازلاً وتتنفس الصعداء (ترجمة بشرية).
- 4- High-level government bodies.
- أ- جثث حكومية رفيعة المستوى (ترجمة آلية).
- ب- هيئات حكومية مرموقة (ترجمة بشرية).
- 5- A red-herring speech.
- أ- خطاب الرنجة الحمراء (ترجمة آلية).
- ب- خطابٌ مضلل (ترجمة بشرية).
- 6- Banding for a position in the organization
- أ- النطاقات للحصول على منصب في المنظمة (ترجمة آلية).
- ب- بيان بالأجر الخاص براتب الوظيفة في الشركة/ المؤسسة. (ترجمة بشرية).
- تتزايد وتيرة أخطاء الترجمة الآلية عند تعدد مرادفات الكلمة الواحدة؛ لأن المترجم الآلي لا يستطيع دائماً اختيار المرادف المناسب كما هو موضح أدناه:
- 1- If everyone understands.
- الكلمة الأولى في الجملة لها عدة مرادفات في اللغة العربية: (لو - إن - إذ).

2- Analogy

هذه الكلمة متعددة المعاني، عند ترجمتها إلى اللغة العربية على النحو التالي: قياس - استخلاص - تمثيل - تناظر جزئي - تناظرٌ وظيفي - تشابهٌ جزئي - تشابه وظيفي - تشبيه - تماثل - أسلوب القياس في اتخاذ القرارات - عدّاد السرعة التماثلي - تحويل تناظري إلى رقمي تناظري - قياس صوتي تعاملي - تشابه مع انبعاث الصفيح - تمريرة النطاق - قياس العلة - الاستحسان بالقياس الخفي - الاستنتاج عن طريق القياس - طريقة التناظر الكهربائي.

3- Ultimate

كذلك هذه الكلمة على سبيل المثال، لها مرادفات عديدة في اللغة العربية كما يلي: أبعد، أخير - أساسي - أعلى - أقصى - أولي - آخر - تام - جوهري - ختامي - ذروة - قمة - مُطل - مطلق - نهاية المطاف.

4- Action

أخيراً هذه الكلمة المألوفة، لها مرادفات متنوعة عند ترجمتها إلى العربية، كما يلي: إجراء - حدث - حركة - سلوك - سَهم مالي - عَكْس - سُكُون - فعل - تَصَرَّف - تَغَوُّط - خُصُوم - حدث - فعلة - قضية أمام المحاكم - مُبَادَاة - مُبَادَرَة - أثر - أداء - تأثير، - تَبَرُّز - تَحَرُّك، - تداعي - تدبير.

تزداد الترجمة الآلية سوءاً في النصوص ذات السياق الأدبي، خاصة الشعر، وكذلك في النصوص العلمية، والتقنية، والطبية التي تضم مصطلحات جديدة، غير متاحة في المعاجم التي تغذي المترجم الآلي، أو النصوص العلمية التي لم يسبق ترجمتها إلى اللغات المستهدفة.

تأثير الترجمة الآلية والترجمة بمساعدة الحاسوب على المترجمين

تسبب اختراع الترجمة الآلية في مخاوف كبيرة لمعظم المترجمين؛ بسبب افتراض أن الهدف من إنشاء وتطوير الترجمة الآلية، كان استبدال المترجم البشري بالآلة. في الحقيقة لا تعتبر الترجمة الآلية تهديدًا للمترجم البشري؛ لأن أجهزة الحاسوب لن تحل محل المترجمين أبدًا، ولكن المترجمين الذين يستخدمون أجهزة الحاسوب سيحلون محل المترجمين الذين لا يفعلون ذلك. منذ البداية أحرزت الترجمة الآلية والترجمة بمساعدة الحاسوب تقدمًا هائلًا، ولكن عند مقارنة الترجمة الآلية والترجمة البشرية، تبين وجود تباين في جودة الترجمة بسبب بعض المعضلات اللغوية والثقافية.

الترجمة الآلية وتعدد المعاني داخل السياق:

تعتبر ظاهرة تعدد المعاني في اللغة الإنجليزية شائعة جدًا، أي أن الكلمة لها معانٍ مختلفة عند ترتيبها مع كلمات مختلفة، أو وضعها في سياقات متباينة. على سبيل المثال، نجد لكلمة "فرع" عدة معانٍ مختلفة، مثل "فرع شجرة"، أو "تقسيم إداري في منظمة أو هيئة"، إلخ. يختار المترجم البشري المعنى وفقًا للسياق. في المقابل لا تستطيع الترجمة الآلية تحديد المعنى من خلال السياق. تترجم الآلة عادة من خلال المعنى الأول الذي يظهر في المعجم المخزن في ذاكرة النظام.

وظائف ومعاني الأفعال:

في اللغة الإنجليزية تضطلع الأفعال بعدة وظائف، مما يطرح مشكلة أخرى للترجمة الآلية. على سبيل المثال كل من الفعلين التاليين، يستخدمان كأفعال رئيسية مفاهيمية أو كأفعال مساعدة حسب موقعهما في الجملة:

"Do" - "have" are used as both notional and auxiliary verbs.

في الترجمة الآلية يتعامل الحاسوب مع هذين الفعلين بوصفهما أفعالاً أساسية فقط، في سياقات عديدة؛ مما يتسبب في الإخلال بالمعنى. وعلى عكس المترجم البشري فإن الآلة أحياناً تكون غير قادرة على تحليل قواعد الجملة بأكملها للحكم على ما هو الفعل الأصلي أو الصفة أو الظرف أو الحال.

الغموض النحوي:

يظهر الغموض النحوي للغة الإنجليزية في مواقع شتى، ومنها كيفية استخدام حروف الجر التي تعتمد على السياق، كذلك يظهر الغموض النحوي في أشكال عديدة أخرى. مثلاً، يمكن أن يستخدم المتحدث باللغة الإنجليزية نفس الجملة أو العبارة للتعبير عن مضامين مختلفة. وبالمثل نجد نفس الوضعية الغامضة بالنسبة لنفس العبارة لدى المستمعين المختلفين. بمعنى آخر: يمكن أن يكون لدى المستمعين المختلفين ردود أفعال مختلفة عند سماع نفس العبارة. الآلة لا تفهم معنى التواصل الإنساني الذي يحدث بين البشر في سياقات محددة. ولذلك لا يمكنها تعديل النص ليواكب الموقف أو المتلقي. على عكس المترجم البشري الذي يمكنه معالجة الغموض النحوي، بعد تحليل السياق، فإن الترجمة الآلية لا تستطيع القيام بذلك مما يؤثر بشكل سلبي على المنتج الترجمي.

يمكن أن يؤدي تطبيق الترجمة الآلية والترجمة بمساعدة الحاسوب إلى تسريع الترجمة وتلبية المتطلبات، ومع ذلك لا تزال الترجمة الآلية، والترجمة بمساعدة الحاسوب، تواجهان الكثير من المشكلات. وبالتالي نحن بحاجة إلى التعامل بشكل صحيح مع العلاقة بين الإنسان والآلة. لا يمكن للمترجمين البشريين الاعتماد كلياً على الآلات والأدوات المحوسبة. في الوقت الحاضر يمكنهم استخدام الترجمة الآلية، باعتبارها مرجعاً أو مسودة تحتاج إلى مزيد من التنقيح من قبل المترجمين أنفسهم.

حجم ومحتوى المعجم الآلي:

اللغة الطبيعية هي نظام هائل من الكلمات التي لا نهاية لها، ويتم ابتكار كلمات جديدة باستمرار. غالبًا ما يتم الاستشهاد بالكلمات والجمل من الأدب والشؤون الدولية والأحداث الكبرى والأفلام، ولذلك فإن أحدث برامج الترجمة الآلية لا يمكنها تضمين جميع الكلمات والعبارات الجديدة واستيعابها. ونظرًا لأن القواميس والمعاجم الآلية غير قادرة على جمع كل الكلمات، فإذا ظهرت الكلمات الجديدة أو غير المؤلفوة التي لم يتم تجميعها في هذه القواميس في النص المصدر، فمن المحتمل أن يؤدي ذلك إلى ارتكاب الأخطاء إبان الترجمة الآلية.

النصوص ذات الدلالات الثقافية:

تُظهر التحليلات أن هناك فرقًا كبيرًا بين المترجمين البشريين وآلات الترجمة، عندما يكون للنص المصدر دلالات ثقافية، أو دينية، أو أدبية. يمكن للمترجمين البشريين تعويض المعلومات المحذوفة بطرق مرنة، مثل كتابة الملاحظات، وإعادة الصياغة، والإضافة، وما إلى ذلك. وبالمقارنة فإن نظام الترجمة الآلية، ونظام الترجمة بمساعدة الحاسوب، يفتقران إلى فهم الثقافات المختلفة، وهما غير قادرين على التغلب على الحواجز الثقافية بين اللغات المختلفة. يتأثر المترجمون بالتقاليد الثقافية المختلفة، وينقلون نفس الفكرة بأساليب متنوعة، عكس الترجمة الآلية. لذلك أثناء الترجمة بين اللغات المختلفة، لا يمكن للمترجمين التقليديين التغاضي عن الجوانب الثقافية وتأثيرها على الترجمة.

على الرغم من ذلك فإن تطبيق برامج الترجمة الآلية والحاسوب في ممارسة الترجمة، يمكن أن يساعد المترجمين الأذكياء من ذوي الخبرات على تحسين ترجماتهم.

تقييم الترجمة الآلية (جوجل ترانزليت)

يمكن لهذه التكنولوجيا أن تزيل حاجز اللغة، وتوفر الفرصة لأي شخص للتواصل مع أي شخص آخر. جوجل ترانزليت هو تطبيق مجاني عبر الإنترنت، تقدمه شركة جوجل يسمح للمستخدمين بترجمة الكلمات والجمل والمستندات، وحتى مواقع الويب في الحال. تشمل ترجمة جوجل أنظمة الحاسوب التي تنتج ترجمات بناءً على أنماط تتضمن أعدادًا كبيرة من النصوص، وتعتمد على سلسلة من القواعد تخص لغات معينة. بمعنى آخر يتم تخزين أعداد كبيرة من النصوص في ذاكرة جوجل؛ ليتم استردادها من قبل المترجمين عند الحاجة. إن ترجمة جوجل قد غيرت نهجها من الترجمة الآلية القائمة على الإحصاء، إلى الترجمة الآلية العصبية؛ لسد الفجوة بين الترجمة البشرية والآلية، والتعامل مع المشكلات التي تواجه الترجمة الآلية، مثل زيادة أعداد الكلمات النادرة والمصطلحات الغريبة غير المألوفة.

إن تقييم الترجمة الآلية مهمة شاقة للغاية. وقد ظل التقييم البشري للترجمة الآلية هو الوسيلة الوحيدة لتوفير التغذية الراجعة اللازمة لمواصلة تقدم هذا التطور التقني المستمر في مجال الترجمة. لقد ظل التقييم البشري الذي يعتمد على القياس اليدوي لجودة الترجمة الآلية عنصرًا أساسيًا في تطوير أنظمة الترجمة الآلية. على الرغم من المخاوف بشأن الأحكام البشرية، فإن التقييم البشري هو معيار ذهبي لتقييم جودة الترجمة الآلية، ويستخدم كخط أساس يتم من خلاله الحكم على مقاييس نجاح هذه الترجمة، ولا يوجد مقياس تقييم آخر ليحل محل الأحكام البشرية عن الترجمة الآلية. ولتقييم جودة مخرجات الترجمة الآلية، يجب مراعاة العديد من المعايير والقواعد، مثل: الملاءمة الدلالية، وسلاسة اللغة، والقابلية للفهم.

بغض النظر عن جودة الترجمة التي ينتجها نظام جوجل ترانزليت، التي تشتمل اليوم على أكثر من مائة لغة، وإلى جانب ترجمة النصوص توفر خدمة الترجمة من جوجل لمستخدميها العديد من التسهيلات، مثل: نطق النصوص المترجمة، وإبراز الكلمات المقابلة في النص المصدر والنص الهدف، والكشف التلقائي عن اللغة، واقتراح المزيد من الترجمات، ولذلك قامت جوجل ببناء مجتمع ترجمة نشط، وقدمت خدمة ترجمة معقولة، تمكن المستخدمين من اختيار ما يصل إلى خمس لغات لتقديم ترجمات شبه دقيقة أحياناً.

في ذات السياق يعد تقييم جودة مخرجات الترجمة الآلية أمراً مكلفاً، ويستغرق وقتاً طويلاً؛ نظراً للعديد من الجوانب التقنية. ذكر الباحثون أن المترجمين البشريين هم المفتاح الرئيسي لتقييم جودة الترجمة الآلية، وأيضاً لمعالجة السؤال الذي لم تتم الإجابة عليه حتى الآن: متى وكيف يتم استخدام الترجمة الآلية لتسهيل عمل الترجمة الاحترافية؟

يجب أن نضع في الاعتبار ثلاثة معايير مهمة لتقييم جودة مخرجات الترجمة الآلية، بما في ذلك الملاءمة الدلالية، وسلامة اللغة، والقابلية للفهم. وينظر إلى معيار الملاءمة الدلالية على نطاق واسع على أنه المعيار الذهبي لتقييم جودة الترجمة الآلية. ويؤدي التقييم إلى معرفة ما، إذا كانت ترجمة جوجل، مثلاً، تعتبر بمثابة مساعدة، أو تهديد للمترجمين البشريين. من ناحية أخرى تساعد نتائج التقييم، الطلبة الدارسين للترجمة، والمترجمين المتدربين، على تحسين معرفتهم باستخدام الترجمة الآلية في ترجماتهم، ولفت انتباه معلمي الترجمة، إلى ضرورة تعديل مناهج الترجمة، وتصميم مساقات جديدة تأخذ في الاعتبار ممارسات الترجمة الآلية، والتقنيات الجديدة. علاوة على ذلك، تشجع نتائج التقييم المترجمين المحترفين على الإسهام في تحسين وظيفة الترجمة وفق برامج وأنظمة جوجل.

تقييم الترجمة الآلية: آيات القرآن نموذجاً:

الكلمة العربية "القرآن" تعني حرفياً "التلاوة"، ويعتبر المسلمون في جميع أنحاء العالم القرآن الكريم كلمة الله الحرفية، التي أملاها الله من خلال كبير الملائكة جبريل، على النبي محمد، وينقسم القرآن الكريم إلى 114 سورة، وتتكون كل سورة من عدة آيات. يختلف طول هذه السور اختلافاً كبيراً، حيث يُقاس طول السور بعدد الآيات التي تحتوي عليها، والتي تختلف من آيات قليلة، كما في السورة الأولى (الفاتحة)، التي تتكون من سبع آيات فقط، إلى مئات الآيات ضمن نفس السور القرآنية، مثل السورة الثانية (البقرة)، التي تتكون من 286 آية. العدد الإجمالي للآيات القرآنية 6236 آية. القرآن أصلاً مدون ومكتوب باللغة العربية، ومن أجل الحفاظ على أصالة اللغة العربية العالية، يسعى مترجمو القرآن إلى اللغات الطبيعية الأخرى؛ ليكونوا متسقين من حيث المحتوى، والقواعد النحوية، والدلالات، والسياق، مع خصائص لغة القرآن البلاغية. ومع ذلك فمن المعروف أنه عند ترجمة الكلمات والعبارات، من لغة إلى أخرى، يمكن استخدام أكثر من كلمة لترجمة، أو تفسير كلمة معينة. بالإضافة إلى ذلك، وبسبب طبيعة التباينات بين اللغات الطبيعية المختلفة في كثير من الحالات، قد لا تُنتج الترجمة كلمة بكلمة، عبارات ذات معنى. على هذا النحو قد يكون المعنى العام للنص، أو العبارة، أو التفسير، ضرورياً لشرح معنى آيات القرآن، على الرغم من حقيقة أن هذا المعنى قد لا يكون ترجمة، أو تفسيراً حرفياً، أو ترجمة كلمة بكلمة متطابقة مع الأصل.

لقد اكتسبت اللغة العربية الاهتمام مؤخراً؛ بسبب العديد من العوامل الدينية والسياسية. ولا يزال الناطقون بغير اللغة العربية بحاجة إلى ترجمة القرآن الكريم؛ لمعرفة معناه، أو لفهمه. لذلك يستخدم العديد من غير الناطقين باللغة

العربية، أنظمة الترجمة الآلية عبر الإنترنت؛ لترجمة القرآن الكريم من اللغة العربية، إلى اللغات الطبيعية الأخرى. وتحتاج هذه الأنظمة إلى التقييم؛ لمعرفة جودة مخرجاتها. في هذه الدراسة، تم إجراء تقييم آلي لاستخدامات بعض أنظمة الترجمة الآلية المجانية، في ترجمة النص القرآني العربي، إلى اللغة الإنجليزية. اعتمدت الدراسة على مجموعة البيانات التي تحتوي على النصوص العربية الأصلية، لعدة سور مختارة من القرآن الكريم.

حتى يتم رصد جودة الترجمة الآلية المجانية، تم اختيار عدة مقاييس لتقييم الخصائص الأساسية للنص المترجم، وهذه المقاييس تختص بالتقييم التلقائي لأنظمة الترجمة الآلية. وترصد وتقارن العلاقة بين ناتج نظام الترجمة الآلية وترجمة المراجع، والعناصر البشرية المهنية، بناءً على كيفية اختيار، وترتيب الكلمات، والعبارات، وتحلل مقاييس التقييم أوجه التشابه بين (مخرجات) الترجمة الآلية والترجمة البشرية.

من الجدير بالذكر أن تقييم أنظمة الترجمة الآلية أمر شخصي، وليس موضوعيًا، سواء تم إجراؤه بواسطة خبير بشري محترف، أو بواسطة أنظمة محوسبة باستخدام مقاييس متعددة، مثل المقياس ثنائي اللغة المستخدم في هذه الدراسة. وقد تختلف جودة الترجمة الآلية؛ بسبب الخوارزميات المختلفة المستخدمة، بواسطة أنظمة الترجمة الآلية، وقد ترجع الاختلافات التلقائية في تقارير المقاييس المتعددة المستخدمة لتقييم فعالية أنظمة الترجمة الآلية المختلفة إلى عدة أسباب على النحو التالي:

أولاً- قد تعتمد المقاييس في عملية التقييم على خوارزميات الترجمة الآلية أو حجم القاموس/المعجم أو صحته، أو طبيعة اللغة وطبيعة نص الإدخال.

ثانيًا- لا يدعي أحد أن تقييم أنظمة الترجمة الآلية بطريقة إلكترونية، هو

أفضل من تقييم المهنيين البشريين. ومع ذلك فإن هذه الأدوات، والأنظمة، والمقاييس الإلكترونية، أقل تكلفة وأسرع عند مقارنتها بالجهود البشرية.

ثالثاً- يعتبر مقياس التقييم ثنائي اللغة المزدوج أحد مقاييس التقييم التلقائي الأقدم والأكثر انتشاراً، والذي تم تقديمه في عام 2002، وتعود شعبية إلى سهولة حساب تقييماته. وتم تقديم العديد من إصدارات هذا المقياس في دراسات مختلفة، ويستخدم المقياس آلية أحادية معدلة لحساب الارتباط بين الترجمات في اللغات المستهدفة ولغات المصدر.

رابعاً- يعتمد المقياس عند تقييم جودة الترجمة، والخصائص الأساسية للنص على اختيار الكلمات، وموضع الكلمة لتقييم أنظمة الترجمة الآلية تلقائياً. خامساً- إن مقياس تقييم الترجمة، هو مقياس موجه نحو الاستدعاء والاسترجاع، ويحاول حل بعض المشكلات مثل التعامل مع مطابقة دقيقة لمواقع الكلمات، ويمكن أن يتعامل النظام مع أكثر من لغة مصدر، إبان عملية التقييم. ويُجري النظام موازنة بين الكلمات والعبارات عندما يختلف ترتيب الكلمات في النصوص الأصلية عن ترتيبها في النصوص الهدف.

سادساً- من خلال تقييم جودة الترجمة الآلية فيما يخص بعض آيات القرآن ذات العبارات الطويلة، تبين أن نسبة الترجمة الصحيحة أقل من 22% وأحياناً تصل نسبة الترجمة الصحيحة إلى أقل من 10% مع وجود مشكلات في برامج الترجمة الآلية، خاصة في قدرتها على التمييز بين أنواع الجمل ووظائفها.

(Declarative, interrogative, exclamatory, and imperative)

(توضيحية، استفهامية، تعجبية، جمل أمر)

تشير نتائج التقييم عدم نجاعة الترجمة الآلية في ترجمة النصوص الدينية بدقة.

الترجمة الآلية والذكاء الاصطناعي:

أصبح الذكاء الاصطناعي موضوعًا ساخنًا في الآونة الأخيرة. ويبدو أنه لا يمكن للمرء قراءة أي منفذ إخباري رئيسي دون الاصطدام بقصة، أو اثنتين على الأقل من قصص الغلاف حول ما يعنيه الذكاء الاصطناعي للمستقبل، ناهيك عن القفزات التي حققها هذه التقنية بالفعل. تخبرنا هذه الدراسات والمقالات، بأن أنظمة التشغيل الآلي للحاسوب ستغير تمامًا نطاق ما هو ممكن، وكيفية تفاعل البشر مع الآلات، بفضل تطبيقات الذكاء الاصطناعي. وكلما سمعنا أكثر عن الذكاء الاصطناعي، بدا أننا على أعتاب حياة تسيطر عليها الآلات بالكامل تقريبًا. وإلى حد ما فإن هذه التوقعات ليست مستحيلة، ومن العدل أن نقول إن بعض الضجيج المحيط بالذكاء الاصطناعي معقول ويمكن قبوله.

بلا ريب فإن التقدم الذي تم إحرازه في مجال الذكاء الاصطناعي في العهد الحديث، مثير للإعجاب بالتأكيد. وفي مجال صناعة اللغة، كل الأنظار تتجه إلى الترجمة الآلية، وما يمكن أن يعنيه الذكاء الاصطناعي لمستقبل الترجمة. على وجه الخصوص كانت هناك مناقشات ومجادلات حول ما إذا كانت الترجمة الآلية العصبية ستلغي الحاجة إلى مترجمين بشريين. وبفضل المقالات الإعلامية السائدة حول نظام الترجمة الآلية العصبية الجديد من جوجل، يعتقد الكثيرون أن غياب المترجمين البشريين في المستقبل القريب هو احتمال قائم. لكن هذا الإصرار على مستقبل بلا مترجمين، لم يتم ترسيخه كما هو مفترض. لقد قطعت الترجمة الآلية شوطًا طويلاً بالتأكيد، لكن التكنولوجيا الجديدة ليست في وضع يمكنها من استبدال البشر تمامًا وطردهم خارج أسوار التاريخ.

أدت تقنية الذكاء الاصطناعي لتحسين منتجات أنظمة الترجمة الآلية العصبية، ولكن هذا التطور لا يشكل نقلة نوعية كما تم الترويج له إعلاميًا. لقد أعلنت

جوجل عن عمليات تطوير كبيرة، وكانت خطوة جذبت اهتمام محترفي الصناعة ووسائل الإعلام. ادّعت جوجل أنها زادت من دقة الترجمات بشكل كبير، من خلال الانتقال إلى نموذج الشبكة العصبية العميقة، والابتعاد عن عمليات الترجمة الإحصائية. بينما يمكن أن تركز الترجمة الآلية الإحصائية، أو الترجمة الآلية القائمة على العبارات فقط، أو على أجزاء من الجمل، يقوم هذا النظام الجديد بفحص الجملة بأكملها عند الترجمة، ويتذكر الجزء الأول من الجملة عندما ينتهي من فك تشفير الكلمة الأخيرة. كما ادعت جوجل أن دقة الترجمة اقتربت من الترجمة البشرية.

دحضت الأبحاث اللاحقة مزاعم جوجل ووصفتها بأنها مضللة ودعائية، بهدف السيطرة على السوق. وأكدت وكالة الفضاء الكندية، أن طرق اختبار دقة نظام جوجل الجديد يشوبها عيوب كبيرة. على وجه التحديد، فإن الأشخاص الذين أشرفوا على تجربة النظام الجديد لم يكونوا من المترجمين المحترفين العاملين في شركة جوجل حسب ادعائها، وإنما كانوا فقط مجموعة من المتحدثين بطلاقة للغة الهدف، والطلاقة ليست كافية لضمان الفهم الكامل لقضايا الترجمة.

تميل هذه الطريقة إلى تمييز الترجمة الصوتية بطلاقة، على الترجمة الدقيقة، ولكن هذه الطريقة لم تحدد عدد الأخطاء التي حدثت عند تطبيق نظام الترجمة الآلية الجديد. أثبتت النتائج، أن النظام الجديد في أحسن حالاته ينتج ترجمة تعادل مستوى الترجمة البشرية السيئة، ومنتجاته الترجمة ليست على مستوى الترجمة البشرية الجيدة، رغم اعتماده على أحدث أنظمة الذكاء الاصطناعي.

الخاتمة

ملاحظات فارقة:

أولاً- لقد اعتبر الكثيرون إعلان وصول نظام الترجمة الآلية الجديد من جوجل إلى مستوى الدقة البشرية حقيقة راسخة. على الرغم من أن هذا يبدو جيداً، إلا أنه بعيد عن الواقع.

ثانياً- إن إحدى المشكلات التي تعيق قدرة الترجمة الآلية على الترجمة بشكل صحيح، هي حقيقة أنها تفك شفرات النص جملة بجملة. هذا يعني أن الأخطاء تحدث إذا كان هناك حاجة إلى سياق من جملة سابقة؛ من أجل ترجمة مقطع لاحق بشكل صحيح.

ثالثاً- أضف إلى ذلك حقيقة أن الترجمة الآلية ليست قادرة بعد على التقاط الفروق الدقيقة، أو الاستعارات الموجودة في الكلام البشري، وهناك هامش كبير للخطأ.

رابعاً- تشير مقالة مجلة الإيكونوميست الشهيرة⁽¹⁾ في هذا السياق، أنه يمكن لعلماء الحاسوب استخدام البيانات لتدريب الآلات، ولكن لا يمكنهم تدريب الآلات على حسن الإدراك والفطرة السليمة. على وجه الخصوص، لا تمتلك أجهزة الحاسوب المعرفة الكاملة بأحداث، وقضايا العالم الحقيقي، من أجل فهم النص الذي قد يكون غامضاً.

خامساً- تستخدم نفس المقالة، مثالاً للافتة عامة، موجودة على نافورة كتب عليها "هذه ليست مياه للشرب". قد يفهم الإنسان الإشارة على أنها تعني أنه لا

(1) Cited in Saddik Gohar. Scandals of Machine Translation. Cairo: Eyes Press, 2019. P.187.

يجب أن يشرب الماء، ومع ذلك قد يأخذ الحاسوب هذه العبارة - عند ترجمتها آليًا- على أنها تعني أن النافورة لا تنتج لنا مياهًا للشرب.

سادسًا- تظهر أوجه القصور في الذكاء الاصطناعي في المساعدين الرقميين، مثل (سيري) عندما تطلب من هاتف (الآي فون) الخاص بك العثور على مطعم لتناول الطعام، فإنه لا يمتلك "الذكاء" لعرض المطاعم التي تعمل فعليًا، وإذا أجاب (سيري) بالفعل على سؤالك بشكل صحيح، فقد تجد مواقع مطاعم محتملة مغلقة منذ سنوات.

سابعًا- الدماغ البشري هو عضو لم يفهمه العلماء تمامًا بعد، وهو غير قادر على فك شفرة جميع العمليات المعقدة التي تسمح لنا بجمع وتحليل كميات من المعلومات في وقت واحد. إذا كنا لا نزال غير قادرين على اكتشاف كل تعقيدات الدماغ البشري، فكيف يمكننا أن نجعل آلة تكرر شكل الدماغ بشكل عمدي؟!

ثامنًا- على الأقل في الوقت الحالي لا تستطيع أنظمة الترجمة الآلية، لا من قريب، أو بعيد استبدال المترجمين البشريين. حتى لو استطاع المطورون إنشاء نماذج من الترجمة الآلية العصبية تحاكي، وتطابق الدماغ البشري في المعرفة العامة، والتعرف على الفروق الدقيقة والقدرة على التفكير، فسوف يستمر البشر في أن يكونوا جزءًا محوريًا من عملية الترجمة.

تاسعًا- إذا كان هناك أي شيء لصالح استمرار تطوير الترجمة الآلية، فقد ثبت أنها أصبحت تقدم مساعدة للبشر، وتسهل عليهم ترجمة بعض المستندات ذات المصطلحات الثابتة.

عاشرًا- كما ورد في مقال الإيكونوميست⁽¹⁾ فإن "مترجم المستقبل"، لن يكون

(1) Cited in Saddik Gohar. Scandals of Machine Translation. Cairo: Eyes Press, 2019. P.65.

نظام الترجمة الآلية المتطورة، بل سيكون إنساناً يعمل كمسؤول عن مراقبة الجودة. وسيكون هذا الشخص مسؤولاً عن تحليل النص المترجم من خلال الترجمة الآلية ومعرفة ما يحتاجه من مراجعة وتحديث.

أخيراً- سوف تغير أنظمة ونماذج الترجمة الآلية الجديدة طريقة أداء المترجمين لعملهم، لكنهم لن يفقدوا وظائفهم باعتبارهم مترجمين، ولن يفقد المدققون اللغويين مكانتهم كذلك. مع المستندات، والكتب، والتقارير، التي يتعذر على الترجمة الآلية التعامل معها، وسوف تظل الحاجة قائمة للاستعانة بالمترجمين في سد هذه الثغرات.

References

- 1- Allen, Jeffery. "Post-editing". Harold Somers, ed. *Computers and Translation. A Handbook for Translators*. Amsterdam: John Benjamins, 2003.
- 2- Austermühl, Frank. *Electronic tools for translators*. Manchester: St. Jerome Publishing, 2001.
- 3- Denkowski, Michael. *Machine Translation for Human Translators*. Unpublished PhD Dissertation. Pittsburgh, PA: Carnegie Mellon University, 2015.
- 4- Dunne, Keiran J. ed. *Perspectives on Localization*. Amsterdam: John Benjamins Publishing, 2006.
- 5- Gambier, Yves and Henrik Gottlieb. eds. *(Multi) Media Translation*. Amsterdam: John Benjamins, 2001.
- 6- Gohar, Saddik. *Scandals of Machine Translation*. Cairo: Eyes Press, 2019.
- 7- Kauchak, David. *Contributions to research on machine translation*. Unpublished PhD Dissertation .California: University of California in San Diego, 2006.
- 8- Krings, Hans P. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Kent, Ohio: The Kent State University Press, 2001.
- 9- Minako, O'Hagan and Ashworth, David. *Translation-Mediated Communication in a Digital World*. Clevedon: Multilingual Matters, 2002.
- 10-Somers, Harold ed. *Computers and Translation. A Handbook for Translators*. Amsterdam: John Benjamins. *Translation Studies*, 2003.Zetzsche, Jost. *The Translator's Tool Box. A Computer Primer for Translators*. Winchester Bay: International Writers' Group, 2005.



**« نَمْدَجَةُ التَّمَثَلَاتِ الصَّرْفِيَّةِ الْعَرَبِيَّةِ،
وَتَعْلِيمِيَّتُهَا بِوَاسِطَةِ الْحَاسُوبِ، فِي ضَوْءِ
تَحْلِيلِ لَأُمْتَلَةِ نَصِيَّةِ تَطْبِيقِيَّةِ »**

**Modeling and teaching Arabic in a Research of
an Automatic Morphological Analysis
Phenomena of Arabic Texts**

إعداد

د. حُسَامُ الدِّينِ سَمِيرِ عَبْدِ الْعَالِ مُحَمَّدٍ
الأستاذ المشارك في جامعة محمد بن زايد للعلوم الإنسانية.

الملخص

يَنَدِجُ هذا البَحْثُ ضِمْنَ المُعَالَجَةِ الآلِيَّةِ لِلغَةِ العَرَبِيَّةِ، فَيَقَدِّمُ طَرِيقَةً وَمُقْتَرَحًا مِنْ خِلَالِ تَصَوُّرٍ آليٍّ يَحْتَوِي مَعْلُومَاتٍ لِإِنجَازِ تَحْلِيلٍ صَرْفِيٍّ لِلنَّصِّ العَرَبِيِّ. لَقَدْ قَامَتْ فِي العَصْرِ الحَاضِرِ بَحُوثٌ صَرْفِيَّةٌ تَقْتَضِي أُسُسًا دِرَاسِيَّةً حَدِيثَةً، كَمَا أَنَّهَا تَدْعُو إِلَى فَحْصِ الظُّوَاهِرِ الصَّرْفِيَّةِ العَرَبِيَّةِ وَاخْتِبَارِ مُكَوِّنَاتِ النِّظَامِ اللُّغَوِيِّ. وَمِنْهَا دِرَاسَةُ Christian Gaubert (كْرِيسْتِيَان جَوْبِير)، حَيْثُ تُمَثِّلُ تِلْكَ الدِّرَاسَاتُ مَرَحَلَةً مُهِمَّةً فِي النِّظَرِيَّةِ الصَّرْفِيَّةِ العَامَّةِ، وَبِخَاصَّةٍ مَنَاجِهَا المَبْتَكِرَةِ فِي وَصْفِ التَّغْيِيرَاتِ الصَّرْفِيَّةِ فِي اللُّغَةِ العَرَبِيَّةِ هَذَا مِنْ نَاحِيَةٍ. وَمِنْ نَاحِيَةٍ أُخْرَى يَكْتَسِبُ هَذَا المَوْضُوعُ أَهْمِيَّةً كَبِيرَةً فِي الدِّرَاسَاتِ اللُّغَوِيَّةِ، حَيْثُ جِدَّةُ المَنْهَجِ الوُظَيْفِيِّ المُطَبَّقِ فِيهَا، وَبَلُورَةُ مَفَاهِيمِ صَرْفِيَّةٍ وَنَحْوِيَّةٍ جَدِيدَةٍ. كَمَا يَهْدَفُ البَحْثُ إِلَى تَقْدِيمِ دِرَاسَةٍ عِلْمِيَّةٍ تَطْبِيقِيَّةٍ لاسْتِخْدَامِ الذِّكَاءِ الاصْطِنَاعِيِّ فِي مُعَالَجَةِ الصَّرْفِ العَرَبِيِّ. وَمِنْ أَهَمِّ نَتَائِجِ الدِّرَاسَةِ الإِجَابَةِ عَنْ جُمْلَةٍ مِنَ التَّسْأُولَاتِ حَوْلَ طَرَائِقِ الوُصُولِ البَنِيَوِيِّ إِلَى مَا هِيَاتِ النُّصُوصِ العَرَبِيَّةِ، وَكَيْفِيَّاتِ مُقَارِبَتِهَا صَرْفِيًّا، وَأَشْكَالِ التَّحْلِيلِ المُسْتَخْدَمَةِ، وَأَنْمَاطِ الأَبْنِيَةِ الصَّرْفِيَّةِ، وَالمَنْهَجِ النَّسْقِيِّ المُوَظَّفِ، وَرَمُوزِهِ، وَعِلَامَاتِهِ، وَتَسْلُسِلِ وَسَائِطِهِ الحَاسُوبِيَّةِ المُتَعَدِّدَةِ. وَمَا مَدَى أَثَرِ رُؤْيَا كْرِيسْتِيَان جَوْبِيرِ فِي إِسْهَامِ المَنْهَجِ التَّطْبِيقِيِّ الحَدِيثِ فِي تَقْرِيبِ ظَوَاهِرِ الصَّرْفِ العَرَبِيِّ لِلْمُعَالَجَةِ الآلِيَّةِ؟ وَكَيْفَ يُنْظَرُ إِلَى تَحْلِيلِ النَّصِّ العَرَبِيِّ المَشْكُولِ وَغَيْرِ المَشْكُولِ، وَالتَّفْرِيقِ بَيْنَ الحَالَاتِ الَّتِي يَعْتَوُّهَا اللَّبْسُ البَنِيَوِيُّ، وَمَا الطَّرَائِقُ المُعْتَمَدَةُ فِي تَحْلِيلِ بَنِيَّةِ المُرَكَّبَاتِ الصَّرْفِيَّةِ الإِفْرَادِيَّةِ وَالسِّيَاقِيَّةِ؟

الكلمات المفتاحية: نَمَذَجَةُ التَّمَثُّلاتِ، التَّحْلِيلُ الصَّرْفِيُّ الآلِيّ، كْرِيسْتِيَان جَوْبِير، نُصُوصُ اللُّغَةِ العَرَبِيَّةِ، نَمَازِجُ تَطْبِيقِيَّة.

Abstract

This study is intended to introduce the attempt by the French linguistic researcher Christian Gaubert in a research of the phenomena of an automatic morphological analysis of Arabic texts through the modern morphological applied theory in analytic geometry of the structure of the Arab word, the extrapolation of our reading of its terms, the framing of its concepts and the discussion of its recent vision of the Arab potential in its morphological structure, offering that vision for the benefit of the Arab reader, seeking to manage a scientific discussion of what came in this attempt, and directing what appeared in it according to the modern morphological field.

Keywords: Modeling and teaching Arabic, Automatic morphological analysis, Christian Gaubert, Arabic texts, analytical descriptive study, applied examples.

المقدمة

أ. موضوع البحث:

مما لا ريب فيه، أنَّ التطورَ التكنولوجي وتقنيات المعارف والمعلومات قد باتت سمات متغلغلة في عالمنا الذي نحيا فيه، ذلك العالم الذي يسير بخطى واسعة، ومتسارعة نحو كشف كل مجهول، عبر نوافذ أكثر رحابة؛ لتحقيق آفاق علمية بعيدة. لقد دخلت هذه التكنولوجيا - وما تفرَّع عنها من ذكاءاتٍ صناعية، وما شملته من عوالم العلوم والفلسفات المختلفة - حياة الإنسان، ولغاته، وآدابه، وحقوقه المعرفية الأخرى، وأضحت محوراً مهماً تدور حوله مظاهر حياتية كثيرة. فالذكاء الاصطناعي اللغوي مثلاً IAL كما يُعرِّفه المتخصصون: هو مقدرة الآلة أو الحاسوب على تمثُّل العقل الإنساني، ومحاكاة وظائفه في سبيل معالجة المكوّنات اللغوية للنصوص، والتعامل مع مستويات تحليلاتها المختلفة: الصوتية، والكتابية / الخطّية، والبنوية / الصرفية، والتركيبية / النحوية، والدلالية / البراجماتية، والتعرُّف الدقيق على الجزئيات والحدود الدُّنيا والمعلومات الكميّة المبنوثة في النصوص والمُدوّنات؛ ومن ثمَّ نشوء ما يمكن أن يُطلق عليه بميلاد الفضاءات الإلكترونية المترابطة الفائقة.

تعدُّ المعالجة الآليّة الصّرفيّة لنصّ بالعربية من أهمّ فروع علم اللّغة، ومن المحاور الأساسيّة التي تُثير لدى الباحثين المُحدثين أهميّة كبيرة حول دراسة النّصوص الصّرفيّة العربيّة، وإمكانية مُعالجتها أُتوماتيكياً في دراسة مُرضية في أسسها وجوانبها. فاخترتُ هذا البحث لمحاولة الرّبط بين القديم المُتمثّل في أبنيّة الصّرف العربيّ، والمعاصر المُتمثّل في رؤية اللساني الفرنسي Christian Gaubert (كريستيان جوبير) لهذه الأبنيّة التّراثية من منظور حاسوبي، واختياره لنماذج

نَصِيَّة لُغَوِيَّة - تَنَوَّعَتْ بَيْنَ الْعَرَبِيَّةِ فِي مَسْتَوَاهَا الْمُعَاَصِرِ، مُتَمَثِّلًا فِي رِوَايَةِ "قَنْدِيلِ أُمِّ هَاشِمٍ" لِلْكَاتِبِ الرَّوَائِيِّ الْكَبِيرِ يَحْيَى حَقِّي، وَبَيْنَ أَمَثَلَةٍ أُخْرَى مُتَفَرِّقَةٍ مُنْتَقَاةٍ مِنَ الْعَرَبِيَّةِ فِي مَسْتَوَاهَا التُّرَاثِيِّ - مَادَّةٌ تَطْبِيقِيَّةٌ لِدِرَاسَتِهِ.

يَنْدَرُجُ هَذَا الْبَحْثُ ضِمْنَ الْمُعَالَجَةِ الْآلِيَّةِ لِللُّغَةِ الْعَرَبِيَّةِ، فَيَقْدِّمُ طَرِيقَةً وَمُقْتَرَحًا مِنْ خِلَالِ تَصَوُّرٍ آلِيٍّ يَحْتَوِي مَعْلُومَاتٍ لِإِنْجَازِ تَحْلِيلٍ صَرْفِيٍّ لِلنَّصِّ الْعَرَبِيِّ. لَقَدْ قَامَتْ فِي الْعَصْرِ الْحَاضِرِ بُحُوثٌ صَرْفِيَّةٌ تَقْتَضِي أُسُسًا دِرَاسِيَّةً حَدِيثَةً، كَمَا أَنَّهَا تَدْعُو إِلَى فَحْصِ الظُّوَاهِرِ الصَّرْفِيَّةِ الْعَرَبِيَّةِ وَاخْتِبَارِ مُكَوِّنَاتِ النِّظَامِ اللُّغَوِيِّ. وَمِنْهَا دِرَاسَةُ Christian Gaubert (كْرِيسْتِيَان جَوْبِر)، حَيْثُ تُمَثِّلُ تِلْكَ الدِّرَاسَاتُ مَرَحَلَةً مُهِمَّةً فِي النَّظَرِيَّةِ الصَّرْفِيَّةِ الْعَامَّةِ، وَبِخَاصَّةٍ مَنَاجِمَهَا الْمَبْتَكِرَةَ فِي وَصْفِ التَّغْيِيرَاتِ الصَّرْفِيَّةِ فِي اللُّغَةِ الْعَرَبِيَّةِ هَذَا مِنْ نَاحِيَةٍ. وَمِنْ نَاحِيَةٍ أُخْرَى يَكْتَسِبُ هَذَا الْمَوْضُوعُ أَهْمِيَّةً كَبِيرَةً فِي الدِّرَاسَاتِ اللُّغَوِيَّةِ، حَيْثُ جِدَّةُ الْمَنْهَجِ الْوُظَيْفِيِّ الْمُنَبِّطِ فِيهَا، وَبَلُورَةُ مَفَاهِيمِ صَرْفِيَّةٍ وَنَحْوِيَّةٍ جَدِيدَةٍ.

بـ أهداف البحث:

- تقديم دراسة علمية تطبيقية لاستخدام الذكاء الاصطناعي في معالجة الصرف العربي.
- تهيئة الصرف العربي وتطويره للمعالجة الآلية جديرًا بأن يَمُنَحَ الْعَرَبِيَّةَ رُقِيًّا وَوُضُوحًا وَانْتِشَارًا، إِلَى جَانِبِ دَعْمِهَا بِعَوَامِلِ الْقُوَّةِ وَالتَّيَبَاتِ، وَذَلِكَ فِي ضَوْءِ الْغَرَضِ مِنْ بِنَائِهَا أَنْيًّا وَمُسْتَقْبَلًا.

جـ إشكالية البحث:

يُجِيبُ الْبَحْثُ عَنْ جُمْلَةٍ مِنَ التَّسْأُلَاتِ حَوْلَ طَرَائِقِ الْوُصُولِ الْبَنِيَوِيِّ إِلَى مَاهِيَاتِ النُّصُوصِ الْعَرَبِيَّةِ، وَكَيْفِيَّاتِ مَقَارِبَتِهَا صَرْفِيًّا، وَأَشْكَالِ التَّحْلِيلِ الْمُسْتَخْدَمَةِ،

وأنماط الأبنية الصرفية، والمنهج النَّسقي المَوْظَّف، ورموزه، وعلاماته، وتسلسل وسائله الحاسوبية المتعدِّدة. وما مدى أثر رؤية كريستيان جوبير في إسهام المنهج التَّطبيقي الحديث في تقريب ظواهر الصَّرف العربي للمُعَالَجَة الآليَّة ؟ وكيف يُنظَرُ إلى تحليْلِ النَّصِّ العربيِّ المشكول وغير المشكول، والتَّفريق بين الحالات التي يعتمدها اللَّبس البنيوي، وما الطَّرائق المُعتمَدة في تحليل بُنية المُركَّبات الصَّرفيَّة الإفراديَّة والسِّيَاقِيَّة ؟

د- منهج البحث:

أمَّا المنهج الذي سلكناه في هذه الدراسة، فَيَتَلَخَّصُ في المنهج الوَصفيِّ التَّحليليِّ، وهو مَنهجٌ يهتمُّ بِوصفِ الظَّواهر، وباستقراء المسائل العلميَّة، ثُمَّ يقومُ بدراستها وتحليلها، مُستخدِمًا الوسائل التي بها تَنشَكُلُ اللغة، وتُمارَسُ بواسطة عددٍ مُحدَّدٍ مِنَ النَّاطِقِينَ بها في زمنٍ مُعيَّن. وهذا المنهج يُعِينُ على تَمَثُّلِ الصُّورة اللُّغويَّة لموضوع البَحْث، كما يُسَهِّمُ في تَوْضيحها بصيغَةٍ وافيةٍ وهيئةٍ جليَّةٍ؛ لِيَصِلَ إلى نتائجٍ علميَّةٍ تكشفُ النَّقَابَ عن جهودِ العالِمِ الفرنسيِّ كريستيان جوبير في مجالِ المُعالِجَة الآليَّة لمنظومةِ الصَّرفِ العربيِّ، وقيمة مثل هذه الجهود في ضوءِ ما تَوَصَّلَتْ إليه أبحاثُ الصَّرفِ في النِّظرياتِ اللُّغويَّة الحديثة.

هـ- الإطار العام لموضوع البحث:

وفي إطارِ الإجابة عن الأسئلة المطروحة في فرضيات البحث، فقد سِرْنَا وفق خُطَّةٍ تُعَرِّضُ الموضوعَ في تمهيدٍ وأربعة مباحث؛ حيث أَوْضَحْنَا في التمهيد التَّعريفَ بالمؤلِّفِ الفرنسيِّ، كما تَطَرَّقْنَا إلى تَحْدِيدِ المفاهيم الأساسيّة مِنْ مثل: بيان الفرق بين "الصَّرف" و"التَّصريف" و"الاشتقاق" لدى القُدَّماء والمُحدِّثين، باعتباره أساس موضوع الدِّراسَة، والمعايير التي أُعْتُمِدَتْ في اللِّسانياتِ الحاسوبيَّة لمفهوم "الصَّرف"

وتطبيقاته على النصوص العربية المُستهدفة آليًا.

وبعد ذلك كانت دراسة معايير التحليل عنده في المباحث الأربعة الأخرى. فالمبحث الأول: يعالج المصطلحات الرئيسة المُستخدمة لطريقة التحليل الصرفي إلى المؤلفات المباشرة أو التحليل التوزيعي. والمبحث الثاني: كان في التوصيف. والمبحث الثالث: في تطبيقات المنهج الآلي. والمبحث الرابع: في تحليل النصوص. وكُنّا نسعى في دراستنا إلى إدراك فاعلية هذه المحاولة في البحث الصرفي العربي، ومدى ارتقاءها له؛ لأنّ دراسة النصّ الصرفي العربي ليست دراسة الأشكال والصيغ والأبنية فحسب، بل يجب فيها دراسة ما وراء البنى والأشكال. وقد وجدنا أنّ دراسة هذه المحاولة تُطابق أسس المنهج الحديث. وختَمنا عملنا بخاتمة ضُمّنت مُجمل النتائج التي توصّلنا إليها.

دخل نظري للمحددات العامة والمفاهيم الأساسية

1- التعريف بالمؤلف:

عُرفَ Christian Gaubert (كريستيان جوبير)^(*) ببحوثه في التفكير اللساني

(*) تتلخص مسيرته البحثية في النقاط الآتية:

- مهندس مدني في المدرسة الوطنية العليا للتقنيات المتقدمة ENSTA، عام 1990 م.
- أستاذ علوم الحاسب والهندسة اللغوية وتطبيقاتها على العربية بالمعهد الفرنسي المصري لتعليم العلوم IFEPEMIS وبجامعة الأزهر في الفترة من عام 1990 م إلى 1992 م.
- مسؤول المعلوماتية واللغة في مؤسسة IFAO منذ عام 1992 م.
- عضو في فريق DATAT الخاص بدراسات وأبحاث المعالجات الآلية للغة العربية منذ عام 1994 م.
- المشاركة في حفريات النقلون- بدير الملاك في الفيوم (المركز البولندي للأثار المتوسطة) منذ عام 1998 م.

مجالات الاهتمام البحثي:

- الهندسة اللغوية والمعالجات الأتوماتيكية للغة العربية.
- دراسات لغوية لمخطوطات البردي.

البرامج المنجزة:

- دراسة لغوية حاسوبية لمجلل صرقي/نحوي للغة العربية المكتوبة التراثية والمعاصرة، استناداً إلى وصفٍ منتظم لمورفولوجيا وبناء التركيب العربي، والاعتماد على مبادئ نظرية الفرضيات الأتوماتيكية AFND.
- بالتعاون مع أندريه جاكاري (CNRS) وكلود أوديرت (جامعة إكس-اون- بروفانس بمارسيليا): تطوير أساليب التحليل اللغوي للبنى السطحية لنصوص العربية، وتطوير وسائل رفع اللبس باستخدام الأدوات الكتابية، والأدوات الرسومية، وتصورات المفاهيم النحوية.
- تطبيق هذا المجلل على مختلف المدونات العربية، ولا سيما تلك التي درست في مؤسسة IFAO.
- دراسة لغوية لوثائق المخطوطات العربية التي تم اكتشافها خلال الحفريات البولندية في النقلون- دير الملاك في الفيوم، التي تتألف من 50 مخطوطة وورقة لأسرة قبطية في الحقبة الفاطمية، فضلاً عن مجموعة من المدونات لأوراق وثائقية عربية.

من مؤلفاته ومؤتمراته العلمية:

- ملصق بعنوان: «تطوير برنامج حاسوبي لتنقية المعلومات النحوية التركيبية للغة العربية باستخدام برمجيات "صرقية"» بالتعاون مع أندريه جاكاري، وغسان مراد، وإبراهيم دجوه (المجلس الوطني للأبحاث بباريس)، 2003 م.

==

الحاسوبي في علوم اللغة العربية. وقد أسهم منذ سنة 1990 م من خلال أعماله العلمية وأطروحاته البحثية في إبراز الجوانب الحاسوبية في هذه العلوم. وقد دعم ذلك بجملة من الدراسات منها:

- *Analyse morphologique d'un texte arabe par ordinateur: méthode d'évaluation, résultats*, in *Annales islamologiques*, tome XXIX, Institut français d'archéologie orientale du Caire, 1995, pp. 283-310⁽¹⁾.

==

- تقديم "برنامج صرفية" باللغة العربية، الجامعة الأمريكية بالقاهرة، القسم العربي، ديسمبر 2004 م.
 - البرنامج التجريبي (صرفية) للهيئة الحاسوبية للغة العربية.
 - كلود أديبرت، كريستيان جوبير، وأندريه جاكربني، "الحد الأدنى من الموارد المتاحة للتحليل العربي: طريقة تفاعلية لبناء أوتوماتيكية قابلة للتطوير"، مقال على الإنترنت، ميدار 2009 م.
 - ك. أديبرت، ك. جوبير، أ. جاكربني، الطبعة المشتركة لسجل الحوليات الإسلامية أنالس: اللسانيات العربية، برنامج المعالجات الآلية للغة العربية (Tala)، 2010 م.
 - "كواكب"، وهو تطبيق على الشبكة العنكبوتية الإلكترونية لهندسة المعالجات الآلية للنصوص العربية، مجلة أنالس للعلوم الإسلامية، ifao، 2010 م.
 - أ. جاكربني، ك. جوبير، ك. أديبرت، "الهياكل التركيبية وإجراءات التعرف الآلي عليها باللغة العربية".
<http://www.medar.info/report-ws-malta.pdf> 2010
 - البرمجيات عبر الإنترنت (النسخة التجريبية): كواكب، 2010-2016
<http://www.ifao.egnet.net/kawakib>
 - الموقع الإلكتروني: أوتومات عربية، 2010-2016
<http://automatesarabes.net>
 - أ. جاكربني، ك. جوبير، "برنامج موجدور في اللسانيات الظاهرية العربية وتطبيقاته في مجالات البحث والترشح الدلالي"، مجلة أنالس للعلوم الإسلامية، 2014 م.
 - أ. جاكربني، ك. جوبير، "برنامج كواكب، أداة لتوفير التغذية الراجعة بين القواعد النحوية ومُدونات نصوص العربية في سبيل إعداد نموذج نظري"، مؤتمر البرمجة اللغوية العصبية، 2014 م.
 - العمل بالتعاون مع جان- ميشيل موتون، ورجال وقرى الفيوم في وثائق ودراسات لغوية لمخطوطات البزدي العربية خلال القرون العاشرة والحادية عشرة، المدرسة العليا للتعليم العالي، 2014 م.
<http://www.ifao.egnet.net/ifao/recherche/chercheurs/>
- (1) هذا هو العمل العلمي الذي نُقِشَ خلال بحثنا الراهن، ونقوم بتحليله وتقديمه للقارئ العربي.

- *Stratégie et règles minimales pour un traitement automatique de l'arabe*, thèse de doctorat, Université Aix-Marseille I, 2001.
- «Présentation du logiciel Sarfiyya » en langue arabe, Université Américaine de Caire (AUC), département d'arabe, décembre 2004.
- «Kawâkib, une application web pour le traitement automatique de textes arabes », in Annales islamologiques, N° 44 (2010), IFAO, pp. 53-60.
- C. Audebert, C. Gaubert, A. Jaccarini, Coédition du dossier des Annales islamologiques, N° 44 (2010): «Linguistique arabe, Programme de traitement par automates de la langue arabe (Tala) », pp. 1-61.
- A. Jaccarini, C. Gaubert, « Le programme Mogador en linguistique formelle arabe et ses applications dans le domaine de la recherche et du filtrage sémantique », pp. 482-497, in Annales islamologiques, N° 47, 2014.
- A. Jaccarini, C. Gaubert, «Démonstration de Kawâkib, outil permettant d'assurer le feedback entre grammaire et corpus arabe pour l'élaboration d'un modèle théorique », Conférence TALN, 2014.
- Étude d'un analyseur morpho-syntaxique de l'arabe écrit classique et moderne, fondé sur une description régulière de la morphologie et de la syntaxe et s'appuyant notamment sur la théorie des automates (AFND).
- En collaboration avec André Jaccarini (CNRS) et Claude Audebert (Université d'Aix-en-Provence), mise au point de méthodes d'analyse surfacique et linguistiquement pertinentes de textes arabes. Développement de méthodes de désambiguïsation et d'outils

graphiques de conception de grammaires.

- Logiciel expérimental Sarfiyya de traitement automatique de l'arabe.

وَمِنْ مَزَايَا أَعْمَالِهِ أَنَّهُ لَمْ يُقْصِرْهَا عَلَى الْمُصَنَّفَاتِ النَّحْوِيَّةِ، وَإِنَّمَا أَقَامَهَا عَلَى اسْتِقْصَاءِ شَمَلِ نَمَازَجٍ مِنَ الْأَسْلُوبِ الْأَدْبِيِّ، وَالتَّارِيخِيِّ، وَنَمَازَجٍ لِدَرَسَاتِ لُغَوِيَّةٍ لِمَخْطُوطَاتِ الْبَرْدِيِّ، وَمَخْطُوطَاتٍ عَرَبِيَّةٍ أُخْرَى، وَطَرَائِقِ الْكِتَابَةِ الْعَرَبِيَّةِ، وَالْعُلُومِ النَّحْوِيَّةِ وَالصَّرْفِيَّةِ. فَبَرَهَنَ بِصَنْيعِهِ هَذَا عَلَى تَنْوُّعِ الْأَبْعَادِ اللَّسَانِيَّةِ لِلتَّفَكِيرِ اللَّغَوِيِّ فِي عُلُومِ الْعَرَبِيَّةِ. وَقَدْ أُتِيحَ لَهُ ذَلِكَ بِفَضْلِ مَعْرِفَتِهِ الدَّقِيقَةِ بِخُصُوصِيَّةِ الْمَنْظُومَةِ الْإِصْطِلَاحِيَّةِ الْعَرَبِيَّةِ، وَمَا يُمَيِّزُهَا مِنَ الْمَفَاهِيمِ الْمُتَدَاوِلَةِ فِي الْفَضَاءِ الْفَرَانْكُوفُونِيِّ وَالْأَنْجُلُوسَكْسُونِيِّ. حَيْثُ تُعَدُّ مُعْظَمُ أَعْمَالِهِ نَمُودَجًا مِنَ التَّقَاطُعِ الْمُثْمَرِ بَيْنَ النَّظَرِيَّاتِ اللَّسَانِيَّةِ/الْحَاسُوبِيَّةِ الْحَدِيثَةِ وَالتُّرَاثِ النَّحْوِيِّ الْعَرَبِيِّ. وَفِي عَمَلِهِ هَذَا يُقَدِّمُ لَنَا كَرِيسْتِيَانُ جُوبِيرَ دَرَسَةً فِي التَّحْلِيلِ الصَّرْفِيِّ الْإِلَهِيِّ لِنَصِّ بِاللُّغَةِ الْعَرَبِيَّةِ. وَهِيَ دَرَسَةٌ تُمَثِّلُ حَوْصَلَةً تَأْلِيفِيَّةً لِبَحْثِهِ السَّابِقَةِ، كَمَا أَنَّهَا عَمَلٌ عَلَيَّيْ مُسْتَعْدَمٌ فِي الْعَدِيدِ مِنَ الْجَامِعَاتِ الْفَرَنْسِيَّةِ الْكُبْرَى. تُعَدُّ هَذِهِ الدِّرَاسَةُ تَطْبِيقًا دَقِيقًا لِكُلِّ مَنْ يَسْعَى إِلَى اقْتِحَامِ غَمَارِ الْبَحْثِ الصَّرْفِيِّ الْإِلَهِيِّ لِلُّغَةِ الْعَرَبِيَّةِ، وَاسْتِخْرَاجِ جُذُورِ الْكَلِمَاتِ، وَالتَّعَامُلِ مَعَ حَالَاتِ اللَّبْسِ اللَّغَوِيِّ وَالْكِتَابِيِّ النَّاجِمَيْنِ عَنِ الْبِنْيَةِ الصَّرْفِيَّةِ، فَهِيَ تَسْلُكُ مَسْلَكِ التَّعَمُّقِ الْعِلْمِيِّ، وَالتَّبَصُّرِ الْمَعْرِفِيِّ، كَمَا تَمْتَازُ بِجَمْعِهَا بَيْنَ نَظَرِيَّاتِ اللَّسَانِيِّينَ الْمُحْدَثِينَ، وَالنَّظَرَاتِ الثَّاقِبَةِ لِعُلَمَاءِ التُّرَاثِ، بِشَرْحِ الْمِصْطَلَحَاتِ، وَالْغُوصِ فِي الْمَفَاهِيمِ، وَتَشَعُّبِ الْأَفْكَارِ، وَغُمُقِ عِلَاقَاتِهَا بِالْمُبَاحَثِ الرِّيَاضِيَّةِ، وَالْمَنْطَقِيَّةِ ذَاتِ الصِّلَةِ. وَلَعَلَّهُ مِنَ الْمُثِيرِ أَنْ نُشِيرَ إِلَى أَنَّ هَذَا الْعَمَلَ عَلَى الرَّغْمِ مِنْ دَقَّةِ الْبَحْثِ وَتَشَعُّبِهِ، وَوَصْفِهِ اللَّسَانِي الْحَاسُوبِي الشَّامِلِ لِأَبْنِيَةِ الْكَلِمَاتِ الْعَرَبِيَّةِ، فَإِنَّ الْأَسْلُوبَ الْمُنَهْجِيَّ الَّذِي كُتِبَ بِهِ سَيَجْذِبُ شَرِيحَةً مِنَ الْقُرَّاءِ الْبَاحِثِينَ عَنِ الدَّقَّةِ الْعِلْمِيَّةِ فِي الْمَعَالِجَةِ الْحَاسُوبِيَّةِ لِأَنْظِمَةِ الصَّرْفِ الْعَرَبِيِّ، وَمُسَاعَدَةِ

الدّارسين اللّسانيين على إدراك كُنْه كِيفِيات التّفَاعُلِ البنيويّ، والمكوّنات اللّغوية مع عوامل السّياق؛ لغرض التّفسير المُنضبط للصّيغة الصرفية للكلمة، ولمعناها الذي جَاءَتْ له.

2- في تحديد المفاهيم الأساسيّة:

تتميّز المحاولات المتماشكة للغويين والحاسوبيين في العلاج الآليّ للغة العربية في الانطلاق من فرضيّة أساسيّة مفادها وضوح المفاهيم واستقلالها، ودقّة النظريات اللّغوية التي يعتمد عليها التّصميم الحاسوبيّ المُعالج، والصّيغة المنطقيّة لبنيّة اللغة. يتّضح لنا ذلك بتمحيص هذه المفاهيم، وتبيان تلك المنطلقات الواردة في بحثنا على ضوء مُقارَبة إيسْتيمولوجية دقيقة لتصوّرات Christian Gaubert (كريستيان جويبر) وطُرُق تحليله، واستنتاج مُعطيات كمية لتوصيفات بنيويّة في الحقل الصّرفيّ العربيّ؛ كونها وجْهًا مِنْ وَجُوهِ نَمْدَجَة اللغة.

2.1 مُسمّيات الصّرف، والتّصريف، والاشتقاق عند العرب وفي اللّسانيات الحديثة والحاسوبية:

2.1.1 عند العرب القدماء:

تدوّر في معاجم اللغة تعريفات "الصّرف" morphologie و"التّصريف" conjugaison على النّحو الآتي: "الصّرف" مَصْدَرٌ لـ (صَرْفَ) الثّلاثي، و"التّصريف" مَصْدَرٌ لِلْفِعْلِ الثّلاثي المزيد (صَرَّفَ). والصّيغتان تدلّان على دلالات التّغيير، والتّبديل، والتّحويل⁽¹⁾. ولعلّ الفارق بينهما يتّضح في أنّ "التّصريف" يُفيد معنى

(1) يُنظر: الجوهريّ، الصّحاح تاج اللغة وصحاح العربية، 1385/4. وانظر أيضًا: ابن فارس، مُعْجَم مقاييس اللغة، 343/3. ابن أبي بكر الرّازي، مُخْتَار الصّحاح، 175/5. ابن منظور، لسان العرب، 189/9. الفيروزآبادي، القاموس المحيط، 826-827.

التغيير أكثر من إفادة "الصَّرف" لهذا المعنى⁽¹⁾. وفي اصطلاح النُّحاة ظهر مصطلح "التَّصريف" مُرَافِقًا وَقَرِينًا لمصطلح "النَّحو" *grammaire* في المؤلفات التُّراثية النُّحوية⁽²⁾، وقد لُوْحِظَ بعد ذلك أَنَّ انتقلت تَسْمِيَةُ "التَّصْرِيف" في كثيرٍ مِنَ الْمُصَنَّفَاتِ إِلَى "عِلْمِ الصَّرف"⁽³⁾. وقد تطرَّقت بعضُ الدراسات إلى بَيَانِ أَنَّ

(1) يُرَاجَع في ذلك: خالد بن عبد الكريم بسندي، الصَّرف والتَّصْرِيف وتداخل المصطلح، مجلة جامعة الملك سعود، المجلد العشرون، 319/، 377.

(2) قال الرُّضِي: «اعلم أَنَّ التَّصْرِيفَ جُزْءٌ مِنْ أَجْزَاءِ النُّحُوِّ بِلَا خِلَافٍ مِنْ أَهْلِ الصَّنَاعَةِ، وَالتَّصْرِيفُ - عَلَى مَا حَكَى سِيبَوَيْهِ عَنْهُمْ - هُوَ أَنْ تَبْنِيَ مِنَ الْكَلِمَةِ بِنَاءً لَمْ تَبْنِهِ الْعَرَبُ عَلَى وَزْنٍ مَا يَنْتُهُ، ثُمَّ تَعْمَلْ فِي الْبِنَاءِ الَّذِي بَنَيْتَهُ مَا يَقْتَضِيهِ قِيَاسُ كَلَامِهِمْ، كَمَا يَتَبَيَّنُ فِي مَسَائِلِ التَّمْرِينِ. وَالْمُتَأَخَّرُونَ عَلَى أَنَّ "التَّصْرِيفَ" عِلْمٌ بِأَبْنِيَةِ الْكَلِمَةِ وَبِمَا يَكُونُ لِحُرُوفِهَا مِنْ أَصَالَةٍ وَزِيَادَةٍ وَحَذْفٍ وَصَحَّةٍ وَإِعْلَالٍ وَإِدْغَامٍ وَإِمَالَةٍ، وَبِمَا يَعْضُرُ لِأَخْرَافِهَا مِمَّا لَيْسَ بِإِعْرَابٍ وَلَا بِنَاءٍ مِنَ الْوَقْفِ وَغَيْرِ ذَلِكَ». انظر: رَضِي الدِّين الإِسْتِرَابَادِي، شرح شَافِيَةِ ابْنِ الْحَاجِبِ، 6/1-7. ولقد أُوْرِدَ التَّهَانَوِي فِي "كَشَافِ اصطلاحاتِ الفنون" أَنَّ مُصْطَلَحِي "الصَّرف" وَ"التَّصْرِيفَ" عِنْدَ الْمُتَأَخَّرِينَ مُتَرَادِفَانِ، وَالتَّصْرِيفُ عَلَى مَا حَكَى سِيبَوَيْهِ عَنْهُمْ جُزْءٌ مِنَ الصَّرفِ الَّذِي هُوَ جُزْءٌ مِنْ أَجْزَاءِ النُّحُوِّ لِأَنَّهُ مِنْ مَبَادِئِ النُّحُوِّ، وَالمَبَادِئُ مِنَ الْأَجْزَاءِ كَمَا نَعْرِفُ. وَيُعَلِّلُ التَّهَانَوِي فِي مَوْضِعٍ آخَرَ مِنْ كَشَافِهِ أَنَّ مُصْطَلَحَ "المَبْنَى" لَا يَقِفُ فِي حَقِيقَتِهِ عِنْدَ مَسْأَلَةِ الدَّلَالَةِ وَارْتِبَاطِ الْأَسْمَاءِ بِالمَعْنَى، بَلْ يَطَالُ دَرَسَاتُ الْمَعْنَى وَالتَّرَكِيبِ اللُّغَوِيِّ. وَلَعَلَّ لَفْظَ "المَبْنَى" مُرْتَبِطٌ بِالْبِنْيَةِ وَمَفَاهِيمِ التَّرَكِيبِ الصُّورِيِّ لِللُّغَةِ. إِذْ إِنَّ اللُّغَةَ أَشْكَالَ عُنَاصِرِهَا الْحُرُوفِ، وَالأَصْوَاتِ تَتَرَكَّبُ عَلَى نَظْمٍ مُعَيَّنٍ، وَفِي أَنْسَاقٍ مُحَدَّدَةٍ، لِتَتَمَيَّزَ بِنَمَطٍ مَا. إِنَّ النُّحُوَّ مَنْطِقَ الْعَرَبِيَّةِ، فَفِيهِ يَتَبَيَّنُ وَجْهٌ تَصَرَّفُ الْأَلْفَاظِ فِي الْمَعَانِي. إِنَّ الْعُلُومَ الْبَيَانِيَّةَ الْعَرَبِيَّةَ قَدْ كَشَفَتْ عَنْ مَنْطِقٍ دَاخِلِيٍّ فِيهَا، وَذَلِكَ حِينَمَا دَفَعَ بِهَا تَطَوُّرُهَا الذَّاتِي إِلَى أَنْ تُصَلَّ إِلَى مَا وَصَلَتْ إِلَيْهِ. فَالتَّعْقِيدَاتُ وَالتَّفْنِينَاتُ الَّتِي جَاءَتْ عَلَى يَدِ اللُّغَوِيِّينَ لَمْ تَكُنْ خَنْقًا لِلْحَيَاةِ فِي اللُّغَةِ، بَلْ كَانَتْ لَا بُدَّ مِنَ الْكَشْفِ عَنِ النَّظْمِ الْبِنْيَوِيِّ لِهَذِهِ اللُّغَةِ. إِنَّ هَذِهِ الْعَمَلِيَّةَ تُسَبِّحُ عَقْلَ اللُّغَةِ وَالبَيَانِ. وَهَذَا مَا حَدَثَ عِنْدَ مَا تَمَّ ضَبْطُ عِلْمِ الصَّرفِ وَالاِشْتِقَاقِ، وَعِلْمِ الْمَعَانِي وَالبَيَانِ وَعِلْمِ النُّحُوِّ. انظر: التَّهَانَوِي، موسوعة كشاف اصطلاحات الفنون والعلوم، المَقْدِمَةُ 34، وَالمَجْلَدُ الْأَوَّلُ 23.

(3) انظر على سَبِيلِ الْمِثَالِ لَا الْحَصْرِ الْمُصَنَّفَاتِ الْآتِيَةِ: عبد القاهر الجرجاني، المفتاح في الصَّرف. المِيدَانِي، نزهة الطَّرْفِ فِي عِلْمِ الصَّرفِ. ابْنُ يَعِيشَ، شرح المَلُوكِي فِي التَّصْرِيفِ. ابْنُ جِيٍّ، المَنْصِفُ فِي التَّصْرِيفِ. الْحَمَلَاوِي، شَذَا الْعَرَفِ فِي فَنِّ الصَّرفِ. رَضِي الدِّينِ الإِسْتِرَابَادِي، شرح شَافِيَةِ ابْنِ الْحَاجِبِ. ابْنُ مَالِكٍ، إيجاز التعريف في علم التَّصْرِيفِ. خَدِيجَةُ الْحَدِيثِي، أَبْنِيَةُ الصَّرفِ فِي كِتَابِ

"التَّصْرِيف" يرتبط بكثرة دوران الأبنية واشتقاقها والعمل فيها، و"الصَّرْف" يرتبط بالأصول الكَلِيَّة التي ينبني عليها معرفة أحوال المفردات⁽¹⁾.

فابْنُ جَنِّي يَشْرَحُ الْفَوَارِقَ الدَّقِيقَةَ بَيْنَ "التَّصْرِيفِ" و"الاشتقاق" و"النَّحْوِ" حينما يُشِيرُ إلى وجود علاقة النَّسَبِ القَرِيبَةِ والِاتِّصَالِ الشَّدِيدِ بَيْنَ التَّصْرِيفِ والِاشتِّقَاقِ؛ لأنَّ التَّصْرِيفَ هو أَنَّ تُصَرِّفَ الْكَلِمَةَ الْوَاحِدَةَ عَلَى وَجْهِ كَثِيرَةٍ، وَمَثَلٌ لَذَلِكَ بِالْفِعْلِ "ضَرَبَ" فَتَبْنِي مِنْهُ مِثْلَ "جَعَفَرَ" فَتَقُولُ: "ضَرَبَ"، وَمِثْلَ "قَمَطَرَ": "ضَرَبَ"، وَمِثْلَ "دَرَّهَمَ": "ضَرَبَ"، ... إلخ. وكذلك الاشتقاق أيضًا، فتشتق الماضي مِنَ الْمَصْدَرِ (الضَّرْبِ) فَتَقُولُ: "ضَرَبَ"، ثُمَّ تَشْتَقِ مِنْهُ الْمَضَارِعَ فَتَقُولُ: "يَضْرِبُ"، ثُمَّ تَقُولُ فِي اسْمِ الْفَاعِلِ: "ضَارِبٌ"، وَعَلَى هَذَا نِظَائِرُ هَذِهِ الْكَلِمَةِ. وَهَذَا مِثْلُ أَنَّ تَقُولُ: "تَتَصَرَّفُ فِي الْبَاطِلِ"، أَيْ: تَأْخُذُ فِي ضَرْبِهِ وَأَفَانِيْنِهِ. فَمِنْ هَا هُنَا تَقَارِبًا وَاشْتَبَكَاءً، إِلَّا أَنَّ التَّصْرِيفَ وَسِطَةٌ بَيْنَ النَّحْوِ وَاللُّغَةِ يَتَجَاذِبَانِهِ، وَالِاشتِّقَاقُ أَقْعَدُ فِي اللُّغَةِ مِنَ التَّصْرِيفِ. كَمَا أَنَّ التَّصْرِيفَ أَقْرَبُ إِلَى النَّحْوِ مِنَ الْاشتِّقَاقِ، يَدُلُّكَ عَلَى ذَلِكَ أَنَّكَ لَا تَكَادُ تَجِدُ كِتَابًا فِي النَّحْوِ إِلَّا وَالتَّصْرِيفَ فِي آخِرِهِ. وَالِاشتِّقَاقُ إِنَّمَا يَمْرُوكَ فِي كُتُبِ النُّحُو، مِنْهُ أَفَاضٌ مُشْرَدَةٌ لَا يَكَادُ يَعْقِدُ لَهَا بَابٌ. فَالتَّصْرِيفُ إِنَّمَا هُوَ لِمَعْرِفَةِ أَنْفُسِ الْكَلِمِ الثَّابِتَةِ، وَالنَّحْوُ إِنَّمَا هُوَ لِمَعْرِفَةِ أَحْوَالِ الْمُنْقَلَةِ⁽²⁾.

==

سيبويه. مهدي بن علي بن مهدي آل ملحان القرني، الترتيب الصَّرْفِي في المؤلفات النُّحَوِيَّة والصَّرْفِيَّة إلى أواخر القرن العاشر الهجري. غنيم الينبعاوي، جهود ابن جَنِّي في الصَّرْفِ وتقويمها في ضوء علم اللغة الحديث. محمد خليفة الدناع، دور الصَّرْفِ في منهجي النَّحْوِ والمعجم. محمود سليمان ياقوت، ظاهرة التَّحْوِيلِ فِي الصَّيْغِ الصَّرْفِيَّة.

(1) يُنْظَرُ: خَالِدُ بْنُ عَبْدِ الْكَرِيمِ بَسَنْدِي، الصَّرْفُ وَالتَّصْرِيفُ وَتَدَاخُلُ الْمِصْطَلَحِ، 377/.

(2) ابْنُ جَنِّي، الْمُتَّصِفُ، شَرَحَ كِتَابَ التَّصْرِيفِ لِأَبِي عِثْمَانَ الْمَازِنِيِّ، الْمُقَدِّمَةُ 3/4. وَانْظُرْ أَيْضًا: ابْنُ عَصْفُورِ الْإِشْبِيلِيِّ، الْمَمْتَعُ الْكَبِيرُ فِي التَّصْرِيفِ، 33-44.

2.1.2 في اللسانيات الحديثة والحاسوبية :

يذهب الطيّب البكوش إلى أنّ الدِّراساتِ الأجنبيّة الحديثة تُعرِّف عِلْمَ الصَّرْفِ بأنّه «البحث في نشأة الكلمات والتّغيرات التي تطرأ على مظهرها الخارجي في الجُمْلَة»⁽¹⁾.
وَيَسْتَعْرِضُ جورج مونان في قاموسه اللِّسَانِيّ بَعْضًا من هذه التّعريفات بقوله:

«Traditionnellement, étude des formes sous lesquelles se présentent les mots dans une langue, des changements dans la forme des mots pour exprimer leurs relations à d'autres mots de la phrase, des processus de formation de mots nouveaux, etc. Chez Vendryes, la morphologie est l'étude des morphèmes qu'il distingue des sémantèmes. Chez Sweet, la morphologie étudie les variations formelles de chacune des catégories grammaticales. On y traite de flexion, de composition et de dérivation ("morphologie lexicale"), d'ordre des mots, des parties du discours. Chez Bloomfield, la morphologie étudie les variations des formes des mots (unités lexicales comme unités grammaticales) dans lesquels les morphèmes liés figurent comme éléments constitutifs. Chez Martinet, la morphologie concerne uniquement les variantes de signifiants des monèmes, ou allomorphes au sens bloomfieldien. La morphologie syntagmatique étudie les variantes de monèmes grammaticaux (les accords, la conjugaison, etc.). Moins proprement nommée, la morphologie lexicale étudie les variantes des monèmes dans la composition ou la dérivation de mots.»⁽²⁾

(1) المرجع السابق، /17.

(2) Voir Georges Mounin, Dictionnaire de la linguistique, pp. 221-222. Voir aussi: Jean Dubois et autres, Dictionnaire de linguistique et des sciences du langage, pp. 311-312.

ومَضْمُون هذا النَّصِّ بالعربية يكشف عن أَنَّ عِلْمَ الصَّرْفِ يدرسُ الأشكال التي تظهر تحتها كلمات لغة ما، والتغيرات في شكل الكلمات من أجل التعبير عن علاقاتها بكلماتٍ أخرى من الجملة، ومسارات تشكيل الكلمات الجديدة، إلخ. فعِنْدَ فندريس، الصَّرْفُ هو دراسة المورفيمات التي يُميزها عن وحدات الدَّلالة أو السِّمَنَتيمات. وعند سويت، يدرس الصَّرْفُ التَّنَوُّعات الشَّكْلِيَّة لِكُلِّ واحدٍ من الأصناف النَّحْوِيَّة. فيتم تناول تصريف الأسماء والتَّركيب والاشتقاق (تصريفات مُعْجَمِيَّة)، وترتيب الكلمات وأجزاء الخطاب كعناصر مُشَكِّلَة. وعند مارتينيه يتعلَّق الصَّرْفُ فقط بتنوُّعات دوال المونيمات، أو البدائل الصَّرْفِيَّة بالمعنى البلومفيلدي. كما يدرس الصَّرْفُ المُركَّبِي تنوُّعات المونيمات النَّحْوِيَّة (الرَّوَابِط، تصريف الأفعال، إلخ). مع تسمية غير دقيقة فَإِنَّ التَّصْرِيفَ المعجمي يدرس تنوُّعات المونيمات في تركيب أو اشتقاق الكلمات⁽¹⁾.

وأبرز ما يُسْتَلْتَجُ مِنْ كُلِّ هذا سعة معنى "التَّصْرِيف"؛ إذ يكاد يشمل "النَّحْو" لولا اختصاص النَّحْو أساساً بالبحث في التَّغْيِير الذي يلحق أواخر الكلمات. لذلك فَإِنَّ الصَّرْفَ يتميَّز عنه بالبحث في التَّغْيِير الذي يطرأ على أبنية الكلمات. وهكذا نرى أَنَّ عِلْمَ الصَّرْفِ في العربية متعدّد الجوانب والأبعاد، فهو يحوي ثلاثة أنواع من التغيرات على صيغةٍ من الصَّيْغ:

(1) يُنْظَر: جمال الحضري، ترجمة معجم اللسانيات بإشراف جورج مونان، 452/ ومِمَّا يَتَّصِلُ بهذه التَّعْرِيفَاتِ الْحَدِيثَةِ ما أَشَارَ إِلَيْهِ نَبِيلٌ عَلِيٌّ مِنْ أَنَّ "الصَّرْفَ" يُعَدُّ رَابِطَةً الْعَقْدِ لِعَنَاصِرِ الْمَنْظُومَةِ اللُّغَوِيَّةِ، فَهُوَ يُمَثِّلُ وَسْطاً لُغَوِيّاً، يَجْمَعُ بَيْنَ هَيْكَلِيَّةِ النَّحْوِ وَتَحْلِيلِيَّةِ الْفُونُولُوجِي، وَبَيْنَ أَطْرَادِ الْاِثْنَيْنِ، وَاصْطِلَاحِيَّةِ الْمَعْجَمِ. وَتَتَجَلَّى عَلَى مَسْتَوَى الصَّرْفِ مَظَاهِرُ التَّبَايُنِ اللَّغَوِيِّ بَيْنَ اللَّغَاتِ؛ لِذَا فَقَدْ كَانَ وَمَا يَزَالُ مِنْ أَمَمٍ الْأَسْوَاسِ الَّتِي تَرْتَكِزُ عَلَيْهَا الدِّرَاسَاتُ اللَّغَوِيَّةُ الْمَقَارَنَةُ وَالتَّقَابِلِيَّة. وَإِذَا كَانَ النَّحْوُ هُوَ أَسَاسُ الْخَاصِيَّةِ الْخَلَاقَةِ لَتَنَوُّعِ الْإِسْتِخْدَامِ اللَّغَوِيِّ بِمَا يُتِيحُهُ مِنْ تَرْكِيبَاتٍ لَا نِهَائِيَّةٍ لِلْجُمْلِ، فَالصَّرْفُ بِلَا شَكٍّ هُوَ مَصْدَرُ التَّوَسُّعِ اللَّغَوِيِّ بِمَا يُوفِّرُهُ مِنْ وَسَائِلَ عَدِيدَةٍ لِتَكْوِينِ كَلِمَاتٍ جَدِيدَةٍ. رَاجِعْ فِي ذَلِكَ: نَبِيلٌ عَلِيٌّ، اللُّغَةُ الْعَرَبِيَّةُ وَالْحَاسُوبُ، 247/.

- 1- تغيير صرفي بحت: يتعلّق أساسًا بالاشتقاق.
 - 2- تغيير صرفي/ صوتي: يتعلّق بتأثير التّغيير الصّوتيّ في بنية الصّيغة صرفيًا.
 - 3- تغيير صوتيّ بحت: يتعلّق بتعامل الأصوات⁽¹⁾.
- وإذا أضفنا إلى هذا أنّ وَجَهَ الخِلاف بين مفهوم "الصّرف" عند العرب القدماء وبين مفهومه الحديث يتضمّن أمرين:
- الأول: استبعاد التّغيّرات الصّوتية غير الدّالة على معنى من الصّرف بالمفهوم الحديث؛ إذ يرى المحدثون أنّ التّغيير الذي يحدث في بنية الكلمة لغرضٍ لفظيٍّ، أو بسببٍ صوتيٍّ لا شأن لعلم الصّرف به، بل هو من اختصاص علم الأصوات. في حين اشتملَ صرف القدماء على التّغييرات في المبنى والمعنى. الثّاني: اقتصار صرف القدماء على الأسماء المتّممّكة والأفعال المتّصرّفة من الكلمات العربية. على حين نرى الصّرف بالمفهوم الحديث يشمل جميع الكلمات التي يحدث فيها تغيير ذو دلالة. فالتّصريف في علم اللغة الحديث يبحث في الوحدات الصّرفية أو (الصّرفيات) التي تؤدّي وظائف محدّدة في الصّيغ⁽²⁾. فالمحدثون من علماء اللغة يرون أنّ "التّصريف" لا يقوم إلّا على ما يُقرّره علم الأصوات من حقائق، فهو يتعامل مع البنية الدّاخلية للكلمات، وأنّ الظّاهرة الاشتقاقية في جوهرها منجّ للجدور مع الأنماط الصّوتية لسلاسل حركات الصّيغة الصّرفية، فالتّغييرات التي تؤدّي إلى تحولاتٍ في المعنى هي من صميم "الصّرف"⁽³⁾.

(1) انظر: الطّيب البكوش، التّصريف العربي من خلال علم الأصوات الحديث، / 17-20. وانظر أيضًا: فردينان دي سوسير، دروس في الألسنية العامة، ترجمة: صالح القرماضي ومحمد الشاوش- محمد عجينة، / 202.

(2) انظر: سلوى حمادة، المعالجة الآلية للغة العربية المشاكل والحلول، / 88-89. وانظر أيضًا: هدى سالم عبد الله آل طه، النّظام الصّرفي للعربية في ضوء اللّسانيات الحاسوبية: مثل من جمع التّكسير، / 29.

(3) انظر: سلوى حمادة، مرجع سابق، / 90.

وواضح من هذه الأنساق التعريفية في المباحث اللسانية الحديثة وتطبيقاتها في اللسانيات الحاسوبية أن مفهوم علم "الصرف" يشير إلى تضمّن الصرف للاشتقاق والتّصريف^(*)، من مثل مقولات العدد، والجنس، والتّعين، والمطابقة، وتنوع الأبنية الاسمية والفعلية في جداول الأزمنة المجردة والمزيدة. على حين نجد أن مفهوم "الاشتقاق" يتضمّن إقامة اللسان العربي على الجذور التي تُعبّر عن اسم الحدث، أو المصدر، كما يُمثّل الجذر أصل الأشكال المشتقة من الأوزان. والأوزان

(*) يرى نبيل علي أن مسألة التفريق بين التّصريف والاشتقاق موضع جدل حاد في اللسانيات الحاسوبية، ومصدر الخلط في نظره يكمن في أن عمليات التّصريف والاشتقاق تستخدم في كثير من اللغات نفس آليات تكوين الكلمات من إلصاق وتغيير في الحركات وحذف وصرفيمات عدمية، وهلمّ جرّاً. بينما ينعت "العربية" في هذا الصّدّد بأنها تمتاز بوضوح الفرق بين عمليات التّصريف والاشتقاق، فالتّصريف لصقيّ في الغالب، والاشتقاق انصهاريّ حيث يُبنى على التّشكيل من خلال الأنماط. وقُدّم في النّقاط التالية ملخصاً لأهمّ مظاهر الاختلاف بين التّصريف والاشتقاق في تطبيقات المعالجة الآلية لمنظومة الصرف العربي وطريقة تناول مجموعة الآليات الصرفية:

(أ)- حيث يشمل التّصريف خصائص بنية الكلمات ذات الصلة بالنّحو، أي التي يتعامل معها النّحو، أو يقوم بتعديلها، أو يضيف عليها، أو يؤلّدها. (ب)- لا يؤثّر التّصريف على الوظائف النحوية والأدوار الدلالية لجذوع الكلمة، ولا يؤثّر على معناها. (ج)- يتّسم التّصريف بالاطراد، في حين يتّصف الاشتقاق بعدم الانتظام، وخضوعه لكثير من القيود الدلالية والمعجمية. (د)- تقع زوائد التّصريف عادةً على حدود الكلمات، في حين يؤثّر الاشتقاق في متن الكلمة ذاته.

كما تتلقّى وحدة التّصريف مدخلاتها من المعجم، أو من وحدة الاشتقاق من خلال وحدة التّعديل "الصرف/صوتي"، وكذلك من منظومة النّحو فيما يخصّ الخصائص النحوية للكلمة المراد تصريفها من حيث موقعها ووظيفتها النحوية. ويلزم التّنويه إلى أن التّصريف يتعامل مع الفعل والاسم والصفة والحرف أحياناً (كما في اللغة العبرية)، ويُعدّ الفعل أكثر أقسام الكلام من حيث اتّساع نطاق التّصريف. ويشير المؤلّف إلى تنوع تطبيقات "الاشتقاق" اللغويّ في اللسانيات الحاسوبية، وتنوع عمليات تكوين الكلمات في الآليات الرئيسية التالية: 1- الاشتقاق بالإلصاق. 2- الاشتقاق بتغيير الحركات وتلوينها. 3- الاشتقاق بتغيير موضع النّبر. 4- الاشتقاق بتغيير نغمة نُطق الكلمات. 5- الاشتقاق بالحذف. 6- الاشتقاق بالصّرفيم العدمي. 7- الاشتقاق بتكرار المقاطع. 8- الاشتقاق بأنماط الصّيغ الصرفية. يُنظر: نبيل علي، اللغة العربية والحاسوب، 258/، 260، 263.

نوعان: فعلية أو اسمية، وتشتق من جذرٍ متكوّن من ثلاثة أو أربعة أحرف. وتنتج الأبنية الفعلية أفعالاً مجردة أو مزيدة، أمّا الأبنية الاسمية فتعطي الصفات أو غير الصفات. ويجري تكثير الأبنية بآليات الجذور والأوزان. ويتشكّل مفهوم "التصريف" بما يظهر فيه من أنّ كلّ ما لا ينتهي إلى المفردة الاسمية أو الفعلية وإلى بنية الوزن المجرد هو عبارة عن زيادة بما يُسمّى "الصّرافيم"، وهي ثلاثة أنواع: السّوابق، والدّواخل، واللّواحق. السّوابق تُوجد في تصريف الأفعال والأسماء، والدّواخل تكون بين مكوّنات الكلمة. أمّا اللّواحق فتأتي بعد الكلمة مباشرة، وتمثّل عادةً علامات معنوية. والهدف من هاته "الصّرافيم" هو الإشارة إلى العدد (مفرد، أو مثنى، أو جمع)، والجنس (مذكر، أو مؤنث)، والتّعيين (معرفة، أو نكرة)⁽¹⁾.

2.2 بنية نظام الصّرف العربيّ والمعالجة الآليّة:

يتطرّق حاتم صالح الضّامن إلى تبيان أدوار منظومة البنية الصّرفيّة العربيّة؛ حيث يُعدّ علم الصّرف من أقدم فروع الدّراسات اللّغويّة، فإذا كان النّحو العربيّ يمثّل أساس الخاصيّة الإبداعية للعربية، فإنّ الصّرف هو مورد التّوسّع والتّباين والانفتاح اللّغويّ بما يوفره من وسائل عديدة لتكوين وخلق كلمات جديدة، وما تسمّى به العربية بارتقاء نظام صرفها، واطراده، وتعاضل مكانته داخل المنظومة اللّغوية، ممّا يسمّح بإعادة تحليل علاقة التحليل/التركيب التي تثري اللّغة وتزيد من خصوبتها المعجمية⁽²⁾. وعلى ضوء دراسته الحديثة « سبل تطوير مُحلّل الصّرف الآليّ المستخدم في حوسبة الموسوعات العربيّة »، خرّج حسن مظفر باستعراض سريع لمفهوم "المعالجة الآلية للصّرف العربيّ"، يُقصد به: «اعتماد

(1) يُنظر: منصف الشرفي، وهاني عمار ضمن أعمال النّدوة الدّوليّة للعلوم وهندسة الحاسوب، 326/.

(2) يُنظر: حاتم صالح الضّامن، الصّرف، 55/، وما بعدها. وانظر أيضًا: خليل أحمد عمارة، في التّحليل اللّغويّ: منهج وصفيّ تحليليّ.

نُظِّم حوسبة متقدِّمة تستند إلى خوارزميات برمجية تستثمر المنطق الصِّرفيَّ العربيَّ في معالجة المفردة العربية، عن طريق استخلاص العناصر الأولية لبنية الكلمة، ومباشرة تحديد سماتها الصِّرفية، والصِّرف/نَحْوِيَّة، والصِّرف/دلالية، القابلة للاستنباط من هذه البنية»⁽¹⁾.

كما لفت نبيل علي في كتابه «اللغة العربية والحاسوب»⁽²⁾ إلى إمكانية تفكيك المنظومة الصِّرفية في لسانيات الحاسوب العربية وتطبيقاتها إلى ثلاثة عناصر: [أ] عناصر تكوين الكلمات، وتشمل عمليات الاشتقاق، والتَّركيب والمزج والاختصار.

[ب] عناصر التَّدَاخُل مع المنظومة النَّحْوِيَّة، وتشمل عمليات التَّصْرِيف، والضَّبْط الإعرابي، ودمج الأدوات والضمائر المتصلة (الدمج "الصِّرف/نحوي"). [ج] عنصر التَّعْدِيل "الصِّرف/صَوْتِي"، وتشمل عمليات التَّغْيِير الفونولوجيَّة المصاحبة للعمليات الصِّرفية المختلفة (كعمليات الإبدال والإعلال على سبيل المثال).

وتتطلَّب كل معالجة آلية للغة دراسة عميقة للغة وقواعدها؛ للتَّمَكُّن مِنْ إنجازِ مرحلة التَّحْلِيلِ الصِّرفيِّ، وهذا هو هدفنا في هذا العمل الذي نقوم بدراسته، حيث نُسَلِّطُ الضَّوْءَ على ما قامَ به Christian Gaubert (كريستيان جوبير) من تصميم وتنفيذ مُحَلِّلِ صرْفِيٍّ للنَّصِّ العربيِّ غير المشكُّول، واعتماده تصوُّراً آلياً خاصاً قامَ بإنجازه. وتنقسم مُقَارَبَتُهُ إلى أربع مراحل: هي تقسيم الجُمْلِ إلى كلمات، وحذف الزَّوَائِد عن الكلمة، ثُمَّ إيجاد أصلها، وأخيراً ترميز الكلمات المُكوِّنة للنَّصِّ

(1) يُنظَر: حسن مظفر الرِّزْو، مرجع سابق (هامش 34). الفقرة 4- المُعَالَجَةُ الآليَّةُ للصِّرفِ العربيِّ.

(2) انظر: نبيل علي، اللغة العربية والحاسوب، 258.

بخصائصها الصرفية والإعرابية التركيبية.

ومن صعوبات هذه الدراسة عدم توافر قاعدة بيانات كاملة وشاملة للغة العربية. فهذه الأخيرة تُعرف بأنها لغة اشتقاقية، حيث تنطلق من جذور وأوزان، وإضافة سوابق ولواحق لتتكوّن الكلمة. كذلك يصعب التفريق في بعض الحالات بين الحرف الأصلي للكلمة والحرف الزائد. فمن أمثلة ذلك ما تكشف عنه بعض ظواهر اللبس القائم، والأخطاء اللغوية الحادثة في مثل وزن (فعل)، حيث يتباين مشهده الصيغي، فيمكن أن يدلّ على اسم أو فعل. ومن الآفاق التي نرجو تحقيقها من مثل هذه الدراسات والمحاولات الجادة، تحليل النصّ العربي المشكول وغير المشكول، وبذلك ندرس التّعامل مع الحروف ومع الشّكل (الفتحة، الضّمة، الكسرة، ... إلخ)، كما نطمح إلى توسيع مثل هذه المقاربات لتشمل معالجة المدوّنات اللّغوية ومُستويات الأبنية التركيبية والدّلالية، وإضافة في مجال البحث اللّسانيّ والتّحليل النّحويّ الأتوماتيكيّ.

المبحث الأول - في المصطلحات الرئيّسة المستخدمة لطريقة التحليل الصرفي إلى المؤلفات المباشرة أو التحليل التوزيعي

1. في مبادئ التّحليل ومفاهيم التّطبيق:

1.1 المبادئ:

يقوم (كريستيان جوير) في نهجه التّحليلي بدراسة توزيع الوحدات التّصريفية التي تُؤلف الكلمات العربيّة. فالفكرة الأساسيّة عنده هي أنّ عمَل المُعالج الآليّ قائمٌ على تصوّر التّراكيب البنيويّة للكلمات المنتمية إلى مُستوياتٍ مُختلفة. ويتحدّد كلّ عنصّر منها:

- حسب توزيعه على الصّعيد الذي ينتمي إليه.
- حسب وحدات الصّعيد الأعلى.

- إدراك جُمْلَةٍ مِنَ التَّحَوُّلاتِ الحُرَّةِ، والاحتمالات المُمْكِنَةِ، والأشكال الفرعية الطَّارِئة لِزُمرَةِ الكلمات المراد تحليلها وتَقْطِيعها.
- تحديد توزيع الوحدة البنيوية على ضوء التَّعرُّف على مَجْمُوعِ القرائن والسمات ودراسة العلامات الإملائية، التي تُساعد المحلِّل على تمييز الوحدات، واستخلاص النتائج الصحيحة لنماذج التحليل.
- التَّعرُّف على العناصر الدَّالة في بنية الوحدات.
- اعتبار الوحدة صورة أو شكل لِسَانِي مُجرَّد قابل للفصل والتَّقْطِيع⁽¹⁾.

1.2 المفاهيم:

تَعْتَمِدُ طَرِيقَةُ جَوْبِرِ على جُمْلَةٍ مِنَ المفاهيم الأساسية التي تُوفِّرُ لنا الحَلَّ المنشود. ويُمكن بيان تلك المفاهيم في الصُّورِ الآتية:

[أ] البناء La construction:

هُوَ ضَمُّ مجموعةٍ مِنَ الأُبنية، أو الوحدات الصَّرْفية (المورفيمات)، أو مجموعة من الصِّيَاغِم، أو الكلمات، بحيثُ تُشكِّلُ على مُستَوًى مِنْ مُستويات التحليل وحدة تركيبية. فَ "الكتاب" بناء صرْفِي/تركيبِي لِأَنَّهُ يَحْوِي مُورفِيمَيْنِ أو مُؤَلَّفَيْنِ أو وَحْدَتَيْنِ: (أل) + (كِتاب). أمَّا الوحدة التركيبية فهي بِنَاءٌ أَيْضًا وَيُمَثِّلُهَا قولنا: "كُتِبُوا" ← (كُتِبَ) ← مورفيم مُستَقِلٌّ + (وا) ← مورفيم مُقَيَّد (لا يُشكِّلُ دلالةً مُستقلةً وَحْدَهُ).

[ب] المؤلَّف/المَقْوَمُ المُباشِرُ:

يَتَنَاولُ "التَّحْلِيلُ" إلى المَقَوِّمَاتِ المُباشِرَةِ "النَّصَّ بِوصْفِهِ سِلْسِلَةً مُتَعاقِبَةً مِنْ الوَحَدَاتِ البِنْيَوِيَّةِ التي يَرْتَبِطُ بعضها ببعضٍ. واستجابةً لأهمية الشُّعُورِ بالنِّظامِ

(1) يُنظَرُ في ذلك: رونالد ايلوار، مَدْخَلٌ إلى اللِّسَانِيَّاتِ، تَرْجُمة: بدر الدِّين القَاسِمِ، 105-117.

الدَّخِلِيَّ لِلنُّصُوصِ، شرعت اللِّسَانِيَّاتُ التَّوْزِيعِيَّةُ منذ بلومفيلد بإعدادِ التَّحْلِيلِ إلى مُقَوِّمَاتٍ مُبَاشِرَةٍ⁽¹⁾، وقد اقْتَبَسَهَا كريستيان جوبير في مُحَاوَلَتِهِ الْعِلْمِيَّةِ، واستفادَ منها في التَّطْبِيقَاتِ الْحَاسُوبِيَّةِ الْمُخْتَلِفَةِ الَّتِي تَمَّ إِعْدَادُهَا.

وَتُصَاغُ الْمُسَلَّمَاتُ الْأَسَاسِيَّةُ لِهَذِهِ الْمُقَوِّمَاتِ كَمَا يَلِي:

- يَجُوزُ تَحْلِيلُ كُلِّ كَلِمَةٍ إِلَى ثَلَاثِ مُقَوِّمَاتٍ مُبَاشِرَةٍ: الْمَدْرَجِ الْأَسْفَلِيِّ، وَالْمَدْرَجِ الْفَعْلِيِّ، وَالْمَدْرَجِ الْحَرْفِيِّ أَوِ الْأَدَاتِيِّ.
- يَجُوزُ اعْتِبَارُ كُلِّ نَصٍّ سِلْسِلَةً مُتَرَابِطَةً مِنَ الْمُقَوِّمَاتِ.

وَالْمَلَاخِظُ أَنَّ هَذَا النَّمُودَجَ يُفِيدُ فِي تَمَثُّلِ عَدَدٍ وَافِرٍ مِنَ الْأَبْنِيَةِ التَّصْرِيفِيَّةِ وَصِغِهَا الصُّورِيَّةِ، وَقُدْرَةِ النَّمُودَجِ عَلَى التَّفْرِيعِ وَالتَّحْلِيلِ إِلَى مُقَوِّمَاتٍ مُبَاشِرَةٍ، وَيُظْهِرُ ذَلِكَ فِي التَّحْلِيلِ إِلَى أَصْغَرِ وَحْدَةٍ ذاتِ مَعْنَى⁽²⁾، أَوْ تَوَلِيدِ التَّنَوُّعَاتِ النُّطْقِيَّةِ لِلْفُونِيمِ الْأَصْلِيِّ بِمَا لَا يُؤَثِّرُ فِي الدَّلَالَةِ⁽³⁾، بَحِثٌ لَا يُمَكِّنُ تَجَزئةَ هَذِهِ الْوَحْدَةِ إِلَى وَحَدَاتٍ أَصْغَرٍ؛ وَمِنْ ثَمَّ فَهِيَ لَيْسَتْ أَبْنِيَّةً تَرْكِيبِيَّةً. وَنَذْكُرُ مِنْ خُصَائِصِ هَذِهِ الْمُؤَلَّفَاتِ أَنَّهَا تَدْخُلُ فِي تَكْوِينِ أَبْنِيَةِ صِغِيَّةٍ أَكْبَرَ مِنْهَا، مِثْلُ: السَّوَابِقِ، وَالِدَوَاخِلِ،

(1) المرجع السابق، /128.

(2) عَرَّفَ "المورفيم" بأنه «أصغر وحدة لغوية ذات معنى». يُنْظَرُ: ماريو باي، أُسُسُ عِلْمِ اللُّغَةِ، ترجمة أحمد مُخْتَارِ عُمَرُ، /53.

(3) عَرَّفَ "الفونيم" بأنه: «وحدة صوتية صغرى يُفْضِي تَغْيِيرُهَا إِلَى تَغْيِيرِ الْمَعْنَى». وَقِيلَ هُوَ «مجموعة، أو تنوع، أو ضَرْبٌ يَضُمُّ أَصْوَاتًا يُنْظَرُ إِلَيْهَا الْمُتَكَلِّمُونَ عَلَى أَنَّهَا تُمَثِّلُ وَحْدَةً وَاحِدَةً». وَيُطْلَقُ مُصْطَلَحُ "الألفون" عَلَى التَّنَوُّعَاتِ الصَّوْتِيَّةِ السِّيَاقِيَّةِ لِلْفُونِيمِ. رَاجِعْ فِي ذَلِكَ:

- J. Kramsky, The phoneme: An Introduction to the History and Theories of A Concept, p. 127.

- ماريو باي، أُسُسُ عِلْمِ اللُّغَةِ، ترجمة أحمد مُخْتَارِ عُمَرُ، /50.

- F. Katamba, An Introduction to Phonology, pp. 18-22.

واللّوآحق التي تأتي على شكل مَبَانٍ زائدة عن الأصل⁽¹⁾. وعلى ذلك ينبغي في التحليل أن نُحدّد المؤلفات المباشرة لأَيّة بنية تصريفية، وأن نلجأ إلى تفكيك الظاهرة الصرفية إلى عناصرها الأولى التي تتألف منها؛ لذا يجب تجزئة كل بناء وتقسيمه إلى أقل عددٍ جائز من المؤلفات البنيوية. كما يجب معرفة ما إذا كان المؤلفان المتجاوزان مُرتبطين في علاقة تركيبية قوية وواضحة ذات تمثيلٍ دلالي، وهذا ما يُورده Christian Gaubert (كريستيان جوير)⁽²⁾ في دراسته لإيجاد تفسيرات للعلاق الحفية بين البنية الظاهرة ودلالاتها المستترة. إن تحليل النصوص إلى مقوماتها المباشرة يزود المعالج بنظرة عامة، تظهر فيها العناصر مرتبة على مستويات بعضها فوق بعض⁽³⁾.

[ج] التطبيق يقوم على مكوّنين رياضيين هما: الجذر (ثلاثي أو رباعي)، والوزن (ف. ع. ل أو ف. ع. ل. ل)، حيث يتولّى الجذر المساهمة في وضع البنية الأساسية للكلمة، ويتولّى الوزن وضع هيكلها العام، وتوزيع الحركات وكذلك المورفيمات بنوعها (الحرة والمقيّدة) داخل بنية الكلمة، بهدف توليد الكلمات. ثمّ التعرّف الآلي على الحروف الزوائد على المباني الأصلية للكلمة. أمّا في مجال حلول الكلمة في مجال التركيب الجمليّ فيلجأ في عملية استخراج البيانات من النصوص إلى النحو الشكليّ؛ حيث يقبل عددًا من الانتقالات بين الحالات اللغوية المختلفة، وهو ما يمكن أن نطلق عليه اسم المحلل الصرف/نحويّ،

(1) راجع في ذلك المؤلفات الآتية:

- F. Katamba, Morphology, pp. 41-46.

(2) Voir Christian Gaubert, Analyse morphologique d'un texte arabe par ordinateur: méthode d'évaluation, résultats, in Annales islamologiques, tome XXIX, pp. 283-310.

(3) يُنظر: رونالد ايلوار، مدخل إلى اللسانيات، 131-128.

الذي يقوم بتفكيك الكلمة قيد التحليل عن طريق إزالة ما يتصل بها من سوابق ولواحق ودواخل، كذلك يعتمد إلى ردّ التعديلات الصوتية الحادثة في الكلمة إلى أصلها. كما يقوم بعملية التفكيك النحوي الوظيفي لأنواع الكلمات.

2. بلورة مفاهيم جديدة ومصطلحات مستخدمة:

2.1 التعريف بالمصطلحات:

يُنحصر الهدف من هذه الفقرة في التعريف بالمصطلحات المستخدمة عند جوبير في محاولته العلمية، وترجمة هذه المصطلحات إلى العربية؛ للاستئناس بها عند وضع المصطلح العلمي، وخاصة عند بناء قاعدة بيانات أو بنوك المصطلحات العربية. والنقطة الأخرى تتعلق بالمساهمة في تقديم مصطلحات المعلوماتية⁽¹⁾، وذلك رغبة في تعميم هذه المصطلحات لفائدة الباحثين العرب، وذلك تيسيراً لجوارٍ مستقبليٍّ وشيك بين اللسانيين والحاسوبيين، حيث يؤكد الباحثون والدارسون⁽²⁾ أنّ الثروة المصطلحية التي يحفل بها التراث العربي في سائر العلوم، لم تكن موحدةً بنظرية مصطلحية معروفة الخطوط والأهداف. وهذا يستدعي

(1) يُنظر: الطيّب البكوش، رضا السويدي، عبد المجيد بن حمادو، مُعجم المصطلحات اللسانية الإعلامية، 139-166. أحمد بو عزي، تعريب المصطلحات المستعملة في الحواسيب الصغرى، 97-117. عبد الرحمن بن عبد العزيز الفاضل، البنك الآلي السُعودي (باسم) تجربة عربية لتوثيق المصطلحات العلمية، مجلة اللسان العربي، ع47، 79-105. ليلى المسعودي، علم المصطلحات وبنوك المعطيات، مجلة اللسان العربي، ع28، 85-93.

(2) يُنظر على سبيل المثال المؤلفات التالية: محمد رشاد الحمزاوي، في سبيل نظرية مصطلحية عربية ممكنة، 86-89. سعد الحاج بكري، عدنان صديق نوح، محمد سمر قندي، مصطلحات المعلومات واللغة العربية (المشاكل والحلول)، 90-93. سهام محمد كعكي، مشكلات التعريب في علوم الحاسوب، 94-96. أحمد بو عزي، تعريب المصطلحات المستعملة في الحواسيب الصغرى، 97-117.

منّا استثمار نتائج الوضع والترجمة لمتابعة المستجدات في مجال اللسانيات الحاسوبية، والتفاعل بين العربية والمصطلحات المعلوماتية. فالمصطلح هو وحدة لغوية تأتي تسميةً لمفاهيم عديدة بين العلامات والأشياء⁽¹⁾، أو بتعبير آخر بين الدال والمدلول⁽²⁾. ولا تخفى خطورة المصطلح العلمي على مجموع الأبحاث والدراسات العلمية إن لم يُوظف بشكلٍ جيّد. كما أنّ إشكالية المصطلح تُلامس قضيةً جوهرية في نقل التكنولوجيا الحديثة، وتكنولوجيا اللغات. وفي هذا الصدد ينبغي مراعاة خصوصية المصطلحات، وبيان مفاهيمها عند تصميم المحلل الصّرفي، بحيث يشمل البحث تحديد الفوارق الدقيقة لها، واختيار التعريف المعياري المناسب لكل مصطلح؛ وذلك بهدف إدراك الأسس التي تستند إليها تقنية التحليل الصّرفي الآلي، وماهية المنطق البرمجي واللغوي الذي تُوظفه. وأهم المصطلحات الواردة التي نحاول أن نُلقّي عليها الضوء ما نتلمسه في السطور الآتية:

(1) Voir Ferdinand de Saussure, Cours de linguistique générale, pp. 97-99. Voir aussi à ce sujet:

- Manuel Celio, Concepts, termes et reformulations, in Travaux du CRTT, pp. 21, 49, 62-64.
- Henri Béjoint et Philippe Thoirion, « Modèle rationnel, définition et dénomination », in Autour de la dénomination, pp. 5-19.
- Nabil Esber, La terminologie de la linguistique: problèmes de traduction, français-arabe, pp. 36-37, 50.

(2) Voir Ferdinand de Saussure, Op. cit., pp. 100-103. Voir aussi:

- François Rastier, « Le terme: entre ontologie et linguistique », C.N.R.S Texte publié dans La banque des mots, n° 7, pp. 35-65.

2.2 قائمة الرموز والاختصارات المُستعملة مُرتبة حسب وُرودها في محاولة كريستيان جوير العَلَمِيَّة:

الرَّمز الاختصاري	التَّحْدِيد المِصْطَلَحِي	التَّرْجَمَة
G2	Grammaire morphologique non déterministe fondée sur l'analyse de textes	نَحْو غير حَتِيّ
Gn3	Grammaire formelle représentée par un automate	نَحْو شَكْلِيّ غير حَتِيّ
n3	Nominal trilitère	أَسْمَاء ثَلَاثِيَّة
c, c, c	Consonnes saines dans le cas présent	الْحُرُوف الْأَسَاسِيَّة فِي الْجِذْرِ
E _i	Un état initial	حَالَة أَصْلِيَّة
E _f	Un état final	حَالَة نِهَائِيَّة
Σ_r	Ensemble des consonnes nommées « solides »	الصَّوَامَت (الصَّلْبَة) الَّتِي تُعَدُّ أَصْلِيَّة فِي الْجِذْرِ
Σ_f	Son complémentaire dans l'ensemble des consonnes « saines »	المجموعة الثَّانِيَّة التَّكْمِيلِيَّة (مجموعة الحروف الصَّامِتة الصَّحِيحَة الَّتِي يُمَكِّنُ أَنْ تَلْتَصِق بِالْأَسْم)
R1	Élément radical 1	الْعُنْصُر الْأَسَاسِيّ الْأَوَّل فِي الْجِذْرِ
R2	Élément radical 2	الْعُنْصُر الْأَسَاسِيّ الثَّانِي فِي الْجِذْرِ

الرَّمز الاختصاري	التَّحْدِيد المصطلجي	التَّرْجَمَة
R3	Élément radical 3	العُنْصُر الأساسي الثَّالِث في الجِذْر
Int	Interrogatif	استفهامي
Coor	Coordonnant	عَاطِف
Prep	Préposition	حرف جَرّ
Art	Article	أداة التَّعْرِيف
Pref	Préfixe de schème	سَوَاقِب الصِّيْغَة
Inf1	Infixe de schème entre R1 et R2	دَوَاحِل الصِّيْغَة بين الجِذْر الأوَّل والجِذْر الثَّانِي
Inf2	Infixe de schème entre R2 et R3	دَوَاحِل الصِّيْغَة بين الجِذْر الثَّانِي والجِذْر الثَّالِث
Suff	Suffixe de schème	لَوَاحِق الصِّيْغَة
Nisb	Nisba	ياء النِّسَب
Sgfm	Singulier féminin	مُفْرَد مُؤَنَّث
Plma	Pluriel masculin	جَمْع مُذَكَّر
Plfm	Pluriel féminin	جَمْع مُؤَنَّث
Duma	Duel masculin	مُتَنِّي مُذَكَّر
Dufm	Duel féminin	مُتَنِّي مُؤَنَّث
Obnd	Cas direct non déterminé	حالة نَصْب مُباشرة غير مُحدَّدة
Post	Posfixe	لاحِقَة
Γ	Ensemble de productions écrites ou mots graphiques de l'arabe	مجموعة مِنَ الكِتَابَات أو الكلمات المطبوعة للغة العربية

الرَّمز الاختصاري	التَّحْدِيد المصطلحي	التَّرْجَمَة
G	Grammaire formelle morphologique conçue dans le but d'accepter les éléments de Γ en détaillant les ambiguïtés	النَّحْو الشَّكْلِي الصَّرْفِي الذي تَمَّ إعداده بهدف قبول العناصر التي تنتمي للمجموعة الكتابية للعربية مع توصيف أشكال اللبس
M	Mot arabe	الكلمة العربية
Nr	Nombre des éléments	عدد العناصر في أشكال التحليل (حالات لبس/حالات أخرى)
C	Consonne	صَوْت صَامِت
S	Semi-consonnes	نِصْف صَوَامِت (حَرْفَا الواو، والياء)
H	La hamza	الهَمْزَة
Ccc	Nom trilitère sain	الجُذُور الصَّحِيحَة للأسماء الثلاثية
Csc	Type racine (S)	الجُذُور الْمُعْتَلَّة الجَوْفَاء
Cc	Type racine (sourds)	الجُذُور الْمُضَعَّفَة
Ccs	Type racine (S)	الجُذُور الْمُعْتَلَّة النَّاقِصَة
Css	Type racine (SS)	الجُذُور ذات الصَّامِت مع حرفي عِلَّة (واو أو ياء)
Sc	Type racine (S)	الجُذُور الْمُعْتَلَّة المِثَال
Ssc	Type racine (SS)	الجُذُور ذات حرفي عِلَّة (واو أو ياء) مع الصَّامِت
Hcc	Type racine (H)	الجُذُور مَهْمُوزَة الحرف الأوَّل

الرمز الاختصاري	التحديد المصطلجي	الترجمة
chc	Type racine (H)	الجذور مَهْمُوزَة الحرف الثاني
cch	Type racine (H)	الجذور مَهْمُوزَة الحرف الثالث
chs	Type racine (HS)	الجذور مَهْمُوزَة الحرف الثاني مع اعتلال الحرف الثالث (واوي أو يائي)
csh	Type racine (SH)	الجذور مَهْمُوزَة الحرف الثالث مع اعتلال الحرف الثاني (واوي أو يائي)
hsc	Type racine (HSC)	الجذور مَهْمُوزَة الحرف الأول مع اعتلال الحرف الثاني (واوي أو يائي)
hcs	Type racine (HCS)	الجذور مَهْمُوزَة الحرف الأول مع اعتلال الحرف الثالث (واوي أو يائي)
hss	Type racine (HSS)	الجذور مَهْمُوزَة الحرف الأول مع اعتلال الحرفين الثاني والثالث (الواو أو الياء)
cc = 1	Noms sourds à consonnes solides (forme assimilée)	أسماء تتكوّن من صوامت (فيها حرف مُضَعَّف)
cc = 2	Noms sourds à consonnes solides (forme dissimilée)	أسماء تتكوّن من صوامت (فيها حرفان مُتَكَرِّرَان)
c1	Racine saine-1	جذر ذو حرف صامت-1

الرَّمز الاختصاري	التَّحْدِيد المصطلجي	التَّرْجَمَة
c2	Racine saine-2	جذر ذو حرف صامت-2
c ₃	Racine saine-3	جذر ذو حرف صامت-3
τ _v	Solutions accompagnées éventuellement de bruit radicalement exact/nombre de noms trilitères étudiés	مقياس التَّحْقُق مِنْ التَّحْلِيل (الحلول التي يصحبها ضجيج مُثَبَّتَة جذوره/عدد الأسماء الثلاثية المدروسة)
τ _{bra}	Bruit radicalement attesté/total des réponses	الضَّجِيج المُثَبَّتَة جذوره/مجموع الحلول

المبحث الثاني- في التوصيف

1- تَوْطِئَة:

1-1 مَيْكَنَة التَّحْلِيلِ الصَّرْفِي:

ثَمَّة عدد من الخوارزميات (*) الْمُعْتَمَدَة في تنفيذ المُحَلَّلَات الصَّرْفِيَّة، فبعضها يعتمد على أحد معجمات اللغة العربية، وبعضها يعتمد على معجم صرفي خاص بالمحلل (وهو قاعدة مُعْطِيَات تُخَزَّن فيها الجذور وجميع صور الكلمات المُشْتَقَّة من هذه الجذور)، وبعضها - كما في حالتنا هذه - لا يعتمد على معجم لغوي، ولا

(*) تُعرَّف الخوارزمية بأنها مجموعة من الخطوات المتسلسلة والمحددة، التي يسلكها المحلل أو المُعالِج، بهدف الوصول إلى إيجاد حلولٍ لِقَضَايَا مُعَيَّنَة، واستخلاص نتائج دقيقة لها. ويتعبّر آخر: هي توصيف دقيق وكامل على شكل خطوات متسلسلة معدودة ومعرفة، تُحدّد طريقة إنجاز عمل ما، أو حلّ مسألة ما. يُنظَر في ذلك: عبد الله عبد العزيز الموسى، مُقَدِّمَة في الحاسب والإنترنت. علي سليمان، مدخل إلى الحاسوب والخوارزميات.

على معجم صرفي، بل من خلال مُدَوَّنة مُرَمَّزة لأجزاء الكلام. ولكلٍّ من هذه الخوارزميات إيجابياتها وسلبياتها تتعلق بسرعة الأداء، والاستيعاب، والدمج في مستويات المعالجة الأخرى⁽¹⁾.

وسَمَّ نبيل علي "مَيْكَنَة التَّحْلِيلِ الصَّرْفِيِّ" بالنَّسَقِيَّةِ وفِكْرَةِ النِّظَامِ، مُقَرِّراً أَنَّهُ «يُقْصَدُ بِمَيْكَنَةِ التَّحْلِيلِ الصَّرْفِيِّ قِيَامُ النِّظَامِ الآلِيِّ باستخلاصِ العنصرِ الأوَّلِيَّةِ لِبُنْيَةِ الكلمة، وتحديدِ سِمَاتِهَا الصَّرْفِيَّةِ، والصَّرْفِ/نَحْوِيَّةِ، والدَّلَالِيَّةِ التي يُمَكِّنُ استنباطها مِنْ بُنْيَةِ الكلمة»⁽²⁾.

2-1 تَنَوُّعُ نَمَازِجِ مُعَالَجَةِ الصَّرْفِ أَلِيًّا:

- هُنَاكَ عِدَّةُ نماذجٍ لمُعَالَجَةِ الصَّرْفِ أَلِيًّا يُمَكِّنُ تَصْنِيفُهَا فِي الصُّوَرِ الآتِيَةِ:
- نَمَازِجُ تَنْظُرُ إِلَى الكلمةِ وَبُنْيَتِهَا بِاعْتِبَارِهَا سِلْسِلَةً مِنَ الرُّمُوزِ.
 - نَمَازِجُ ذاتِ مُستويين فِي التَّحْلِيلِ، تَعْتَمِدُ عَلَى إدْرَاكِ عِلَاقَةِ التَّنَازُّرِ التي تَرْبِطُ بين رموزِ البُنْيَةِ العَمِيقَةِ والسَّطْحِيَّةِ.
 - نَمَازِجُ لِلتَّحْلِيلِ تَقُومُ عَلَى نَظَرِيَّةِ التَّقْطِيعِ الدَّائِي.
 - نَمَازِجُ تُبْنَى عَلَى أَساسِ فَصْلِ قَوَاعِدِ الصَّرْفِ عن برنامجِ التَّحْلِيلِ، فَتَتِمُّ صِيَاغَةُ القَوَاعِدِ عَلَى هِيئَةِ نَحْوٍ ذِي حَسَاسِيَّةٍ سِيَّاقِيَّةٍ؛ لاستخلاصِ عناصرِ بُنْيَةِ الكلمات.
 - نَمَازِجُ التَّحْلِيلِ بِالترْكِيْبِ، وهو ما أُطْلِقَ عَلَيْهِ اسمُ "المُعَالِجِ الصَّرْفِيِّ مُتَعَدِّدِ الأطْوَارِ"، بحيثِ يَسْتَطِيعُ مُعَالَجَةُ العناصرِ الصَّرْفِ/نَحْوِيَّ، والعناصرِ

(1) رَاجِعْ فِي ذَلِكَ: مَنْصَفُ الشَّرْفِيِّ، وَهَانِي عِمَارِ ضَمْنِ أَعْمَالِ النَّدْوَةِ الدَّوْلِيَّةِ لِلْعِلْمِ وَهَنْدَسَةِ الْحَاسُوبِ: الدَّوْرَةُ السَّادِسَةُ، /365.

(2) نَبِيلُ عَلِي، اللُّغَةُ الْعَرَبِيَّةُ وَالْحَاسُوبُ، /306.

الاشتقاقية، والإعرابية. كما أنه قادرٌ على التعامل مع أطوار التشكيل المختلفة للكلمة العربية⁽¹⁾.

يُستشرفُ ممَّا تقدَّم أنَّ عمليَّات التحليل الصرفي تختلفُ من حيث درجة مُلاءمتها لمطالب الصرف العربي، وأنَّها تتميَّزُ بصُعوباتٍ في مراحل التطبيق عن بعضها الآخر؛ وذلك نظرًا لطبيعة التصريف، وما يحتاجه في التحليل من ردِّ الفرع إلى أصله، أو تضمُّن البنى السطحية لبنى عميقة، أو تحويل البنية العميقة إلى مُقابلها السطحي، أو استعادة البنية العميقة واستردادها من البنية السطحية.

2- جهود كلود أديبرت وأندريه چاكاريني متطلقات حاسوبية لدفع المحاولة العلمية لكريستيان جوبير:

إنَّ إمكانية التحليل الصرفي الآلي للغة العربية analyse morphologique automatique de l'arabe، واستخراج جذور الكلمات التي أوضحتها دراساتُ الباحثين Claude Audebert (كلود أديبرت) و André Jaccarini (أندريه چاكاريني)⁽²⁾ تفتحُ آفاقًا عريضةً في مجال المعالجة الآلية للغة العربية. وتكمن قوة

(1) المرجع السابق، /306-314.

(2) هذه الدراسة في جوهرها تسلكُ نهجَ المقالات التي نشرها كلود أديبرت وأندريه چاكاريني منذ 1987م، التي يستقي منها هذا البحثُ المصطلحات، والنتائج الأساسية. وسنكتفي هنا باستعراضٍ سريعٍ لطريقة عمل أبحاث النحو الشكلي بدون تفاصيل الأسس النظرية. ولمزيدٍ من بسط القول في المسألة راجع المؤلفات الآتية:

- Claude Audebert et André Jaccarini, (1986), « À la recherche du lābar. Outils en vue de la réalisation d'un programme d'enseignement de l'arabe assisté par ordinateur », in Annales Islamologiques, tome XXII, pp. 217-256.
- (1988), « De la reconnaissance des mots-outils et des tokens », in Annales islamologiques, tome XXIV, pp. 269-293.
- (1994), « Méthode de variations de grammaire et algorithme morphologique. Vers un

المنهج الذي تقوم عليه الدراسة في نقطتين هما: الإقرار بحالات اللبس اللغوي الطبيعي واللبس الكتابي الناجمين عن الصِّرف⁽¹⁾، بالإضافة إلى كفاءة الخوارزمية (المنهاج) المتبعة. وَمِنْ الْمُمكنِ أَنْ يَكُونَ النَّحْوُ الصَّرْفِيُّ قَابِلًا لِلتَّعْدِيلِ وَلِلتَّحْقِيقِ وَالبُلُورَةِ، وذلك عند أخذِ التَّنَوُّعَاتِ الْمُخْتَلِفَةِ لِلظُّوَاهِرِ النَّحْوِيَّةِ الْعَرَبِيَّةِ variations de grammaire في الاعتبار⁽²⁾. إِنَّ تصريفَ الاسمِ الثَّلَاثِيِّ جَدِيرٌ بالاهتمام، حيثُ تَحْتَلُّ هذه الفِئَةُ مِنَ الكَلِمَاتِ مكانة الصَّدَارَةِ بالمقارنة بالفئات الأخرى، مثل:

==

- extracteur de racine en arabe », in Bulletin d'études orientales, tome XLVI, pp. 79-98.
- André Jaccarini, (1989), "Automatic Generation of Unvocalized Arabic Recognition Programs: Theory and Applications", in Proceeding of the 2nd Conference en Arabic Computational Linguistic, pp. 66-87.
 - (1997), Grammaires modulaires de l'arabe, modélisations, mise en œuvre informatique et stratégies, thèse de doctorat, Université Paris IV.
 - (1999), « Vers une théorie du moniteur syntaxique », in Annales islamologiques, tome XXXIII, pp. 75-103.
 - (2000), « Quelques opérations sur les langages associés aux opérateurs syntaxiques », in Annales islamologiques, tome XXXIV, pp. 179-192.
 - (2001), "A modifiable structural editor of grammars for arabic processing", in Proceedings of the ACL/EACL Workshop, ARABIC Language Processing, Toulouse.

(1) رَاجِعْ فِي ذَلِكَ: مَهْدِي أَسْعَدُ عَزَّار، ظَاهِرَةُ اللَّبْسِ فِي الْعَرَبِيَّةِ جَدَلُ التَّوَاصُلِ وَالتَّفَاصُلِ، /99-121. وانظر أيضًا: حلمي خليل، العربية والغموض: دراسة لغوية في دلالة المبني على المعنى، /179-187. تَمَامَ حَسَّان، أَمْنُ اللَّبْسِ وَوَسَائِلُ الْوَصُولِ إِلَيْهِ فِي اللُّغَةِ الْعَرَبِيَّةِ، حَوْلِيَّةُ كَلِيَّةِ دَارِ الْعُلُومِ. بَرِيكَانُ بْنُ سَعْدِ الشَّلُوبِيِّ، اللَّبْسُ، أَسْبَابُهُ وَطُرُقُ اجْتِنَابِهِ فِي التَّفْقِيدِ الصَّرْفِيِّ، مجلة جامعة أم القرى لعلوم اللغات وأدائها، ع6، /45-90.

(2) قام Claude Audebert (كلود أوديبرت) و André Jaccarini (أندريه جاكارينى) بتوضيح هذا المفهوم في مقالٍ لهما بعنوان:

« Méthode de variations de grammaire et algorithme morphologique. Vers un extracteur de racine en arabe », in Bulletin d'études orientales, tome XLVI, pp. 79-98.

« منهج تَغْيِراتِ الظُّوَاهِرِ النَّحْوِيَّةِ والخوارزميات الصَّرْفِيَّةِ نحو إعداد برنامج استخراج الجذور في اللغة العربية ».

الأفعال أو الأدوات كما بيَّنها أوديبورت وچاكاريني⁽¹⁾، وكذلك الأسماء الأخرى الرباعية والثنائية والأعلام.

وقد طُوِّرَ أولُ نَحْوٍ⁽²⁾ غير حتمي⁽³⁾ *grammaire non déterministe*، وأُطلقَ عليه اسم G2، وهو الذي سمح بالتعرُّف على أغلب الصُّعُوباتِ النَّاجِمة عن مُرَاعَاةِ حَالَاتِ اللَّبْسِ اللُّغَوِيِّ، الذي يَنْشَأُ عن النموذج المبرمج، وكذلك الذي يخصّ طريقة كتابة الصَّرْف العربي.

يَقْتَرِحُ كريستيان جوبير في دِرَاسَتِهِ هنا مَنَهْجًا لتقييم وتطويرِ نَحْوٍ/صرفي شكلي، مَبْنِيٍّ على تحليلِ النُّصُوصِ⁽⁴⁾. وقد أُعِدَّ هذا المنهج، بفضل تصميمِ مسبق لبرنامج تحليل، مِّن السَّهْلِ تحويله إلى وحدة قياس لغويٍّ مستقل، أو يندرج ضمن برنامج أشمل وأوسع.

(1) انظر لنفس الباحثين: مقال "التَّعرُّف على الكلمات الوُظَيْفِيَّة".

- Claude Audebert et André Jaccarini, «De la reconnaissance des mots-outils et des tokens», in Annales islamologiques, tome XXIV, pp. 269-293.

(2) Voir Claude Audebert et André Jaccarini, « Méthode de variations de grammaire et algorithme morphologique. Vers un extracteur de racine en arabe », in Bulletin d'études orientales, tome XLVI, pp. 82-86.

(3) النَحْو غير الحتميِّ هو الذي يَسْمَحُ أثناء تحليل الكلمة بوجود عدَّة احتمالاتٍ قابلة للتَّحْلِيلِ النِّظَامِيِّ تلقائيًا.

(4) يُنْظَرُ في ذلك:

- Christian Gaubert, Op. cit., pp. 283-310.

وانظر أيضًا:

- N. Chomsky, & G.A MILLER, L'analyse formelle des langues naturelles.

- Alexander Meduna, Formal Languages and Computation: Models and Their Applications.

3- استخدام النحو الشكلي Gn3 في محاولة جوبير للتوصيف المرن للأسماء ثلاثية الجذر:

يَنبَغِي هنا التَّعْرِيجُ باقتضابٍ على ما يَشْتَمِلُ عليه مفهوم "النَّحو الصُّوري أو الشَّكْلِي"، بِحَيْثُ يَكُونُ بِمُكْنَةِ اللُّغَوِيِّ أَنْ يُنْزَلَ التَّعْبِيرُ الاصْطِلَاحِيَّ مَنْزِلَتُهُ الدَّلَالِيَّةُ الَّتِي يُسْتَوْدَعُ فِيهَا. وَلَعَلَّهُ مِمَّا يُجَلِّي هذا المفهوم ما أَشَارَ إِلَيْهِ طه عبد الرحمن⁽¹⁾ على صَعِيدِ تَطْبِيقِ الوَسَائِلِ الصُّورِيَّةِ وَالرِّيَاضِيَّةِ فِي مُعَالَجَةِ الظَّوَاهِرِ اللِّسَانِيَّةِ، وَصِيَاغَةِ الْمَنْطِقِ لِلتَّعَابِيرِ اللُّغَوِيَّةِ وَالْأَبْنِيَّةِ التَّرْكِيبِيَّةِ، وَالِاسْتِفَادَةِ مِنْ أُسَالِيهِهِ الْجِسَامِيَّةِ فِي وَصْفِ وَكَذَلِكَ فِي وَضْعِ أَنْحَاءٍ دَقِيقَةٍ لِللِّسَانِ، وَتَأْوِيلَاتٍ لَهُ بِذَاتِ الدِّقَّةِ، مَعَ بَيَانِ حُدُودِ هَذِهِ التَّأْوِيلَاتِ وَتِلْكَ الْأَنْحَاءِ، عَنْ طَرِيقِ قَوَاعِدِ مُطْلَقَةٍ لَا تَتَّقِيدُ بِالسِّيَاقِ، وَأُخْرَى مُقَيَّدَةً بِهِ.

وَيَتَّبِعُ أَنْدَرِيه جَاكَارِينِي بَيَانَهُ لِمَفْهُومِ "النَّحو الشَّكْلِي" وَضُرُورَتِهِ فِي التَّحْلِيلِ الْحَاسُوبِيِّ لِللُّغَةِ الْعَرَبِيَّةِ بِقَوْلِهِ:

« En effet, lors de la construction d'une grammaire informatique, le linguiste est constamment confronté à deux exigences qui le plus

(1) يُنْظَرُ: طه عبد الرحمن، المنطق والنَّحو الصُّوري، 9-22، 45-97، 143-146. وَيُرَاجَعُ أَيْضًا الْمَوْقُفَاتِ الْآتِيَّةُ:

- Pierrette Bouillon, Traitement automatique des langues naturelles, p. 78.
- A. Colmerauer, H. Kanoui, P. Roussel et R. Pasero, Un Système de Communication Homme-Machine en Français, Groupe de Recherche en Intelligence Artificielle.
- N. Chomsky, The logical structure of linguistic theory.
- N. Chomsky et M.P. Schutzenberger, "The algebraic theory of context-free languages", Computer programming and formal systems, Studies in logic series, pp. 119-161.
- N. Goodman, The structure of appearance.
- Eric Wehrli, L'analyse syntaxique des langues naturelles, p. 23.
- Maurice Gross et André Lentin, Notion sur les grammaires formelles, p. 30.

souvent (mais pas toujours) opèrent en sens inverse:

1. d'une part, il doit tendre vers l'exhaustivité de la description linguistique ;
2. d'autre part, il doit veiller à ce que sa grammaire n'atteigne pas un niveau de complexité qui serait cause d'un ralentissement important de l'analyse. Il doit notamment veiller à ce qu'elle contienne le plus grand nombre de règles à haut niveau de productivité, c'est-à-dire qui rendent compte des régularités les plus importantes et les plus fréquentes, et les privilégier par rapport aux autres. Or il se trouve, le plus souvent, que dans les grammaires de taille importante on est en présence d'un petit noyau très productif dans un ensemble de règles dont la productivité va en décroissant rapidement.»⁽¹⁾

إِحَالُ أَنْ مَا تَقَدَّمَ يُعَوِّزُهُ فَضْلُ بَيَانٍ؛ ذَلِكَ أَنَّنا يَنْبَغِي أَنْ نَنْظُرَ إِلَى قَوْلِ چَاكَارِينِي عَلَى أَنَّهُ لَا يُمَكِّنُ فَصْلَ التَّحْلِيلِ اللَّسَّانِي الْحَاسُوبِي بِدُونِ تَطْبِيقِ آيَاتِ النَّحْوِ الصُّورِيِّ، وَفِي هَذَا دَلَالَةٌ وَاضِحَةٌ عَلَى تَبَيُّنِ الْأَدْوَارِ الْوُضُفِيَّةِ الَّتِي يُؤَدِّيهَا النَّحْوُ الشَّكْلِي.

وَمِنْ الْمَعَانِي الَّتِي أَضَافَهَا چَاكَارِينِي وَأَلْمَحَ إِلَيْهَا فِي نَصِّهِ السَّابِقِ، أَنَّ اللَّجُوءَ إِلَى النَّحْوِ الصُّورِيِّ، مِنْ شَأْنِهِ أَنْ يُقَلِّلَ مِنْ صُعُوبَاتِ التَّحْلِيلِ التَّلْقَائِيِّ لِبِنْيَةِ الْأَنْظِمَةِ الْكِتَابِيَّةِ، وَالتَّصْرِيفِيَّةِ، وَالنَّحْوِيَّةِ لِلْغَةِ الْعَرَبِيَّةِ، وَيَجِبُ أَنْ تَظَلَّ وَسَائِلُ هَذَا النَّحْوِ مَفْتُوحَةً دَائِمًا لِلسَّمَّاحِ بِتَحْسِينِهِ.

(1) Voir André Jaccarini, Grammaires modulaires de l'arabe, modélisations, mise en œuvre informatique et stratégies, tome 1, pp. 7, 11.

- وَمِنْ خِلَالِ بِنَاءِ النَّحْوِ الْحَاسُوبِيِّ، يُوَاجِهُ اللُّغَوِيُّ بِاسْتِمْرَارِ اثْنَيْنِ مِنَ الْمُتَطَلِّبَاتِ التي في معظم الأحيان تعمل في الاتجاه المعاكس:
- 1- السَّعْيُ إِلَى تَحْقِيقِ الشُّمُولِيَّةِ فِي وَصْفِ اللُّغَةِ.
 - 2- الإلمام بالقواعد العامة والجُزئية لبنية اللغة، بحيث لا تصل هذه القواعد إلى مستوى مِنَ التَّعْقِيدِ، الذي مِنْ شَأْنِهِ أَنْ يُسَبِّبَ تَبَاطُؤًا كَبِيرًا فِي التَّحْلِيلِ. كَمَا يَجِبُ عَلَيْنَا - علي وجه الخصوص - أَنْ نَضْمَنَ لِلنَّحْوِ الصُّورِيِّ احْتِوَاءَهُ عَلَى أَكْبَرِ عَدَدٍ مِنَ قَوَاعِدِ الْإِنْتاجِيَّةِ الْعَالِيَةِ، أي التي تَعَكِّسُ أَهَمَّ الْقَوَانِينِ وَأَكْثَرَهَا تَوَاتُرًا.
- وحسب كريستيان جوبير يُمكنُ تعريف النَّحْوِ الصُّورِيِّ التَّطْبِيقِيِّ الذي يُمَثِّلُهُ المَعَالِجُ الِآلِيَّةُ بِأَنَّهُ:

« Une grammaire représentée par un automate peut être décrite comme un ensemble d'états dont un état initial E_i et un état final E_f . Les variables d'entrée sont des mots ou suites de caractères pris dans un ensemble désigné par *alphabet*. Chaque état constitue une étape de l'analyse du mot, l'étape suivante étant atteinte par la lecture d'un caractère selon une catégorie linguistique précise: ce sont là les transitions ou règles de réécriture, aisément représentables par arcs orientés dans un diagramme. Chaque règle ou arc correspond à une catégorie linguistique précise à l'intérieur du mot. Des transitions particulières appelées *epsilon-transitions* permettent le passage direct d'un état à un autre lorsque la ou les catégories intermédiaires ne figurent pas dans le mot. Si le dernier caractère du mot analysé emprunte un arc qui peut mener directement à l'état final, le mot est accepté et donc conforme à la grammaire. »⁽¹⁾

(1) Voir Christian Gaubert, Op. cit., p. 284.

إنَّ المتأملَ في حديثِ جوبير السابق يجد أنَّ ما يعرضه من دلائلٍ لقواعد النُّحو الشُّكليِّ يَتمثَّلُ في مجموعةٍ من الحالات التي تتألف من حالةٍ أصلية E_i ، وحالةٍ نهائية E_f ، والمدخلات المتغيرة تكون عبارة عن كلماتٍ، أو عدَّة حُرُوفٍ تنتمي إلى مجموعة تُسمَّى الألف باء، وتُشكِّلُ كُلَّ حالةٍ خُطوةً في تحليل الكلمة، والحالة التالية تكون عند قراءة حرفٍ وفقاً لفئةٍ لغويةٍ مُحدَّدة، وهذه هي مراحل الانتقال، أو قواعد إعادة الكتابة التي يسهلُ تمثيلها بواسطة أقواسٍ في رسمٍ توضيحيٍّ. كُلُّ قاعدةٍ أو قوسٍ يختصُّ بفئةٍ لغويةٍ محدَّدة للكلمة. وهناك انتقالاتٌ خاصَّة تُسمَّى epsilon-transitions تسمح بالمرور المباشر من حالةٍ إلى أخرى، عندما لا تظهر الفئة أو الفئات الوسيطة (البينية) في الكلمة. فإذا كان آخرُ حرفٍ من الكلمة التي يتم تحليلها يتبع قوساً يمكنه من الوصول مباشرةً إلى الحالة النهائية، فيتم قبول الكلمة، وتُصبحُ مُتَوَافِقةً مع النُّحو الشُّكليِّ. ونلاحظ حينئذٍ أنَّ الكلمة تكونُ مُتَوَافِقةً لعدَّة أسبابٍ نظراً لعدم حتمية النموذج. وعلى العكس إذا لم يُمكن ربط الحرف الأخير بالحالة النهائية تكون الكلمة مرفوضةً، وبالتالي غير مبنية وفقاً لقواعد النُّحو الشُّكليِّ.

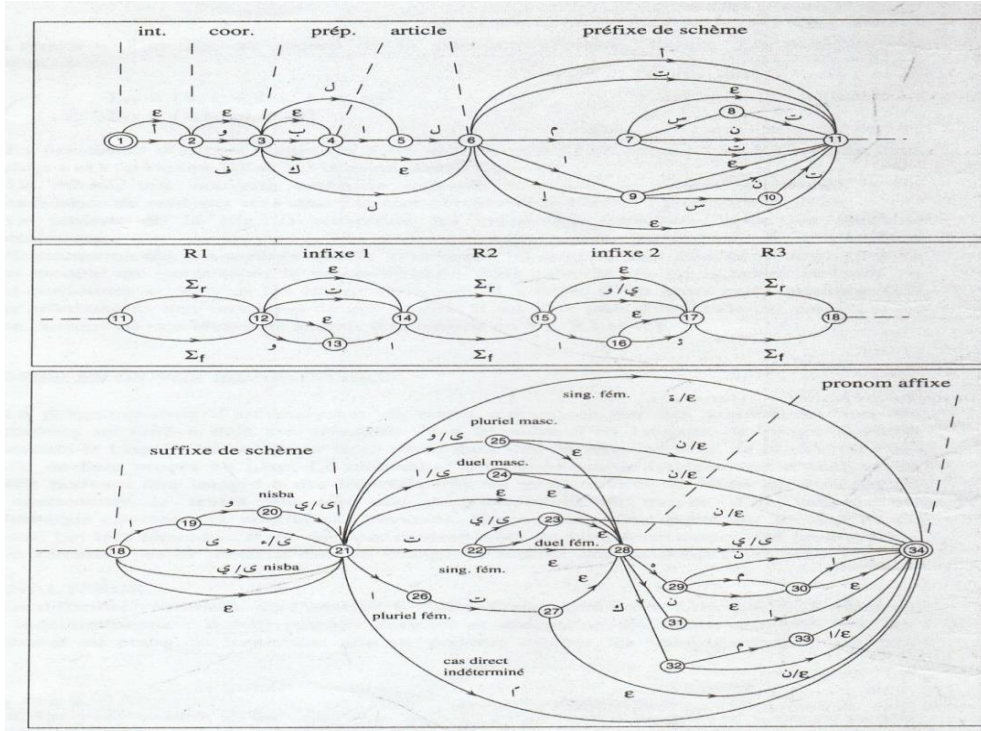
إنَّ هذا النُّحو الشُّكليِّ غير الحتمي Gn3 يُعدُّ نُسخةً مُعدَّلةً⁽¹⁾ بدرجةٍ قليلةٍ من النُّحو الشُّكليِّ G2، وهذا النُّحو يَعْمِدُ إلى تحليهِ دَقِيقٍ، لكنَّه لا يهتمُّ بأنَّ يشملَ كافة السَّوَاقِ واللَّوَاقِ المُحتمَلةَ للاسم العربيِّ، وذلك في حالة المصدر الثلاثيِّ الصَّحيح. وينطبق هذا النُّحو على الكلمات التي تتكوَّن من مجموعةٍ من الحُرُوفِ العربية باستثناء علامات التَّشكيلِ الخَارِجِيَّة (الفتحة، والضَّمة، والكسرة، والسُّكُون، والوَصلَة، والتَّنوين، والشَّدة). أمَّا الهمزة فهي على العكس تُؤخَذُ في

(1) هذه التَّعديلات المشار إليها والمُحمَّولة على النُّحو الشُّكليِّ سَتَتَّضِحُ للقارئ بصورةٍ تَفْصِيلِيَّةٍ في الصَّفحات التالية من هذا البَحْث.

الاعتبار بكل أشكالها. وبناءً على ذلك تتضح جميع أشكال الاسم العربي (المصدر) بشكل جلي، وعند استخراجها تُساعد على تقدير الجذور المحتملة للاسم (أو المصدر)، وكذلك على صياغة الفروض حول إعراب الكلمة داخل الجملة في بعض الحالات.

إن النموذج ¹⁾Gn3 (c, c, c) الثلاثي الصحيح (فعل) (ص، ص، ص)، حيث: (ص = اختصار صامت)، الذي يوضحه الشكل الأول Fig.1 من رسم بيان الحالات Diagramme d'états على هيئة Graph de transition يُمثل إحدى حالات تطوير هذا النحو، وهو المرجع الرئيسي الذي يؤخذ به في هذا البحث.

(1) يستخدّم كريستيان جوبير هذا التّحديد الوظيفيّ عند الضرورة. و « n3 » عنده هي اختصار لـ (اسم ثلاثي)، والحروف (c, c, c) تُقابل بالعربية (ص، ص، ص)، - حيث "ص" تعني اختصار "صامت" - هي الحروف الأساسية أو الأصلية في الجذر، وليس بها حروف علة، ويُرمز لها بـ R1/, R2/, R3، أي: بنية الجذر الأول، بنية الجذر الثاني، بنية الجذر الثالث.



الشكل الأول: بيان حالات تطوير النحو الشكلي

كما يجدر بنا أن نهتم بالجزء الجذري في هذا النحو، حيث إن وجود حروف مزيّدة له تأثير سلبي، يدفعنا إلى التفريق بين نوعين من الصّوامت. وهناك المجموعة الأولى Σ_r وهي الصّوامت (الصّلبة) التي تعدّ أصلية في الجذر، وهي لا يمكن فصلها، فيما عدا حالات استثنائية نادرة، تنتج من إضافة حروف مزيّدة إلى عمليات الاشتقاق وفقاً للشكل الثامن VIII⁽¹⁾.

(1) كما في حالة حرف الدال "د" في كلمة "ازدهار"، وحرف الطاء "ط" في كلمة "مُصْطَنع". ويُريد بالشكل الثامن هنا أحد أشكال الصّبيغ الصرفية لأبنية المشتقات وأصول أفعالها المشهورة عند العرب كما تقول تغريد عنبر: « [...] أوزان الفعل التي حددها اللغويون العرب، ويبلغ عددها تسع عشرة صيغة. »

أما المجموعة الثانية التكميلية $\sum_f$ ، فهي مجموعة الحروف الصامتة الصحيحة، وهي الحروف التي يُمكن أن تلتصق بالاسم مثل: أداة التعريف، وحروف الجرّ، والأعجاز أو اللواحق، مثل: الضمائر المتصلة: واو الفاعلين، تاء الفاعل، نون النسوة، ياء المؤنثة المخاطبة، ألف الاثنين: قاموا، قُمتُ، قُمن، قومي، قاما. ونون الوقاية = وَفَقْنِي، وحركات الإعراب وحروفه، وعلامات التانيث: كَتَبَتْ، وعلامات التثنية والجمع: كِتَابَان، مُعَلِّمُونَ. وكذلك العناصر المكوّنة للأوزان الاشتقاقية⁽¹⁾.

المجموعة الأولى $\sum_r$ = (ث، ج، ح، خ، د، ذ، ر، ز، ش، ص، ض، ط، ظ، ع، غ، ق).

==

راجع: تغريد عنبر، الفعل الماضي مُسنَدًا إلى ضَمَائِرِ الرُّفْعِ المتصلة دراسة صرفصوتية، المجلة العربية للدراسات اللغوية، 69/.

(1) لقد وَقَفَ بعض الدارسين المعاصرين عند بَحْثِهِمْ لطبيعة البَظَام اللغوي العربي وَفَقَهُ التَّعَامُلَ مع هذه اللغة إلى القول بأن العربية تتميز بلزوم "طَبَائِعَ" خاصة مِنْهَا: (أ)- الطَّبِيعَةُ الاشتقاقية: وهي تعني أَنَّ اللغةَ تعتمدُ نظامًا قائمًا على الاشتقاق مِنْ جَذَرٍ صامِتٍ واحد، وتَتَعَدَّدُ اشتقاقاته دون أَنَّ تكون مُحَدَّدَةً. (ب)- الطَّبِيعَةُ الإِلْصَاقِيَّةُ: وهي تعني أَنَّ اللغةَ تعتمدُ في توليداتها الصَّرْفِيَّةِ على مبدأ يَخْتَلِفُ عن الاشتقاق مِنْ جَذَرٍ واحد؛ لَأَنَّهَا في واقع الأمر لا تَمْتَلِكُ جذورًا، بل تَمْتَلِكُ مورفيمات مُسْتَقَلَّةً قابِلةً للتَكَثُّرِ عن طريق اللواصِق، وهذه اللواصِق قد تكون في أول الكلمة، وعند ذلك تُسَمَّى بالسَّوَابِقِ prefixes؛ لَأَنَّهَا تسبِقُ الكلمة لتُعْطِيها معنى جديدًا، وقد تكون هذه اللواصِق في آخِر الكلمة، فتمنحها معنى جديدًا، وعند ذلك تُسَمَّى باللَّوَاخِقِ suffixes، وأما إذا كانت في داخِلِ البَيْئَةِ، فإنَّهَا تُسَمَّى عند ذلك بالحِشَوَاتِ أو (الحشو) infixes. (ج)- الطَّبِيعَةُ الاشتقاقِيَّةُ الإِلْصَاقِيَّةُ: وهي تعني أَنَّ اللغةَ العربيةَ تَجْمَعُ بَيْنَ الطَّبِيعَتَيْنِ السَّابِقَتَيْنِ، فيمكنُ أَنْ تعتمدَ في التَكَثُّرِ الصَّرْفِيِّ وتوليد المورفيمات على الاشتقاق اعتمادًا كبيرًا، كما يُمكنُ أَنْ تعتمدَ على نظام الإِلْصَاقِ أيضًا. وتأسيسًا على ما تَقَدَّمَ لا يُمكنُ الاعتماد على الأداء المباشر للكلمة، بل لا بُدَّ مِنَ النَّظَرِ إلى الجَذَرِ الصَّامِتِ للمفردة، وإدراج اشتقاقاته المختلفة تحته. وَضِمْنَ التَّصْنِيفِ تحت الجَذَرِ، يُمكنُ إضافة جميع الأداءات اللغوية القائمة على اعتماد نظام الإِلْصَاقِ أيضًا دون نظَرٍ إلى المورفيم الكَلِّي؛ إذ يبقى الجَذَرُ الصَّامِتِ هو الأصل في ترتيب المواد اللغوية. يُرَاجَعُ في ذلك: يحيى عباينة، الصَّرف العربي التَّحْلِيلِيّ نظرات مُعَاَصِرَة، 14-13/.

المجموعة الثانية $\sum_f = (ب، ت، س، ف، ك، ل، م، ن، هـ)$.
 إنَّ أقواس $\sum_d$ و $\sum_f$ في حالة الحُرُوف الأصول⁽¹⁾ التي تُمثِّل الجذور R1/, R2/, R3 تُجسِّد احتمالات التَّوصُّل لاكتشاف الجذر الثلاثي الصَّحيح بواسطة النَّحو الشَّكلي Gn3. ويُطلَق على الحُرُوف المَزِيْدَة عَلَى مَبْنَى الكَلِمَة - التي لا تُشكِّل جُزْءًا أَصِيْلًا مِنَ الصِّيْغَة الصَّرْفِيَّة، أو مِنَ الجذر - حُرُوفٌ لاصِقَة سَابِقَة (قَبْلَ السَّلْسَلَة)، أو لَاحِقَة (بَعْدَ السَّلْسَلَة) وَفَقًا لِمَكَانِهَا.

أَمَّا المِجْمُوعَةُ الثَّانِيَّةُ $\sum_f$ ، فَمِنْ تَشَكُّلِ اِحْتِمَالَاتٍ لِلِخَطَأِ، كَمَا تُعَدُّ فِي الْوَقْتِ نَفْسِهِ اتِّحَادًا لِنَوْعَيْنِ مِنَ المِجْمُوعَاتِ الْفَرْعِيَّةِ لِلصَّوَامِتِ: النَّوعُ الْأَوَّلُ $\sum_{f1}$ يَخْتَصُّ بِاللَّوَاحِقِ (بَعْدَ السَّلْسَلَة): $\sum_{f1} = (ب، ت، س، ف، ك، ل، م، ن)$. والنَّوعُ الْآخَرُ $f3 \sum$ وَيَخْتَصُّ بِالسَّوَابِقِ (قَبْلَ السَّلْسَلَة)، وَيَضُمُّ: (ت، ك، م، ن، هـ).

وَقَدْ تَمَّتْ إِضَافَةُ الْحُرُوفِ (الدَّوَاحِلِ) الْمَزِيْدَة الْمُرَكَّبَة، مِثْل: الْوَاوُ + الْأَلْفُ (و ا) بَيْنَ الْحَرْفِ الْأَصْلِيِّ الْأَوَّلِ (الْجِذْرِ الْأَوَّلِ R1)، وَالْحَرْفِ الْأَصْلِيِّ الثَّانِي (الْجِذْرِ الثَّانِي R2)، كَمَا فِي كَلِمَةِ (عَوَاطِفَ)، وَكَذَلِكَ (ا، ئ) بَيْنَ الْجِذْرَيْنِ الثَّانِي وَالثَّالِثِ فِي نَحْوِ كَلِمَةِ (مَلَائِكَة). وَقَدْ يَحْدُثُ أَيْضًا أَنْ يَتِمَّ إِدْرَاجُ فِئَةٍ جَدِيدَةٍ سُمِّيَتْ « suffixe du schème » "الحُرُوفِ اللَّاحِقَة الْخَاصَّة بِالْأَوْزَانِ"، مِمَّا يَسْمَحُ بِالتَّعَرُّفِ عَلَى أَوْزَانِ مِثْل: (فَعْلَاءَ)، وَمِرَاقِبَة تَصْرِيفِهَا بَعْدَ إِضَافَةِ النِّسْبَةِ⁽²⁾.

(1) نُقَسِمُ الْحُرُوفَ فِي الْكَلِمَةِ إِلَى أَصُولٍ وَزَوَائِدَ. فَالْأَصُولُ هِيَ مَا يَلْزِمُ الْكَلِمَةَ عِنْدَ تَصْرِيفِهَا مِنْ هَيْئَةٍ إِلَى هَيْئَةٍ، إِلَّا مَا حُذِفَ وَهُوَ مُعْتَلٌّ فَأَصْلِيٌّ أَيْضًا بِأَشْكَالِهِ. وَالزَّوَائِدُ هِيَ مَا زِيدَ عَلَى الْأَصُولِ عِنْدَ التَّصْرِيفِ، وَانْتَقَالَ الْكَلِمَةُ مِنْ هَيْئَةٍ إِلَى هَيْئَةٍ، وَتُجْمَعُ فِي أَحْرَفٍ كَلِمَة: سَأَلْتُمُونَهَا، بِالْإِضَافَةِ إِلَى التَّضْعِيفِ (كَمَا هِيَ الْحَالُ فِي الْوِزْنِ الْمَزِيدِ افْعُوْعَلْ ؛ إِذْ تَكَرَّرَ عَيْنُ الْكَلِمَةِ، أَوْ فِي افْعِيعَالِ، ... إلخ). يُنْظَرُ: دِيزِرِه سَقَال، الصَّرْفُ وَعِلْمُ الْأَصْوَاتِ، 16/.

(2) Voir Christian Gaubert, Op. cit., pp. 285-287.

وَيُوضَّحُ جدول الشَّكل الثَّاني Fig. 2 التَّالي جميع الفئات المُدرَّجَة، مَعَ ذِكرِ مِثَالٍ لِيُظْهِرَ كُلٌّ مِنْهَا فِي النِّصِّ.

الْفِئَة	أَمْثَلَة	سِيَّاقُ النُّصُوصِ العَرَبِيَّة	صَفَحَاتُ تَوْثِيقِ النُّصُوصِ (مِنْ فُصُولٍ مُتَّفَرِّقَة لِيُحْيِي حَقِّي، قَنْدِيلُ أُمِّ هَاشِمٍ) ⁽¹⁾
- استفهام	← - أ أُبقلمه ؟	- « [...] سِيرَة ذَاتِيَّة بِقَلَمٍ [...] ».	- ص: 9.
- حرف عطف	← - و وغيره	- « [...] وَغَرِيْزَةُ التَّقْلِيدِ تُغْنِي عَنِ الدَّفْعِ [...] ».	- ص: 59. - ص: 59.
- حرف جرّ	← - ب بزيارة	- « [...] إِذَا قَدِمَ الْقَاهِرَة وهو صبي مع رجال الأسرة ونساءها للتبرُّك بزيارة أهل البيت [...] ».	- ص: 59، 61، 63، 75، 76، 81، 84، 93، 97، 98، 103، 105، 118.
- أداة التعريف	← - ال الشيخ	- « كَانَ جَدِّي الشَّيْخُ رجب عبد الله إِذَا قَدِمَ القاهرة [...] ».	

(1) رواية "قنديل أم هاشم" مِنْ بَيْنِ أشهر الروايات التي كتبها الروائي الكبير يحيى حقي، بين عامي 1939م، و1949م، ونُشِرَتْ لأول مرة في سلسلة "اقرأ"، 18 يونيو 1944م، ثُمَّ نُشِرَتْ فيما بعد بطبعاتٍ أُخرى، أَضَافَ إِلَيْهَا يحيى حقي سيرته الذاتية، فَطُبِعَتْ مَعَ رواية "قنديل أم هاشم". وَرَوَايَة "قنديل أم هاشم" تُعْتَبَرُ مِنْ بَيْنِ أهمِّ روايات يحيى حقي، وتحكي عن شاب مصري يعيش في وسط أناس قرويين، ثم يذهب لدراسة الطب في أوروبا، ويعود بعد سنواتٍ ليعمل طبيباً للعيون في مصر، وهنا تبدأ أحداثُ القصة، ويعيش هذا الشاب في تذوق معاناة الخرافة، ويجاهدها بطريقةٍ عنيفة، ثم تستمر أحداثُ قصة هذا الشاب الذي كانت أسرته تحترق ليعيش هو، لعله يحقق بعض آمالها ... رَاجِع: يحيى حقي، قنديل أم هاشم، 59-122.

الْفِئَة	أَمْثَلَة	سِيَاقُ النُّصُوصِ الْعَرَبِيَّةِ	صَفَحَاتُ تَوْثِيقِ النُّصُوصِ (مِنْ فُصُولٍ مُتَفَرِّقَةٍ لِيَحْيِي حَقِّي، قُنْدِيلُ أُمِّ هَاشِمٍ)
الجذر والوزن/ال صِيغَة	- سَوَابِقِ الصِّيغَة/الْوِزْنِ - الحرف الأصليّ الأول في الجذر - حرف مزيد (مِنْ الدَّوَاخِلِ) بين الأصلين الأول والثاني - الحرف الأصليّ الثاني في الجذر - حرف مزيد (مِنْ الدَّوَاخِلِ) بين الأصلين الثاني والثالث - الحرف الأصليّ الثالث في الجذر - لَوَاحِقِ الصِّيغَة/الْوِزْنِ	- « [...] وَغَرِيزَةُ التَّقْلِيدِ تُغْنِي عَنِ الدَّفْعِ [...] » - « [...] وَغَرِيزَةُ التَّقْلِيدِ تُغْنِي عَنِ الدَّفْعِ [...] » - « كَانِ جَدِّي الشَّيْخُ رَجَبُ عَبْدِ اللَّهِ إِذَا قَدِمَ الْقَاهِرَةَ [...] » - « [...] وَغَرِيزَةُ التَّقْلِيدِ تُغْنِي عَنِ الدَّفْعِ [...] » - « [...] وَغَرِيزَةُ التَّقْلِيدِ تُغْنِي عَنِ الدَّفْعِ [...] » - « [...] وَغَرِيزَةُ التَّقْلِيدِ تُغْنِي عَنِ الدَّفْعِ [...] » - « [...] وَتَخَلَّصَ مِنْ الرُّؤَاوِ وَالْغُرَبَاءِ » - « [...] أَمَّا أَغْلَبِيَّةُ الشَّعْبِ فَتَبَسَّمَتْ لِسَدَاجَةِ هَؤُلَاءِ الْقُرَوِيِّينَ » - « [...] فَمَهْوِيٌّ مَعَهُمْ عَلَى عَتَبَتِهِ الرُّحَامِيَّةِ يَرْشُقُهَا	- ص: 59. - ص: 59. - ص: 59. - ص: 59. - ص: 59. - ص: 66. - ص: 59-60. - ص: 59. - ص: 60.
بَعْدَ السَّلْسَلَةِ (إِضَافَة لَوَاحِقِ)	- مُفْرَدٌ مُؤَنَّثٌ بقبلته،	- ت التقليد - ق التقليد - ا القاهرة - ل التقليد - ي التقليد - د التقليد - اء والغرباء - ي القرويين - ت بقبلته،	

صفحات توثيق النصوص (من فصول مُتفرقة ليعي حقي، قنديل أم هاشم)	سياق النصوص العربية	أمثلة	الفئة	
- ص: 59.	بِقُبَلَاتِهِ [بِقُبَلَتِهِ] «...» - وَهَكَذَا عَاشَتْ الْأُسْرَةُ فِي رِكَابِ "السِّتِ" وفي جَمَاهَا «...» - وَأَقْدَامُ الدَّاخِلِينَ وَالخَارِجِينَ تَكَادُ تَصْدِمُ رَأْسَهُ «...» - [...] وَهَذَا مِنْ كَرَامَاتِ أُمِّ هَاشِمٍ «...» - [...] تَمْتَلِئُ الْيَدَانِ مِنْ جَدِيدٍ شَيْئًا فَشَيْئًا «...» - [...] صَوْنُهَا الصَّارِخِ يَجْذِبُ الْوُجُوهَ النَّوَافِدَ، وَعَيْنَاهَا السَّاجِرَتَانِ تَسْتَهْوِيَانِ الْمُطَلَّاتِ [...] «...» - [...] وَإِذَا شَاهَدَ فِعْلَهُمْ أَحَدُ رِجَالِ الدِّينِ الْمُتَعَالِينَ، أَشَاحَ بِوُجْهِهِ نَاقِمًا عَلَى الرِّمَنِ «...» - [...] إِذَا قَدِمَ الْقَاهِرَةُ وهو صبي مع رجال الأسرة ونسائها للتَّبَرُّكِ بِزِيَارَةِ أَهْلِ الْبَيْتِ [...] «...»	عَاشَتْ ← - ين الداخِلين ← - ات كَرَامَاتِ ← - ان الرَّائِرَانِ، الْيَدَانِ - تان ← السَّاجِرَتَانِ ← - ا نَاقِمًا ← - ها ونسائها	- جمع مُذَكَّر سالم - جمع مُؤنَّث - مُثَنَّى مُذَكَّر - مُثَنَّى مُؤنَّث - حالة نَصْب مباشرة غير مُحَدَّدَة - لاصِقَة/ لَوَاصِق (ضمير متصل في آخر الكلمة)	

الشكل الثاني : جدول فئات النحو الشكلي

نلاحظ إذن أن النحو الشكلي الثلاثي $Gn3(c, c, c)$ ، يُمكن أن يتضمنه نموذج يسمح بالتعرّف في بعض الحالات على جذور تحتوي حُرُوفًا نصف صامتة -semi consonnes أو ما نُسَمِّيه في الدرس الصوتي الحديث بأنصاف الحُرُوف أو أنصاف الحركات العربية⁽¹⁾ مثل: (الواو، أو الياء)، والحُرُوف التي بها همزة. وقد ذهب جوبير في هذا السياق إلى إدراج تغيرات الظواهر النحوية في هذا الاتجاه، حيث يتّضح من ذلك أنه إذا كان الجذر صحيحًا فلا يسمَح النظام الآلي بتصوّر وجود حرفي (الواو، أو الياء) بين الحُرُوف الأصلية.

المبحث الثالث- في التطبيق الآلي

1- قاعدة المعطيات اللغوية والتعامل مع برامج معلوماتية:

تتألف قاعدة المعطيات من كلمات اللغة العربية، وفي مختلف أشكالها التصريفية والصرفية: الاسم، والفعل، والأداة. وعلى صعيد التعامل مع برامج معلوماتية تهدف إلى التحليل الآلي للغة العربية، فقد تمّ بالفعل في محاولة كريستيان جوبير العلمية برمجته مُحلّل للنصوص، يعتمد على القواعد الشكلية

(1) أنصاف الحُرُوف أو أنصاف الحركات العربية: هي حركات بسيطة أو مركبة تقوم بدور الحرف أحياناً، ونجد منها في العربية نوعين هما: الواو والياء. هذا الازدواج في دور كلٍ منهما جعل العربية تستعمل للحالتين رمزاً واحداً (و-ي): فالواو حرف في مثل: وحد، وحركة في مثل: يوجد (أي ضمة طويلة). والياء حرف في مثل: يس، وحركة في مثل: يبيع (أي كسرة طويلة). واستعمال نفس الرمز للحرف والحركة، يؤدي إلى اللبس أحياناً؛ لذلك قد يكون من الأوضح استعمال الرموز العالمية عند الكتابة الصوتية: (الواو حرف: w، والواو الحركة: ū، والياء الحرف: y، والياء الحركة: ī). أمّا الواو في مثل: يوم، والياء في مثل: ميل، فتكونان مع الفتحة السابقة لهما حركتين مُزدوجتين هما: (و: aw - ي: ay). فهذا الدور المزدوج يجعل هذين الصوتين ضعيفي الاستقرار: كثيري الحذف، سهلي الإدغام، وهو ما جعل النحاة العرب يصفونهما بالاعتلال أي المرض. يُراجع في ذلك: الطيّب البكوش، التصريف العربي من خلال علم الأصوات الحديث، /53-54.

أو النَّحْوِ الشَّكْلِيِّ غير الحَثِّي (النَّحْوِ الشَّكْلِيِّ الثَّلَاثِيِّ) Gn3، في إطارٍ مَا يُعْرَفُ بـ "لغة البرمجة المفسرة، أو لغة البرمجة التَّفْصِيلِيَّةِ Lisp"⁽¹⁾. وقد اسْتَفَادَ هَذَا الْمُحَلِّلُ مِنَ الْبِنْيَةِ الطَّبِيعِيَّةِ لِلْقَائِمَةِ الْخَاصَّةِ بِلُغَةِ Lisp إِلَى حَدٍّ بَعِيدٍ.

إِنَّ إِعْدَادَ بَرَامِجٍ مُسْتَقِلٍّ لِتَحْلِيلِ اللُّغَةِ الْعَرَبِيَّةِ - يُمَكِّنُ دَمَجَهُ فِي بَرَامِجٍ حَاسُوبِيَّةٍ أُخْرَى مِثْلَ بَرَامِجِ التَّحْكُمِ فِي قَوَاعِدِ الْبَيِّنَاتِ أَوْ مُعَالَجَةِ النُّصُوصِ - قَدْ فَرَضَ ضَرُورَةَ الْقِيَامِ بِالْبَرْمَجَةِ بِوَسْطَةِ لُغَةِ خَوَارِزْمِيَّةٍ تَقْبَلُ الْإِحْتِمَالَاتِ الْمُسْتَقْبَلِيَّةَ.

لَقَدْ وَقَعَ اخْتِيَارُ جَوِيرٍ عَلَى لُغَةِ الْبَرْمَجَةِ «C» الْوَاسِعَةِ الْإِنْتِشَارِ حَالِيًّا، وَعَلَيْهِ فَقَدْ أَوْضَحَ مَيِّزَةَ هَذِهِ اللُّغَةِ مُصَرِّحًا بِالْقَوْلِ:

« Notre choix s'est porté sur le langage C, aujourd'hui très répandu ; il est dès lors envisageable de faire fonctionner nos programmes sur différents types de matériel sans en changer le moteur que constitue l'analyseur. »⁽²⁾

حَيْثُ يَبْقَى ثَمَّ امْتِيَازٌ لِهَذِهِ اللُّغَةِ، وَهِيَ يَسْتَقِيمُ وَصْفُ الظَّوَاهِرِ الْبِنْيَوِيَّةِ، كَمَا يُمَكِّنُنَا تَشْغِيلَ بَرَامِجِنَا بِوَسْطَةِ مُخْتَلَفِ أَنْوَاعِ الْأَجْهَازَةِ، بِدُونِ تَغْيِيرِ الْمُحَرِّكِ الَّذِي يُشَكِّلُهُ الْمُحَلِّلُ.

2- تقنية المحلل:

تَلْبُعُ صُعُوبَةُ اسْتِخْدَامِ التَّحْلِيلِ بِوَسْطَةِ الْقَوَاعِدِ الشَّكْلِيَّةِ الثَّلَاثِيَّةِ Gn3، أَوْ

(1) Voir Christian Gaubert, Op. cit., p. 287. Voir aussi: Claude Audebert et André Jaccarini, « Méthode de variations de grammaire et algorithme morphologique. Vers un extracteur de racine en arabe », in Bulletin d'études orientales, tome XLVI, pp. 77-91.

(2) Voir Christian Gaubert, Op. cit., p. 287.

تَجْدُرُ الْإِشَارَةُ إِلَى أَنَّهُ قَدْ تَزِيدُ فَائِدَةُ تَعْدِيلِ الْمُحَلِّلِ وَبَرْمَجَتِهِ بِلُغَةِ Lisp فِي الْإِحْتِبَارَاتِ، وَالتَّحْكُمِ مِمَّا يُعَزِّزُ مِنْ تَحْسِينِ الْبَرَامِجِ الْمَوْرِفُولُوجِيَّةِ، وَتَطَوُّرِهَا، وَتَعْظِيمِ أَدَاءِهَا.

مَا يُسَمَّى بِـ "النَّحْوِ الشَّكْلِيِّ الثَّلَاثِيِّ" لِعَدَمِ حَتْمِيَّتِهِ؛ حَيْثُ مِنَ الضَّرُورِيِّ أَنْ يَتِمَّ التَّعَرُّفُ عَلَى الاختياراتِ الْمُتَعَدِّدَةِ الَّتِي تَظْهَرُ خِلَالَ الْمُعَالَجَةِ، وتسجيلها بهدف عَرْضِ التَّفْسِيرَاتِ⁽¹⁾ المختلفة للكلمة وفقاً للقواعد المتبعة.

تَقُومُ مُحَاوَلَةُ جَوَابِ الْعِلْمِيَّةِ بِتَوْضِيحِ التَّفْسِيرَاتِ الْمَشْرُوعَةِ لِكُلِّ كَلِمَةٍ⁽²⁾ عَلَى شَكْلِ شَجَرَةِ بَيَانَاتٍ. وَالْمِثَالُ الْمُقْتَرَحُ فِي الشَّكْلِ الثَّلَاثِ التَّالِي يُوضِّحُ كُلَّ تَفْرِيعَاتِ الشَّجَرَةِ الَّتِي تَمَثِّلُ فِي نَمُودَجَيْنِ، يُمَكِّنُ تَفْسِيرُهُمَا بِطَرِيقَتَيْنِ، أَوْ ثَلَاثَةً. حَيْثُ يَتِمُّ تَصْنِيفُ التَّفْسِيرَاتِ وَفَقّاً لَطَرِيقَةِ الْبَحْثِ، وَبِدُونِ آيَّةٍ دَلَالَةٍ لُغَوِيَّةٍ، أَوْ تَرْتِيبِ أَلْفَبَائِيٍّ. فَكُلُّ عُنْصُرٍ أَوْ حَرْفٍ مِنَ الْكَلِمَةِ يَتِمُّ تَفْصِيلُهُ إِلَى فِتْنَةٍ بِشَكْلِ مُخْتَصَرٍ⁽³⁾.

فَرَاشَةُ

R3:هـ / R2:ش / /a:inf1 / R1:ر /:coor/ 1/

post:هـ / R3:ش / /a:inf2 / R2:ر / R1:ف / 2/

لِلتَّبَرُّكِ

R3:ك / R2:ر / R1:ب /:pref/ ت /:Jart/ 1/

post:ك / R3:ر / R2:ب /:inf1/ ت / R1:ل /:Jprep/ 2/

post:ك / R3:ر / R2:ب / R1:ت /:Jart/ 3/

الشَّكْلُ الثَّلَاثِ: مِثَالٌ لِإِجَابَةِ حَاسُوبِيَّةٍ نَاجِةٍ بِاسْتِخْدَامِ الْمُحَلِّلِ الْأَلِيِّ

(1) يَتَّضِحُ مَفْهُومُ مُصْطَلَحِ "التَّفْسِيرَاتِ" بِدِقَّةٍ خِلَالَ الْأَمَثِلَةِ التَّحْلِيلِيَّةِ التَّالِيَةِ.

(2) لَيْسَ نَمَّ رَيْبٌ أَنَّ الْمَجَالَ لَا يَسْمَحُ فِي هَذَا الْمَوْضِعِ مِنَ الْبَحْثِ بِاسْتِحْضَارِ مُنَاقَشَةٍ مُفْصَّلَةٍ لِلخَوَازِمِيَّةِ الْمُسْتَعْدَمَةِ. وَلِذَلِكَ يُعْتَبَرُ الْمُحَلِّلُ فِي هَذِهِ الدِّرَاسَةِ بِمِثَابَةِ "الصُّنْدُوقِ الْأَسْوَدِ" الَّذِي تَتَكَوَّنُ مُدْخَلَاتُهُ مِنَ النُّصُوصِ وَالْقَوَاعِدِ النَّحْوِيَّةِ الْمُسْتَعْمَلَةِ، وَتَكُونُ مُخْرَجَاتُهُ بِمِثَابَةِ تَفْصِيلَاتٍ تَحْلِيلِيَّةٍ لِكُلِّ كَلِمَةٍ يُرَادُ الْكَشْفُ عَنْهَا وَيَدْرَأُ اللَّبْسَ الَّذِي يَكْتَنِفُهَا.

(3) Voir Christian Gaubert, Op. cit., p. 288.

تفسيرات المثال الأول: (فراشة):

[أ] التفسير الأول:

ف: حرف عطف / ر: الجذر الأول / ا: (حشو) حرف مزيد 1 / ش: الجذر الثاني / هـ: الجذر الثالث.

[ب] التفسير الثاني:

ف: الجذر الأول / ر: الجذر الثاني / ا: (حشو) حرف مزيد 2 / ش: الجذر الثالث / هـ: حرف لاحق.

تفسيرات المثال الثاني: (للتبرك):

[أ] التفسير الأول:

ل: حرف جرّ / ل: التعريف⁽¹⁾ / ت: حرف سابق / ب: الجذر الأول /

(1) يَنْضَافُ مَلْحَظٌ مُهِمٌّ بِخُصُوصِ (أَلِ) التَّعْرِيفِ، حَيْثُ تُشَكِّلُ بَنِيَّتُهَا مُشْكِلةً فِي النِّظَامِ الْحَاسُوبِيِّ، وَتَحْدِيدَ كَيْفِيَّةِ التَّعَامُلِ مَعَ هَذِهِ الْأَدَاةِ؛ لِأَنَّ مِنْ أَوْضَاعِهَا الْمُلَبَّسَةِ:

(أ) أَنَّ تَأْتِي زَائِدَةٌ وَمُنْفَصِلَةٌ فِي أَوَائِلِ بَعْضِ الْكَلِمَاتِ، وَلَيْسَ لَهَا أَيُّ دَلَالَةٍ مُهِمَّةٍ، وَبِذَلِكَ يَسْهَلُ إِهْمَالُهَا فِي التَّرْتِيبِ، أَوْ فِي الْبَحْثِ عَنْ مَفَاهِيمَ مُعَيَّنَةٍ، مِثْلَ وَرُودِهَا فِي الْكَلِمَاتِ التَّالِيَةِ: (الإدارة، الاقتصاد).

(ب) أَنَّ تَأْتِي زَائِدَةٌ فِي بَعْضِ الْكَلِمَاتِ، وَلَكِنَّهَا مُتَّصِلَةٌ فِي أَوَائِلِهَا، مِثْلَ: (القانون، البحار، المحيطات)، وَبِذَلِكَ يُمَكِّنُ إِهْمَالُهَا وَوَضْعُهَا ضَمْنَ قَائِمَةِ التَّوَقُّفِ، إِلَّا أَنَّهُ يَجِبُ فِي هَذِهِ الْحَالَةِ التَّمْيِيزُ بَيْنَهَا وَبَيْنَ الْكَلِمَاتِ الَّتِي تَأْتِي بِأَلْفٍ وَلَامٍ أَصْلِيَّتَيْنِ، مِثْلَ: (الله، التزام، التباس)؛ لِأَنَّ الْحَاسِبَ الْأَلِيَّ لَا يُفَرِّقُ بَيْنَ (أَلِ) التَّعْرِيفِيَّةِ، وَالْأَلْفِ وَاللَامِ الْأَصْلِيَّتَيْنِ، كَمَا أَنَّ (أَلِ) التَّعْرِيفِيَّةَ قَدْ تَدَخَّلَ عَلَى الْمَبْدُوءِ بِأَلْفٍ وَلَامٍ أَصْلِيَّتَيْنِ فِي أَوَّلِ الْكَلِمَةِ، وَلَكِنْ نَجِدُ أَنَّ الْأَلْفَ الْأَصْلِيَّةَ قَدْ التَّصَقَّتْ بِاللَامِ فِي (أَلِ) التَّعْرِيفِ، مِثْلَ: (الإله، الالتزام)، وَبِذَلِكَ يُمَكِّنُ إِهْمَالُ (أَلِ) التَّعْرِيفِ عِنْدَ الْبَحْثِ عَنْ هَذَا النَّوعِ مِنَ الْكَلِمَاتِ، وَيُمَكِّنُ إِعْدَادَ قَائِمَةٍ بِالْكَلِمَاتِ وَاحْتِسَابِهَا آليًّا عِنْدَ الاسْتِرْجَاعِ، عَلَى أَنَّهُ يَجِبُ التَّفَرِيقُ بَيْنَ الْكَلِمَاتِ الَّتِي تَتَشَابَهُ فِي الشَّكْلِ وَتَخْتَلِفُ فِي الْمَعْنَى، مِنْ حَيْثُ اشْتِمَالُهَا عَلَى (أَلِ) التَّعْرِيفِ، أَوْ وَجُودِ الْأَلْفِ وَاللَامِ ضَمْنَ حُرُوفِهَا كَمَا فِي الْكَلِمَاتِ التَّالِيَةِ: (إلهام) مِنْ (أَلْهَمَ)،

ر: الجذر الثاني/ ك: الجذر الثالث.

[ب] التفسير الثاني:

ل: حرف جَرٍّ⁽¹⁾/ ل: الجذر الأول/ ت: (حَشُو) حرف مزيد1/ ب: الجذر الثاني/ ر: الجذر الثالث/ ك: حرف لاحق.

[ج] التفسير الثالث:

ل: حرف جَرٍّ/ ل: التعريف/ ت: الجذر الأول/ ب: الجذر الثاني/ ر: الجذر الثالث/ ك: حرف لاحق.

3- استخدام المحلل الصرفي:

إنَّ إعدادَ أشكالِ النِّحْوِ الاسميِّ والفِعْليِّ في هيئةِ برامجِ تحليل، يَصْعُبُ تَصَوُّرُهُ بدونِ اللُّجُوءِ إلى مُعَالَجَةٍ تِلْقَائِيَّةٍ نِظَامِيَّةٍ لِلنُّصُوصِ الْمُخْتَلِفَةِ، وتحقيقِ النَّتَاجِ الْمُتَحَصِّلَةِ.

لَمْ يَكْتَفِ كريستيان جوبير في مُحَاوَلَتِهِ بِتَصْمِيمِ الْمُحَلِّلِ الْبِنْيَوِيِّ اعْتِمَادًا عَلَى

==

و(الهام) مِنْ (مُهِم)، و(ألبان) جمع (البن) و(البان) نَوْعٌ مِنَ الشَّجَرِ ... إلخ. وبذلك يَجِبُ الانْتِبَاهُ عند إهمالِ (أل) التَّعْرِيفِ الرَّائِدَةِ بحيثُ لا يُؤَثَّرُ ذَلِكَ عَلَى الْكَلِمَاتِ الْمَبْدُوءَةِ بِأَلْفٍ وَلاَمٍ أَصْلِيَّتَيْنِ. رَاجِعٌ فِي ذَلِكَ: سَرِيعٌ مُحَمَّدٌ السَّرِيعُ، نِظَامُ ابْنِ النَّدِيمِ (مَعْهَدُ الْإِدَارَةِ الْعَامَّةِ)، ضِمْنَ أَعْمَالِ السِّجَلِ الْعِلْمِيِّ لِنَدْوَةِ اسْتِخْدَامِ اللُّغَةِ الْعَرَبِيَّةِ فِي تَقْنِيَةِ الْمَعْلُومَاتِ، /325-326.

(1) تَشْمَلُ الْحُرُوفُ الْمُتَّصِلَةُ حُرُوفُ الْجَرِّ مِثْلُ: (اللام، والباء، والكاف) وَقَدْ شَكَّلْتُ مِثْلَ هَذِهِ الْحُرُوفِ عَقِبَهُ أُخْرَى فِي تَحْلِيلِ الْكَلِمَاتِ، فَكَمَا نَعْرِفُ أَنَّ كَثِيرًا مِنَ الْكَلِمَاتِ تَتَّصِلُ بِحُرُوفِ الْجَرِّ سِوَا أَكَانَتِ هَذِهِ الْكَلِمَاتِ ذَاتِ دَلَالَةٍ مِثْلُ: (لِلزَّرَاعَةِ، بِالتَّرْبِيَةِ، كَالْمَعَادِنِ)، أَوْ كَانَتْ مُتَّصِلَةً بِمُفْرَدَاتٍ غَيْرِ دَالَّةٍ، مِثْلُ: (بِهِ، لِذَلِكَ، كَذَلِكَ). وَهَذِهِ الْحُرُوفُ شَائِعَةٌ الْوُرُودِ فِي كَثِيرٍ مِنَ الْكَلِمَاتِ فِي الْعَرَبِيَّةِ؛ حَيْثُ إِنَّ الْحَاسِبَ لَا يُفَرِّقُ بَيْنَ حَرْفِ الْجَرِّ الرَّائِدِ الْمَبْدُوءِ بِهِ الْكَلِمَةُ، وَالْحَرْفِ الَّذِي مِنْ أَصْلِ الْكَلِمَةِ، مِثْلَ الْحَالَاتِ الْآتِيَةِ: (لِلإِدَارَةِ، لِلزَّرَاعَةِ، لِبَاسٍ، بَالِغٍ، بِتَعْلِيمٍ، لِتَفْسِيرٍ، كَعَلَمٍ، ... إلخ). لِمَزِيدِ بَسْطِ الْقَوْلِ فِي الْمَسْأَلَةِ يُنْظَرُ: سَرِيعٌ مُحَمَّدٌ السَّرِيعُ، مَرْجِعٌ سَابِقٌ، /326-327.

عِدَّة كلماتٍ مُنْعَزِلَةٍ، بل قَامَ بإعدادِ برنامجٍ تجريبيٍّ « logiciel » أسماه "صَرْفِيَّة" « Sarfeyya، يسمحُ بتحليلِ نصٍّ كَامِلٍ يَتَكَوَّنُ مِنْ عِدَّةٍ صفحات (1).

ويعني مبدأ "تغيرات الظواهر النحوية": تَوْفُّرُ أَنْ يَقُومَ الْمُعَالِجُ بِتَكْوِينِ الْجُزْءِ الْجِذْرِيِّ (2) لِلنَّحْوِ الشَّكْلِيِّ، مع قبول أو رفض الصَّوَامِتِ، والهِمَزَاتِ، وَأَنْصَافِ الصَّوَامِتِ semi-consonnes، ثُمَّ يَتِمُّ إِظْهَارُ التَّحْلِيلِ الْفِعْلِيِّ، أو تسجيله في مَلَقَاتٍ مُحَايِدَةٍ، تحوي جميع تفسيرات كلمات النَّصِّ، والجذور، والأوزان التي تَمَّ استخراجُهَا.

وَمِنْ الْمُهِّمِّ ملاحظة أَنَّ التَّقَدَّمَ الأخير للحاسباتِ الصَّغِيرَةِ قد جَعَلَ التَّكْلِيفَةَ الزَّمَنِيَّةَ لِلتَّطْبِيقَاتِ الْخَاصَّةِ بِمِثْلِ هذه المعالجات مقبولا. والدَّلِيلُ على ذلك أَنَّ زَمَنَ تَحْلِيلِ نَصٍّ فِي مِثْلِ حَجْمِ هذه الدِّرَاسَةِ لَا يَتَعَدَّى الثَّانِيَةَ الْوَاحِدَةَ (3).

(1) Voir Christian Gaubert, Op. cit., p. 288.

(2) يقومُ الجِذْرُ في اللغةِ العربية بوضعِ البِنْيَةِ الأساسية للكلمة، ويتولَّى الوزن بناء هيكليها العام، عن طريق توزيع الحَرَكَاتِ على مختلف أصوات الكلمة، كما يقوم بتوزيع السَّوَابِقِ واللَّوَاحِقِ إلى مُكَوِّنَاتِ الجِذْرِ بهدف توليد الكلمات. فالجِذْرُ يُمَثِّلُ المدخلات في البرنامج اللُّغَوِيَّ في كِفَايَةِ المتكلم العربي، ففي مرحلة التوليد يتم اختيار الجِذْرِ، ثم تقوم الكِفَايَةُ اللُّغَوِيَّةُ في تطبيق القوانين، قوانين المطابقة بين مادة الجِذْرِ اللُّغَوِيَّ والمادة الصُّورِيَّةُ (ف، ع، ل)، ثم يَتِمُّ تفعيل الوزن عن طريق قوانين وخوارزميات السَّوَابِقِ واللَّوَاحِقِ والحَرَكَاتِ. وفي مرحلة التَّحْلِيلِ تكون العملية معكوسة؛ حيث يَتِمُّ تجريد الكلمة مِنَ السَّوَابِقِ واللَّوَاحِقِ والحَرَكَاتِ وإعادتها إلى الجِذْرِ. ولكلِّ صيغة صرفية مقابل دلاليٌّ موجود في الكِفَايَةِ اللُّغَوِيَّةِ، وهذا يعني ارتباط المستوى الصَّرْفِيِّ بالمعجم والدَّلَالَةِ. ولعلَّ أول ما يواجه المُعَالِجَةَ الآليَّةُ هو التَّدَاخُلُ بين المستويات اللُّغَوِيَّةِ، ممَّا يُوْجِبُ استعمال أنظمة متعدّدة لمعالجة تلك المستويات، وأن يرتبط كُلُّ منها بِالْآخَرِ. يُنْظَرُ في ذلك: عبد العزيز بن عبد الله المهيوبي، اللسانيات العربية والحاسوب (مُعَالِجَةُ الصَّرْفِ آليًّا)، منشورات مجلة الاتِّصَالَاتِ والعالم الرِّقْمِيِّ.

(3) اعْتَمِدَ هَذِهِ الْأَدَاءَاتُ اعْتِمَادًا كَبِيرًا عَلَى كَفَاءَةِ الْمُعْدَّاتِ الْمُسْتَحْدَمَةِ؛ وَاسْتِشْرَافِ التَّفَاصِيلِ الْخَاصَّةِ وَالدَّقِيقَةِ بِهَذِهِ الْمُعْدَّاتِ يُسَهِّمُ فِي بِنَاءِ النَّمُودَجِ الْأَمْثَلِ لِلْمُعَالِجَةِ الْآلِيَّةِ وَاسْتِقَامَةِ تَقْنِيَاتِ الْبَرْمَجَةِ.

المبحث الرابع - في تحليل النصوص

1- كَيْفِيَّةُ تَقْيِيمِ طَرِيقَةِ التَّحْلِيلِ الصَّرْفِيِّ:

1-1 تَصْنِيفُ النَّتَائِجِ:

علينا أَنْ نَذْكُرَ هُنَا الفرضيات التي تَخُصُّ الكلمة المكتوبة، التي يُطَبَّقُ عليها قَوَاعِدُ النَّحْوِ الشَّكْلِيِّ الثَّلَاثِيِّ Gn3:

1- الكلمة تَنْتَهِي لِلُغَةِ الْعَرَبِيَةِ الْفُصْحَى، وليست قديمة أو مَهْجُورَة (خارجة عن الاستخدام الحالي).

2- الكلمة غير مَشْكُورَة بِالْعَلَامَاتِ، سَوَاءً كَلِمًا أَوْ جُزْئِيًّا، وَهَذَا التَّشْكِيلُ لَا يُؤَبِّهُ لَهُ فِي مَسْأَلَةِ التَّحْلِيلِ.

3- الهمزة تُكْتَبُ بِشَكْلِ صَحِيحٍ، فَمَثَلًا: (أحد)، وليس (إحد)، و (إدارة)، وليس (أدارة).

4- الكلمة لَا تَحْوِي آيَةَ أَخْطَاءٍ إِمْلَائِيَّةٍ أَوْ مَطْبَعِيَّةٍ، عَلَى الرَّغْمِ مِنْ أَنَّ هَدَفَ هَذِهِ الْمَحَاوَلَةِ هُنَا لَيْسَ التَّصْحِيحُ الْإِمْلَائِيُّ.

تُعَدُّ هَذِهِ الشُّرُوطُ مُقَيَّدَةً - إِلَى حَدٍّ كَبِيرٍ - لِإِجْرَاءَاتِ التَّحْلِيلِ الْآلِيِّ؛ حَيْثُ يَعْمَلُ الْمُعَالِجُ عَلَى تَمْيِيزِ مِثْلِ هَذِهِ الْكَلِمَاتِ، وَيُسَاعِدُ فِي تَبْسِيطِهَا وَالتَّعَرُّفِ عَلَى فَصَائِلِهَا لِاسِيَّمَا تِلْكَ الَّتِي تَحْوِي هَذِهِ التَّأْلِيفَ؛ فَتَمَّ عَدَدٌ لَا بَأْسَ بِهِ مِنَ النُّصُوصِ الْمُعَاَصِرَةِ لَا تَتَسَاوَقُ مَعَ أُخْرَى فِي سِيَاقَاتِهَا الْبِنْيَوِيَّةِ التَّرْكِيْبِيَّةِ، وَمِنْهَا مَا يَتَعَامَلُ بِصُورَةٍ سَلْبِيَّةٍ مَعَ الصُّعُوبَاتِ الْإِمْلَائِيَّةِ، وَخَاصَّةً مَا نَجِدُهُ فِي لُغَةِ الصَّحَافَةِ⁽¹⁾.

أَمَّا النَّحْوُ الشَّكْلِيُّ الثَّلَاثِيُّ Gn3 فَهُوَ قَائِمٌ عَلَى مَلَحَظِ أَنَّهُ يَعْمَلُ كَجِهَازٍ تَرْشِيحٍ (فلتر) فِي مُوَاجَهَةِ آيَةِ مَجْمُوعَةٍ مُتَتَالِيَةٍ مِنَ الْحُرُوفِ، وَعِنْدئذٍ يُمَكِّنُ رَفْضَ كَلِمَةٍ أَوْ

(1) Voir Christian Gaubert, Op. cit., p. 289.

قبولها. ويعني الرّفْض أنّ الكلمة لا ينطبق عليها القواعد الخاصّة بالنّحو الشّكليّ المُبرّج. وكذلك يَتَكَرَّرُ في بعضِ الحالاتِ قبول الكلمات، لكن ذلك لا يعني تحليلًا صَحِيحًا للكلمة. فالحَقِيقَةُ إِذْنِ مَا أُقِرَّ في الاستعمال على أنّ هذا النّحو لا يزال حتى يومنا هذا يعملُ بدون أيِّ مُعْجَمٍ أو مَسْرَدٍ للمُفْرَدَاتِ؛ وَمِنْ ثَمَّ لا يُوجَدُ أيُّ تأكيدٍ على وجودٍ واضحٍ للجذر المُستخرج⁽¹⁾. وليس ثَمَّةَ تَحَكُّمٍ آلي مُباشِرٍ وناجعٍ في استِخراج الصّيغة الاشتقاقِيَّة للكَلِمَةِ، إلّا بالحُصُولِ عليها وبالتّالي طُرُق تمثيلها مِنْ خِلال اللُّجُوءِ إلى إعادة تشكيلها بواسطة الفِئَات التّالية: سوابق، دَوَاحِل/ حَشُو 1، دَوَاحِل/ حَشُو 2، لَوَاحِق. ودَلِيلُنَا على ذلك أنّ السّابِقَةَ التي يَتِمُّ استِخراجُها، يُمكنُ أَلّا تَنَسِجُ أو تَتَفَقَّ مع الحَرف المَزِيد الدّاخل الثّاني، ممّا يَتَرَتَّبُ عليه خَطَأُ التّفْسير المُقَرَّح. والمُعَالِج هو الحَكْمُ الوَحِيدُ على دِقَّة الحُلُول. وفي هذا الصّدَد يَتَضَيّحُ لَنَا ضرورة مُراعاةِ الاعْتِمَادِ المُصْطَلَحِيّ، أو الالتزام بمجموعةِ مُصْطَلَحَاتٍ تَسْمَحُ بتصنيفِ النّتائِجِ وفقًا لِصِحَّتِها.

2- مفاهيم الصّمت/السّكون، الضّجيج/التشوّش في تطبيقات النّظم الآليّة للصّرف العربيّ:

1-2 المصطلّحات:

اعْتَمَدَ كريستيان جوبير⁽²⁾ في مُحَاوَلَتِهِ على تَحْدِيدِ بعضِ التّعريفات لِدلالة الرُّمُوزِ المُسْتَخْدَمَةِ، نُجْمِلُهَا في النِّقَاطِ الآتية:

(1) Ibid., p. 289.

وانظر أيضًا: حازم يوسُف عبد العظيم، وأحمد عبد المجيد محمد، القراءة الآليّة للنّصّ العربيّ بِمَسَاعَدَةِ المُصَحِّحِ الهِجَائِيّ، 196-185/ محمد غزالي خيَّاط، تَمَثِيلُ الدلالة الصّرفيّة في النّظْمِ الآليّة لِقَهْمِ اللغة العربيّة، 312-299/.

(2) Voir Christian Gaubert, Op. cit., p. 289.

♦ Γ : عبارة عن مجموعة من الكتابات أو الكلمات المطبوعة للغة العربية، التي تُجيب عن عدد من السمات اللسانية الآتية: مجموعة الأسماء الثلاثية الصحيحة. مجموعة الأفعال المهُموزة أو التي تحوي الهمزة كحرف أصيل في الجذر. ويُطلق على هذه المجموعة "حقل التطبيق" في مقابل المجال الخاص بالكلمات المكتوبة.

♦ G : هو النحو الشكلي/الصرفي، الذي تمّ إعداده بهدف قبول العناصر التي تنتمي للمجموعة Γ ، مع توصيف أشكال اللبس، ورفض العناصر التي لا تنتمي للمجموعة.

يسمخ الجدول الآتي في الشكل الرابع بتعيين التعريفات مقرونة بأمثلة حقيقية، حيث الحقل Γ هو مجموعة الأسماء المكوّنة من جذر ثلاثي صحيح. و G : هو النحو الشكلي الثلاثي $Gn3(c, c, c)$:

مِثَال	الكَلِمَة	النَّوع	عدد العناصر أو العمليات في أشكال التحليل	ناتج التحليل	الجذر	الوزن	الطبقة	أسباب الضجيج	المُحصلة والاستنتاجات	ملاحظات
1	اصطدام	اسم صحيح	0	مُفَرَّغٌ بِلا نتيجة			سُكُون - هُدوء		(ط) مُتَعَدِّرٌ أَنْ تكون (دَاخِلٌ أَوَّل)	
3	رِجَال	اسم صحيح	0	1/ر:R1ج:R2/؛R3:inf2	رجل	فعال	حُلُول			
4	أَشْغَال	اسم صحيح	2	1/أ:prefش:R1غ:R2:inf3:JL	شغل	أفعال	حُلُول			
				2/أ:intش:R1غ:R2:inf3:JL	شغل	فعال	تَشْوُش - ضَجِيج	لَبَسٌ دَاخِلِيٌّ	سَبَبُ اللَّبَسِ يَكْمُنُ فِي أَنَّ الجذر واحدٌ	
				1/ك:prepن:R1ف:R2ه:R3:aduma	نفه	فعل	تَشْوُش - ضَجِيج	Post-concat. ها:	سَبَبُ اللَّبَسِ يَكْمُنُ فِي أَنَّ الصَّيْغَة (نفه) ليست موجودة	(ك) فِي الجذر الأول يُمكن أَنْ تُعَامَلَ كحرف جَرِّ
5	كنفها	اسم صحيح	3	2/ك:prepن:R1ف:R2ه:R3:adobnd	نفه	فعل	تَشْوُش - ضَجِيج	Post-concat.	سَبَبُ اللَّبَسِ يَكْمُنُ فِي اللاحقة	
				3/ك:R1ن:R2ف:R3ه:post/post:aduma	كنف	فعل	حَلَّ مَعَ بَاعِث ضَجِيج			
				1/ل:prepن:art:prefب:R1/ر:R2:ك:R3	برك	تفعل	حَلَّ مَعَ بَاعِث ضَجِيج			

مِثَال	الكَلِمَة	النُّوع	عدد العناصر أو العمليات في أشكال التحليل	ناتج التحليل	الجذر	الوزن	الطَّبَقَة	أَسْبَاب الضَّجِيج	المُحَصِّلَة والاستِثْناءات	مُلاحظات
6	لِلتَّابِرِك	اسْمٌ صَحِيحٌ	3	R2/:ل2/ prep/:ل1/ت:R1/inf1/ب:R2/ post:ر:R3/ك	لبر	فتعل	تَشَوُّش - ضَجِيج	Pré-concat. لل:	سَبَبُ اللَّبْسِ يَكْمُنُ فِي أَنَّ الصِّيغَةَ (لبر) ليست موجودة	ضَجِيج بسبب الصِّيغَة الخَاطئة أو المغلوطة
				R2/:ل3/ prep/:ل1/ت:R1/art:ب:R2/ post:ر:R3/ك	تبر	فعل	تَشَوُّش - ضَجِيج	لَبَسٌ دَاخِلِيٌّ	سَبَبُ اللَّبْسِ يَكْمُنُ فِي حُضُور الجذر	اللُّغْن أو الخطأ ناجم عن: (ل) والضَّمير الألق الرائد
2	جَدِّي	اسْمٌ مُضَعَّف	0	مُفَرَّغٌ بِلاَ نَتِيجَة			رَفُض/طَرَح			(ي) مُتَعَدِّرٌ أَنْ تكون الجذر التَّالِث
7	البَيْت	اسْمٌ مُعْتَلٌّ	1	inf2/:ل1/ prep/:ل1/ت:R1/ب:R2/ي:inf2/ R3:ت	لبت	فعل	ضَجِيج بدون حَلْ	Pré-concat. ال:	سَبَبُ اللَّبْسِ يَكْمُنُ فِي أَنَّ الصِّيغَةَ (لبت) ليست موجودة	
8	يَخْرُج	فِعْلٌ	0	مُفَرَّغٌ بِلاَ نَتِيجَة			رَفُض/طَرَح			(ي) مُتَعَدِّرٌ أَنْ تكون سَابِقَة في الصِّيغَة
9	اسْتَقَرَّ	فِعْلٌ	1	R2/:ل1/ pref/:س:R1/ت:R1/inf1/ق:R2/ R3:ر	سقر	افتعل	ضَجِيج بدون حَلْ	لَبَسٌ دَاخِلِيٌّ	سَبَبُ اللَّبْسِ يَكْمُنُ فِي حُضُور الجذر	ضَجِيج بسبب الصِّيغَة الخَاطئة أو المغلوطة

مِثَال	الكَلِمَة	النُّوع	عدد العناصر أو العمليات في أشكال التحليل	ناتج التحليل	الجذر	الوزن	الطبقة	أسباب الضجيج	المُحَصِّلَة والاسْتِنَاجَات	ملاحظات
10	تَفْطَنَ	فِعْلٌ	1	/ت1:pref/ف:R1ط:/:R2ن:R3	فطن	تفعل	ضَجِيج بدون حَلْ	لَبَسَ دَاخِلِيٌّ	سَبَبُ اللَّبَسِ يَكْمُنُ فِي أَنَّ الجذرَ وَاحِدٌ	
11	فَعِنْدَيْهِ	أَدَاةٌ	0	مُقَرَّغٌ بِلاَ نتيجة			رَفُض/طَرَح			(ند) مُتَعَدِّرٌ أَنْ تكونَ لَاحِقَةً في الصِّيغَةِ
12	قبل	أَدَاةٌ	1	/ق:/:R1ب:/:JR2:R3	قبل	فعل	تَسْوُس - ضَجِيج	لَبَسَ دَاخِلِيٌّ	سَبَبُ اللَّبَسِ يَكْمُنُ فِي أَنَّ الجذرَ وَاحِدٌ	

الشَّكْلُ الرَّابِعُ: أَمثلةٌ لتصنيفِ النتائجِ الخاصَّةِ بالنَّحْوِ الشَّكْلِيِّ الثُّلَاثِيِّ (c, c, c) Gn3 عندَ تَطْبِيقِهِ على الأَسْمَاءِ الثُّلَاثِيَّةِ الصَّحِيحَةِ والكَلِمَاتِ الأُخْرَى

إِذَا مَا جَالَ الطَّرْفُ مَجَالَهُ فِي هَذَا الْجَدُولِ فَإِنَّهُ سَمَّيْتَنِي إِلَى ضُرُوبٍ مِنَ
الْوَشَائِحِ فِي تَحْلِيلِ الْكَلِمَاتِ وَعِلَاقَةِ الْجُزْءِ بِالْكُلِّ، وَإِمْكَانِيَّةِ اسْتِنْبَاطِ مَجْمُوعَةٍ
قَوَاعِدٍ وَفَقِ قَوَانِينِ ثَلَاثَةٍ مُعَالَجَاتِ النَّحْوِ الشَّكْلِيِّ الْإِلَى.

فَإِذَا كَانَتْ الْإِجَابَةُ الْخَاصَّةُ بِالنَّحْوِ الشَّكْلِيِّ لِكَلِمَةٍ مَا فِي اللُّغَةِ الْعَرَبِيَّةِ، مُكَوَّنَةً
مِنْ عَدَدٍ مُعَيَّنٍ مِنَ الْعَنَاصِرِ، فَإِنَّهُ يَتَوَلَّدُ لَدَيْنَا مَجْمُوعَةٌ مِنَ الْإِحْتِمَالَاتِ الْآتِيَةِ⁽¹⁾:

- إِذَا كَانَ عَدَدُ الْعَنَاصِرِ = 0، فَإِنَّ نَاتِجَ التَّحْلِيلِ يُمَكِّنُ أَنْ يُمَثَّلَ بِ:
(أ) سُكُونٍ/هُدُوءٍ، فِي حَالَةِ الْكَلِمَةِ الَّتِي تَنْتَهِي إِلَى Γ (مِثَال 1 فِي جَدُولِ الشَّكْلِ
الرَّابِع).

(ب) رَفُضٍ/طَرَحٍ، فِي حَالَةِ الْكَلِمَةِ الَّتِي لَا تَنْتَهِي إِلَى Γ (مِثَال 2).
- إِذَا كَانَ عَدَدُ الْعَنَاصِرِ < 0 ، فَإِنَّ نَاتِجَ التَّحْلِيلِ يُمَكِّنُ أَنْ يَسْتَقْبَلَ حَالَةَ لَبْسٍ
اسْمِيٍّ: وَحَاصِلُ الْإِجَابَةِ إِذَنْ يَكُونُ مُؤَلَّفًا مِنْ عَدَدِ الْعَنَاصِرِ الْمُفْسَّرَةِ أَوْ عَدَدِ
الْإِحْتِمَالَاتِ التَّأْوِيلِيَّةِ الرَّمْزِيَّةِ لِلنُّصُوصِ، الَّتِي تُوصَفُ بِ:
(أ) الْحَلِّ، وَذَلِكَ إِذَا كَانَتْ الْكَلِمَةُ تَنْتَهِي إِلَى Γ ، وَمِنْ ثَمَّ يُصْبِحُ التَّفْسِيرُ أَوْ التَّأْوِيلُ
نَاجِحًا، كَمَا فِي حَالَةِ (التَّأْوِيلِ الْأَوَّلِ لِلْمِثَالِ 3، وَ3.1).

(ب) ضَحِيحٍ/تَشَوُّشٍ، وَذَلِكَ إِذَا كَانَتْ الْكَلِمَةُ تَنْتَهِي إِلَى G ، وَمِنْ ثَمَّ يَصِيرُ نَاتِجُ
التَّحْلِيلِ خَاطِئًا، كَمَا فِي حَالَةِ (الْمِثَالِ رَقْم 4.2)، أَوْ إِذَا كَانَتْ الْكَلِمَةُ لَا تَنْتَهِي
إِلَى Γ ، كَمَا فِي حَالَةِ (الْمِثَالِ رَقْم 7.1).

وَالْحَلُّ، إِذَا كَانَ مُوجُودًا، فَهُوَ بِطَبِيعَةِ الْحَالِ حَلٌّ وَاحِدٌ: حَيْثُ لَا يَأْخُذُ فِي الْإِعْتِبَارِ
الْحَالَاتِ النَّادِرَةِ جَدًّا، الَّتِي يَقُومُ فِيهَا الْمُعَالِجُ الْإِلَى بِالتَّعَامُلِ مَعَ الْكِتَابَةِ الْعَرَبِيَّةِ فِي
الْكَلِمَاتِ الَّتِي لَا تَسْتَعْمِلُ التَّشْكِيلَ عُمُومًا، وَمِنْ ثَمَّ يُمَكِّنُ أَنْ نَتَحَصَّلَ عَلَى عِدَّةٍ مَعَانٍ

(1) Ibid., p. 291.

لنمادج هذه الكلمات. ولذلك فمن الممكن الحديث عن حلٍ صَاحِب، كما في حالة (المثال رقم 5.3)، إذا كَانَ نَاتِجُ التَّحْلِيلِ يحتوي على حَلٍّ بين الضَّجِيجِ، وبالمثل إذا كانت الإجابة حَالَتُهَا فقط ضَجِيجًا، فَإِنَّهَا تُوسَمُ حينئذٍ بِضَجِيجٍ دون حَلٍّ، كما في حالة (المثال رقم 7.1).

2-2 الأسباب الطَّبِيعِيَّةُ للضَّجِيجِ (النتائج المُشَوَّشَة):

إنَّ دِرَاسَةَ أَسْبَابِ الضَّجِيجِ لَهَا أَهَمِّيَّةٌ مِحوَرِيَّةٌ لِفَهْمِ ظَوَاهِرِ اللَّبْسِ التي يُسَبِّبُهَا نِظَامُ الكِتَابَةِ العَرَبِيَّةِ. ويعني الضَّجِيجُ: نتائج التَّحْلِيلِ الخَاطِئُ للكلمة، الذي يَرْجِعُ إلى أَسْبَابٍ مُخْتَلِفَةٍ. ولتَحْدِيدِ هذه الأسبابِ نَتَّبِعُ الأسلوبَ التَّالِيَّ⁽¹⁾:

■ إذا أَدَّى حَذْفُ جَمِيعِ الحُرُوفِ السَّابِقَةِ أو اللَّاحِقَةِ لِلجِذْرِ المُرادِ تَحْلِيلُهُ إلى الحُصُولِ على ذَاتِ النَتَائِجِ الخَاطِئَةِ، التي ظَهَرَتْ عند وُجُودِ هذه الحُرُوفِ، يَكُونُ الضَّجِيجُ نَاتِجًا عَنِ اللَّبْسِ الدَّاخِلِيِّ كما في الأمثلة (4.2، 6.3، 9.1).

■ إذا اسْتَمَرَّ الضَّجِيجُ، عِنْدَمَا تُنْزَعُ الكَلِمَةُ لِتَحْلِيلِ عَنَاصِرِ مَا بَعْدَ السَّلْسَلَةِ، فَإِنَّا نَتَحَدَّثُ عَنِ الضَّجِيجِ قَبْلَ السَّلْسَلَةِ. وَهَذَا مَا يُعْرَفُ بِظَوَاهِرِ التَّحَوُّلِ الجِذْرِيِّ، وَبِتَغْيِيرٍ آخَرَ: إذا حَدَثَ نَفْسُ الخَطَأِ فِي تَحْلِيلِ الكَلِمَاتِ عِنْدَ حَذْفِ جَمِيعِ الحُرُوفِ اللَّاحِقَةِ لِلجِذْرِ، فَقَطْ يُصْبِحُ الضَّجِيجُ نَاتِجًا عَنِ إِضَافَةِ السَّابِقَةِ إِلَى الجِذْرِ. وهذه الظَّاهِرَةُ تَدُلُّ على تَأَخُّرِ الجِذْرِ، أي أَنَّهُ قَدْ تَمَّ اعتَبَارُ الحَرْفِ الذي يَسْبِقُ الجِذْرَ الأوَّلَ على أَنَّهُ الحَرْفُ الأوَّلُ مِنَ الحُرُوفِ الأَصْلِيَّةِ الثَّلَاثَةِ لِلجِذْرِ. وكذلك تَمَّ تَفْسِيرُ الحَرْفِ الثَّالِثِ لِلجِذْرِ على أَنَّهُ حَرْفٌ مَزِيدٌ؛ لَأَنَّهُ يَنْتَبِئُ إلى مَجْمُوعَةِ الاحْتِمَالَاتِ الخَطِئَةِ f3 كما في حَالَةِ الأمثلة (6.2، 7.1).

■ وَعَلَى العَكْسِ، فَإِنَّ الضَّجِيجَ النَّاتِجَ عَنِ "مَا بَعْدَ السَّلْسَلَةِ" يُشِيرُ إلى التَّفْسِيرَاتِ

(1) Id., ibid., p. 291.

الخاطئة، التي تظل باقية رغم حذف عناصر "ما قبل السلسلة"⁽¹⁾؛ حيث تنعكس الأدوار، فينتهي الجذر الأول إلى المجموعة الخطرة $\sum f1$ كما في الأمثلة (5.1، 5.2).

■ يُمكن أن نضمن حالات أخرى مثل إدراج حالي أداة التعريف "أل"، وحرف الجر "ل" السابق لأداة التعريف⁽²⁾، التي يُمكن أن تنسب إلى فصيلة الضجيج الناتج عن وجود واحد من الأسباب الموضحة في الحالات السابقة، وما يُمكن أن تُنجزه العربية لتمييز التاليف التي تقبلها في تحليل أبنية كلماتها وأوزانها، والتعرف على بعض الروائد الطارئة، وتحليل أنواع الكلمات الاستثنائية.

3-2 الضجيج الخاص بالنموذج:

إن حقيقة عدم خضوع الصيغة الاشتقاقية لاختبار حضورها في اللغة، يسمح أحياناً بظهور تفسيرات خاطئة لها، لكن مع ذلك تكون مقبولة في جانب النحو الشكلي. وهذه الظاهرة ليست دائماً مصدراً مباشراً للضجيج، إنما ترجع بالأحرى إلى عيب في بناء النحو الشكلي *grammaire formelle*، الذي يُمكن معالجته بإدراج قائمة من الأوزان الصرفية. وهذا الضجيج يُسمى ضجيج الوزن الخاطئ *bruit de schème erroné* كما في الأمثلة (6.1، 6.2).

ونمة عيب آخر في هذا النموذج، هو السماح بوجود حرف التعريف (أل)، والضمير المتصل في نفس الوقت، وهذه حالة حاسوبية مستحيلة؛ لأنها تؤدي إلى تعريف أو تحديد مزدوج للاسم. وهذا النوع من الضجيج شأنه شأن باقي الأنواع التي تنشأ عن عيوب ذات طبيعة لغوية، سيتم حسابه ضمن ضجيج النتائج غير الصحيحة صرفياً كما في المثال (6.3). ولا يُمكن القضاء على هذا الضجيج إلا

(1) راجع على سبيل المثال: تحليل كلمة «كنفها» من جدول الشكل الرابع.

(2) انظر: هامش رقم «55»، و«56» من موضوعات المبحث الثالث من هذه الدراسة.

بتعديل قواعد إعادة كتابة النحو الشكلي أو توسيعها⁽¹⁾.

4-2 نتائج الضجيج:

إذا كان هدف التحليل هو استخراج الجذر، فيجب تصنيف الضجيج وفقًا لأهمية حدوثه في هذه العملية. وقد يحدث فعلًا أن يكون الجذر هو نفس جذر الحل، سواء ظهر الحل في مكان آخر أم لم يظهر. ونطلق على هذا الضجيج اسم "الضجيج ذو الجذر الصحيح" bruit radicalement exact كما في المثال (4.2).

وهناك الضجيج غير صحيح الجذر، ويظهر في التفسيرات التي تكون جذورها ثابتة وحاضرة، وهذه الحالات محيرة بصفة خاصة. فهذه النتائج المشوشة ناتجة عن حالات اللبس اللغوي الخاصة بأسلوب كتابة العربية بدون علامات التشكيل، وذلك قد يتسبب أحيانًا في تردد القارئ كما في حالة الأمثلة رقم (6.3، 9.1).

يشمل اللبس الخاص بالجذر عدد الجذور الحاضرة المختلفة في ناتج التحليل.

إن "الضجيج المتبقي/الكامن" bruit résiduel يتكون من التفسيرات التي لا تكون جذورها مثبتة أو حاضرة، ولحد من هذه النتائج يكفي اللجوء إلى قائمة الجذور المثبتة لحذف الأخطاء، كما في الأمثلة (أرقام 5.1، 5.2، 6.2، 7.1)⁽²⁾.

ويتجلى معنى المصطلحات المتبعة عند دراسة النتائج الخاصة بالأفعال وبالأدوات والحروف التي لا تنتمي إلى المجموعة Γ، حيث يتم استبعاد بعض هذه النتائج بيسر وسهولة، لكن بعضها الآخر يتم قبوله، وتعتبر النتائج إذن ضجيجًا واضحًا. وهذه النتائج التي تعد ضجيجًا، يكون جذورها صحيحًا إذا اختلط وزن

(1) يُسمى النحو "الموسع" « augmentée » ليعني الخروج من إطار اللغات التقليدية لإدراج بضع قوائم مفردات، أو أقواس لا يُسمح باستيعابها إلا تحت شروط معينة، حتى يمكن "نمذجة" الحالات الشاذة عن القواعد.

(2) Voir Christian Gaubert, Op. cit., p. 292.

الكلمة بدون تشكيل بوزن الجذر الثلاثي الاسمي بلا تشكيل، الذي يُحدِّدُه النَّحْوُ الشَّكْلِيّ الذي يتمُّ تطبيقُه أثناء المُعالِجَة الالِيَّة. وَيَتِمُّ إدراج هذا النوع من الضَّجِيجِ ضَمْنَ حَالَاتِ اللَّبْسِ اللُّغَوِيِّ الدَّاخِلِيِّ *ambiguïté interne* كما في المثال رقم (10.1)⁽¹⁾. وَبِنَاءً عَلَى تَقَدُّمٍ، فَإِنَّ عَمَلِيَّةَ اسْتِخْرَاجِ الْبَيِّنَاتِ مِنَ النُّصُوصِ بِوَاسِطَةِ النَّحْوِ الْمُبْنِيِّ عَلَى تَحْلِيلِ الْكَلِمَاتِ، وَاسْتِخْرَاجِ جُذُورِهَا، يَجِبُ أَنْ تَقُومَ عَلَى حَصْرِ دَقِيقٍ لِحَالَاتِ الضَّجِيجِ وَالصَّمْتِ. وَمِنْ جَانِبٍ آخَرٍ يَتَّضِحُ أَنَّ مِثْلَ هَذِهِ الدِّرَاسَةِ يَجِبُ أَنْ تُطَبَّقَ عَلَى جَمِيعِ كَلِمَاتِ النَّصِّ مَا عَدَا الْأَسْمَاءَ الْأَعْلَامَ، وَالْكَلِمَاتِ الْأَجْنِبِيَّةَ فِي اللُّغَةِ؛ وَذَلِكَ حَتَّى يَتَسَنَّى حِسَابُ قُدْرَةِ الْمُحَلِّلِ الصَّرْفِيِّ عَلَى الْاسْتِبْعَادِ أَوْ الْفَصْلِ، فِيمَا يَخُصُّ الْفِئَاتِ الْآخَرَى، الَّتِي تَخْتَلِفُ عَنْ فِئَةِ الْكَلِمَاتِ الْمُفْتَرَضِ اسْتِخْرَاجُهَا (الفئة المُستَهْدَفَة لِلتَّحْلِيلِ)⁽²⁾.

. سِمَة مُهِمَّةٌ لِلنَّحْوِ الشَّكْلِيِّ:

هُنَاكَ نَوْعٌ مُحَدَّدٌ مِنْ تَغْيِيرَاتِ الظَّوَاهِرِ النَّحْوِيَّةِ يَرْتَبِطُ بِحُدُوثِ تَغْيِيرَاتٍ فِي تَرْتِيبِ الْكَلِمَاتِ وَتَسْلُسُلِهَا فِي النَّتَائِجِ (إِجَابَةِ الْمُحَلِّلِ). وَبِالْفِعْلِ فَإِذَا افْتَرَضْنَا أَنَّ النَّحْوَ الشَّكْلِيَّ G يَقْبَلُ عِدَّةً مِنَ الْإِنْتِقَالَاتِ T بَيْنَ حَالَتَيْنِ E و F فَإِنَّ الْإِجَابَةَ R لِتَحْلِيلِ كَلِمَةٍ مَا تَنْقَسِمُ إِلَى نَوْعَيْنِ:

- «R1» الْإِجَابَةُ الْأُولَى: تَحْوِي التَّفْسِيرَاتِ الَّتِي تَسْتَعِيرُ T بَيْنَ الْأَقْوَاسِ الْمَسْمُوحِ بِهَا.

- «R2» الْإِجَابَةُ الثَّانِيَّةُ: تَحْوِي التَّفْسِيرَاتِ الَّتِي تَسْلُكُ طَرِيقًا آخَرَ (- un arc epsilon -

transition) بَيْنَ حَالَتَيْنِ تَوْجَدَانِ مِثْلًا عَلَى يَمِينِ وَيسارِ E و F.

وَمِنْ مِثْلِ مَا تَقَدَّمَ، إِذَا تَمَّتْ إِضَافَةُ إِنْتِقَالٍ تَكْمِيلِيٍّ مُحْتَمَلٍ بَيْنَ E و F، فَسَتَكُونُ الْإِجَابَةُ الْإِلِيَّةُ الْمُتَحَصِّلَةُ G' لِنَفْسِ الْكَلِمَةِ الْمُرَادِ تَحْلِيلُهَا هِيَ ذَاتُهَا الْمُسْتَخْرَجَةُ

(1) Ibid., p. 292.

(2) Id., ibid., p. 293.

بِنَفْسِ النَّمَطِ بَيْنَ $R'1$ ، و $R'2$. وَمِنْ الْوَاضِحِ أَنَّ $R'1$ ، هُوَ $R1$ ، وَأَنَّ $R'2$ ، هُوَ $R2$ ، وَأَنَّهَا وَحَدَاتٌ تَأْتِي لِاسْتِنْبَاطِ تَفْسِيرَاتٍ لِإِمْكَانَاتٍ جَدِيدَةٍ مِنَ التَّحَوُّلِ وَالانتِقَالِ بَيْنَ الْحَالَاتِ E و $F^{(1)}$.

إِنَّ نِطَاقَ هَذِهِ الْخَصَائِصِ مُهِمٌّ فِي حَالَتِنَا: فَهُوَ يَعْنِي أَنَّهُ قَدْ يَكْفِي بِالنِّسْبَةِ لِفَهْمِ تَغْيِيرِ بَنِيَوِيٍّ وَاحِدٍ أَنْ تَدْرُسَ اسْتِجَابَةَ الْحَاسُوبِ مِنْ خِلَالِ الْإِنْتِقَالِ بَيْنَ هَذِهِ الْحَالَاتِ عَنْ طَرِيقِ إِضَافَةِ قَوْسٍ بَيْنَ حَالَتَيْنِ مِنْ حَالَاتِ التَّحْلِيلِ. لِذَا فَإِنَّ الْاسْتِنْتِاجَ الْكَامِلَ وَالْمُحَصَّلَةَ الْنَهَائِيَّةَ هُمَا مَجْمُوعُ الْأَجْوِبَةِ الْأَصْلِيَّةِ وَالْمُسَاعِدَةِ. وَعَلَى ذَلِكَ يُمَكِّنُ تَعْمِيمُ هَذَا الْإِجْرَاءِ عَلَى عَدَدٍ مِنَ الْأَقْوَاسِ شَرِيطَةً أَنْ تَتَّحِدَ أَوْ تَرْتَبِطَ دَائِمًا نَفْسُ الْحَالَتَيْنِ. وَيُمْكِنُ أَيْضًا اللُّجُوءُ إِلَى تَطْوِيعِ هَذِهِ الظَّاهِرَةِ وَبَسْطِهَا وَفَقِ التَّطْبِيقَاتِ اللَّسَانِيَّةِ الْحَاسُوبِيَّةِ الْمُخْتَلِفَةِ، مِنْ مُنْطَلَقِ الْوَعْيِ الدَّقِيقِ بِخَصَائِصِ اللُّغَةِ الْعَرَبِيَّةِ⁽²⁾.

3- عملية التتقيب (استخراج المعلومات) في النص:

إِنَّ مَنْ يَرْعَبُ فِي تَقْيِيمِ فَائِدَةِ هَذَا النَّحْوِ، فَعَلَيْهِ أَنْ يَتَّبِعَ الْخُطُواتِ التَّالِيَةَ:

(أ) الْخُطْوَةُ الْأُولَى: وَهِيَ تَحْلِيلُ مُسَبِّقٍ، يَقُومُ عَلَى اسْتِخْرَاجِ يَدَوِيٍّ لِلْجُذُورِ وَإِدْرَاجِهَا فِي قَاعِدَةِ بَيَانَاتٍ، تَشْمَلُ عَلَى الْأَقْلَى الْكَلِمَةَ، وَتَصْنِيفُهَا إِلَى اسْمٍ، أَوْ فِعْلٍ، أَوْ أَدَاةٍ، ... وَتَصْنِيفَاتٍ أُخْرَى مُلْحَقَةً بِالْقِسْمَةِ الثَّلَاثِيَّةِ، وَجُذُورِهَا وَكَذَلِكَ نَوْعِ الْجَذْرِ صَحِيحًا كَانَ أَمْ أَجُوفًا أَمْ غَيْرِهِ.

(1) Idem.

(2) Idem.

تَتَجَاوَزُ الدِّرَاسَةُ الْكَامِلَةُ لِهَذِهِ الْخَصَائِصِ نِطَاقَ هَذَا الْبَحْثِ. وَمَعَ ذَلِكَ، يُلَاحَظُ أَنَّ اسْتِخْدَامَ الصِّيغَةِ الْجَبْرِيَّةِ *formulation algébrique* يَجْعَلُ مِنَ الْمُمْكِنِ التَّذْلِيلَ بِسُرْعَةٍ عَلَى هَذِهِ الْحَالَةِ مِنَ الْخَطِيئَةِ (عِلَاقَةٍ بَيْنَ كِمِّيَّتَيْنِ حِينَ يُؤَدِّي تَغْيِيرُ أَحَدَهُمَا إِلَى تَغْيِيرٍ فِي الْأُخْرَى يَكُونُ مُتَنَاسِبًا مَعَ تَغْيِيرِ الْكِمِّيَّةِ الْأُولَى). انْظُرْ فِي ذَلِكَ:

- BEO XLVI, annexe I.

- عَبْدُ الرَّحْمَنِ الْحَاجِ صَالِحٍ، مُنْطَلَقُ النَّحْوِ الْعَرَبِيِّ وَالْعِلَاجُ الْحَاسُوبِيُّ لِللُّغَاتِ، / 27-41.

- (ب) الخطوة الثانية: وهي استخدام برنامج يقوم بالتحليل الآلي للنص.
- (ج) الخطوة الثالثة: وهي التأكد من صحة نتائج البرنامج، بمطابقة كل التفسيرات، وإثراء قاعدة الكلمات وفقاً للمصطلحات التي سبق توضيحها أعلاه.
- (د) الخطوة الرابعة: وهي دراسة الإحصاءات الناتجة عن الخطوة السابقة، ثم تحليلها، وذلك عن طريق النسب، حتى يتم قياس نجاح نتائج عمليات استخراج المعلومات ونسب التردد وكذلك أسباب الضجيج ... إلخ⁽¹⁾.

إنّ ظواهر الاختلافات النحوية تهدف إلى الحدّ من مصادِر الضجيج، أو منع الصّمت الذي تمّ رصده في نهاية الخطوة الرابعة. ومحاولات تعديل مجال دراسة القواعد النحوية تتم أيضاً بواسطة ظواهر الاختلافات. وهذه الاختلافات تؤدي إلى إعادة الخطوات: الثانية، والثالثة، والرابعة، ومقارنة النتائج الحاصلة عن الخطوة الرابعة. وفي حالة دخول إحدى هذه الظواهر في إطار التسلسل الكتابي، فإنّه يكفي القيام بدراسة تحليل النص الناجم فقط عن وجود القوس أو الأقواس الجديدة بين الحالات المطلوبة. إنّ دراسة مصادر الضجيج يمكن أن تؤدي إلى تحديد متغيرات في عملية التنقيب في النص، بحيث يمكن اكتشاف الأسباب الرئيسة، واحتمالية التداخل فيما بينها. ويجري حوار بين الخطوتين: الثالثة والرابعة؛ حتى نصل إلى توصيف مرضٍ للظواهر المرصودة أثناء عملية التحليل. ويمكن عن طريق النسب ملاءمة قياس مدى النجاح الكلي الذي حققه النحو الشكلي في مهمة التحليل الصرفي، وهذه النسب قد تصبح فيما بعد عناصر لمقارنة نمطين من النحو في التعامل مع ذات النص، أو لمقارنة نتائج نفس النحو في معالجة نصين لهما طبيعة مختلفة، مثلاً مقال صحفي وأبيات من الشعر.

(1) Voir Christian Gaubert, Op. cit., p. 294.

وهذه هي الطريقة التي ارتكز عليها جوبير وتتبعها في محاولته هنا. ولعلّه ممّا سهّل عملية التنقيب، أنّ تحليل النصّ قد تمّ في شكل مُستند معلوماتي قياسي، فيمكنُ تبعاً لذلك لأيّ برنامج إعداد قواعد البيانات، أن يُعيد قراءته، ثمّ تسمّح وظائف البحث والانتقاء بإجراء الإحصاءات المطلوبة.

4- عملية تحليل نصّ أدبيّ معاصر:

1-4 الخطوة الأولى: استخراج المعلومات من النصّ بشكلٍ يدويّ:

لقد قام جوبير بدراسة تطبيقية لنصّ أدبيّ معاصرٍ، بهدف إجراء تحليل صرّفيّ أوليّ للأسماء. والنصّ يتكوّن من الفصلين الأول والثاني، مع مواضع متفرّقة من رواية "قنديل أم هاشم" للكاتب يحيى حقي. فهذا النصّ تتوافر فيه كلّ الفروض التي أوضحناها أعلاه، لكنّه من الطّبيعيّ أن يحتوي على عددٍ من الأسماء الأعلام، والأسماء المُقتبسة من اللّهجة المصريّة القاهريّة. أمّا بالنسبة إلى الجوّارات التي تُمثّل جزءاً ضئيلاً من حجم الكلمات، فقد تمّ حذفها من النصّ، وفي هذا الصّدّد كان ينبغي إجراء إحصاءٍ دقيقٍ لقياس القيمة الحقيقيّة للتحليل الآليّ.

يوضّح الجدول الأول (من الشكل الخامس) التّوزيع النّسبيّ لتصنيف الكلمة بالمقارنة بإجمالي ظهورها في النصّ. مثلاً يتمّ عدّ مجموعتي الحروف (جدي)، و(لجدي)، على أنّهما كلمتان مختلفتان. وليس غريباً أن يتّضح لنا أنّ النصّ الذي يتكوّن من 1300 كلمة تقريباً، يتشكّل نصفه من أسماء ثلاثية، وتكون الأدوات أو المفردات 30 بالمائة من الكلمات، وتتواجد الأفعال الثلاثية بنسبة أقلّ من 20%. وتُمثّل الأعلام وكلّ الفئات الأخرى نسبة لا يُعتدّ بها؛ نظراً لأنّ مجموعها لا يتعدّى نسبة 5%. وبناءً عليه تُصبح أولويّة إعداد النّحو الخاصّ بالأسماء مُبرّرة تماماً. ويبيّن الجدول الثاني نتائج التنقيب، واستخراج الجذور يدويّاً لكلّ اسمٍ ثلاثي. وهذه هي الاختصارات التي تمّ استخدامها:

« c »: حرف صامت، وهو عنصر ينتمي إلى إحدى المجموعتين Σd أو Σf .

« s »: حرف نصف صامت، يعنى حرف علة واو، أو ياء.

« h »: يعنى حرف الهمزة.

« = »: هذا الرمز إذا وَقَعَ في مكان الحرف/الأصل الثالث، فهذا يدلُّ على أنَّ الجذر يحتوي على حرفين مُكَرَّرَيْن الثاني والثالث: $R3 = R2$.

وَنَقْسِمُ الأسماءُ المُضَعَّفة التي تتكوَّن من صَوَامِت صلبة consonnes solides إلى نوعين: الشَّكْل المُضَعَّف أو المُشَدَّد، ومِثَالُه كلمة (مَرَّة)، وَيُرْمَز له ب cc
1 =، والشَّكْل المُتَبَايِن أو المُنْبَسِط، ومِثَالُه كلمة (مُرور)، وَيُرْمَز له ب $cc = 2$ ⁽¹⁾.

وَيُمْكِنُنَا أَنْ نُلَاحِظَ النِّسْبَةَ الْكَبِيرَةَ لِلْجُذُورِ الصَّحِيحَةِ (ccc)، فهي تُمَثِّلُ مَا يَفْرُبُ مِنْ نِصْفِ الْحَالَاتِ، وحوالي 22 % مِنْ إِجْمَالِي النَّص. أَمَّا الْحَالَاتُ الْآخَرَى الْأَكْثَرُ شُيُوعًا فهي الجذورُ الْمُعْتَلَّةُ الْجَوْفَاءُ (csc)، ثُمَّ الجذورُ الْمُضَاعَفَةُ (=cc)، والجذورُ الْمُعْتَلَّةُ: المِثَالُ (scc)، والمُهمُوزة (hcc). وَالْحَالَاتُ غَيْرُ الْقِيَاسِيَّةِ لَا تَتَعَدَّى نِسْبَةَ 3 % مِنْ إِجْمَالِي الْجُذُورِ الْاِسْمِيَّةِ الثَّلَاثِيَّةِ لِهَذَا النَّص. وَالمُتَأَمِّلُ فِي الْجَدُولِ الثَّلَاثِ يَجِدُ أَنَّ مُحْتَوَاهُ يُشِيرُ إِلَى نَتَائِجِ التَّنْقِيبِ، واستخراج الجذور للأفعال الثلاثية، وَأَنَّ الجذورَ الصَّحِيحَةَ تُمَثِّلُ نِسْبَةَ النِّصْفِ، بَيْنَمَا تَبْلُغُ الْأَفْعَالُ الْمُعْتَلَّةُ الْوَسْطِ نِسْبَةَ الرَّبْعِ. أَمَّا الْحَالَاتُ الشَّاذَّةُ وَالْغَيْرُ قِيَاسِيَّةٌ فَتَنْسَبُهَا تَقَارِبُ نِسْبَةِ الْاِسْمِ. فَإِذَا أَرَادَ الْمَرْءُ أَنْ يَعْقِدَ مُقَارَنَةً هَذِهِ الظَّوَاهِرِ بِالْإِحْصَائِيَّاتِ الَّتِي أُجْرِيتْ عَلَى 4814 جِذْرًا ثَلَاثِيًّا مُثَبَّتًا فِي أَحَدِ الْمَعَامِجِ الْعَرَبِيَّةِ الْكَلَّاسِيكِيَّةِ⁽²⁾، الَّتِي أَوْضَحَتْ أَنَّ نِسْبَةَ الْجُذُورِ الثَّلَاثِيَّةِ الصَّحِيحَةِ تَرْتَفِعُ إِلَى

(1) Ibid., p. 295.

(2) وَفَقًا لِإِدْرَاسَةٍ قَامَ بِهَا عَلِي حَلَمِي مُوسَى مِنْ خِلَالِ مَعْجَمِ الصِّحَاحِ لِلْجَوْهَرِيِّ بِعَنْوَانِ: "دراسة إحصائية لجذور معجم الصِّحَاح باستخدام الكمبيوتر". وقد أَكْمَلَ هَذِهِ الدِّرَاسَةَ حَسَن حَبِيلِي فِي رِسَالَةِ دَكْتُورَا بِعَنْوَانِ:

Phonologie générative et morphologie flexionnelle et dérivationnelle de l'arabe.

وَلَمَّا زِيدَ بَسْطُ الْقَوْلِ يُنْظَرُ: عَلِي حَلَمِي مُوسَى، (1989م)، دِرَاسَةٌ تَقْنِيَّةٌ مُقَارَنَةٌ لِمَعَامِجِ الصِّحَاحِ وَلِسَانِ

62 %. أمّا باقي الحالات الشاذة، فهي تتبع نفس الترتيب التنازلي، الذي نتج عن دراستنا هذه. وتأسيساً على ما سبق فإن أهمية هذا البحث تستوجب فرض السؤال التالي: هل من الضروري تصميم قواعد النحو الصرّي المتعدّدة، بحيث تلائم عدد الحالات الصرّفيّة الشاذة سواء الأسماء أم الأفعال ؟ أم أنّه من الأجدي أن نحاول الجمع بين الحالات المحيّزة في نموذج واحد عن طريق حلّ وسط بلا شك ؟ ثمّ ما هي هذه الحالات المحيّزة في المعالجة الآليّة ؟

سوف نحاول فيما يلي الإجابة بتوضيح إلى أيّ مدى يؤدي مفهوم تغيرات الظواهر النحوية إلى التغلّب على مشاكل المعالجة الآلية.

الجدول الأول: التوزيع النسي لتصنيف الكلمات إلى فئات في النص:

الفئة	نوع الجذر	أول ظهور للكلمة	التكرار	المجموع	النسبة	التركيبي
الاسم	ثلاثي	540	60	600	46.7	48.5
	ثنائي	10	2	12	0.9	
	رباعي	12		12	0.9	
	علم	20	14	34	2.6	
الفعل	لهجة عامية	16		16	1.2	3.9
	ثلاثي	221	13	234	18.2	18.4
المفردات	رباعي	2		2	0.2	
	ثلاثي	37	40	77	6.0	29.2
	أخرى	88	211	299	23.3	
الإجمالي		946	340	1286	100.0	

==

العرب وتاج العروس، /147-158. علي حلمي موسى وعبد الصبور شاهين، دراسة لجذور معجم تاج العروس (باستخدام الكمبيوتر). المنصف عاشور، ظاهرة الاسم في الفكر النحوي، منشورات كتيّة الآداب، سلسلة اللسانيات، المجلد XII.

الجدول الثاني: التوزيع النسبي لأنواع جذور الأسماء الثلاثية:

نوع الجذر الثلاثي	أول ظهور للكلمة	التكرار	المجموع	النسبة	التراكمي
الصحيح	ccc	274	10	284	47.3
المضعف	cc = 2	26		26	4.3
	cc = 1	25	7	32	5.3
نصف صامت	csc	89	19	108	18.0
(حرفاً الواو، والياء)	ccs	42	6	48	8.0
	scc	25	3	28	4.7
همزة	hcc	20	6	26	4.3
	cch	6		6	1.0
	chc	4		4	0.7
غير قياسي	cs =	6	3	9	1.5
(معتلّ وبه	ssc	5	1	6	1.0
همزة أو شدة)	csh	5		5	0.8
	hc =	5	5	10	1.7
	css	3		3	0.5
	hsc	2		2	0.3
	scs	1		1	0.2
	sc =	1		1	0.2
	hcs	1		1	0.2
		540	60	600	100.0

الجدول الثالث: التوزيع النسبي لأنواع جذور الأفعال الثلاثية:

نوع الجذر	أول ظهور للكلمة	التكرار	المجموع	النسبة	التركيبي
الصحيح	ccc	2	104	44.4	94.9
المضعف	cc = 2				
نصف صامت (حرفاً الواو، والياء)	cc = 1		14	6.0	
	csc	7	55	23.5	
	ccs	3	25	10.7	
	scc		12	5.1	
همزة	hcc				
	cch		11	4.7	
	chc		1	0.4	
غير قياسي (معتل وبه همزة أو شدة)	css		5	2.1	5.1
	chs	1	3	1.3	
	csH		2	0.9	
	hss		1	0.4	
	hcs		1	0.4	
		13	234	100.0	

الشكل الخامس: نتائج عملية التنقيب في عينات النصوص المدروسة

2-4 المرحلة الأولى للتحليل: النحو الشكلي Gn3 (c,c,c) للاسم الثلاثي الصحيح:

بما أن كل تحليل يظهر في شكل قائمة من التفسيرات لإجمالي الأسماء التي يتكوّن منها النص، فقد اكتفينا بإدراج تحليل لمقتطفات ممثلة للنص في ملاحق هذا البحث⁽¹⁾.

(1) انظر: ملاحق الدراسة (ملحق-1، وملحق-2)؛ حيث نتحصل على تحليلين للمعالجة الحاسوبية باستخدام النحو الشكلي الثلاثي Gn3 (c, c, c)، و Gn3 (c, s, c).

أَمَّا بِالنَّسْبَةِ لِلإِجَابَاتِ الْخَالِيَةِ، أَوْ حَالَاتِ الصَّمْتِ وَالرَّفْضِ فِيهَا لَا تَرِدُ فِي قَوَائِمِ التَّفْسِيرَاتِ⁽¹⁾. إِنَّ الْجَدَاوِلَ (أ) تُبَيِّنُ تَوَزِيعَ الإِجَابَاتِ فِيهَا يَتَعَلَّقُ بِالصَّمْتِ وَالِاسْتِبْعَادِ وَالرَّفْضِ. أَمَّا النَّجَاحُ وَالضَّجِيجُ الْمُتَعَلِّقَانِ بِنَتَائِجِ تَحْلِيلِ الْأَسْمَاءِ وَالْأَفْعَالِ الثَّلَاثِيَةِ وَالْمُفْرَدَاتِ/ الْأَدَوَاتِ، فَهُمَا يُمَثِّلَانِ نِهَايَةَ الْخُطْوَةِ الثَّلَاثَةِ مِنَ التَّحْلِيلِ⁽²⁾.

1- النَّحْوُ الشَّكْلِيُّ (c,c,c) Gn3 والأسماء الثلثية الصحيحة:

لَقَدْ وَقَفَ جَوْبِيرُ⁽³⁾ فِي مُعَالَجَتِهِ لِلْفُظَةِ الْاِسْمِيَّةِ الصَّحِيحَةِ عِنْدَ مِلْحَظِ التَّوْزِيعِ التَّكَامُلِيِّ، حَيْثُ تَتِمُّ فِيهِ الْمُعَالَجَةُ وَفْقَ الْمَرَاكِلِ الْآتِيَةِ:

- تَفْسِيمُ الْكَلِمَاتِ الْمُرَادِ تَحْلِيلُهَا إِلَى ثَلَاثِيَّةٍ مُرْتَبَةِ بِنِظَامِ 1- الزَّوَائِدِ الْيُمْنَى. 2- الْأَصْلُ الْجِذْرِيّ. 3- الزَّوَائِدِ الْيُسْرَى.

- تَلَاوُمُ الزَّوَائِدِ مَعَ الْأَصْلِ.

- مُعَالَجَةُ الْأَصْلِ الْجِذْرِيِّ الصَّحِيحِ، وَتَلَاوُمُ الْجِذْرِ مَعَ الْوِزْنِ.

- مُقَارَنَةُ الْأَصْلِ بِالْأَوْزَانِ الْعَمِيقَةِ أَوْ السَّطْحِيَّةِ.

وَيَنْضَافُ بِنَاءً عَلَى مَا تَقَدَّمَ أَنَّ النَّمُودَجَ اللَّسَانِيَّ لِلتَّعَرُّفِ الْآلِيِّ عَلَى الْأَسْمَاءِ الثَّلَاثِيَّةِ الصَّحِيحَةِ يَبْقَى نَاجِحًا دَائِمًا فِي التَّوَصُّلِ إِلَى الْحَلِّ، مَا عَدَا حَالَاتِ "الصَّمْتِ" النَّاجِمَةِ عَنْ إِغْفَالِ ظَاهِرَةِ تَحَوُّلَاتِ الدَّوَاخِلِ الصَّرْفِيَّةِ، بِإِضَافَةِ حَرْفِ مَزِيدٍ/حَشْوٍ (الصِّيغَةُ الثَّامِنَةُ VIII)⁽⁴⁾، فَهَذِهِ الْأَخِيرَةُ تَحْمِلُ مَعْلُومَاتٍ بِنْيَوِيَّةَ مُهِمَّةَ

(1) Voir Christian Gaubert, Op. cit., p. 297.

وانظر أيضًا: شَافِيَةُ بْنُ طَامَةَ، الْاِسْتِكْشَافُ الْآلِيُّ لِلْفُظَةِ الْاِسْمِيَّةِ، /403-427.

(2) Voir Christian Gaubert, Op. cit., p. 297.

(3) Idem.

(4) رَاجِعْ هَامِشَ رَقْمِ «44» مِنْ هَذَا الْبَحْثِ. فَهَذِهِ الظَّاهِرَةُ رُبَّمَا تَكُونُ مُعَدَّلَةً، لَكِنَّهَا تَسْتَلْزِمُ تَوْضِيحَاتٍ نَحْوِيَّةَ زَائِدَةٍ لِنَلَّا يَنْشَأُ التَّشْوُّشُ وَالضَّجِيجُ.

؛ لذا فإنه يُمكن إدراجها في الوحدات التي تشكّل حالاتٍ يصعبُ تمييزها آليًا، على أنّ هذه النسبة الضئيلة من حالات الصّمت، هي نتاج تطوير النموذج الأصلي لهذا المنهج في التحليل الصّرفي أي النّحو الثنائي (G2). أمّا الضّجيج الناتج فهو يرجع في كثيرٍ من الأحيان إلى وجود حروفٍ لاحقة ليست ضمائر متصلة مثل: "ا"، "ين"، "ي" بعد الجذر الثالث ؛ وفي هذه الحالة يُمكن أن تصل تفسيرات الاسم إلى خمس تفسيرات (انظر تحليل كلمة "شحاذي" التي تظهر في النصّ المقتطف)⁽¹⁾. إنّ تأثير وجود "أل" التعريف يكاد يكون مُنعدمًا. كما أنّ أغلب حالات اللبس الداخلي تنتج عن ظهور حرف الهمزة "أ" قبل الجذر الأول R1، الذي يتمّ دائمًا تصنيفه في هذا النصّ على أنّه سابقٌ تُضاف للوزن الصّرفي، لكنّه يُمكن تفسيره على أنّه حرف استفهام، وهذا يحدث نادرًا. لكنّ الضّجيج الذي يُمثّل أكبر قدرٍ من الحيرة، هو الضّجيج المثبت/الحاضر الجذر، فحدوثه هنا ضئيلٌ للغاية⁽²⁾.

(1) يُنظر ملحق-1، مثال تحليلي رقم 37.

فالملاحظ عند خروج النتائج من المحلّل الصّرفي أنّه يُمكن لوحدة لغوية أن تقبل أكثر من تحليل ؛ لأنّ المُعالجة تتمّ خارج السياق. ويُعود هذا الالتباس أساسًا إلى:

- نسق الكتابة العربية التي لا تستعمل الشّكل عُمومًا.
- الاشتراك اللَّفْظي. وتُمثّل حالات الالتباس بالأشكال الآتية: (أ)- اسم/فعل. (ب)- اسم/أداة. (ج)- فعل/أداة. (د)- اسم/اسم. (ه)- فعل/فعل. (و)- أداة/أداة. راجع في ذلك: شافية بن طامة، الاستكشاف الآلي للفظة الاسمية، 411. وانظر أيضًا: محمد علي الزّركان، اللسانيات وبرمجة اللغة العربية في الحاسوب، 70-53.

(2) Voir Christian Gaubert, Op. cit., p. 297.

3-4 النَحْوُ الشَّكْلِيّ (c,c,c) Gn3 والكَلِمَاتِ الاسْمِيَّةِ الأُخْرَى:

إِنَّ النَّظْرَةَ الدَّقِيقَةَ إِلَى حَالَاتِ الْأَسْمَاءِ الَّتِي تَتَكَوَّنُ جُذُورُهَا مِنْ حُرُوفٍ مُضَعَّفَةٍ soursds، وَلَكِنْ مِنَ النَّوعِ الْمُتَبَايِنِ، تُبَيِّنُ لَنَا أَنَّ نَتَائِجَهَا تَقْبَلُ الْمُقَارَنَةَ؛ لِأَنَّهَا تَنْدَرُجُ تَحْتَ فِئَةِ الْجَذُورِ الثَّلَاثِيَّةِ الصَّحِيحَةِ. أَمَّا بَاقِي الْأَسْمَاءِ فَيَجِبُ عَلَى النَّحْوِ الثَّلَاثِي (c,c,c) Gn3 اسْتِبْعَادُهَا. غَيْرَ أَنَّ هَذَا لَمْ يَحْدُثْ، حَيْثُ يُمَثِّلُ الضَّجِيجُ أَكْثَرَ مِنْ ثُلُثِ التَّحْلِيلَاتِ الصَّرْفِيَّةِ. وَذَلِكَ يَرْجِعُ إِلَى وُجُودِ حَرْفِ التَّعْرِيفِ "أل"، الَّذِي يَتَسَبَّبُ فِي مَشَاكِلَ وَإِجَابَاتٍ خَاطِئَةٍ عِنْدَ تَحْلِيلِ الْجَذْرِ الثَّلَاثِي الْأَجُوفِ وَالنَّاقِصِ (csc)، وَ (ccs)، لَكِنَّهُ يُمَكِّنُ التَّحَكُّمَ فِي مَصْدَرِ هَذَا الضَّجِيجِ؛ لِأَنَّهُ يُفَرِّزُ جُذُورًا طَفِيلِيَّةً، تَبْدَأُ بِحَرْفِ "اللَّامِ"، وَيَنْدُرُ إِثْبَاتَ وَجُودِهَا.

إِنَّ الْحُرُوفَ الَّتِي تَلْحَقُ بِالْجَذُورِ، وَعَلَى رَأْسِهَا الضَّمَائِرُ الْمُتَّصِلَةُ مِثْلُ: "ه"، و"ها"، وَ"ك" يَنْتُجُ عَنْهَا نِسْبَةٌ كَبِيرَةٌ مِنَ الضَّجِيجِ الَّذِي لَهُ جِذْرٌ مُثَبَّتٌ، وَهَذَا يَخُصُّ الْأَسْمَاءَ الَّتِي تَكُونُ جُذُورُهَا مُعْتَلَّةً جَوْفَاءً وَمُضَعَّفَةً. وَيَتِمُّ اسْتِبْعَادُ فِئَاتِ الْجَذُورِ الأُخْرَى، فَحِجْمُ تَكَرُّارِهَا الْمَحْدُودِ وَنِسْبَةُ الضَّجِيجِ الْقَلِيلَةِ تُنْقِصُ مِنْ قِيَمَةِ الْإِحْصَاءَاتِ ⁽¹⁾. إِنَّ الْاسْتِبْعَادَ الْمَتَوَقَّعَ لِلْأَفْعَالِ لَيْسَ صَارِحًا: 40 % مِنَ الْأَفْعَالِ تَقْبَلُ تَفْسِيرَاتٍ تَنْتَمِي صَرَاحَةً إِلَى الضَّجِيجِ، وَبِالتَّحْدِيدِ الضَّجِيجِ الْحَاضِرَةِ جُذُورُهُ، وَهُوَ يَنْتُجُ عَنْ تَطَابُقِ الْأَوْزَانِ الصَّرْفِيَّةِ لِلْجَذُورِ الصَّحِيحَةِ، الَّتِي سَبَقَ تَوْضِيحُهَا. وَالْجَدَاوِلُ الْآتِيَّةُ تُبَيِّنُ مَلَامِحَ التَّحْلِيلِ الْآلِيَّ لِأَبْنِيَّةِ الْكَلِمِ فِي الْعَرَبِيَّةِ ⁽²⁾.

(1) Ibid., p. 299.

(2) Id., ibid., pp. 298-299.

جداول الفرع (أ): تجريد وفحص إجابات النحو الشكلي الثلاثي Gn3 (c, c, c):

التقييم		الأسباب				نوع التشوش أو الضجيج				الفئة/النوع				إجمالي عدد الإجابات	مجموع حالات الظهور	نوع الجذر	الأسماء/ النحو الثلاثي الشكلي Gn 3 (c, c, c)
نسب الضجيج الخاضرة	استخراجات ناجحة	لبس داخلي	ما بعد السلسلة	لل و ال	قبل السلسلة (ما عدا أداة التعريف)	وزن/ صيغة غير صحيحة	ضجيج (كامن)	ضجيج (حضور الجذر)	ضجيج (نفس الجذر)	ضجيج	حل	رفض	سكون				
16.7	0	17				4	15	3		18		267		18	284	ccc	الصجيج
-	0	1					1			1		25		1	26	cc = 2	المضعف
-	0	2					2			2		30		2	32	cc = 1	
2.4	74.1	30	10		1	1	56	3	52	45	80			125	108	csc	
-	0	6	1			1	5	2		7		44		7	48	ccs	جذر يتضمّن حرف علّة
-	0	6		4		4	9	2		11		19		11	28	scc	
-	0	1					1			1		25		1	26	hcc	
-	0											4		0	4	chc	هَمْزَة
-	0											6		0	6	cch	
-	0			4		4	4			4		5		4	6	ssc	
-	0											5		0	5	csh	المتبقي
-	0											10		0	10	hc=	
-	0		1				1			1		2		1	3	css	
-	0	2					2			2		7		2	9	cs=	
-	0											2		0	2	hsc	
-	0				1		1			1		1		1	1	scs	
-	0				1		1			1		1		1	1	sc=	
-	0		1				1			1		1		1	1	hcs	
5.71		65	13	8	3	14	99	10	52	95	80	454	0	175	600		الإجمالي
14.00		35	3	8	2	13	43	7	0	50	0	454	0	50	492		ما عدا صيغة csc

Tableau II.

التَّقيِيم		الأسباب					نوع التَّشْوِش أو الصَّحِيح			الفئة/النَّوع							
يَسْمَب الصَّحِيح الخاصِرة	استخراجات ناجحة	لَيْسَ دَاخِلِيٌّ	ما يَعدُّ السَّلسَلَة	لل و ال	قبل السَّلسَلَة (ما عدا أداة التَّعْرِيف)	وَرَن/ صِيغة غير صحيحة	صَّحِيح (كاملين)	صَّحِيح (حُضُور الجذر)	صَّحِيح (نفس الجذر)	صَّحِيح	حَلْ	رَفُض	سُكُون	إجمالي عدد الإجابات	مجموع حالات الظُّهُور	نوع الجذر	الأفعال / النَّحو الثَّلَاثِي الشَّكْلِي 3 Gn (c, c, c)
5.08	48.1	49	2			5	6	3	50	59		53		59	104	ccc	الصَّحِيح
															0	cc = 2	المُضَعَّف
27.3	0	8	2			1	8	3		11		5		11	14	cc = 1	جذر يَتَضَمَّن حرف عِلَّة
8	0	24			1	2	23	2		25		33		25	55	csc	
0	0	25				2	25			25		13		25	25	ccs	
-	0	4				1	2	2		4		8		4	12	scc	
-															0	hcc	جذر يَتَضَمَّن هَمْزَة
-	0											1		0	1	chc	
-	0											11		0	11	cch	
-	0											5		0	5	css	المتَّبِقي
-	0											3		0	3	chs	
-	0											2		0	2	csh	
-	0											1		0	1	hss	
-	0											1		0	1	hcs	
8.06		110	4	0	1	11	64	10	50	124	0	136	0	124	234		الإجمالي
10.8		61	2	0	1	6	58	7	0	65	0	83	0	65	130		ما عدا صيغة ccc و cc = 2

Tableau III.

التَّقيِيم		الأسباب					نوع التَّشْوِش أو الضَّجِيج			الفئة/النَّوع							المُفْرَدَات / النَّحْو الثَّلَاثِيّ Gn 3 الشَّكْلِيّ (c, c, c)
يُسَبِّب الضَّجِيج الحَاضِرَة	استِخْرَاجَات ناجِحَة	لَيْسَ دَاخِلِيّ	مَا يَغْدُ السَّلْسَلَة	لل و ال	قَبْل السَّلْسَلَة (ما عدا أداة التَّعْرِيف)	وَرَن/ صِيغَة غَيْر صَحِيحَة	ضَّجِيج (كَايْن)	ضَّجِيج (حُضُور الجِذَر)	ضَّجِيج (نَفْس الجِذَر)	ضَّجِيج	حَلْ	رَفُض	سُكُون	إِجْمَالِي عِدَد الإِجَابَات	مَجْمُوع حَالَات الظُّهُور	نَوْع الجِذَر	
0	64.7	11	2			1			11	11		6		11	17	ccc	الصَّجِيج المُضَعَّف
																cc = 2	
52.9	0	9	12		4		8	9		17		11		17	20	cc = 1	
-	0	1			1		1			1		9		1	10	csc	جِذَر يَتَضَمَّن حَرْف عِلَّة
33.3	0	3	1				2	1		3		19		3	21	ccs	
-	0											1		0	1	scc	
-	0	4	4				4			4		2		4	4	hcc	جِذَر يَتَضَمَّن هَمْزَة
-																chc	
-	0											1		0	1	cch	
-	0											3		0	3	hsc	الْمُتَبَقِّي مِّنَ الثَّلَاثِيّ
35.3	0	32	19		2		22	12		34		273		34	299		الْغَيْر ثَلَاثِيّ
31.4		60	38	0	7	1	37	22	11	70	0	325	0	70	376		الإِجْمَالِيّ
122		49	36	0	7	0	37	22	0	59	0	319	0	59	359		لَيْسَ صَحِيحًا وَلَيْسَ مُضَعَّفًا

الشَّكْلُ السَّادِس: جداول الفَرْع

(أ): تَجْرِيد وَفَحْصُ إِجَابَاتِ النَّحْوِ الشَّكْلِيّ الثَّلَاثِيّ. 6. Gn3 Fig. (c, c, c)

كَمَا أَنَّ الْجَدُولَ التَّالِيَّ يُظْهِرُ الْأَوْزَانَ الْأَسْمِيَّةَ الْأَكْثَرَ تَوْزِيْعًا، وَهِيَ مُرْتَبَةٌ حَسَبَ نِسْبَةِ تَكَرُّارِهَا.

صِيغَةُ بَدُونِ تَشْكِيلٍ	فَعْلٌ	تَفَعَّلُ	أَفْعَلُ	فَاعِلٌ
مِثَالُ اسْمِيٍّ:	فَعْلٌ	تَفَعَّلُ	أَفْعَلُ	فَاعِلٌ
مِثَالُ فِعْلِيٍّ:	فَعَلَ	تَفَعَّلَ	أَفْعَلَ	فَاعَلَ

الشَّكْلُ السَّابِعُ: صِيغُ بَدُونِ تَشْكِيلٍ مُشْتَرَكَةٌ بَيْنَ الْأَسْمِ وَالْفِعْلِ

وَمِمَّا وَرَدَ عِنْدَ جَوْبِرٍ فِي تَحْلِيلِهِ أَنَّ رَفَضَ الْأَفْعَالِ الْمُعْتَلَةِ الْوَسَطِ (csc)، وَالْآخِرِ أَلِيًّا (ccs)، وَكَذَلِكَ تِلْكَ الَّتِي مِنَ النَّوْعِ الْمُضَعَّفِ الْأَوَّلِ (cc=1) لَيْسَ أَكْثَرَ مُبَاشَرَةً، وَهُوَ مَا يُفَسِّرُهُ فِي حَالَةِ وُجُودِ اللَّاحِقَةِ "التَّاءَ" كَحَرْفٍ مَزِيدٍ عَلَى الْبِنْيَةِ الْأَصْلِيَّةِ لِلْفِعْلِ الْمَاضِي. فَهَذِهِ اللَّاحِقَةُ يُمَكِّنُ اتِّخَاذَهَا أَوْ تَفْسِيرَهَا لِأَجْلِ الْجِذْرِ الثَّالِثِ. وَهَذَا يَعْنِي حُدُوثَ الضَّجِيجِ أَوْ التَّشَوُّشِ الْكَامِنِ فِي بَعْضِ الْأَحْيَانِ. كَمَا أَنَّ بَنَى الْمُفْرَدَاتِ أَوْ الْأَدَوَاتِ ⁽¹⁾ لَيْسَتْ مَرْفُوضَةٌ أَلِيًّا تَمَامًا. وَالْوَاقِعُ أَنَّ الْعَدِيدَ مِنْهَا

(1) تَشْمَلُ الْأَدَاةَ جَمِيعَ مَا يُعْرَفُ بِالْكَلِمَاتِ الْوُظَيْفِيَّةِ، وَقَدْ أَطْلَقَ عَلَيْهَا جَوْبِرٌ مُصْطَلَحَ « tokens » مُتَأَثِّرًا بِتَسْمِيَةِ Claude Audebert و André Jaccarini فِي بَحْثٍ لِهَما بِعَنْوَانِ: «De la reconnaissance des mots-outils et des tokens», in Annales islamologiques, tome XXIV, pp. 269-293. (Voir Christian Gaubert, Op. cit., pp. 299, 300, 304).

حَيْثُ أَشَارَا إِلَى ضَبْطِ هَذَا الْمُصْطَلَحِ وَبَيَانِ الْمَقْصُودِ مِنْهُ، وَتَحْدِيدِهِ عَلَى أَنَّهُ لَا يَكُونُ مَبْنِيًّا فِي الْكَلِمَاتِ non agglutiné à des éléments pré ou posfixés. كَمَا نَجِدُ فِي الثَّرَاثِ النَّحْوِيِّ تَسْمِيَةً أُخْرَى تُكْسِبُ الْمُصْطَلَحَ قِيَمَةً وَتَوْسَعًا دَلَالِيًّا، مَثَلَتْ مَبْحَثًا أَسَاسِيًّا، وَلَعَلَّ مِنْ أَهَمِّ هَذِهِ النُّصُوصِ مَا وَجَدْنَاهُ عِنْدَ ابْنِ هِشَامِ الْأَنْصَارِيِّ فِي "الْمُعْنِي" عِنْدَمَا أَطْلَقَ عَلَى هَذِهِ الرَّمُوزِ مُصْطَلَحَ "المُفْرَدَاتِ"، وَقَدْ تَجَلَّى ذَلِكَ فِي قَوْلِهِ: « وَأَعْنِي [بِهَا، أَيِ الْمُفْرَدَاتِ] الْحُرُوفَ وَمَا تَضَمَّنَ مَعْنَاهَا مِنَ الْأَسْمَاءِ وَالظُرُوفِ؛ فَإِنَّهَا الْمَحْتَاجَةُ إِلَى ذَلِكَ ». رَاجِعْ: ابْنُ هِشَامٍ، مُعْنِي اللَّيْبِ عَنْ كُتُبِ الْأَعْرَابِ، المجلد الأول، 33/، وَسَمَّاهَا السُّيُوطِي "الأدوات" فِيمَا سَاقَهُ فِي النَّصِّ التَّالِي: « وَأَعْنِي بِالْأَدَوَاتِ الْحُرُوفَ وَمَا شَاكَلَهَا مِنَ الْأَسْمَاءِ وَالْأَفْعَالِ وَالظُرُوفِ ». انْظُرْ: السُّيُوطِي، الْإِتْقَانُ فِي عُلُومِ الْقُرْآنِ، الْجُزْءُ الْأَوَّلُ، 190/. يُنْظَرُ فِي ذَلِكَ أَيْضًا:

يَتَشَكَّلُ علي أساسِ الجِذْرِ الثَّلَاثِيِّ الصَّحِيحِ، ووفقًا للوزن الغامض "فعل". إنَّ الجُذُورَ الْمُضَعَّفَةَ مِنْ فِتَّةِ الحَرْفِ المُشَدَّدِ الوَاحِدِ الحَادِثَةِ في الأدوات، كما في حالة تحليل كلمة "كُلَّ"، قد تَمَّ استيعابه بشكلٍ جيِّدٍ في عمليات التَّحْلِيلِ. كما أنَّ هذه المفردات تخضعُ لِنِظَامِ التَّقْطِيعِ الآلِيِّ ذاتِه الحَادِثِ في الأسماء والأفعال مِنْ إمكَانِيَّةِ اتِّصَالِ اللَّوَاحِقِ بِهَا. وَمِنْ الوَاضِحِ تَمَامًا في العمليات الحَاسُوبِيَّةِ أَنَّ هذه المفردات غير الثَّلَاثِيَّةِ مرفوضة؛ لِأَنَّهَا غَالِبًا ما تكونُ مُؤَلَّفَةً مِنْ حَرَفَيْنِ اثْنَيْنِ فقط⁽¹⁾. إِنَّهَا تَسْلُكُ مَسْلَكِ الجُذُورِ الْمُضَعَّفَةِ أو المُشَدَّدَةِ بِمَجَرَّدِ أَنْ تَكُونَ مُتَسَلْسِلَةً⁽²⁾. وَمِنْ يَبْنِ الضَّجِيجِ الخَاصَّ بِالنَّمُودِجِ الآلِيِّ، ضَجِيجِ الوِزْنِ أو

==

محمود إسماعيل الصَّيْنِي، نحو مُعْجَمِ عَرَبِيٍّ لِلتَّطْبِيقَاتِ الحَاسُوبِيَّةِ، /516، 519. محمد بن أحمد- المُنْجِي جعلي- جرار دريفوس- ستيفان كنير، التَّعَرُّفُ على الحُرُوفِ العَرَبِيَّةِ عن طريقِ شَبَكَاتِ الخَلَايا العَصَبِيَّةِ الاصطناعِيَّةِ ضِمْنَ استراتيجِيَّةِ مُعَالَجَةِ الوثَائِقِ إلكترونيًّا، /701-710.

(1) هذه الثَّنَائِيَّةُ قد اتَّخَذَتْ في أَذْهَانِ القَائِلِينَ بِهَا صُورًا مُخْتَلَفَةً، وَأَشْكَالًا مُتَنَوِّعَةً، فَكَانَتْ الثَّنَائِيَّةُ التَّأْرِخِيَّةُ ذاتِ المقطع الواحد، والثَّنَائِيَّةُ المُعْجَمِيَّةُ التي ضَعُفَ حَرْفُهَا الثَّانِي فَأَصْبَحَتْ ثَلَاثِيَّةً بِوساطَةِ الشَّدَّةِ، والثَّنَائِيَّةُ التي كُرِّرَ مَقْطَعُهَا بِكُلِّ حَرْفِيٍّ فَأَصْبَحَتْ رُبَاعِيَّةً بِطَرِيقِ المُضَاعَفَةِ والتَّكْرَارِ. وَنَحْسَبُ أَنَّهُ لَا يَغِيبُ عَنْ أَحَدٍ أَنَّ الثَّلَاثِيَّ الْمُضَعَّفَ، والرُّبَاعِيَّ المُضَاعَفَ، إِنَّمَا يَرْتَدُّانِ حِينئِذٍ إِلَى الأَصْلِ الثَّنَائِيِّ لِلْفِظِ العَرَبِيِّ، وَأَنَّ هَذَا الأَصْلَ الثَّنَائِيَّ يَرْتَدُّ إِلَى الصَّوْتَيْنِ البَّسِيطَيْنِ اللَّذَيْنِ رَكَّبَا مَقْطَعَهُ، وَأَنَّ كُلًّا مِنْ هَذَيْنِ الصَّوْتَيْنِ مَا يَفْتَأُ يُوجِي عِنْدَ التَّرْكِيبِ والامْتِزَاجِ بِمَا كَانَ يُوجِي بِهِ فِي حَالِ البَسَاطَةِ والإِفْرَادِ. رَاجِعْ: صُبْحِي الصَّالِح، دِرَاسَاتٌ فِي فِئَةِ اللُّغَةِ، /147-148. وَنُظَرُ أَيْضًا المُؤَلَّفَاتِ الآتِيَّةُ: عبد الله أَمِين، الاشتقاق، /155-159. رِيَّاض قَاسِم، اتِّجَاهَاتُ البَحْثِ اللُّغَوِيِّ الحَدِيثِ فِي العَالَمِ العَرَبِيِّ، 61/2-78. رَمَضَانَ عبد التَّوَّاب، فُصُولٌ فِي فِئَةِ اللُّغَةِ، /264-266. مُحَمَّدٌ قَهْبِي جِجَازِي، عِلْمُ اللُّغَةِ العَرَبِيَّةِ، 205-208. حَامِدُ عبد القَادِر، ثُنَائِيَّةُ الأَصُولِ اللُّغَوِيَّةِ، بَحْثٌ مُنْشُورٌ بِمَجَلَّةِ مَجْمَعِ اللُّغَةِ العَرَبِيَّةِ بِالقَاهِرَةِ، 11/113-133. أَمِين فَاحِر، ثُنَائِيَّةُ الأَلْفَاظِ فِي المَعَايِمِ العَرَبِيَّةِ وعَلاَقَتُهَا بِالأَصُولِ الثَّلَاثِيَّةِ، 3/14.

(2) قَرَّرَ تَوْفِيقُ العَلَوِيِّ أَنَّ «الصَّرَافِمَ الحَرْفِيَّةَ لَا تَخْضَعُ لِنِظَامِ صَرْفِيٍّ يَتَحَكَّمُ فِيهَا تَحَكُّمُهُ فِي الأَسْمِ والفِعْلِ، فَبِنَاهَا لَا يُمَكِّنُ لِعِلْمِ الصَّرْفِ أَنْ يَدْرُسَهَا بِأَدَوَاتِهِ المَعْهُودَةِ، المُمَثَّلَةِ أَسَاسًا فِي الجِذْرِ والزَّوَائِدِ والوزنِ والصَّبِغَةِ ... إلخ؛ لِهَذَا تَبْدُو الصَّوْنُمِيَّةُ الرَّمْزِيَّةُ مَبْحَثًا أَسَاسِيًّا فِي دِرَاسَةِ الصَّرَافِمِ الحَرْفِيَّةِ، يُبْرَزُ مَا فِيهَا مِنْ انْتِظَامٍ يُمَكِّنُ أَنْ يُعْطِيَ مَا افْتَقَرَتْ إِلَيْهِ فِي عِلْمِ الصَّرْفِ». رَاجِعْ: تَوْفِيقُ العَلَوِيِّ،

==

الصيغة الخاطئة. ويُعدُّ مثلُ هذا النوع من الضجيج مُهمًّا جدًّا: فهو يبلغ 14% من الإجابات. ومن ناحيةٍ أخرى، فيُعدُّ الضجيج الناجم عن عدم الصِّحة الصرفية هامشيًّا جدًّا: فلم تحدث سوى ثلاث حالاتٍ في جميع أجزاء النصِّ.

5- إجراءات التحقق من التحليل:

يُمكننا قياسُ مدى النِّجاح في استخراج الجذور عن طريق النسبة $\tau\sqrt{}$ ، ويتمُّ تعريفها بحسبِ المعادلة الآتية:

$\tau\sqrt{}$ = (الحلول التي يصحُّها ضجيج مثبتة أو حاضرة جذوره) / عدد الأسماء الثلاثية محور الدراسة. ويتمُّ حساب تأثير الضجيج المثبتة جذوره، وهو الضجيج الأكثر إزعاجًا عن طريق τ_{bra} ، حيثُ τ_{bra} = الضجيج المثبتة جذوره / مجموع الحلول.

إنَّ حسابَ هاتين النسبتين لكلِّ أنواع الجذور يدلُّ على قدرة النحو الشكلي Gn3 (c,c,c) على استخراج الجذور الصحيحة لهذا النصِّ، كما يُشيرُ كذلك إلى الخبرة في الحالات الأخرى، التي لا يتِمَّكنُ التحليل من رفضها فعلاً؛ لأنَّ نسبة τ_{bra} تتراوح بين 20 و50 بالمائة. وهذه النسب تقلُّ إلى 15% إذا أخذنا في الاعتبار التحكم في الأوزان الصرفية الموجودة حقيقةً، وذلك من شأنه أن يَمْنَعَ بِفاعليَّة الضجيج الناتج عن وجود أداة التعريف⁽¹⁾.

==

الرُمزية الصوتية في حُرُوف المعاني، 208/. ويُنظر أيضًا لِدَات المُؤَلِّف: المعايير الشَّكليَّة المُحدَّدة لعناصر السِّلْسِلَة النُّطْقِيَّة في العربية: دراسةٌ تأليفِيَّةٌ عامَّة مع تطبيقٍ على الفعل، أطروحة دُكْتُورًا. يُنظر أيضًا: مَنْصُور محمد الغامدي، الإذراك الآلي للتَّضْعِيف، 90-83/.

(1) Voir Christian Gaubert, Op. cit., p. 300.

5-1 مثال تطبيقي: البحث عن طريق الجذر⁽¹⁾:

أول خطوة في هذا العمل تمثلت في بناء مسرد للجذور العربية، وذلك بتصنيفها بالشكل الذي سيجعلها تستغل من قبل كريستيان جوبير في هرمية محاولته العلمية وبنائها التصنيفي. إن استخراج الجذر (المجرد) حتى مع وجود الضجيج يسمح بإجراء البحث في النص عن الكلمات باستخدام جذورها، وهذا يعني إمكانية إعداد مسرد للكلمات الصحيحة في النص بواسطة الجذور، حيث يستطيع المعالج عند إدخال جذر صحيح مثبت c1c2c3 إيجاد جميع الكلمات الموجودة في النص، التي تكونت وفقاً للجذر. وكذلك يمكنه تكوين النحو الشكلي الثلاثي Gn3 (c1c2c3) وفقاً للأقواس المقبولة فقط بالنسبة للحروف الثلاثة المكونة للجذر الأول R1، والثاني R2، والثالث R3 على الترتيب أي الحروف التي تم إدخالها c1c2c3⁽²⁾. على أن

(1) يذهب بعض الدارسين إلى أن السر في اختيار "الجذر" كمُنطلق لمنهج التحليل الصرفي الآلي، يكمن في محاولة الوصول إلى تجريد "وحدة" بنيوية، لها صفة الثبات والانضباط، غير خاضعة للجدل والخلاف حول أصل الكلمة: أهو الفعل أم الاسم؟ يضاف إلى هذا أن هذه الوحدة "الجذر" معروفة المكونات من حيث العدد، فهي إما ثنائية (وهي قليلة)، أو ثلاثية (وهي تمثل الأغلبية)، أو رباعية، أو خماسية. كما أن هذا التحديد للوحدة المجردة المختارة أساساً للعمل من شأنه أن يساعد على توزيع النصوص إلى فئات وفقاً لعدد حروف الجذر المعين. وهذا التحديد والتوزيع يُرشدان المعالج إلى التعرف على مجموعة أوزان الفعل الممكنة طبقاً لمكونات الوحدة المجردة "الجذر". ومن هذا المنطلق تتولد التصريفات المختلفة للكلمات العربية كالأفعال المشتقات والمصادر والأعلام والجوامد. ويجدير بالذكر أن بعض الأعلام ترتد إلى الأوزان. والقيام بعمليات التصريف والتحليل للكلمات العربية يُوجب انتقاء السمات المميزة لكل جذر، والقواعد المُصاحبة لها، وهذا يتم تغطية إمكانات الكلمات العربية من حيث الوقوع أو عدم الوقوع. يُنظر: محمد عبد المنعم حشيش، معالجة اللغة العربية بالحاسوب، 82-73.

(2) إن طريقة الوصف التصنيفي التي أتبعها جوبير هنا جعلته يصطدم بمجموعة من الصعوبات التقنية ذات الطبيعة العلاجية. منها مُشكل التفرقة بين الكلمات التي تتجانس في الصيغة الصرفية الاشتقاقية مع انتمائها إلى نوعين من أقسام الكلم، مثلاً: (ما) التي تكون تارة من الأسماء، وتارة أخرى من الحروف. ومن الصعوبات الأخرى نذكر مُشكل إزفاق كل جذر بجميع احتمالاته الاشتقاقية، وكيفية تسنين الأدوات والحروف التي لا تعرف جذوراً مُطرّدة. يبقى نوع آخر من المشاكل المتعلقة

==

الإجابات غير الخاوية non vides لهذه الدراسة التحليلية تضم التفسيرات المبنية على هذا الجذر، وبحساب يسير للاحتمالات يمكن القول: إنه بالنسبة لنص معتدل الحجم ليس من المحتمل أن يؤدي الضجيج المثبت الجذر إلى الإضرار بهذه الدراسة.

2-5 امتداد تطبيق النحو الشكلي الثلاثي (Gn3) على التحليل الصرفي للاسم غير الصحيح:

أولى جوبير⁽¹⁾ عناية خاصة في التحليل الصرفي بدراسة التعرف الآلي على خصائص اللفظة الاسمية غير الصحيحة، حيث تكون الجذور التي تمثل حالات صرفية غير قياسية غالبية الأشكال الصرفية غير الصحيحة، وتظهر هذه الجذور في شكلين:

- الشكل الصحيح: حيث يكون فيه حرف العلة حرفاً صامتاً أصيلاً في جذر الكلمة.
- الشكل غير القياسي: حيث يكون فيه حرف العلة مدغماً في حرف آخر، أو يكون مخذوفاً.

ومن الملاحظ أن النحو الشكلي (Gn3) يستطيع التعرف على الكلمة في النوع الأول، أي في الأفعال التي تشتمل على أنصاف الحروف أو أنصاف الحركات

==

بالتدقيق الإملائي في الكلمات العربية. لقد اتبع جوبير هذا النسق لاعتقاده بأنه هو الأسلم، « فاللغة العربية خلافاً للغات الأجنبية لا تستجيب للنظام الخطي التسلسلي في الزيادة، فهذه الأخيرة لا تعرف لها موضعاً محدداً في بنية الكلمة كما هو الشأن في اللغة الفرنسية مثلاً، حيث تتمركز جميع العلامات التصريفية إما في نهاية الكلمة أو في بدايتها، ولا يتموضع شيء منها في وسط الكلمة، فحروف الزيادة في اللغة العربية يختلف موضعها في الكلمة، حسب المقولة المولدة، من الزيادة إلى الواسطة إلى اللاحقة. وهذا ما يدعو إلى طريقة تعامل خاصة، تراعى فيها أولاً وقبل كل شيء خصوصيات هذه اللغة ». يُنظر: محمد الحناش، نظرية حاسوب - لسانية لبناء المعاجم الآلية للغة العربية: محاولة في التأصيل، 363-401، 372-373.

(1) Voir Christian Gaubert, Op. cit., p. 301.

العربية (الواو أو الياء)، عن طريق إضافة القوس الذي يُجيز هذه الحروف لأجل الجذرين cwc و c.yc. كما يُمكن في النوع الثاني التعرف على الصيغة المحتملة التي ستأخذها الكلمة. وسنركز اهتمامنا على حالات الجذر الأجوف (c, s, c)؛ لأنه يؤدي في التصريف الاسمي غير القياسي إلى حالات من الحذف والإدغام في الجذور.

وفيما يلي تحديد مصطلحي ضروريي للتفرقة بين مصطلح "الجذر" racine الذي سبق استخدامه، وتمثيل الجذر في الكلمة، سواء أكانت اسماً أم فعلاً. ويجب أن نُشير إلى أننا سنتحدث عن الحرف الثاني من الجذر R2، الذي يُمثله حرف الواو أو الياء في حالة الجذر الأجوف، ومن ناحية أخرى فهناك الحروف التي تُمثل الحرف الثاني من الجذر R'2 التي قد تكون أحد الحروف الأربعة التالية: "و"، "ي"، "ا"، "ذ". ويُظهر الجدول التالي⁽¹⁾ الحالات المحتملة لهذا التمثيل:

الصيغ المشتقة	أمثلة: الشكل الأول	الجذر الثاني R2	الحرف الذي يمثل الجذر R'2
تكوين	جوع	و	و
مكيف	أطياب جيل	ي	ي
مستعبد مقيم	ثياب جيران	و	و
إعارة	باب	و	ا
إمالة مهابة	عار	ي	ي
	سائق	و	ذ
	بائع	ي	ي

الشكل الثامن: الحالات المحتملة لتمثيلات الجذور. Fig. 8.

(1) في الحالة الأخيرة قد يعني الحرف "ذ" همزة أصلية، تُمثل الحرف الثاني R2 من الجذر الثلاثي مثل كلمة "سائل".

وَيُسْتَنْتَجُ مِنْ قِرَاءَةِ مُعْطَيَاتِ الْجَدُولِ السَّابِقِ أَنَّ حَالَاتِ التَّمْثِيلِ الْمُخْتَلِفَةَ تَتَّبِعُ قَوَاعِدَ مُحَدَّدَةً تَبَعًا لِنَوْعِ الْاِشْتِقَاقِ. وَعَلَى الرَّغْمِ مِنْ أَنَّهُ لَا يُمَكِّنُ حَالِيًّا لِلنَّحْوِ الشَّكْلِيِّ Gn3 أَنْ يَتَوَقَّعَ هَذِهِ الْقَوَاعِدَ، بَيِّنَدُ أَنَّهُ بِإِضَافَةِ الْحُرُوفِ الْأَرْبَعَةِ الْمُثَلَّةِ لِلحَرْفِ الثَّانِي مِنَ الْجَذْرِ R2 فَإِنَّهُ يَكُونُ بِمُكْنَةِ الْمُحَلِّلِ أَنْ يَحْسِبَ النَّتَائِجَ الْمُرْتَبَةِ عَلَى تَوْسِيعِ مَجَالِ تَطْبِيقِ هَذَا النَّحْوِ الشَّكْلِيِّ، بِالنِّسْبَةِ لِلصَّحِيحِ الْخَاصِّ بِتَحْلِيلِ الْجَذُورِ غَيْرِ مُعْتَلَّةِ الْوَسَطِ.

إِنَّ تَقْيِيمَ النَّحْوِ الشَّكْلِيِّ Gn3 (c, cs, c) لِتَحْلِيلِ تَصْرِيفِ الْجَذُورِ، يُوضِّحُ أَنَّ اخْتِذَ مَفْهُومِ تَغْيِرَاتِ الظُّوَاهِرِ النَّحْوِيَّةِ variation de Gn3 فِي الْاِعْتِبَارِ يَتِمَاشَى مَعَ شُرُوطِ التَّسْلُسِ الْخَطِّيِّ⁽¹⁾؛ وَبِالتَّالِي يُمَكِّنُنَا عِنْدَ تَحْدِيدِ الْفَنَةِ الْخَاصَّةِ بِالْحَرْفِ الثَّانِي مِنَ الْجَذْرِ R2 فِي النَّحْوِ الثَّلَاثِيِّ Gn3 (c, s, c) الْاِكْتِفَاءَ بِالْاِحْتِكَامِ إِلَى الْأَقْوَاسِ الَّتِي تَمَّ إِدْرَاجُهَا مُؤَخَّرًا، وَذَلِكَ لِتَحْلِيلِ أَثَرِ إِدْخَالِ أَنْصَافِ الْحُرُوفِ فِي الْجَذْرِ الثَّانِي⁽²⁾.

3-5 المرحلة الثانية: التَّحْلِيلُ بِاسْتِخْدَامِ النَّحْوِ الشَّكْلِيِّ لِلْاِسْمِ الثَّلَاثِيِّ الصَّحِيحِ أَوْ الْأَجُوفِ:

إِنَّ جَدَاوِلَ الْفَرْعِ (ب) التَّالِيَةِ تُبَيِّنُ نَتَائِجَ تَحْلِيلِ النَّصِّ بِاسْتِخْدَامِ النَّحْوِ الشَّكْلِيِّ الْمَحْدُودِ Gn3 (c, s, c). لَقَدْ تَمَّ تَصْنِيفُ الْإِجَابَاتِ بِاعْتِبَارِ أَنَّ مَجَالِ التَّطْبِيقِ اقْتَصَرَ عَلَى تَحْلِيلِ الْأَسْمَاءِ الثَّلَاثِيَّةِ الْمُعْتَلَّةِ الْوَسَطِ دُونَ غَيْرِهَا؛ لِذَلِكَ تَمَّ تَرْتِيبُ الْإِجَابَاتِ الْخَاصَّةِ بِالْأَسْمَاءِ الثَّلَاثِيَّةِ الصَّحِيحَةِ (c, c, c) عَلَى أَنَّهَا إِمَّا صَحِيحًا، أَوْ مَرْفُوضَةً، مَعَ

(1) يُعَبَّرُ عَنْ هَذِهِ الْحَالَةِ مِنْ خَطِيَّةِ التَّنَوُّعِ بِوَاسِطَةِ الْمُعَادَلَةِ الْآتِيَةِ:

$$\text{Gn3 (c, cs, c)} = \text{Gn3 (c, c, c)} + \text{Gn3 (c, s, c)}.$$

رَاجِعْ هَامِشَ رَقْمِ «72» مِنْ هَذِهِ الدِّرَاسَةِ.

(2) Voir Christian Gaubert, Op. cit., p. 302.

أَنَّهَا مقبولة من الجزء الصحيح في النحو الشكلي Gn3 (c, cs, c) ⁽¹⁾.

إذا كان جميع الأسماء المعتلة الوسط (c, s, c) تُعطي إجابة واحدة على الأقل، فإن ثلاثة أرباع الإجابات يكون فيها الحرف الممثل للحرف الثاني من الجذر R'2 هو نفس الحرف الثاني R2. وفي هذه الحالة نصل مباشرة إلى الجذر الأجوف الواوي cwc أو اليائي cyc. أما الحالات الباقية التي تتكوّن من الأسماء التي يكون فيها الحرف الممثل للحرف الثاني من الجذر R'2 هو "ا" أو "ن"، ولا يسبب أيّ منهما لبساً أو غموضاً حقيقياً؛ لأنه إذا كان الجذر مثبتاً بالواو cwc أو بالياء cyc، فإن الصيغة المستخرجة تنتج عن أحد الجذرين بالواو أو بالياء. ولذلك تمّ الالتزام في عملية التنقيب هذه بعدم الاكتفاء بالكلمات التي تشتمل على الجذور الجوفاء، ولكن يجب أيضاً إدخال أشكال الصيغة الأولى les formes I لهذه الجذور، كما ينبني على ذلك القول بالتحقق عند إيجاد التفسيرات الخاصة بالكلمات من أنّها مشتقة من الجذر الثلاثي الأجوف بالواو cwc أو بالياء cyc، مع تحديد الضجيج المناسب الذي يُصاحِبُ كلاً منهما ⁽²⁾.

لنأخذ مثلاً كلمة "بيع" فهي تُثير اللبس؛ حيث إنّ النحو الشكلي قد تعرّف عليها مباشرة، وحدّد الحرف الثاني من الجذر هو حرف الياء y R2، ومع ذلك فإنّ الرجوع إلى مسرد الصيغ الأولى les formes I يذكّرنا بوجود الجذر "بيع" المأخوذ من "بوع". وتعدّ هذه الكلمة حلاً، لكنّها أيضاً نوع من الضجيج المثبت الجذر، على الرغم من أنّ إجابة المحلل لا تحوي إلا تفسيراً واحداً. وعلى النقيض فإنّ كلمة "ساق" تعتبر ضجيجاً؛ لأنّ الجذر لا تتم معرفته إلا من خلال الحرف الممثل له وهو "الألف"، الذي ينشأ من الجذر "سوق" فقط. وهذا الجذر ينشأ عنه "ضجيج

(1) Ibid., p. 302.

(2) Id., ibid., p. 302.

صحيح الجذر "bruit radicalement exact"، أمّا الإمكانية الأخرى فهي "سيق"، فهي غير حاضرة، وتُعدُّ ضَجِيجًا مُتَبَقِّيًا أو كَامِنًا⁽¹⁾.

جداول الفرع (ب): تجريد وفحص إجابات النحو الشكلي الثلاثي Gn3 (c, s, c):

(1) Idem.

Tableau I

التَّقيِيم		الأسباب					نوع التَّشَوُّش أو الضَّجِيج			الفئة/النَّوع							الأسماء/ النَّحو الثلاثي الشَّكليّ Gn 3 (c, c, c)
نِسَب الضَّجِيج الخاصِّرة	استخراجات ناجحة	لُبْسٌ دَاخِلِيٌّ	ما يَعدُّ السَّلسَلَة	لل و ال	قبل السَّلسَلَة (ما عدا أداة التَّعريف)	وَزْن/ صِيفَة غير صحيحة	ضَّجِيج (كَامِن)	ضَّجِيج (خُضُور الجِذر)	ضَّجِيج (نفس الجِذر)	ضَّجِيج	حَلْ	رَفُض	سُكُون	إجمالي عدد الإجابات	مجموع حالات الظُّهور	نوع الجِذر	
16.7	0	17				4	15	3		18		267		18	284	ccc	الصَّجِيج
-	0	1					1			1		25		1	26	cc = 2	المُصَغَّف
-	0	2					2			2		30		2	32	cc = 1	
2.4	74.1	30	10		1	1	56	3	52	45	80			125	108	csc	
-	0	6	1			1	5	2		7		44		7	48	ccs	
-	0	6		4		4	9	2		11		19		11	28	Scc	
-	0	1					1			1		25		1	26	hcc	جِذرٌ يَتَضَمَّن هَمْزَة
-	0											4		0	4	chc	
-	0											6		0	6	cch	
-	0			4		4	4			4		5		4	6	ssc	المتَّبِقي
-	0											5		0	5	csh	
-	0											10		0	10	hc=	
-	0		1				1			1		2		1	3	css	
-	0	2					2			2		7		2	9	cs=	
-	0											2		0	2	hsc	
-	0				1		1			1		1		1	1	scs	
-	0				1		1			1		1		1	1	sc=	
-	0		1				1			1		1		1	1	hcs	
5.71		65	13	8	3	14	99	10	52	95	80	454	0	175	600		الإجمالي
14.00		35	3	8	2	13	43	7	0	50	0	454	0	50	492		ما عدا صيغة csc

Tableau II.

التَقْيِيم		الأسباب					نوع التَّشْوِش أو الضَّجِيج			الفئة/النَّوع							
نِسَب الضَّجِيج الخاضِرة	استِخراجات ناجِحة	لَيْسَ دَاخِلِيًّا	ما بَعْدَ السَّلْسَلَة	لل و ال	قِبل السَّلْسَلَة (ما عدا أَدَاة التَّعْرِيف)	وَزْن/ صِيفَة غِير صِحيحة	ضَّجِيج (كَاِمِن)	ضَّجِيج (حُضُور الجِذر)	ضَّجِيج (نَفْس الجِذر)	ضَّجِيج	حَل	رَفُض	سُكُون	إِجمالي عدد الإِجابات	مجموع حالات الظُّهور	نَوع الجِذر	الأفعال/ النَّحْو الثَّلَاثِي الشَّكْلِي Gn 3 (c, c, c)
-	0	1				1	3	1		2		102		2	104	ccc	الصَّجِيج
															0	cc = 2	المُضَعَّف
0	0											14		0	14	cc = 1	
10	49.1	30					30	3	27	30		29		30	55	csc	جِذر يَتَضَمَّن حرف عِلَّة
-	0											25		0	25	ccs	
0	0											12		0	12	scc	جِذر يَتَضَمَّن هَمْزَة
															0	hcc	
-	0											1		0	1	chc	
0	0											11		0	11	cch	
-	0					1	1			1		4		1	5	css	الْمُتَبَقِّي
-	0											3		0	3	chs	
-	0						3			1		1		0	2	csh	
-	0											1		0	1	hss	
-	0											1		0	1	hcs	
-		31	0	0	0	2	37	4	27	34	0	204	0	33	234		الإِجمالي
-		1	0	0	0	2	7	1	0	4	0	175	0	3	179		ما عدا صِيفَة csc

Tableau III.

التَقْيِيم		الأسباب					نوع التَّشْوِش أو الضَّجِيج			الفئة/النَّوع							المفردات/ النَّحو الثَّلَاثِي الشَّكْلِي Gn 3 (c, c, c)
نِسَب الضَّجِيج الخَاصِرَة	استِخراجات ناجحة	لَيْسَ ذَاخِلِيًّا	ما يَبْدُو السَّلسَلَة	لل و ال	قَبْل السَّلسَلَة (ما عدا أداة التَّعْرِيف)	وَزْن/ صِيفَة غَير صَحِيحَة	ضَّجِيج (كَايْن)	ضَّجِيج (خُضُور الجِذَر)	ضَّجِيج (نَفْس الجِذَر)	ضَّجِيج	حَل	رَفُض	سُكُون	إِجْمَالِي عِدَد الإِجَابَات	مَجْمُوع حَالَات الظُّهُور	نَوْع الجِذَر	
	0											17		0	17	ccc	الصَّحِيح
																cc = 2	المُضَعَّف
	0											20		0	20	cc = 1	
	ns	10					10		10	10				10	10	csc	المُعْتَل
	0											21		0	21	ccs	
	0											1		1	1	scc	
0	0						8			4		2		4	4	hcc	المُهْمُوز
																chc	
	0											1		0	1	cch	
	0											3		0	3	hsc	الْمُتَبَقِّي مِنَ الثَّلَاثِي
	0		12			2	20	2		12		290		12	299		لَيْسَ ثَلَاثِيًّا
7.41	2.66	10	12	0	0	2	38	2	10	26	0	355	0	27	376		الإِجْمَالِي
0	0	0	12	0	0	2	28	2	0	16	0	355	0	17	366		ما عدا صِيفَة csc

الشكل التاسع: جداول الفرع (ب): تجريد وفحص إجابات النحو الشكلي الثلاثي. Gn3 Fig. 9. (c, s, c)

إِنَّ تَعَدُّدَ حَالَاتِ الضَّجِيجِ الَّتِي تَظْهَرُ بوضوحٍ في الجداول السَّابِقَةِ لَا تُعَدُّ مُخِيفَةً وَلَا تَبْعَثُ عَلَى الْقَلْقِ الْخَطَرُ، حَيْثُ تُشِيرُ النَّتَائِجُ إِلَى نِسْبَةِ ضئيلةٍ مِنَ الضَّجِيجِ الْمَثْبُتِ جَذَرَهُ، بَيِّنُ أَنَّ الضَّجِيجَ الْأَكْبَرَ الَّذِي يَخُصُّ الْجِذْرَ يُعْتَبَرُ نَجَاحًا لِلتَّحْلِيلِ؛ لِأَنَّهُ لَنْ يَتَبَقَّى لَنَا حِينئذٍ سِوَى إِعَادَةِ تَكْوِينِ الْجِذْرِ⁽¹⁾.

إِنَّ تَعَامُلَ النَّحْوِ الشَّكْلِيِّ الْأَجُوفِ (c, s, c) Gn3 مع الْجِذُورِ الثَّلَاثِيَّةِ الصَّحِيحَةِ ccc يَتِمُّ وَفَقًا لِلشُّرُوطِ نَفْسِهَا، وَيَنْدَرُ وَجُودُ حَالَاتٍ حَقِيقِيَّةٍ لِلْبَّسِ، إِذَا كَانَ هُنَاكَ تَحْكَمُ فِي الْوِزْنِ أَوِ الصِّيغَةِ الْمُسْتَخْرَجَةِ. وَذَلِكَ يَسْرِي عَلَى كَافَةِ فِئَاتِ الْكَلِمَاتِ الْأُخْرَى عِدا الْأَفْعَالِ الْمُشْتَقَّةِ مِنَ الْجِذُورِ الْجُوفَاءِ csc. وَفِي هَذِهِ الْحَالَةِ الْأَخِيرَةِ يَجِبُ الْبَحْثُ بِشَكْلِ أَوْسَعٍ فِي الْمُعْجَمِ لِاسْتِعَادِ الْأَشْكَالِ الدَّخِيلَةِ. فَمَثَلًا كَلِمَةُ "عَاد" لَا تَصْلُحُ إِلَّا أَنْ تَكُونَ فِعْلًا فَقَطْ؛ لِأَنَّ هَذِهِ الصِّيغَةَ لَمْ تَثْبُتْ فِي الْاسْتِعْمَالِ اللَّغَوِيِّ كَاسْمٍ. فَالْفِعْلُ "تَعُود" لَا يُمَكِّنُ لِلْمُعَالِجِ الْآلِيِّ التَّعَرُّفَ عَلَيْهِ كَفِعْلٍ؛ لِأَنَّهُ يُمَكِّنُ أَنْ يَكُونَ اسْمًا مَصْدَرًا مُشْتَقًّا مِنَ الصِّيغَةِ الْخَامِسَةِ (V). إِنَّ حَرْفَ "النَّاءِ" كَحَرْفٍ سَابِقٍ عَلَى الْجِذْرِ، هُوَ الَّذِي يُؤَدِّي إِلَى هَذَا اللَّبْسِ، حَيْثُ يُمَثِّلُ ثَلَاثَ حَالَاتٍ الْأَفْعَالِ الَّتِي تَعَرَّفَ عَلَيْهَا هَذَا النَّحْوُ الشَّكْلِيُّ الْإِسْمِي. لَكِنْ هُنَاكَ حَالَاتٌ ثَلَاثَةٌ فَقَطْ تَأْخُذُ الصِّيغَةَ الْخَامِسَةَ، وَبِنَاءً عَلَيْهِ يُصْبِحُ مِنَ الضَّرُورِيِّ - كَمَا سَبَقَ أَنْ أَوْضَحْنَا نَتَائِجَ التَّحْلِيلِ الشَّكْلِيِّ الثَّلَاثِيِّ لِلأَفْعَالِ الصَّحِيحَةِ (c, c, c) Gn3 - أَنْ تَتِمَّ زِيَادَةُ قَائِمَةِ الْجِذُورِ الْمَوْجُودَةِ، بِإِضَافَةِ الصِّيغِ الْأَكْثَرِ لَبْسًا بِالنِّسْبَةِ لِلْجِذُورِ الْمَثْبُتَةِ. إِنَّ الْجَمْعَ بَيْنَ قَوَاعِدِ النَّحْوِ الشَّكْلِيِّ الْخَاصِّ بِالْجِذُورِ الصَّحِيحَةِ (c, c, c) Gn3، وَالنَّحْوِ الْخَاصِّ بِالْجِذُورِ الْجُوفَاءِ (c, s, c) Gn3، سَوْفَ يَمْنَعُ أَيَّ تَأْثِيرٍ مُحَيِّرٍ فِي حَالَةِ وُجُودِ تَحَكُّمٍ كَافٍ. وَأَهَمُّ مَا يَتَمَيَّزُ بِهِ هَذَا الْجَمْعُ يَتَعَلَّقُ بِزَمَنِ الْمُعَالَجَةِ، فَقَدْ أَدَّى مَدَّ تَطْبِيقِ النَّحْوِ الشَّكْلِيِّ لِلْجِذُورِ الثَّلَاثِيَّةِ الصَّحِيحَةِ (c, c, c) Gn3 إِلَى

(1) Id., ibid., p. 304.

الجدور الجوفاء (c, cs, c) Gn3 إلى زيادة عشر وقت التحليل، بينما إذا كانت المعالجة بواسطة النحو الشكلي للجدور المعتلة الوسط (c, s, c) Gn3 فقط فإنها (أي المعالجة) تتم في فترة زمنية متقاربة عند تطبيق النحو الشكلي الخاص بالجدور الثلاثية الصحيحة (c, c, c) Gn3. إن عملية التقطيع segmentation التي يجريها النحو الشكلي للجدور الثلاثية Gn3 لا تضيع هباءً عند الجمع بين نوعي النحو الشكلي في نحو واحد⁽¹⁾.

(1) Idem.

ملاحق الدراسة نماذج تطبيقية من التحليل الصرفي الآلي لنصوص عربية

ملحق (1)

1.1 نموذج لتكوين مستخرج من النص:

« [...] فهذه الأزعفة تعدُّ وتُمَلَأُ بِالْفُؤْلِ النَّابِتِ، وَتَخْرُجُ بِهَا أُمُّ مُحَمَّدٍ، تَحْمِلُهَا فِي مِقْطَفٍ عَلَى رَأْسِهَا. مَا تُهَلُّ فِي الْمَيْدَانِ حَتَّى تُخْتَطَفَ الْأَزْغَفَةُ، وَيَخْتَفِيَ الْمِقْطَفُ، وَتَطِيرُ مَلَأَتْهَا، وَتَرْجِعُ خَجَلَةً، تَتَعَثَّرُ فِي أَذْيَالِهَا غَاضِبَةً ضَاحِكَةً مِنْ جَشَعِ شَحَّاذِي السَّيِّدَةِ، وَتَصِيرُ حَادِثَتُهَا فَكَاهَةً الْأُسْرَةَ بِضَعَةِ أَيَّامٍ يَتَنَدَّرُونَ [...] »⁽¹⁾.

2-1 نموذج لتحليل النص بواسطة النحو الشكلي الثلاثي Gn3 (c, c, c):

يُوضِّحُ الشَّكْلُ التَّالِي العنصر الرئيسة لتحليل الكلمات، حيثُ يَقُومُ الْمُعَالِجُ الْمَذْكُورُ بِدَوْرِ الْمُفَكِّكِ لِيَفْصَلَ جَذْرَ الْكَلِمَةِ عَمَّا يَتَّصِلُ بِهَا مِنَ السَّوَابِقِ وَاللَّوَاقِحِ، وَزَوَائِدِ التَّصْرِيفِ.

وَلَا يَخْفَى أَنَّهُ لَا يَتَّسِعُ مَقَامُ الْبَحْثِ وَلَا مَجَالُهُ هُنَا لِمَطْلَبِ الْحَدِيثِ عَنْ شَرْحِ التَّفَاصِيلِ الْفَنِيَّةِ لِشَبَكَاتِ الْإِنْتِقَالِ الْمُعَزَّزَةِ، وَنَكْتَفِي بِتَاكِيدِ أَهْمِيَّتِهَا فِي تَمْثِيلِ الْقَوَاعِدِ التَّصْرِيفِيَّةِ فِي صِيَاغَةٍ صُورِيَّةٍ تُمَثِّلُ مَسَارَاتِ التَّحْلِيلِ الْمُخْتَلِفَةِ.

(1) راجع: يحيى حقي، قنديل أم هاشم، 65/.

الرقم	الكلمة العربية	التحليل	الجذر	الوزن/الصيغة
1	فَهَذِه	R3: ه/ R2: ذ/ R1: ه/ coor: ف/ هَذِه فعل		
1	فَهَذِه	post: ه/ R3: ذ/ R2: ه/ R1: ف/ فَهَذِه فعل		
2	الأَرْغِفَة	R3: sgfm: ر/ ف/ R2: غ/ R1: ر/ pref: أ/ art: ل/ art: ا/ رَغِفَ أفعلة		
3	تُعَدُّ	R3: د/ R2: ع/ R1: ت/ تعد فعل		
5	بِالْفُؤْل	R3: ل/ inf2: و/ R2: ف/ R1: ل/ pref: ا/ prep: ب/ لِفْل أفعول		
6	النَّابِت	R3: ت/ R2: ب/ inf1: ا/ R1: ن/ art: ل/ art: ا/ نَبِت فاعل		
7	وَتَخْرُجُ	R3: ج/ R2: ر/ R1: خ/ pref: ت/ coor: و/ خَرَجَ تفعّل		
10	مُحَمَّد	R3: د/ R2: م/ R1: ح/ pref: م/ حمد مفعّل		
11	تَحْمِلُهَا	post: ا/ post: ه/ R3: ل/ R2: م/ R1: ح/ pref: ت/ حمل تفعّل		
13	مِقْطَف	R3: ف/ R2: ط/ R1: ق/ pref: م/ قَطَفَ مفعّل		
17	تُهَلِّل	R3: ل/ R2: ه/ R1: ت/ تهلّ فعل		
19	الميدان	duma: ن/ duma: ا/ R3: د/ inf2: ي/ R2: م/ R1: ل/ pref: ا/ لَمَدَ أفعيل		
21	تُخْتَطَفُ	R3: ف/ R2: / ط/ inf1: ت/ R1: خ/ pref: ت/ خُطِفَ تفتعل		
22	الأَرْغِفَة	R3: sgfm: ر/ ف/ R2: غ/ R1: ر/ pref: أ/ art: ل/ art: ا/ رَغِفَ أفعلة		
24	المِقْطَف	R3: ف/ R2: ط/ R1: ق/ pref: م/ art: ل/ art: ا/ قَطَفَ مفعّل		
26	تَطِير	R3: ر/ inf2: ي/ R2: ط/ R1: ت/ تطرّ فاعيل		
28	وَتَرْجِعُ	R3: ع/ R2: ج/ R1: ر/ pref: ت/ coor: و/ رَجَعَ تفعّل		
29	خَجَلَة	R3: ر/ R2: ل/ R1: خ/ خَجَلَ فعلة		
33	غَاظِبَة	R3: ر/ R2: ب/ R1: ض/ inf1: ا/ R1: غ/ غَضِبَ فاعلة		
34	ضَاكِكَة	R3: ر/ R2: ك/ R1: ض/ inf1: ا/ R1: ض/ ضَحِكَ فاعلة		

36 جَشَعَ R3 ع: R2/ ش: R1/ ج/ جشع فعل
37 شَحَّاذِي suff ي: R3/ ذ: inf2/ ا: R2/ ح: R1/ ش/ شحد فعالي
37 شَحَّاذِي duma ي: R3/ ذ: inf2/ ا: R2/ ح: R1/ ش/ شحد فعال
37 شَحَّاذِي post ي: R3/ ذ: inf3/ ا: R2/ ح: R1/ ش/ شحد فعال
37 شَحَّاذِي nisb ي: R3/ ذ: inf2/ ا: R2/ ح: R1/ ش/ شحد فعال
37 شَحَّاذِي plma ي: R3/ ذ: inf2/ ا: R2/ ح: R1/ ش/ شحد فعال
38 السَّيِّدَة sgfm ة: R3/ د: inf2/ ي: R2/ س: R1/ ل: pref/ ا/ لسد أفعيلة
39 وَتَصِيرُ R3 ر: inf2/ ي: R2/ ص: R1/ ت: coor/ و/ تصر فعيل
40 حَادِثَتُهَا post: ا/ post: ه: sgfm ت: R3/ ث: R2/ د: inf1/ ا: R1/ ح/ حدث فاعلة
41 فَكَاهَة sgfm ة: R3/ ه: inf2/ ا: R2/ ك: R1/ ف/ فكه فعالة
43 بَضْعَة sgfm ة: R3/ ع: R2/ ض: R1/ ب/ بضع فعلة

ملحق (2)

1-2 تحليل جُمْلَةٍ بِوَاسِطَةِ النَّحْوِ الشَّكْلِيِّ الثَّلَاثِيِّ Gn3 (c, s, c):

الرَّقْم	الكَلِمَةُ العَرَبِيَّةُ	التَّحْلِيل	الجذر	الوِزْن/الصِّيغَةُ
5	بِالْفُؤْل	R1/و: R2/ل: R3/ف: art/ل: art/ا: prep/ب: فول فعل		
19	المِيدَان دوما	duma/ن: R3/ل: duma/د: R2/ي: R1/م: art/ل: art/ا: ميد فعل		
26	طَير	R3/ر: R2/ي: R1/ط: pref/ت: طير تفعّل		
32	أذْيَالهَا	post/ا: post/ه: R3/ل: inf2/ا: R2/ي: R1/ذ: int/أ: ذيل فعال		
32	أذْيَالهَا	post/ا: post/ه: R3/ل: inf2/ا: R2/ي: R1/ذ: pref/أ: ذيل أفعال		
38	السَّيِّدَة	sgfm/ة: R3/د: R2/ي: R1/س: art/ل: art/ا: سيد فعلة		
39	وَنَصِير	R3/ر: R2/ي: R1/ص: pref/ت: coor/و: صير تفعّل		
41	فَكَاهَا	sgfm/ة: R3/ه: R2/ل: R1/ك: coor/ف: كاه فعلة		

الخاتمة

1.1 النتائج:

- إن الميزة الحقيقية لهذا البحث تتمثل في أنه يُزِيح السِتَارَ لِلْمَرَّةِ الأولى عن ملامح تحليل المشهد اللساني الصّرفي/المعلوماتي لنصوص عربية متنوعة لعالم فرنسي، يتوخّى الدراسة الجادة للفكرة والمحتوى، ثمّ تقديم هذا العمل من منظور عربيّ، بعيداً عن هيمنة الأيديولوجيا، ونزعات الانكفاء على الذات.
- أثبتَ البحثُ أنّ "الصّرف" بمفهومه الحديث الذي اعتمدته كريستيان جوبير في أعماله العلمية هو فرعٌ من فروع اللسانيات الذي يتعامل مع البنية الداخلية للكلمات، من حيث تكوين عناصرها الأولية، والتفاعل بين هذه العناصر، وأنّ كلمة "مورفولوجي" - وتعني باليونانية دراسة الشكل - لا تُعطي دلالة دقيقة لمهمة الصّرف داخل المنظومة اللغوية، وهي المهمة التي لا تقتصر على الأمور المتعلقة بـ "شكل" الكلمات، بل تتعدّها إلى دراسة الصلّة بين مباني الكلمات ومعانيها، وخصائصها النحوية والفونولوجية. وأنّ وجه الخلاف بين مفهوم "الصّرف" عند العرب القدماء وبين مفهوم الحديث يتضمّن أمرين: الأول: استبعاد التغيّرات الصوتية غير الدالة على معنى من الصّرف بالمفهوم الحديث؛ إذ يرى المحدثون أنّ التغيّير الذي يحدث في بنية الكلمة لغرضٍ لفظيٍّ، أو بسببٍ صوتيٍّ لا شأن لعلم الصّرف به، بل هو من اختصاص علم الأصوات. في حين اشتمل صرف القدماء على التغيّيرات في المبنى والمعنى. الثاني: اقتصار صرف القدماء على الأسماء المتمكنة والأفعال المتصرفّة من الكلمات العربية. على حين نرى الصّرف بالمفهوم الحديث يشمل جميع الكلمات التي يحدث فيها تغيير ذو دلالة. فالتّصريف في علم اللغة الحديث يبحث في الوحدات الصرفية أو (الصّرفيات) التي تؤدي وظائف محدّدة في الصّيغ. فالمحدثون من علماء

اللغة يرون أنَّ "التَّصْرِيف" لا يقوم إلا على ما يُقَرِّره علم الأصوات من حقائق، فهو يتعامل مع البنية الداخلية للكلمات، وأنَّ الظَّاهرة الاشتقاقية في جوهرها مزجٌ للجذور مع الأنماط الصَّوتية لسلاسل حركات الصَّيْغة الصَّرْفِيَّة، فالتَّغييرات التي تؤدي إلى تحولاتٍ في المعنى هي من صميم "الصَّرْف". وواضحٌ من هذه الأنساق التعريفية في المباحث اللسانية الحديثة وتطبيقاتها في اللسانيات الحاسوبية أنَّ مفهوم علم "الصَّرْف" يُشيرُ إلى تَضَمُّنِ الصَّرْفِ للاشتقاق والتَّصْرِيف.

- لقد طرَحَ البَحْثُ كَثِيرًا مِنَ الأسئلة التي أفرزتها ثقافة عصر التَّنْمِيَةِ والمُتَغَيِّرَاتِ المعلوماتية، والاحتياج إلى مُستويين مِنَ القِرَاءَةِ، قراءة شاملة تتلوها قراءة مُتَأَنِيَّة؛ حيثُ تتقدَّمُ فيهما إقامة النَّمَاذِجِ اللُّغَوِيَّةِ، وتحليل التَّطْبِيقَاتِ الحاسوبية في مجال الصَّرْفِ العربي؛ بهدف اقتراح إجاباتٍ للعديد من تلك الأسئلة؛ لتحقيق نتائج عملية، رُبَّمَا يلجأ إليها المُنْظَرُونَ اللغويون أنفسهم فيما بعد لبلورة نظرياتهم، واختبار صحَّةِ فُرُوضِهِم.

- في هذا البَحْثِ قَدَّمْنَا نموذجًا مبدئيًّا، يُمكنُ استعماله لتحليل خصائص صرْفِيَّةٍ لِنَصٍّ باللغة العربية، وذلك بالاعتماد على خصائص مُكوِّنَاتِ البِنْيَةِ الصَّرْفِيَّةِ. أمَّا الرِّوَابِطُ التي تُكوِّنُها الفكرة الرئيسة لهذا النِّقَاشِ، هو أنَّه يُمكنُ استنتاج خصائص نظام من خلال خصائص مُكوِّنَاتِهِ وروابطه، وكذلك من خلال العِلاَقَاتِ التي تربط المُكوِّنَاتِ والرِّوَابِطِ باستعمال أسلوب الاستقراء.

- ولقد أَوْضَحْنَا فكرتنا، وذلك بالاعتماد على المنهج الذي اتبعه Christian Gaubert (كريستيان جوبير) في تقييم قواعد الصَّرْفِ العربي وتطويرها، تأسيسًا على تحليل النُّصُوصِ بهيكلية دقيقة: مُكوِّنَةٌ مصدر، ومُكوِّنَةٌ وَجْهَةٌ، كُلُّ مُكوِّنَةٍ داخلية تشمل على وَصْلٍ إيرادٍ واحدة، وعلى وَصْلٍ إصدارٍ واحدة، وكُلُّ رَابِطٍ

يتكوّن من دور مصدرٍ واحدٍ، ومن دور وجهة واحد. كُلُّ ذلك يُمكننا لاحقاً من تصوّر مِعماريّاتٍ حاسوبيّة ذات هياكل صرفية مختلفة.

- لقد ناقشنا قواعد اقتراح جوبير في التّحليل الصّرفيّ الآليّ لنصوص اللغة العربية، وأنّها مُنبَية على أعمال سابقه وبخاصة [Claude Audebert (كُلود أوديرت)، و André Jaccarini (أنديريه جاكارينى)]، وأنّه قد استفاد على نطاقٍ كبيرٍ من نموذَجَيْهما.

- تعرّضَ البحثُ لِعَدَدٍ من المفاهيم سواء لدى القُدماء العرب، أو لدى المعاصرين، هذا ما جعلَ الدِّراسَة مُتَعَدِّدة المِصطلحات، كما تَطَرَّقَ البَحْثُ إلى جَرَدٍ للمِصطلحات التي لَجَأَ إليها كريسْتيان جوبير للتّعبير عن نظراتِه اللِّسانية الحاسوبيّة في قراءة الظواهر الصّرفية العربية.

- تُعْتَبَرُ المُحاوَلَة العِلْمِيّة الصّرفية لِكْرِيسْتيان جوبير مَسْعَى وَظِيفِيّ يقوم على تحديد العلاقات التي تربط بين عناصر المورفيمات الصّرفية، والدّور الذي يُؤدِّيه كل عنصر في تركيب البنى التّصريفية، كما يستعين هذا المَسْعَى باعتماده طريقة تحليل النّصّ إلى المؤلّقات المباشرة كوسيلة لِحَليْلِ بِنْيَة المُركّبات والصّيغ التّصريفية. فهذه المُحاوَلَة تنظر إلى الصّيغة الصّرفية كِبْناءٍ مُتَدَرِّجٍ في مستوياتٍ متتالية، بدءاً من الوحدة الصّوتية إلى الوحدات الدُّنيا (الصّيغِمْ)، وتهتمُّ بالبحث عن العناصر اللغوية المُستترة أو غير الظّاهرة على المستوى الشّكليّ.

- أمّا دِرَاسَة أَهَمِّ الأُسُس والمعايير في التّحليل فقد تَمَثَّلَتْ في التّوصيف، والتّطَبُّق، وتحليل النُّصوص. وقد رَكَّزَتْ على أَهمّيّة البِنْيَة العَميقة للكلمة الصّرفية، والتّخديدات الدّلالية والتّركيبية، وإبراز سمات الكلمات (الأسماء، والأفعال، والأدوات).

- تَبَيَّنَ في البحث طَرَحُ كْرِيسْتيان جوبير مسألة تطبيق النّحو الشّكليّ غير الحَثِيّ

الذي يَسمح أثناء تحليل الكلمة بوجود عدّة احتمالات قابلة للتّحليل تلقائيًا، كما يَستطيع التّعرّف على بنية الكلمة التّصريفية في النّوع الأوّل عن طريق إضافة الأقواس، التي تُؤدّن بوجود حروف العِلّة (الواو، أو الياء). كما يَسمح في النّوع الثّاني من التّعرّف على الصّيغة المُحتمّلة التي ستأخذها الكلمة.

- لقد تَوَضّحَ أَنَّ الطّريقة الحديثة التي اتّبعها كريستيان جوبير في محاولته العلميّة يُمْكِنُهَا أَنْ تُبَسِّطَ خوارزمية التّصريف، وذلك بالفصل بين المعالجة الصّوتية، والمعالجة الكتابيّة الإملائية، كما أنّها تعتمد على بناء قاعدة بيانات مُرَمّزة لأجزاء النّصوص العربيّة من خلال مُدَوّنة مُرَمّزة أيضًا. وعلى ذلك فإنّ مُقترح جوبير يُعدّ لبنّة قويّة مُحدّدة المصدر في إطار دعم المعالجة الآليّة لظواهر الصّرف العربيّ.

- إنّ الوصف المُقترح لكريستيان جوبير في سبيل إنجاز المُحلّل الصّرفيّ للعربيّة يعتمد في مُقارنته على ثلاث مراحل: فهو يقوم في مشروعه أولاً بتقسيم النّصّ، ثمّ إيجاد أصل الكلمة مثلما يُوجد في المعجم، وأخيرًا يقوم بالترميز.

- تقسيم النّصّ العربيّ: هو أول مرحلة في تحليل النّصّ، ويتمثّل في تقطيع النّصّ إلى وحداتٍ صغيرة (فقرات، جُمَل، كلمات، ... إلخ)، وتقسيم النّصّ حسب التّنقيط والفراغات بين الكلمات.

- ثمّ تعيين أصل الكلمة: وذلك بإرجاع الكلمة إلى أصلها أي إلى شكلها الأساسيّ.

- ثمّ التّرميز لأجزاء الكلام ؛ ممّا يُؤدّي إلى تحسين أداء الحاسب في فهم عمليات التّحليل الصّرفيّ، والقُدرة على:

- تحليل الكلمات العربيّة المشكّولة كُلّيًّا أو جُزئيًّا أو غير المشكّولة.
- تحليل الأفعال والأسماء والأدوات بجميع أنواعها ومختلف جذورها.
- تمييز الكلمات غير الصّحيحة.

- تحديد نوع الكلمة (اسم، فعل، صفة، مصدر، ... إلخ).
- تحديد التوزيعات النسيبة لتصنيف الكلمات إلى فئات في النص العربي.
- تحديد سوابق الكلمة.
- تحديد لواحق الكلمة.
- تحديد أواسط الكلمة.
- تحديد جذر الكلمة.
- تحديد وزن الكلمة.
- تحديد أنواع الجذور (صحيح سالم، صحيح مَهْمُوز، مُضَعَّف، مُعْتَلّ، مُعْتَلّ وبه همزة، مُعْتَلّ وبه شدة، ... إلخ).
- تحديد الزيادات في الجذر:
- جذر بدون إضافات.
- جذر يحتوي على لواحق فقط.
- جذر يحتوي على بَوَادِي فقط.
- جذر يحتوي على بَوَادِي ولواحق.
- تحليل الكلمات المشتركة في الاسمية والفعلية والحرفية.
- تحليل الأفعال المشتركة في صيغة الماضي والمضارع والأمر.
- تحليل الأسماء المشتركة في الإفراد والتثنية والجمع.
- إنَّ منهج التَّنْقِيبِ والتَّحْقِيقِ الذي استخدمه كريستيان جوير يُعْطِي إجاباتٍ دقيقة إلى حَدِّ كافٍ، بَيَدَ أَنَّهُ - في نَظَرِنَا - يستوجبُ إجراءات وعمليات تحقُّق وتأكُّد مُستمرّة ومكلفة، وذلك مشروطٌ بأنَّ يكونَ حجم النصِّ الذي يتمّ

تحليله متوسّطاً. والأمر ينبغي أن يقوم على الدّراسات الإحصائية المتكاملة للظواهر اللغوية التي تُفرزها المعالجة الآلية. إنّ هذا الإحصاء لا يُمكن تصوّره إلاّ عندما تصلّ قواعد النّحو الشّكليّ إلى درجةٍ عاليةٍ من الدّقة، بحيث تُغطّي نُصوصاً متنوّعة، من حيث طبيعتها والمفردات المُستخدَمة فيها.

- إنّ نتائج هذه الدّراسة التّطبيقية من خلال نماذج من نصوص العربية المعاصرة، تُشيرُ بوضوحٍ إلى ضرورة توسيع قواعد النّحو الشّكليّ، حيث تأخذ صورة أقواس يُمكن اتّباعها وفقاً لمُعَيّنة مباشرة أو غير مباشرة لقوائم الصّيغ الصّرفيّة المُثبّتة. إنّ إضافة معجم للجذور المُثبّتة ولبعض الصّيغ الصّرفيّة المُمكنة من شأنه أن يُقلّل من عددِ حالات اللّبس اللغويّ، وكذلك يسمح بالتّقليل من زمن البَحْث والتّنقيب في النّصوص، واستخراج المعلومات بصورةٍ أسرع. وهذا التّوجّه يُؤيّدُه كلّ محاولات التّعرّف على الأسماء التي تتضمّن جذورها أحرف عِلّة، أو حروف مُشدّدة، أو همزات.

- أظهرَ البَحْثُ أنّه من المُهم والضروريّ أن تتجمّع في نموذجٍ واحدٍ كلّ الأسماء النَّاتجة عن مختلف حالات الجذور عندما تتّسمُ بمنهجٍ شبيه مُتكرّر، وهذا يُؤدّي بالفعل إلى تقليل مُدّة تقطيع النّصّ، والتّعرّف على الوزن، والمُعَيّنة المُحتملة للمسارد، حتى عند وجود بعض الحالات الاستثنائية التي لا يكتشفها إلاّ مُعالج خاص.

- إنّ التّعرّف على الأفعال يتمُّ بذات الطّريقة، لكنّ نسب توزيع الأفعال قليلة مُقارنَةً بفئة الأسماء ؛ لذلك فهي تستوجبُ عمل دراساتٍ على نصوصٍ أكبر حَجْماً وأكثر تنوعاً من أجل التّوصّل إلى إحصائيات ذات مغزى، وعندئذٍ يتَحَتّمُ علينا توقّع مظاهر كبيرة للضّجيج، ولا بدّ من التّحكّم فيها بواسطة أساليب وإجراءات التّحقّق من التّحليل، ووضْع النّمودج الذي يتفادى

- العيوب اللغوية، ويتجنب النتائج غير السليمة صرفياً، واللجوء إلى تعديل قواعد إعادة كتابة النحو الشكلي أو توسيعها.
- أَوْضَحَ الْبَحْثُ أَنَّهُ عِنْدَ الْقَضَاءِ عَلَى حَالَاتِ اللَّبْسِ تَكُونُ الدِّرَاسَةُ التَّرْكِيبِيَّةُ لِلنَّصِّ الْخَاصَّةُ بِالْحُرُوفِ وَالْأَدَوَاتِ وَبَعْضُ الْأَفْعَالِ هِيَ وَحْدَهَا الْقَادِرَةُ عَلَى تَحْدِيدِ مَدَى صِحَّةِ التَّفْسِيرَاتِ الْمَتَّبِقَةِ.

2-1 التوصيات:

- مِنْ خِلَالِ وَاقِعِ جَمْعِ الْمَعْطِيَاتِ اللُّغَوِيَّةِ فِي هَذِهِ الدِّرَاسَةِ نَرَى اقْتِرَاحَ التَّوَصِيَّاتِ الْآتِيَةِ:
- ضَرُورَةُ اسْتِثْمَارِ خِصَائِصِ الصَّرْفِ الْعَرَبِيِّ فِي النُّظْمِ الْآلِيَّةِ لِتَخْمِينِ النَّقْصِ فِي عُنَاوَرِ التَّشْكِيلِ، وَدَعْمِ عَمَلِيَّةِ التَّشْكِيلِ التَّلْقَائِيِّ، وَتَغْطِيَةِ الْإِحْتِمَالَاتِ الْمُلْبِسَةِ، وَتَحْدِيدِ أَوْلَوِيَّاتِ التَّطْبِيقِ الْقَوَاعِدِيِّ فِي مِثْلِ تِلْكَ الْحَالَاتِ.
- مَا زَالَتْ مِنْظُومَةُ الصَّرْفِ الْعَرَبِيِّ بِحَاجَةٍ مَاسَّةً إِلَى قِرَاءَاتٍ حَاسُوبِيَّةٍ جَدِيدَةٍ تَهْتَمُّ بِشَقِّ الْمَعْنَى (الْبُنْيَةُ الْعَمِيقَةُ وَالْبُنْيَةُ السَّطْحِيَّةُ) وَعِلَاقَتِهِ بِالْمَعْجَمِ، وَتَطَوُّرِ الْعِلَاقَاتِ الصَّرْفِيَّةِ/النَّحْوِيَّةِ عَلَى مَسْتَوَى التَّرْكِيبِ.
- يَجِبُ أَنْ تَهْدَفَ الدِّرَاسَاتُ التَّطْبِيقِيَّةُ فِي مَجَالِ الْمَعَالِجَةِ الْآلِيَّةِ لِلصَّرْفِ الْعَرَبِيِّ إِلَى مِرَاعَاةِ تَحْلِيلِ النُّصُوصِ الْعَرَبِيَّةِ بِالْأَسْلُوبِ الصَّرْفِيِّ، وَاسْتِنْتَاجِ قَوَاعِدِ الْإِسْتِقْرَاءِ الْعَامَّةِ لِمُخْتَلَفِ الْخِصَائِصِ التَّصْرِيفِيَّةِ، وَأَنْ يَكُونَ الْمُعَالِجُ الصَّرْفِيُّ الْآلِيَّ غَيْرَ مُقَيَّدٍ بِشَرِيحَةٍ مُعَيَّنَةٍ مِنَ النُّصُوصِ، بَلْ يَنْبَغِي النَّظَرُ إِجْمَالاً إِلَى تَمَثِيلِ الْقَوَاعِدِ الصَّرْفِيَّةِ بِكُلِّ حَالَاتِهَا الْعَامَّةِ وَالْخَاصَّةِ فِي الْبَرْنَامِجِ، وَالتَّرْتِيبِ الْمُنطَقِيِّ لِأَنْوَاعٍ مُخْتَلِفَةٍ مِنَ الْبَيَانَاتِ الصَّرْفِيَّةِ الْإِلْزَامَةِ لِفَهْمِ بُنْيَةِ الْكَلِمَةِ الْعَرَبِيَّةِ.
- التَّفْكِيرُ فِي إِقَامَةِ هِيََاكِلٍ لِتَفْسِيرِ الدَّلَالَاتِ وَالْمَعَانِي الصَّرْفِيَّةِ وَتَقْيِيسِهَا عَلَى مَسْتَوَى جَمِيعِ الْوَحْدَاتِ اللُّغَوِيَّةِ، وَحَصْرُ جَمِيعِ صُورِ اللَّبْسِ التَّصْرِيفِيِّ وَرَبْطُهُ

بالصُّور المعجمية والتركيبية، والإلمام بالسِّيَاقَات الكُلِّيَّة لِكُلِّ صُورَةٍ، وَوَضْعُ القَوَاعِدِ اللَّازِمَةِ لِفَكِّ اللَّبْسِ؛ وذلك بتوفير المُنْتُونِ النَّصِّيَّةِ والمُدَوَّنَاتِ المُرَمَّزَةِ كَعَيِّنَاتٍ لُغَوِيَّةٍ يُجْرَى عَلَيْهَا بَعْدَ ذَلِكَ عِدَّةُ نَمَازِجِ حَاسُوبِيَّةٍ لِفَهْمِ الأَدَاءَاتِ الصَّرْفِيَّةِ.

قائمة المراجع

أ- العربية:

- 1- الإستراباذي، (رضي الدين محمد بن الحسن)، شرح شافية ابن الحَاجِب، مع شَرْح شَوَاهِدِه لِ: عبد القادر البغدادي صَاحِب خزانة الأدب المتوفى في عام (1093هـ)، حَقَّقَهُمَا، وَضَبَطَ غَرِيْبَهُمَا، وَشَرَحَ مُهِمَّهُمَا: محمد نور الحسن، ومحمد الزفزاف، ومحمد محيي الدين عبد الحميد، دار الكتب العلمية، بيروت - لبنان، 1395هـ-1975م.
- 2- الأَشْمُونِي، (أَبُو الحَسَن نُور الدِّين عَلِيّ بن محمد بن عيسى)، شَرْح الأَشْمُونِي عَلَى أَلْفِيَةِ ابْن مَالِك، قَدَّمَ لَهُ وَوَضَعَ هَوَامِشَهُ وَفَهْرَسَهُ: حَسَن حَمْد، إِشْرَاف: إميل بديع يعقوب، منشورات: محمد علي بيضون، دار الكتب العلمية، بيروت - لبنان، 1419هـ-1998م.
- 3- أَمِين، (عبد الله)، الاشتقاق، مَطْبَعَةُ لَجْنَةِ التَّأْلِيفِ وَالتَّرْجَمَةِ وَالنَّشْرِ، القاهرة، 1376هـ-1956م.
- 4- الْبَكُوش، (الطَّيِّب)، التَّصْرِيفُ الْعَرَبِيُّ مِنْ خِلَالِ عِلْمِ الْأَصْوَاتِ الْحَدِيثِ، الْمَطْبَعَةُ الْعَرَبِيَّة، تونس، ط3، 1992م.
- 5- التَّهَانَوِي، (محمد علي ابن قاضي محمد حامد بن محمد صابر الفاروقي)، موسوعة كَشَافِ اصطلاحات الفنون والعلوم، تقديم وإشْرَاف ومُراجعة: رفيق العجم، تحقيق: علي دحروج، نَقَلَ النَّصَّ الْفَارِسِيَّ إِلَى الْعَرَبِيَّة: عبد الله الخالدي، التَّرْجَمَةُ الْأَجْنَبِيَّة: جورج زينات، مكتبة لبنان ناشرون، بيروت - لبنان، 1996م.
- 6- الْجَرَجَانِي، (أبو بكر عبد القاهر بن عبد الرحمن بن محمد)، المفتاح في الصَّرْف، حققه وقدم له: الدكتور علي توفيق الحَمْد، كلية الآداب - جامعة اليرموك - إربد - عمان، النَّاشِر: مؤسسة الرسالة - بيروت، (1407هـ-1987م).

- 7- ابن جني، (أبو الفتح عثمان)، المنصف، شرح كتاب التصريف لأبي عثمان المازني، الناشر: دار إحياء التراث القديم، ذي الحجة 1373هـ - أغسطس 1954م. ونسخة أخرى: المنصف في التصريف، تحقيق: إبراهيم مصطفى وعبد الله أمين، مطبعة ومكتبة عيسى البابي الحلبي، القاهرة، 1954م.
- 8- الجوهري الفارابي، (أبو نصر إسماعيل بن حماد)، الصحاح تاج اللغة وصحاح العربية، تحقيق: أحمد عبد الغفور عطار، دار العلم للملايين، بيروت - لبنان، ط4، 1407هـ - 1987م.
- 9- ابن الحاجب، (أبو عمرو عثمان بن أبي بكر بن يونس)، الشافية في علمي التصريف والخط، تحقيق: صالح عبد العظيم الشاعر، الناشر: مكتبة الآداب، القاهرة، 2010م.
- 10- حجازي، (محمود فهمي)، علم اللغة العربية - مدخل تاريخي مقارن في ضوء التراث واللغات السامية، وكالة المطبوعات، الكويت، (د.ت).
- 11- الحديثي، (خديجة)، أبنية الصرف في كتاب سيبويه، مكتبة النهضة، بغداد، 1965م.
- 12- أبو الحسين القزويني الرازي، (أحمد بن فارس بن زكريا)، معجم مقاييس اللغة، تحقيق: عبد السلام محمد هارون، دار الفكر، 1399هـ - 1979م.
- 13- حقي، (يحيى)، قنديل أم هاشم، مطابع الهيئة المصرية العامة للكتاب، 2000م.
- 14- حمادة، (سلوى)، (2009م)، المحللات الصرفية للغة العربية، اجتماع خبراء المحللات الحاسوبية الصرفية للغة العربية، دمشق.
- 15- —، (2009م)، المعالجة الآلية للغة العربية المشاكل والحلول، دار غريب للطباعة والنشر والتوزيع، القاهرة - مصر.
- 16- الحملوي، (أحمد بن محمد)، شذا العرف في فن الصرف، تحقيق: نصر الله عبد الرحمن نصر الله، الناشر: مكتبة الرشد الرياض، (د.ت).

- 17- سُلَيْمَان، (علي)، مَدْخَل إلى الحَاسُوب والخوارزميات، جامعة تشرين، 2009م-2010م.
- 18- شريف، (محمد أبو الفتوح)، (1976م)، نظرة وصفية في تصريف الأفعال، الناشر: مكتبة الشباب بالمنيرة، مطبعة قاصد خير، القاهرة - مصر.
- 19- —، (1980م)، نظرة وصفية في تصريف الأسماء، مكتبة الشباب، القاهرة، مصر.
- 20- صابر، (محمود محمد) - عبد المنعم، (عبد المنعم فريد)، « برنامج "مداد" للتحليل الصرفي للكلمات العربية »، شركة مداد لتقنية المعلومات، اجتماع خبراء المحللات الحاسوبية الصرفية للغة العربية، دمشق، 2009م.
- 21- الصَّالِح، (صُبْحِي)، دِرَاسَات في فِقه اللغة، دار العِلْم للملايين، بيروت - لبنان، 1379هـ-1960م، ط14، 2000م.
- 22- الضَّامَن، (حاتم صالح)، الصَّرف، وزارة التعليم العالي والبحث العلمي، جامعة بغداد، مطابع دار الحكمة للطباعة والنشر، بغداد، 1991م.
- 23- عَاشُور، (إنصاف) - بن عبد السلام، (هيبه)، «المُحَلِّل الصَّرْفِي للغة العربية»، المعهد العالي للتَّصْرِيف، تونس، (د.ت).
- 24- عبابنة، (يحيى)، الصَّرف العربيّ التَّحليليّ نظرات مُعاصرة، دار الكتاب الثقافي، الأردن، 2016م.
- 25- عبد التَّوَّاب، (رَمَضَان)، فُصُول في فِقه العربيَّة، مكتبة دار التُّراث، القاهرة - مصر، 1977م.
- 26- عبد الرَّحْمَن، (طه)، المنطق والنَّحو الصُّوري، دار الطَّليعة للطبَّاعة والنَّشر، بيروت - لبنان، 1983م.
- 27- ابن عصفور الإشبيلي، (أبو الحَسَن علي بن مؤمن بن محمد الحَضْرَمي)، الممتع الكبير في التَّصْرِيف، الناشر: مكتبة لبنان، 1996م.

- 28- عليّ، (نبيل)، (1988م)، اللغة العربية والحاسوب (دراسة بحثية)، مطابع الخط، تغريب.
- 29- —، (ديسمبر 2001م)، الثقافة العربية وعصر المعلومات، سلسلة عالم المعرفة، ع276، الكويت.
- 30- عمايرة، (خليل أحمد)، في التحليل اللغوي: منهج وصفي تحليلي، مكتبة المنار، الزرقاء - الأردن، 1407هـ.
- 31- فاخر، (أمين)، ثنائية الألفاظ في المعاجم العربية وعلاقتها بالأصول الثلاثية، مكتبة الكليات الأزهرية، القاهرة - مصر، 1398هـ-1978م.
- 32- قاسم، (رياض)، اتجاهات البحث اللغوي الحديث في العالم العربي (لبنان في القرن التاسع عشر « 1801-1900، 1901-1960 »)، مؤسسة نوفل، بيروت - لبنان، 1982م.
- 33- موسى، (علي حلمي)، (1973م)، دراسة إحصائية لجذور معجم الصحاح (باستخدام الكمبيوتر)، مطبوعات جامعة الكويت، الكويت.
- 34- موسى، (علي حلمي) - شاهين، (عبد الصبور)، دراسة لجذور معجم تاج العروس (باستخدام الكمبيوتر)، مطبوعات جامعة الكويت، الكويت، بلا تاريخ.
- 35- الموسى، (نهاد)، (2000م)، العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، بيروت - لبنان.
- 36- —، (2003م)، مقدمة في تمثيل الكفاية اللغوية للحاسوب، ندوة "الهوية اللغوية والعمولة"، 2003م، نُشرت في كتاب "اللغة العربية وتحديات العصر"، جامعة البترا، الأردن، 2005م.
- 37- الموسى، (نهاد)، (2003م)، من الوصف إلى التوصيف، مقارنة في حوسبة العربية، (د.ط.).

- 38- هلال، (عبد الغفار حامد)، أُبْنِيَةُ الْعَرَبِيَّةِ فِي ضَوْءِ عِلْمِ الشَّكْلِ الصَّوْتِيّ، دار الطَّبَاعَةِ الْمُحَمَّدِيَّة، القاهرة، 1399هـ-1979م.
- 39- ياقوت، (محمود سليمان)، ظاهرة التَّحْوِيلِ فِي الصِّيغِ الصَّرْفِيَّةِ، دار المعرفة الجامعية، الإسكندرية، 1985م.
- 40- الينبعاوي، (غنيم)، جهود ابن جَيِّ فِي الصَّرْفِ وَتَقْوِيمِهَا فِي ضَوْءِ عِلْمِ اللُّغَةِ الْحَدِيثِ، المكتبة التُّجَّارِيَّة، مَكَّة الْمُكَرَّمَة، 1995م.

بـ المترجمة:

- 1- ايلوار، (رونالد)، مَدْخُلٌ إِلَى اللِّسَانِيَّاتِ، تَرْجُمة: بدر الدِّين القَّاسِمِ، مَنَشُورَاتِ وَزَارَةِ التَّعْلِيمِ الْعَالِي، مَطْبَعَة جَامِعَة دِمَشْق، الْجُمْهُورِيَّة الْعَرَبِيَّة السُّورِيَّة، 1400هـ-1980م.
- 2- باي، (ماريو)، أُسُسُ عِلْمِ اللُّغَةِ، تَرْجُمة: أَحْمَدُ مُخْتَارُ عُمَر، مَنَشُورَاتِ جَامِعَة طَرَابْلُس، طَرَابْلُس، 1973م.
- 3- الحضري، (جمال الحَضْرِي)، تَرْجُمة معجم اللسانيات بإشراف جورج موان، المؤسسة الجامعية للدراسات والنشر والتوزيع، بيروت - لبنان، 1433هـ-2012م.
- 4- دي سويسير، (فردينان)، دُرُوس فِي الْأَلْسِنِيَّة الْعَامَّة، تَرْجُمة: صَالِح الْقَرْمَادِي - مُحَمَّد الشَّوْش - مُحَمَّد عَجِينَة، الدَّار الْعَرَبِيَّة لِلْكِتَاب، تُونِس، 1985م.

جـ. الرسائل والدوريات العلمية:

- 1- ابن أحمد، (محمد) - جعلي، (المنجي) - دريفوس، (جرار) - كنير، (ستيفان)، التَّعَرُّفُ عَلَى الْحُرُوفِ الْعَرَبِيَّةِ عَنْ طَرِيقِ شَبَكَاتِ الْخَلَايا الْعَصَبِيَّةِ الْإِصْطِنَاعِيَّةِ ضِمْنَ اسْتِرَاطِيْجِيَّةِ مُعَالَجَةِ الْوُثَائِقِ الْإِلِكْتُرُونِيَّاءِ، ضِمْنَ أَعْمَالِ السَّجَلِ الْعِلْمِيِّ لِنَدْوَةِ اسْتِخْدَامِ اللُّغَةِ الْعَرَبِيَّةِ فِي تَقْنِيَةِ الْمَعْلُومَاتِ، مَطْبُوعَاتِ مَكْتَبَةِ الْمَلِكِ عَبْدِ الْعَزِيزِ الْعَامَّةِ، الرِّيَاض - الْمَمْلَكَةِ الْعَرَبِيَّةِ السَّعُودِيَّةِ،

- 1414هـ-1993م، 701-710.
- 2- أنيس، (إبراهيم)، الحاسبات الإلكترونية في البحوث اللغوية، ضمن أعمال المجمع المصري للثقافة العلمية، ع42، 198-201.
- 3- باربو، (أمين عبد الكريم)، من أصالة الجذور إلى جذور الأصالة، ضمن أعمال ندوة "مجادلة السائد في اللغة والأدب والفكر"، كلية العلوم الإنسانية والاجتماعية، سلسلة 7، المجلد 12، تونس، 2002، 147-159.
- 4- بسندي، (خالد بن عبد الكريم)، "الصرف والتصريف وتداخل المصطلح"، مجلة جامعة الملك سعود، المجلد العشرون، كلية الآداب، المملكة العربية السعودية، 1429هـ-2008م، 319-386.
- 5- بكري، (سعد علي الحاج) - القصير، (توفيق أحمد)، التفاعل بين المعلوماتية واللغة العربية: نظرة متكاملة وخطة مستقبلية، مجلة مكتبة الملك فهد الوطنية، المجلد الأول، ع2، (رجب- ذو الحجة 1416هـ/ ديسمبر 1995- مايو 1996م).
- 6- البكوش، (الطيب) - السويسي، (رضا) - بن حمادو، (عبد المجيد)، معجم المصطلحات اللسانية الإعلامية، ضمن منشورات الجامعة التونسية، مركز الدراسات والأبحاث الاقتصادية والاجتماعية، تونس، اللسانيات العربية والإعلامية، سلسلة اللسانيات 7-، 1989م، 139-166.
- 7- البواب، (مروان) - علم، (يحيى مير) - الطيان، (محمد حسان)، بإشراف: مراياتي، (محمد)، « نظام اشتقاق الكلمة العربية بالحاسب »، مركز الدراسات والبحوث العلمية، دمشق، أشغال الملتقى الرابع للسانيات، 9-12 نوفمبر 1987م، ضمن منشورات الجامعة التونسية، مركز الدراسات والأبحاث الاقتصادية والاجتماعية، تونس، اللسانيات العربية والإعلامية، سلسلة اللسانيات 7-، 1989م، 25-63.
- 8- حشيش، (محمد عبد المنعم)، معالجة اللغة العربية بالحاسوب، ضمن

أعمال السَّجَلِ العِلْمِيِّ لندوة استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامة، الرياض - المملكة العربية السعودية، 1414هـ-1993م، /73-82.

9- الحطاب، (مأمون) - عبد المنان، (حسن)، التحليل الصرفي للغة العربية باستخدام الحاسوب، ضمن منشورات مجمع اللغة العربية، الأردن، 1996م.

10- بن حمادو، (عبد المجيد) - بلغيث، (لمياء هديش) - شعبان، (نهي)، المحلل الصرفي للغة العربية، مخبر "ميراكل"، صفاقس - الجمهورية التونسية، 2009م.

11- الحناش، (محمد)، (1414هـ-1993م)، نظرية حاسوب - لسانية لبناء المعاجم الآلية للغة العربية: محاولة في التأصيل، ضمن أعمال السَّجَلِ العِلْمِيِّ لندوة استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامة، الرياض - المملكة العربية السعودية، /363-401.

12- —، (2002a)، اللغة العربية والحاسوب: قراءة سريعة في الهندسة اللسانية العربية، أو مقارنة في محاكاة الدماغ العربي لغويًا، جامعة الإمارات العربية المتحدة، كلية العلوم الإنسانية والاجتماعية، قسم اللغة العربية وآدابها.

13- —، (2002b)، المقارنة الحاسوبية للغة العربية، منشورات التواصل اللساني، المغرب.

14- —، (2013م)، لسانيات الجيل الرابع ومجتمع المعرفة، مجلة التواصل اللساني، المجلد الخامس عشر، فاس - المغرب.

15- خياط، (محمد غزالي)، تمثيل الدلالة الصرفية في النظم الآلية لفهم اللغة العربية، ضمن أعمال السَّجَلِ العِلْمِيِّ لندوة استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامة، الرياض - المملكة العربية السعودية، 1414هـ-1993م، /299-312.

16- راغين، (بوشعيب)، الحوسبة التوليدية للصرف العربي، ضمن أشغال الندوة

- الدولية الأولى عن الحاسب واللغة العربية، مدينة الملك عبد العزيز للعلوم والتقنية، الرياض - المملكة العربية السعودية، 2007م.
- 17- الرامي، (سالم)، تعليم الصّرف العربيّ بدعْم الحاسوب، ضمن وقائع مؤتمر "التّوليد والنّسقية والتّرجمة الآلية"، معهد الدِّراسات والأبحاث للتعريب، الرباط - المغرب، 2001م.
- 18- الرزّو، (حسن مظفر)، سبل تطوير مُحلّل الصّرف الآلي المُستخدَم في حوسبة الموسوعات العربية، مجلّة الرّافدين الجامعة، عدد رقم 6، منشورات كُليّة الرّافدين الجامعة، العراق، 2001م.
- 19- الرزوق، (آمال عبد اللطيف)، مُحلّل صرفي للكلمات العربية المشتقة، معهد الكويت للأبحاث العلمية، الكويت، 1989م، /47-60.
- 20- زروقي، (طه)، الجوانب البرمجية لدعم العربية في المدقّق الإملائيّ مفتوح المصدر هانسبال، ضمن الأيام الدراسية حول المعالجة الآلية للغة العربية، المركز الجامعيّ بشار، مايو 2007م.
- 21- الرّامل، (عبد الله)، القاعدة الصّرفية لمفردات اللغة العربية، معهد بحوث الحاسب والإلكترونيات، مدينة الملك عبد العزيز للعلوم والتقنية، الرياض - المملكة العربية السعودية، 2001م.
- 22- الزّركان، (محمد علي)، اللّسانيّات وبرمجة اللغة العربيّة في الحاسوب، ضمن أعمال السّجلّ العِلْميّ لندوة استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامّة، الرياض - المملكة العربية السعودية، 1414هـ-1993م، /53-70.
- 23- السّريع، (سريع محمد)، نظام ابن النّديم (معهد الإدارة العامّة)، ضمن أعمال السّجلّ العِلْميّ لندوة استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامّة، الرياض - المملكة العربية

- السعودية، 1414هـ-1993م، /315-336.
- 24- السيد عنبّر، (تغريد)، الفعل الماضي مُسندًا إلى ضمائر الرفع المتصلة دراسة صرفصوتية، المجلة العربية للدراسات اللغوية، المجلد الرابع، ع2، جمادى الأولى 1406هـ - فبراير 1986م.
- 25- الشرفي، (منصف) - عمّار (هاني)، ضمن أعمال الندوة الدولية للعلوم وهندسة الحاسوب: الدورة السادسة، التّأشير: مؤسسة فيليبس، تونس، ياسمين الحمامات، مايو 2010م.
- 26- الشلوي، (بريكان بن سعد)، اللبس، أسبابه وطرق تجنبه في التّفعيد الصّرفي، مجلة جامعة أمّ القرى لعلوم اللغات وآدابها، ع6، رجب 1432هـ-يونيو 2011م، /45-90.
- 27- صالح، (عبد الرّحمن الحاج)، منطوق النّحو العربيّ والعلاج الحاسوبيّ للغات، ضمن أعمال السّجلّ العلميّ لندوة استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامّة، الرياض - المملكة العربية السعودية، 1414هـ-1993م، /27-41.
- 28- آل طه، (هدى سالم عبد الله)، النّظام الصّرفيّ للعربية في ضوء اللّسانيات الحاسوبية: مثل من جمع التّكسير، رسالة دكتوراه، كلّية الدّراسات العُليا، الجامعة الأردنيّة، عمّان - الأردن، 2005م.
- 29- العناتي، (وليد أحمد)، اللّسانيات الحاسوبية العربية (المفهوم، التّطبيقات، الجدوى)، مجلة الزرقاء للبحوث والدراسات، المجلد السابع، ع2، الأردن، 2005م، /61-81.
- 30- مجلّة لُغة العَصْر، عدد 92، أغسطس، 2008م.
- 31- المنظّمة العربية للتربية والثّقافة والعلوم، استخدام اللغة العربية في المعلوماتية، تونس، 1996م.

32- مهديوي، (عمر)، (1999م)، توليد الأسماء من الجذور الثلاثية المعتلة: مقاربة لسانية حاسوبية، رسالة دكتوراه، كلية الآداب والعلوم الإنسانية، فاس - المملكة المغربية.

33- —، (2005م)، خوارزميات توليد الأسماء في اللغة العربية، مركز البحوث العلمية والتقنية لترقية اللغة العربية بالتعاون مع أكاديمية الجزائر بالجزائر.

34- —، (15-17 مايو 2014م)، الهندسة اللغوية والترجمة الآلية: المفهوم والوظيفة، المؤتمر السنوي للمنظمة العربية للترجمة، حول الترجمة والحاسوب.

35- المهيوبي، (عبد العزيز بن عبد الله)، اللسانيات العربية والحاسوب (معالجة الصّرف آلياً)، منشورات مجلة الاتّصالات والعالم الرّققي، ع238، المملكة العربية السعودية، 2008م.

36- الموسى، (نهاد)، (1990م)، "اللغة العربية والحاسوب" لنبيل علي، المجلة العربية للعلوم الإنسانية، جامعة الكويت، ع38، المجلد العاشر.

37- النّاجم، (صلاح راشد)، مَنَهجُ حَاسُوبِيٍّ لِلتَّعَامُلِ مَعَ إِسْنَادِ الْأَفْعَالِ إِلَى الضَّمَائِرِ، دراسات الخليج والجزيرة العربية، ع114، الكويت، 2004، /40-90.

38- نهر، (هادي)، اللغة العربية والحاسوب، مجلّة التّواصل، جامعة عدن، ع4، 2000م، /72-76.

39- الوعر، (مازن)، اللسانيات والعلم والتكنولوجيا نحو تعريب موحّد للسانيات التطبيقية العربية وبرمجتها في الحاسبات الإلكترونية، اللسان العربي، الرّباط، ع22، 1983م.

د- أعمال المؤتمرات والندوات العلمية:

1- السّجّل العِلْمِيّ لندوة استخدام اللغة العربية في تقنية المعلومات، مكتبة الملك عبد العزيز العامة، الرياض - المملكة العربية السعودية، 1414هـ - 1993م.

- 2- اللغة العربية والتقنيات المعلوماتية المتقدمة، ضمن أعمال المناظرة المنظمة بمؤسسة الملك عبد العزيز آل سعود للدراسات الإسلامية والعلوم الإنسانية، الدار البيضاء، 1993م.
- 3- مؤتمر اللغة العربية أمام تحديات العولمة، معهد الدعوة الجامعي للدراسات الإسلامية، بيروت، 2002م.
- 4- مؤتمر اللغة العربية والهوية القومية، جامعة البترا الخاصة، الأردن، 21 آذار 2005م.
- 5- مؤتمر الهوية اللغوية والعولمة، جامعة البترا الأردنية، 2003م.

هـ- الأجنبية:

- 1- ABEILLÉ (Anne), Une grammaire électronique du français, (éd.) CNRS, Paris, 2002.
- 2- ABU AL-CHAY (N.), Un Système expert pour l'analyse et la production des verbes arabes dans une perspective d'Enseignement Assisté par Ordinateur, thèse en Sciences de l'information, Université Lyon 1, 1988.
- 3- ACHOUR (Hadhemi), Contribution à l'étude du problème de la voyellation automatique de l'arabe, thèse de doctorat, Université Paris VII, 1998.
- 4- BEESLEY, (Kenneth), « Finite-State morphological analysis and generation of Arabic at Xerox research: status and plans in 2001 ». In: <http://www.elsnet.org/acl2001-arabic.html>.(2001).
- 5- BEJAOUÏ (M.), Étude et réalisation d'un système expert appliqué à l'analyse morpho-syntaxique de phrases en langue arabe, thèse 3ème cycle, CERFIA, Toulouse, mai. 1985 .

- 6- BÉJOINT (Henri), THOIRON (Philippe), « Modèle rationnel, définition et dénomination », in Autour de la dénomination, Presses Universitaires de Lyon, 2000, pp. 5-19.
- 7- BEN HAMADOU (A.), SAIDI (L.), Analyse Morphosyntaxique de la Langue Arabe-Approche Système Expert, in 8ème Semaine, Tuniso-française en Informatique: Intelligence Artificielle, Faculté des Sciences, Tunis, Mai 1986 .
- 8- BERNARD (G.), « Simuler la compréhension de textes: représentations objets », in Actes du 10ème colloque d'Albi, Lecture et Informatique, document PDF, Toulouse-Le Mirail, 1990, pp. 1-9.
- 9- BERNARD (G.), & FEAT (J.), « Langage naturel et intelligence artificielle », in Stratégies théoriques, collection Histoire et Théories Linguistiques, 10/1, document PDF, presses Universitaires de Vincennes, 1988, pp. 1-12.
- 10- BOUCHÉ (R.), DICHY (J.), HASSOUN (M.), « Enseignement assisté par ordinateur de l'arabe: Simulation à l'aide d'un modèle linguistique la morphologie », in Simulation de modèles linguistiques et enseignement assisté par ordinateur de l'arabe travaux SAMIA I, 1989, pp. 41-62.
- 11- BOUILLON (Pierrette), Traitement automatique des langues naturelles, Aupefl-Uref Duculot, Paris, Bruxelles, 1998.
- 12- BRUSSET (Jean), ABDELGHANI (Saroh), « Création d'une base de données lexicale de l'arabe écrit utilisable par un système morpho-syntaxique », in Actes du IV Colloque International de linguistique, Linguistique arabe et informatique, Tunis, 9-12 Novembre, 1989,

- Publié dans: Université de Tunis, Centre d'Études et de Recherches Économiques et Sociales, Série linguistique n° 7, Tunis, Décembre 1989, pp. 9-28.
- 13- CELIO (Manuel), Concepts, termes et reformulations, in Travaux du CRTT, Presses Universitaires de Lyon, 2005, pp. 21-64.
- 14- CHOMSKY (Noam), (1955-1975), The logical structure of linguistic theory, Chicago: The University of Chicago Press.
- 15- CHOMSKY (N.) & MILLER (G.A.), L'analyse formelle des langues naturelles, Gauthier-Villars/Mouton, Paris, 1968.
- 16- COHEN (David), (1961), « Essai d'une analyse automatique de l'arabe », in T.A informations .
- 17- (1967) « La linguistique chamito-sémitique », in Revue de l'Enseignement supérieur, pp. 3-8 et (éd.) du CNRS, Paris, 1988.
- 18- DE SAUSSURE (Ferdinand), Cours de linguistique générale, Éditions Payot & Rivages, Paris, 1995.
- 19- DESCLÉS (Jean-Pierre), (1989), « La linguistique informatique et le programme SAMIA », in Simulation de modèles linguistiques et enseignement assisté par ordinateur de l'arabe travaux SAMIA I, pp. 12-25.
- 20- DICHY (Joseph), (1993), "Knowledge-system simulation and the computer-aided learning of Arabic verb-form synthesis and analysis", in Processing Arabic, Report n° 6/7, T.C.M.O., Université Catholique de Nimègue, pp. 67-84 et pp. 92-95.
- 21- DUBOIS (Jean), GIACOMO (Mathée), GUESPIN (Louis), MARCELLESI (Christiane), MARCELLESI (Jean-Baptiste), MEVEL (Jean-Pierre), Dictionnaire de linguistique et des sciences du

- langage, Larousse, Paris, 1994.
- 22- GAUBERT (Christian), (1995), « Analyse morphologique d'un texte arabe par ordinateur: méthode d'évaluation, résultats », in Annales islamologiques, tome XXIX, « s.l. », pp. 283-310.
- 23- (1998) « Étude des documents manuscrits arabes découverts lors des fouilles polonaises de Naqlun - Deir Al-Malak au Fayyoun, consistant en 50 parchemins et papiers d'une famille copte de l'époque fatimide, ainsi qu'un corpus de papiers arabes documentaires.
- 24- (décembre 2004), « Présentation du logiciel Sarfiyya » en langue arabe, Université Américaine de Caire (AUC), département d'arabe.
- 25- (mars 2006), « Comptabilité au monastère de Naqlun, exemples de comptes en chiffres epact », atelier du congrès ISAP (International Society for Arabic Papyrology), Alexandrie.
- 26- (mai 2007), « Archives arabes d'une famille copte du Fayyoun (monastère de Naqlun) à l'époque fatimide », École doctorale, IFAO.
- 27- (2009) en collaboration avec AUDEBERT (Claude) and JACCARINI (André), « Minimal Resources for Arabic Parsing: an Interactive Method for the Construction of Evolutive Automata », article en ligne, MEDAR 2009 conference, Le Caire.
- 28- (2010a), « Kawâkib, une application web pour le traitement automatique de textes arabes », in Annales islamologiques, N° 44, IFAO, pp. 53-60.
- 29- (2010b), en collaboration avec AUDEBERT (Claude) and JACCARINI (André), Coédition du dossier des Annales islamologiques, N° 44: «

- Linguistique arabe, Programme de traitement par automates de la langue arabe (Tala) », pp. 1-61.
- 30- (2010c), en collaboration avec JACCARINI (André) and AUDEBERT (Claude), « Structures and Procedures in Arabic Language ». <http://www.medar.info/report-ws-malta.pdf>. Proceedings of LREC, Valetta, Malta.
- 31- (25 27-juin 2010d), Avec HENEIN (N.). « Les bateaux du lac Manzala et quelques exemples d'embarcations nilotiques égyptiennes - technologie, lexicographie », Colloque La batellerie égyptienne, CEAlex, Alexandrie.
- 32- (2014b), en collaboration avec JACCARINI (André), « Le programme Mogador en linguistique formelle arabe et ses applications dans le domaine de la recherche et du filtrage sémantique », pp. 482-497, in Annales islamologiques, N° 47.
- 33- (2014c), en collaboration avec JACCARINI (André), « Démonstration de Kawâkib, outil permettant d'assurer le feedback entre grammaire et corpus arabe pour l'élaboration d'un modèle théorique », Conférence TALN.
- 34- (« s. d. » a), En collaboration avec JACCARINI (André) (CNRS) et AUDEBERT (Claude), (Université d'Aix-en-Provence), Mise au point de méthodes d'analyse surfacique et linguistiquement pertinentes de textes arabes. Développement de méthodes de désambiguïsation et d'outils graphiques de conception de grammaires.
- 35- (« s. d. » b), Application de cet analyseur à divers corpus arabes, en particulier celui des Waqfs étudiés à l'IFAO.
- 36- (« s. d. » c), Conception du site Web de l'IFAO et de bases de

données documentaires en collaboration avec les chercheurs et services concernés.

- 37- (« s. d. » d), Étude d'un analyseur morpho-syntaxique de l'arabe écrit classique et moderne, fondé sur une description régulière de la morphologie et de la syntaxe et s'appuyant notamment sur la théorie des automates (AFND).
- 38- (« s. d. »e), Logiciel expérimental Sarfiyya de traitement automatique de l'arabe.
- 39- HABAILI (Hussein), Phonologie générative et morphologie flexionnelle et dérivationnelle de l'arabe, thèse de doctorat de 3ème cycle, Université de Paris III, 1990.
- 40- HASSOUN (M.O.), DICHY (J.), AWAD (G.), « Vers un dictionnaire des bases conçu pour les besoins de la synthèse et de l'analyse morphologiques informatisées de l'arabe », in Simulation de modèles linguistiques et enseignement assisté par ordinateur de l'arabe travaux SAMIA I, 1989, pp. 79-91.
- 41- JACCARINI (André), (1989), "Automatic Generation of Unvocalized Arabic Recognition Programs: Theory and Applications", in Proceeding of the 2nd Conference en Arabic Computational Linguistic, Kuwait, pp. 66-87.
- 42- KATAMBA (F.), (1989), An Introduction to Phonology, Longman, New York.
- 43- KRAMSKY (J.), The phoneme: An Introduction to the History and Theories of A Concept, Munchen, Willhelm, Finkverlag, 1974.
- 44- LAZREK (A.), Système basé sur les connaissances pour l'analyse morpho-syntaxique optimisée de la langue arabe, thèse 3ème

- cycle, CERFIA, Toulouse, sept. 1986.
- 45- MAINGUENEAU (Dominique), Les termes clés de l'analyse du discours, Seuil, Paris, février 1996.
- 46- MANKAI (Chafia), MILI (Ali), « Un système de compréhension automatique de la langue arabe », in Actes du IV Colloque International de linguistique, Linguistique arabe et informatique, Tunis, 9-12 Novembre, 1989, Publié dans: Université de Tunis, Centre d'Études et de Recherches Économiques et Sociales, Série linguistique n° 7, Tunis, Décembre 1989, pp. 161-195.
- 47- MARCHAND (Pascal), L'analyse du discours assistée par ordinateur concepts, méthode, outils, Armand Colin, Paris, 1998 .
- 48- MEDUNA (Alexander), Formal Languages and Computation: Models and Their Applications, CRC Press, 2014.
- 49- MOUNIN (Georges), Dictionnaire de la linguistique, Paris, 4e édition « Quadrige » PUF, 2004.
- 50- NEYRENEUF (Michel), AL-HAKKAK (Ghalib), Grammaire active de l'arabe littéral, Librairie générale française, « s.l. », 1996.
- 51- PIERREL (Jean-Marie), «Bilan des grandes orientations de recherche en traitement automatique du langage parlé en France », in T.a.l., vol. 38 « Etat de l'art », n° 2, 1997, pp. 7-46.
- 52- PIOLLE (Jacques) et ROMAN (André), «Pour un traitement automatique des textes arabes: présentations d'un outil et leçon d'une expérience», in Cercle linguistique d'Aix-en-Provence Travaux 3 les relations syntaxiques, Université de Provence, 1985, pp. 209-251.
- 53- POTTIER (Bernard), Sur le système des prépositions, in Le Français

- Moderne, n° 29, pp. 1-6, « s.l. », 1961.
- 54- RADY (Mohamed), L'ambiguïté du langage naturel est-elle la source du non-déterminisme des procédures de traitement, thèse de l'Université de Paris XI (Jussieu), 1983.
- 55- RASTIER (François), « Le terme: entre ontologie et linguistique », C.N.R.S Texte publié dans La banque des mots, n° 7, 1995, pp. 35-65.
- 56- ROMAN (André), (1986), « Analyse morpho-syntaxique générale des constituants de la phrase arabe en vue d'une traduction en arabe assistée par ordinateur », in T. A. informations, vol. 27, n° 1, « s.l. », pp. 17-26.
- 57- STEFANINI (M.H.), Talisman: Une architecture Multi-agent pour l'analyse du français écrit, thèse à l'université pierre Mendès-France, Grenoble, 1993.
- 58- TAIBI MELLITI (Nacéra), Contribution à l'étude du traitement automatique des erreurs dans un texte écrit en arabe, Mémoire de magistère, ENSLSH, « s.l. », décembre 1997.
- 59- THALOUTH (B.), AL DANNAN (A.), A Comprehensive Arabic Morphological Analyser Gennrator, Arab School on Science and Technology. Syria, 1985.
- 60- TUERLINCKX (L.), « La lemmatisation de l'arabe non-classique », in Le poids des mots. Actes des 7ème Journées Internationales d'Analyse statistique des Données Textuelles, G. Purnelle, C. Fairon et A. Dister eds, Louvain-la-Neuve, II, 10-12 mars 2004, pp. 1069-1078.
- 61- WEHRLI (Éric), L'analyse syntaxique des langues naturelles,

problèmes et méthodes, Masson, Paris, 1997.

62- ZEMIRLI (Zouhir), KHABET (Salima), « TAGGAR: Un analyseur morphosyntaxique destiné à la synthèse vocale de textes arabes voyellés », Institut National d'Informatique, Alger-Algérie, JEP-TALN, Traitement Automatique de l'Arabe, Fès, 2004.

ز- مواقع وروابط إلكترونية:

- 1- <http://automatesarabes.net>
- 2- [http://www.ifao.egnet.net/ifao/recherche/chercheurs /](http://www.ifao.egnet.net/ifao/recherche/chercheurs/)
- 3- <http://www.ifao.egnet.net/kawakib> -
<http://www.ils.uw.edu.pl,Globe2002,Abstracts,part:14>
- 4- <http://www.m-a-arabia.com>
- 5- [http://www.medar.info/report-ws-malta.](http://www.medar.info/report-ws-malta)



تصنيف صوتيمات اللغة العربية باستخدام نموذج طرف إلى طرف في سبيل كشف أخطاء النطق

**Arabic Phonemes Classification Using End-
to-End Model in View of Mispronunciation
Detection**

إعداد

م. إلهام المفشي

ماجستير، نظم معطيات كبيرة،
المعهد العالي للعلوم التطبيقية والتكنولوجيا

د. أميمة الدكاك

رئيس قسم الاتصالات،
المعهد العالي للعلوم التطبيقية والتكنولوجيا

الملخص

أصبح تعلم اللغات بمساعدة الحاسوب Computer-Aided Language Learning (CALL) موضوعاً هاماً؛ بسبب الحاجة المتزايدة لتعلم لغات جديدة، ويعد كشف وتحديد النطق الخاطئ جزءاً مكملاً لأنظمة تعلم اللغات. لاقت هذه النظم اهتماماً كبيراً؛ لأنها تمكّن متعلّمي اللغة من تحسين قدراتهم اللغوية دون الحاجة للتواصل مع المختصين اللغويين بشكل مباشر، وذلك بالاستفادة من وسائل التعلّم الحديثة والتقنيات المتطورة. يهدف هذا البحث لتحسين تعرف الصوتيمات phonemes بالاعتماد على نموذج طرف إلى طرف، والذي يلغي الحاجة لعملية المحاذاة القسرية والتقطيع للأصوات يدوياً. قمنا باختبار أداء المنهجية المقترحة والمقارنة مع نتائج بحثنا السابق، وحصلنا على نتائج أفضل بالنسبة لتعرف الصوتيمات. إن نسبة التعرف الجيدة تمكّن من توظيف هذا النموذج في أنظمة تعليم اللغات، وتفتح المجال لاستخدامها في تطبيقات أخرى مثل الإملاء الآلي. ركزنا في البحث على اللغة العربية بهدف تقليص الهوة البحثية الموجودة بين التقانات اللغوية الداعمة للغة العربية ومثيلاتها في اللغات العالمية، والتي حققت تقدماً كبيراً في العديد من المجالات.

الكلمات المفتاحية: تعلّم اللغة بمساعدة الحاسوب، كشف أخطاء النطق، التعرف على الصوتيمات، شبكات التعلّم العميق، نماذج طرف إلى طرف.

Abstract

Computer-aided language learning (CALL) has become an important issue due to the increasing need to learn new languages. Mispronunciation detection and identification are considered important components of language learning systems. These systems have received great attention because they enable language learners to improve their language proficiency without the need for direct communication with language specialists, using modern learning methods and advanced technologies. This research aims to improve automatic phoneme recognition using an end-to-end model that eliminates the need for phoneme alignment and segmentation. The performance of the proposed methodology was tested and compared with our previous system results, and we obtained better phoneme recognition rates. These good recognition results enable the use of this model in language learning systems and allow its use in other applications, such as speech-to-text systems. The focus of the research was on Arabic, aiming to reduce the research gap between the linguistic technologies supporting Arabic and those supporting other international languages, which have achieved great progress in many fields.

Keywords: Computer-Assisted Language Learning CALL, Mispronunciation Detection ،Phoneme recognition, Deep neural networks, End-to-End Models.

1- المقدمة

حازت نظم تعلّم اللغات بمساعدة الحاسوب على قدر كبير من الاهتمام في السنوات الأخيرة؛ بسبب التقدم الملحوظ في تقنيات الذكاء الاصطناعي والتعلّم الآلي، حيث تهدف هذه الأنظمة إلى استخدام الموارد الحاسوبية والتقنيات الحديثة، في تسهيل عملية تعلّم اللغة، وحل مشكلة صعوبة التواصل بشكل دائم بين مدرس اللغة ومتعلّمها. وقد تم إجراء العديد من الأبحاث في هذا المجال من أجل العديد من اللغات (الإنكليزية (K. Li et al., 2017)، والصينية (W. Li et al., 2016)، والألمانية، واليابانية، والعربية (Nazir et al., 2019)...) تناولت بعض هذه الأبحاث موضوع كشف النطق الخاطئ وتصحيحه؛ بهدف تحسين النطق عند المتحدثين غير الأصليين للغة، وتم العمل على التحقق من النطق وفق عدة مستويات بدءًا من مستوى المتحدث؛ بهدف تقييم طلاقة الكلام عند النطق واستخدامه في اختبارات الكفاءة اللغوية المنطوقة، وانتهاءً بمستوى الصوتيم (phoneme) على مستوى كل وحدة صوتية في الكلام، إذ يقدم هذا المستوى معلومات أكثر دقة عن مكان ونوع الخطأ الذي ارتكبه المستخدم، مما يرفع من سوية عملية التعلّم. يهدف هذا البحث إلى دراسة إمكان تحسين تعرف الصوتيمات اعتمادًا على نموذج طرف إلى طرف، بحيث ندخل له الإشارة الكلامية مع النص ويعطي في خرجه الصوتيمات الموافقة. هذا النموذج أبسط من الطرق التقليدية لتعرف الصوت التي تعتمد على تدريب نماذج لغوية على كمية كبيرة من المعطيات جرى سابقًا محاذاة الصوت فيها مع النص الموافق لها؛ فإنه يصعب توفير هذه الموارد اللغوية لأي لغة، لأنها تتطلب وقتًا وجهدًا لتحقيق دقة تعرف جيدة للنماذج المدربة. نتناول في تنمة هذه الورقة في قسم 2- الدراسة

المرجعية أعمالاً مشابهة في كشف أخطاء النطق وتعرف الصوتيمات، ثم نشرح 3- منهجية البحث المتبعة وما فيها من مراحل، ونعرض في قسم 4- الجزء العملي والنتائج التجارب التي قمنا بها والنتائج التي حصلنا عليها، وأخيراً نعرض في قسم الخاتمة والتوصيات خلاصة البحث والآفاق المستقبلية له.

2- الدراسة المرجعية

نعرض في هذه الفقرة بعضاً من أهم الأعمال التي تناولت كشف أخطاء النطق، إضافة إلى الأعمال الخاصة بتعرف الكلام بشكل عام.

1-2 كشف أخطاء النطق

بالنسبة لكشف الأخطاء الصوتية في النطق تم العمل وفق عدة منهجيات، ويمكن تصنيف هذه المنهجيات حسب عدة عوامل (Chen & Li, 2017)؛ كاعتمادها على وجود نظام تعرف آلي على الكلام، والحاجة لوجود معرفة باللغة الأصلية، ولغة المتعلم، إضافة إلى قدرة المنهجية على كشف وجود خطأ في النطق وتحديد طبيعته ومكانه. وبحسب هذه العوامل، تقسم المنهجيات إلى منهجيات تعتمد على درجة الثقة Confidence score-based، ومنهجيات تعتمد على القواعد Rule-based، ومنهجيات تعتمد على المصنفات Classifier-based، ومنهجيات تعتمد على التعلم العميق Deep neural network-based. ونعتمد في بحثنا على التعلم العميق.

1-1-2 المنهجيات المعتمدة على درجة الثقة

يتم في هذه المنهجيات حساب درجة للثقة تعبر عن مدى قرب اللفظ المنطوق من اللفظ الصحيح، ثم تتم مقارنة النتيجة مع عتبة محددة لأخذ القرار فيما إذا كان النطق صحيحاً أم لا. تم اعتماد عدة مقاييس للمقارنة، ويُعدّ مقياس صحة

النطق (Goodness Of Pronunciation (GOP الأكثر استخدامًا (Witt & Young, 2000)، إذ يعتمد على حساب نسبة احتمال likelihood ratio توافق الصوت المنطوق مع الصوت الصحيح. وقد تم العمل على تحسين جودة النموذج الصوتي المستخدم في هذه المنهجية باعتماد شبكات التعلم العميق (Deep Belief Nets بدلاً من النموذج الغوسي (Gaussian mixture model (GMM (Hu et al., 2015). إن مقياس GOP حساس جدًا لجودة النموذج الصوتي (Witt & Young, 2000)، كما أنه يجري تحديد عتبة القرار اعتمادًا على بيانات التدريب فقط، مما يجعل من الصعب اعتماد هذه المنهجية للتعميم من أجل طيف واسع من أخطاء النطق. ومع أنّ هذه المنهجية يمكنها الكشف عن الأخطاء لكنها غير قادرة على تشخيص نوع الخطأ الذي ارتكبه المتعلم.

2-1-2 المنهجيات المعتمدة على القواعد

تتطلب هذه المنهجيات معرفة مسبقة بقواعد النطق الخاطئ، وتتميز بقدرتها على تحديد مكان ونوع الخطأ الذي قام به المتكلم، إلا أنّها محصورة بالأخطاء المعرفة مسبقًا ضمن القواعد، ولا يمكن للنظام التعرف على أخطاء جديدة غير معرفة مسبقًا. يتم بناء قواعد النطق الخاطئ يدويًا من قبل خبراء لغويين (Abdou et al., 2006; Harrison et al., 2008; Mao et al., 2018; Meng et al., 2007)، ويتم بناء نموذج صوتي للنطق الصحيح، ثم يجري تطبيق القواعد لتوليد نماذج للنطق الخاطئ. بعد ذلك، تجري المقارنة بين خرج النموذجين -عند اختبار النطق للمتعلّم- لتحديد صحة النطق. يتطلب بناء قواعد النطق وجود خبرة لغوية، إضافة إلى أن هذه القواعد قد لا تغطي كل الحالات؛ لذا تم تحسين هذه الطرق باستنتاج قواعد النطق الخاطئ آليًا باستخدام قواعد بيانات متعلّم اللغة L2 (Lo et al., 2010).

3-1-2 المنهجيات المعتمدة على المصنفات

تعتمد هذه المنهجيات على تصنيف الوحدات الصوتية، باستخدام مصنفات مختلفة: كمصنف آلة شعاع الدعم Support Vector Machine، أو أشجار التصنيف Decision Trees، أو الشبكات العصبونية Neural Networks. يتم استخلاص سمات مختلفة من الإشارات الصوتية في المرحلة الأولى لتكون دخلاً للمصنفات في المرحلة التالية. أثبتت مصنفات SVM فعاليتها في العديد من الأبحاث (Maqsood et al., 2016) (Wei et al., 2009). وفي (Hu et al., 2014) تم اقتراح استخدام مصنف a Neural Network (NN) based, Logistic Regression بالاعتماد على الشبكات العصبونية، ومقارنته مع مصنف SVM ومنهجية GOP حيث أعطى نتائج أفضل في الحالتين. كما تم استخدام شبكات التعلم العميق؛ لتحسين جودة النموذج الصوتي المستخدم في استخلاص سمات الوحدات الصوتية قبل تدريب المصنف. وبصورة عامة، فإن هذه المنهجية تعتبر أفضل من كل من المنهجيتين السابقتين، إلا أنها بالمقابل تحتاج إلى كمٍ كافٍ من بيانات التدريب.

4-1-2 المنهجيات المعتمدة على التعلم العميق

في هذه المنهجيات يتم استخدام نوع خاص من الشبكات العصبونية، وهي شبكات التعلم العميق، لنمذجة التجريدات العالية المستوى في البيانات، واستخلاص سمات تحسّن دقة كشف أخطاء النطق كما في الأبحاث (Algabri et al., 2021; Hussain et al., 2020; K. Li et al., 2017). في بعض الأبحاث جرى استخدام الشبكات العصبونية التلافيفية Deep Convolutional Neural Network (CNN)؛ لاستخلاص السمات الممثلة للإشارات الصوتية من الصور الطيفية (spectrograms) (Nazir et al., 2019). ويهدف تحسين كشف خطأ

النطق بالاستفادة من شبكات تعلم مدربة مسبقاً تم العمل وفق طريقة نقل التعلم transfer learning؛ إذ أثبتت التجارب أن هذه النماذج المدربة يمكن تعميمها على مهام عديدة، كاستخدام نموذج لغة مدرب على لغة معينة، كمرحلة أولية في التصنيف في لغة أخرى (Duan, Kawahara, Dantsuji, & Nanjo, 2017) وهذا يوفر الوقت ويساعد بشكل كبير في حال وجود موارد لغوية محدودة (Duan, Kawahara, Dantsuji, & Zhang, 2017). جرى أيضاً استخدام شبكات التعلم العميق في بعض الأبحاث لتحسين دقة النموذج الصوتي (Al-Marri et al., 2018; Shahin et al., 2014).

2-2 تعرف الصوتيات

جرى العمل في الأبحاث السابقة لتعرف الأصوات وفق عدة نماذج (Malik et al., 2021)، ويمكن تصنيف هذه النماذج إلى:

1-2-2 نماذج معتمدة على نموذج ماركوف الخفي HMM

يمثل النموذج الإحصائي الأكثر استخداماً، وتعتمد معظم نظم تعرف الكلام عليه (Deng & Yu, 2015). يقوم هذا النموذج بنمذجة الإشارة الصوتية المتغيرة مع الزمن، من خلال سلسلة من الحالات الخفية، كل حالة لها توزيع احتمالي مختلف لأشعة السمات، تعتمد على الحالات السابقة فقط. تم تعديل بنية هذا النموذج بدمجه مع نموذج الخليط الغاوسي الإحصائي GMM، الذي ينمذج سمات طيفية هي cepstral coefficients بشكل فعال (Thampi et al., 2014). بعدها جرى تحسين النموذج بإدخال شبكات التعليم العميق؛ ليصبح نموذجاً هجيناً (DNN-HMMs) (Al-Marri et al., 2018). ورغم أن هذه النماذج تحقق أداءً عالياً في الشروط الجيدة للمتكلم، إلا أنها تعاني من مشاكل عند وجود الضجيج وتداخلات الأصوات في البيئة (Hagen & Morris, 2005). إضافة إلى الحاجة

لتدريب مجموعة نماذج مختلفة (معجم النطق والنمذجة الصوتية ونمذجة اللغة) بشكل منفصل.

2-2-2 نماذج طرف إلى طرف End-to-End:

نذكر هنا، استخدام الشبكات العصبونية التكرارية RNN، والشبكات العصبونية التلافيفية CNN، ونماذج المرمز-مفكك الترميز، ونستخدم الأخير في بحثنا الحالي.

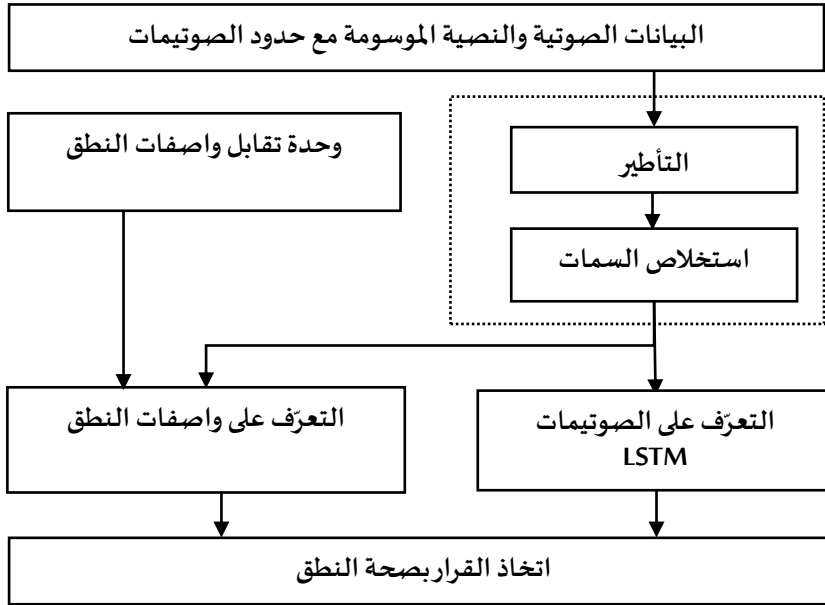
- الشبكات العصبونية التكرارية RNN تتميز هذه الشبكات بقدرتها على التكيف مع إشارات صوتية بأطوال مختلفة، إضافة إلى أنها لا تحتاج إلى قواميس لغوية، أو إلى جهد بشري لتنميط مجموعات المعطيات المستخدمة. تستخدم الشبكات العصبونية التكرارية الثنائية الاتجاه لتضمين السياقات السابقة والتالية للكلام. تقتصر قدرة ذاكرة RNNs على حفظ ما يقارب عشر خطوات زمنية، وذلك بسبب أن إشارة التغذية الخلفية، إما أن تتلاشى، أو أن تنفجر. وقد جرت معالجة هذه المشكلة من خلال الشبكات العصبونية المتكررة للذاكرة الطويلة القصيرة الأمد (LSTM-RNN)؛ إذ إن شبكات LSTM قادرة على تعلم أكثر من 1000 خطوة زمنية، وذلك بحسب تعقيد الشبكة المبنية. وقد أظهرت الأبحاث نتائج جيدة لتعرف الصوت باستخدام هذا الشبكات (Sak et al., 2014). وعلى الرغم من فعالية هذا النوع المرتفعة في التعرف؛ إلا أنه يتطلب وجود محاذاة بين الإشارة الصوتية المدخلة والنص الموافق لها.

- الشبكات العصبونية التلافيفية CNN: استخدمت هذه الشبكات في عدة مهام؛ منها، استخلاص السمات الممثلة للإشارات الصوتية من الصور الطيفية (spectrograms) (Nazir et al., 2019)؛ بهدف كشف خطأ النطق،

إضافة لاستخدام CNN في مهام تعرف الكلام (Abdel-hamid et al., 2014).
- نماذج مرمز- مفكك ترميز Encoder-Decoder: في هذا النموذج تجري ضمناً عملية المحاذاة القسرية بين تسلسلي الدخل والخرج، مما يحل مشكلة الحاجة لعملية تقطيع الصوت والمحاذاة. ويمكن لهذا النموذج التعامل مع أصوات بأطوال زمنية مختلفة. استخدم هذا النموذج في المرجع (Zhou et al., 2018)؛ لتعرف الكلام المتعدد اللغات بموارد لغوية منخفضة، وحصلوا على خطأ تعرف كلمة أقل من حالة النموذج المعتمد على شبكة LSTM. وفي عمل آخر (Hussein et al., 2022)، تم اقتراح نموذج طرف إلى طرف على اللغة العربية، بلهجات مختلفة، وأعطى نتائج أفضل من نتائج النماذج السابقة المعتمدة على نموذج ماركوف الخفي.

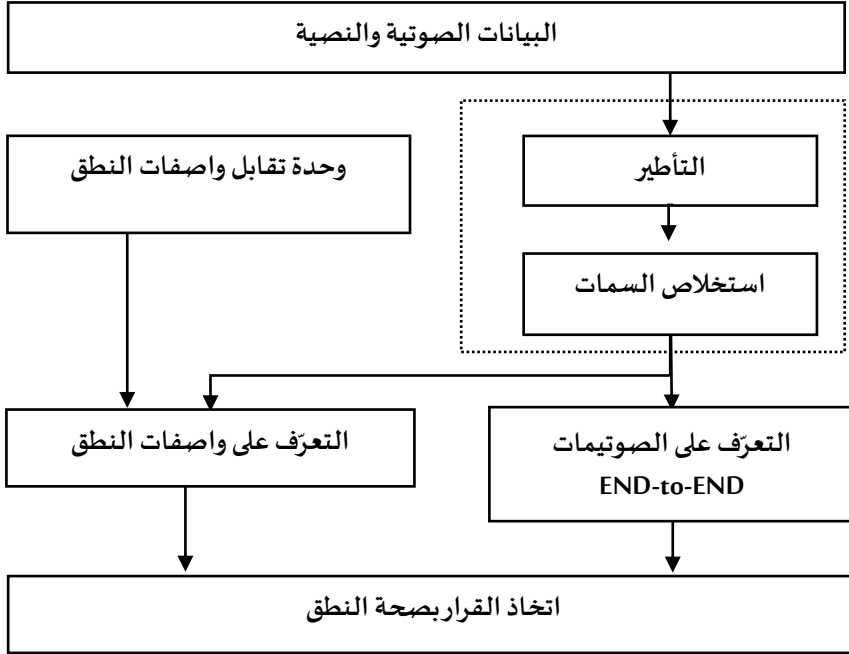
3- منهجية البحث المتبعة

كنا قد اقترحنا في بحث سابق، تحديد موضع الخطأ في النطق، باستخدام واصفات النطق الكلامية، إذ تتميز هذه الواصفات بفاعليتها في تطبيقات التعرف المنبئة تجاه تغيرات المتكلم ووجود ضجيج في البيئة المحيطة، وذلك مقارنة بالسمات الصوتية التقليدية. إضافة إلى ذلك، فهي تساعد بشكل فعال في تحديد خطأ النطق، حيث يتم اعتبار الصوت خاطئاً عند وجود تغير في أحد واصفاته الأساسية. وقد جرى العمل سابقاً على هذا الموضوع، بالاعتماد على بيانات صوتية مقسمة يدوياً على مستوى الصوتيم مع الحدود الزمنية الموافقة لها كما يوضح الشكل 1.



الشكل 1 نموذج العمل المقترح في عمل سابق

نركز في البحث الحالي على القيام بعملية التقسيم آلياً، بالاعتماد على نماذج End-to-End، إذ توفر هذه النماذج عملية المحاذاة والتقسيم آلياً، دون الحاجة إلى بيانات موسومة، والتي تحتاج إلى خبراء مختصين باللغة، إضافة إلى تكلفتها العالية من حيث الوقت والجهد الذي تستغرقه للقيام بهذه العملية. ويوضح الشكل 2 النموذج المقترح في البحث الحالي.



الشكل 2 نموذج العمل المقترح في العمل الحالي

أما فيما يتعلق بواصفات النطق، فهو نفسه المستخدم في عملنا السابق (E Almfashi; O Dakkak; N Ghneim, 2022). كنا قد جدولنا مخارج الحروف وصفاتها حسب الجداول (1-2). وصممنا مصنفات ثنائية، تأخذ الإشارة الكلامية لكل صوتيم، وتعطي المخرج وتوفر الصفات فيه. وكانت نسبة تعرف هذه الواصفات لا تقل عن 90% على مجموعة المعطيات التي جرت الدراسة عليها (Seddiq et al., 2016). KAPD.

الجدول 1 مخارج الحروف في اللغة العربية (Morsy et al., 2018)

الحروف	مخارج الحروف
ع ح	بلعومي (حلقي) Pharyngeal
ء ه ع ح غ خ	حلقي Pharynx
ء ه	حنجرية (أقصى الحلق) Glottal
ق	لهوية Uvular
ج ش	فوق اللثة Post-alveolar
ج ش ي	وسط اللسان Middle_tongue
ك ق	أقصى اللسان Deep_tongue
ض ل	حافة اللسان Tongue_border
ط د ت ص س ز ظ ذ ن ر	طرف اللسان Tongue_tip
ظ ث ذ	بين الأسنان (لثوي) Interdental

الجدول 2 صفات الحروف في اللغة العربية (Morsy et al., 2018)

صفات الحروف	الحروف
Whisper الهمس	فحثه شخص سكت
Whistle الصفير	ص س ز
Adhesion الإطباق	ص ض ط ظ
Affricates مركبة	ج
Deviate الانحراف	ل ر
Elevation الاستعلاء	خص ضغط قظ

الحروف	صفات الحروف
ص س ز ظ ذ ه ح ع غ خ ش ف ث	Fricatives احتكاكي
وي ض ف غ خ ح ه ص س ش ذ ظ ز ا وي المدية	Softness الرخاوة
ش	Spreading التفشي
أجد قط بكت	Strength الشدة
هاوي	Hiding الخفاء
لن عمر	Moderate المتوسط
ر	Repetition التكرار
ض	Prolongation الاستطالة
ا وي المدية	Vowels الصوتيات
-	Silence الصمت

1-3 تحضير المعطيات ومعالجتها

جرى العمل على معطيات من عدة مدونات سنتكلم عنها في فقرة المعطيات المستخدمة للتدريب. تتم في هذه المرحلة قراءة الملفات الصوتية للجمل المنطوقة، إضافة لقراءة البيانات النصية الموافقة. بعد ذلك، يجري استخلاص السمات الصوتية لكل إطار زمني وفق نافذة زمنية بطول 25ms (تتضمن وسطياً صوتياً واحداً) وانزياح زمني 10ms (الضرورة معالجة الإشارة نستخدم نوافذ على شكل جرس تخفف من أثر القطع عند الحواف)؛ حيث تعد الإشارة الكلامية مستقرة ضمن النافذة الزمنية السابقة.

جرى استخلاص معاملات طيفية معروفة في معالجة الإشارة الكلامية: MFCC (Mel Frequency Cepstral Coefficient) إضافة إلى معامل لوغاريتم الطاقة،

ومشتقات هذه المعاملات من الدرجة الأولى والثانية ($\Delta, \Delta\Delta$) لكل إطار صوتي. يتم حساب معاملات MFCC بالاعتماد على تحويل فورييه السريع (FFT) باستخدام مجموعة مرشحات تمرير حزمة كما في المعادلة (1) (Lu et al., 2003)

$$(1) \quad MFCC(t, k) = \sqrt{\frac{2}{N} \sum_{n=1}^N \log p_n \cos \left(k(n - 0.5) \frac{\pi}{N} \right)}$$

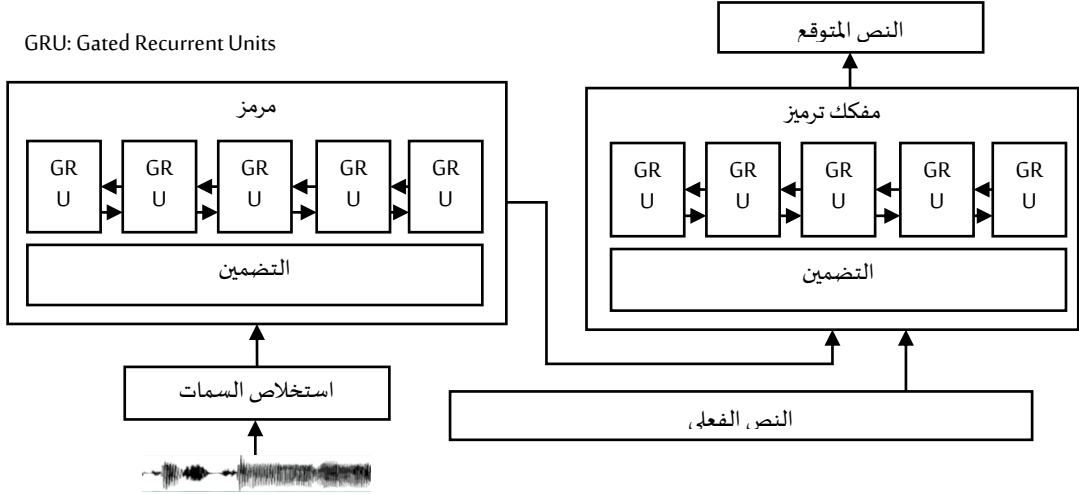
تمثل N عدد مرشحات تمرير الحزمة، p_n تمثل طاقة الخرج من المرشح رقم n في الزمن t ، $k = 1, 2, 3, \dots, L_k$ حيث L عدد معاملات MFCC. تتمثل إحدى المزايا المهمة لمعاملات MFCC أنها مميّزة للكلام المنطوق وللمتكلم، وهذا ما دفع لاستخدامها في أنظمة تعرّف الكلام ونعرف المتكلم.

بالنسبة لوحدة تقابل واصفات النطق، فهي مأخوذة من خبراء اللغة، ويجري فيها تحضير جداول تقابل كل واصفة مع الصوتيمات التابعة لها في قاعدة البيانات المدروسة وفق جدول تقابل محدد. نحصل في خرج هذه الوحدة على وسم لكل صوتيم يحدّد إذا كانت الواصفة موجودة أم لا في جملة الواصفات التي يتم تدريب النظام عليها.

2-3 تعرّف الصوتيمات

بعد استخلاص السمات الصوتية، يتم إدخالها لوحدة التعرّف على الصوتيمات، وتصنيفها باستخدام شبكة End-to-End لها بنية (مرمز- مفكك ترميز) تحتوي على شبكات عصبونية تكرارية، تتكون من وحدات بوابات متكررة (GRU's). تتميز هذه الوحدات بأنها تحتاج إلى موارد حسابية أقل من LSTM، كما تستخدم الشبكة ثنائية الاتجاه؛ لتضمين معلومات السياق بالاتجاهين السابق والتالي، وهذا يغني عن إضافة السمات الصوتية للإطارات المجاورة.

يتكون النموذج من مكونين: المرمز، ومفك الترميز. ويوضح الشكل 3 بنية الشبكة المستخدمة.



الشكل 3 بنية شبكة Encoder- Decoder المستخدمة

3-3 تعرف واصفات الكلام

تعد واصفات الكلام من السمات المميزة والهامة؛ كونها تمثل عملية نطق الكلام السليم بشكل غني ومفيد، ويعبر عنها بمخارج الحروف وصفاتها. ما يميز هذه الصفات أنها تساعد في كشف النطق الخاطئ، حيث يؤدي وجود تغير في إحدى هذه الواصفات إلى حدوث تغير في صوت الحرف؛ وبالتالي إلى خطأ في نطقه. كما أنها مشتركة بين اللغات، وهذا يُمكن من الاستفادة من قواعد بيانات صوتية من لغات مختلفة؛ لتدريب نموذج صوتي لهذه الواصفات. ويمكن الرجوع إلى بحثنا السابق لمزيد من التفاصيل عنها (E Almfashi; O Dakkak; N Ghneim, 2022).

4- الجزء العملي والنتائج

تم إجراء التجارب باستخدام المخدم السحابي Google Colab (Google Colaboratory, n.d.)؛ وذلك بهدف الاستفادة من قدرات التخزين والمعالجة الحسابية GPU التي تقدمها Google مجاناً، والتي تساعد في التنفيذ السريع عند تدريب شبكات التعلّم العميق. كما جرت بعض عمليات المعالجة والتجارب على حاسوب بمعالج intel core i7 وذاكرة 8GB. وقد استخدمت لغة بايثون لعمليات معالجة البيانات وتدريب الشبكات. كما جرت بعض عمليات المعالجة والتجارب على حاسوب بمعالج intel core i7 وذاكرة 8GB.

1-4 المعطيات المستخدمة للتدريب Speech Corpora

KACST Arabic Phonetic Database (KAPD)

هي مجموعة بيانات صوتية تم إنشاؤها من قبل مدينة الملك عبد العزيز للعلوم والتقنية King Abdul-Aziz City for Science and Technology (KACST) (Alghmadi, 2003). تحوي المجموعة تسجيلات صوتية لسبع متحدثين أصليين للغة العربية لما يقارب 1.2 ساعة، تضم 35 صوتيماً، ويوضح الجدول 3 هذه الصوتيمات. جرى العمل في هذا البحث على النسخة المطورة من القاعدة، حيث طُوّر جزء منها وقُطِع يدوياً على مستوى الصوتيم، بهدف استخدامها في تطبيقات التعلم الآلي والتنقيب عن المعطيات. تحوي المجموعة المطورة 4749 تسجيلاً صوتياً لسبعة متحدثين، لما يقارب 52 دقيقة. ويظهر البحث توازن مجموعة المعطيات نسبة لما يقارب 80% من الصوتيمات الموجودة فيها.

الجدول 3 صوتيمات مجموعة المعطيات KAPD

رقم الصوتيم	الصوتيم	رمز الصوتيم	الصوتيمات المقابلة
1	ب	bs10	bs10,bs20
2	ت	ts10	ts10,ts20
3	ث	vs10	vs10,vs20
4	ج	jb10	jb10,jb20
5	ح	hb10	hb10,hb20
6	خ	xs10	xs10,xs20
7	د	ds10	ds10,ds20
8	ذ	vb10	vb10,vb20
9	ر	rs10	rs10,rs20
10	ز	zs10	zs10,zs20
11	س	ss10	ss10,ss20
12	ش	js10	js10,js20
13	ص	sb10	sb10,sb20
14	ض	db10	db10,db20
15	ط	tb10	tb10,tb20
16	ظ	zb10	zb10,zb20
17	ع	cs10	cs10,cs20
18	غ	gs10	gs10,gs20
19	ف	fs10	fs10,fs20
20	ق	qs10	qs10,qs20
21	ك	ks10	ks10,ks20
22	ل	ls10	ls10,ls20,lb20

رقم الصوتيم	الصوتيم	رمز الصوتيم	الصوتيمات المقابلة
23	م	ms10	ms10,ms20
24	ن	ns10	ns10,ns20
25	هـ	hs10	hs10,hs20
26	و	ws10	ws10,ws20
27	ي	ys10	ys10,ys20
28	ء	hz10	hz10,hz20
29	َ	as10	as10
30	ا	as20	as20
31	ِ	is10	is10
32	ي	is20	is20
33	ُ	us10	us10
34	و	us20	us20
35	صمت	sil	sil

2-4 مقاييس التقييم

جرى تقييم تعرف الصوتيمات باستخدام معدل خطأ الصوتيم PER، حيث يتم حسابه كما في المعادلة التالية:

$$PER = \frac{S + D + I}{N} \times 100$$

يمثل N عدد الصوتيمات المرجعية، S عدد الصوتيمات التي تم إبدالها بصوتيمات أخرى، D عدد الصوتيمات المحذوفة، وتمثل I عدد الصوتيمات التي تم إدخالها.

بالنسبة لتعرف الواصفات، تم إجراء التقييم على مستويين: الأول: هو مستوى الإطار الزمني frame (النافذة الزمنية 25ms). والثاني: هو مستوى الصوتيم phoneme. يتم التقييم على مستوى الصوتيم بالاعتماد على نتيجة تصنيف إطاراته المكونة له، حيث يتم اعتماد صنف الإطارات الأغلبية majority. تم اعتماد مقاييس الضبط accuracy، والدقة precision، والاسترجاع recall، والتقييم الشامل f1، ومعدل خطأ الصوتيم (PER) Phoneme error rate لتقييم النتائج.

3-4 تعرف الصوتيمات

تم في هذه المرحلة بناء نموذج تعرف الصوتيمات باستخدام بنية (مرمز - مفكك ترميز) وتطبيقها على قاعدة بيانات KAPD

تعرف الصوتيمات في قاعدة بيانات KAPD

تنقسم المجموعة KAPD إلى مجموعة تدريب بنسبة 71.4% ومجموعة اختبار بنسبة 28.6%.

تم اختيار 10% من بيانات التدريب لعملية التحقق كما في (Algabri et al., 2021)، واستخلاص السمات الصوتية (13 معامل MFCC) إضافة إلى معامل لوغاريتم الطاقة، ومشتقات المعاملات من الدرجة الأولى والثانية (Δ , $\Delta\Delta$).

تحتوي مجموعة بيانات التدريب على 1360 جملة صوتية تتكون من 8134 صوتيماً.

حصلنا على نسبة خطأ في التعرف 6.97% اشتملت على أخطاء تبديل وحذف فقط بأغلبية أعلى لأخطاء التبديل حيث شكلت نسبة أخطاء الحذف 1%.

بتحليل الأخطاء وجدنا أن معظمها أخطاء تبديل بين حروف الجوف (الألف المدية والواو المدية والياء المدية)، والحركات المقابلة لها (الألف مع الفتحة، الياء

مع الكسرة...)، كون مخرج هذه الأصوات مشترك وتختلف بعضها عن بعض بزمان الصوت فقط، إضافة إلى أخطاء تبديل بين الحروف المشتركة في المخرج مثل (ز-س / ح-ه..)، إذ تشترك هذا الحروف في عدة صفات، وهذا يجعل صوتها متقاربًا، ويزيد من احتمال خطأ التعرف فيها.

الجدول 4 معدل خطأ الصوتيم في KAPD

المنهجية	PER (%)	Accuracy (%)
AFD-Obj (YOLOv3-tiny-1S)	10.84	89.35
PD-Obj (YOLOv3-tiny-2S)	5.63	94.56
Sparse coding	14.73	85.3
عملنا السابق LSTM	8.52	91.48
العمل الحالي BiGRU encoder + Attention decoder	6.97	93.02

يبين الجدول 4 معدل خطأ الصوتيم في KAPD مقارنة مع بعض الأعمال: بالمقارنة مع (Algabri et al., 2021)، نجد أن النظام PD-Obj الذي اعتمد تقنيات كشف الأغراض object detection للتعرف على الصوتيمات بالاعتماد على الصور الطيفية للإشارة الصوتية، حقق معدل خطأ أقل. وحصلنا في هذا العمل باستخدام بنية النموذج المقترح على أداء أفضل مقارنة مع العمل السابق باستخدام شبكة LSTM، وأداء أفضل من كل من نظامي AFD-Obj الذي يعتمد على واصفات النطق، للكشف عن الأغراض بهدف تصنيف الصوتيمات. ونظام SCPCS الذي يستخدم مصنف التمثيل المخلخل التقليدي، ويوظف سمات الترميز المخلخل في عملية التعرف. ما يميز نموذجنا، هو عدم الحاجة لحدود

الصوتيمات في عملية التدريب على عكس الأعمال السابقة، فهو قابل للتطبيق على معطيات بحجم أكبر إذ تتم فيه عملية التقسيم والمحاذاة للأصوات بشكل ضمني.

4-4 الخاتمة والتوصيات

عرضنا في البحث الحالي استخدام نموذج طرف إلى طرف End to End لتعرف الصوتيمات باستخدام آلية مرمر - مفكك ترميز Encoder-Decoder، يلغي هذا النموذج الحاجة لوجود بيانات صوتية مقطعة يدوياً، ويوفر عملية المحاذاة القسرية بين الحروف والصوتيمات الموافقة، فقد شكلت محدودية هذه الموارد اللغوية عائقاً في تطوير أنظمة مساعدة لتعليم اللغة العربية. حصلنا على نتائج تعرف أفضل بالنسبة، باستخدام قاعدة بيانات KAPD (نسبة الضبط في التعرف 93.02% Accuracy) مقارنة بالبحث السابق الذي قمنأ به والمعتمد على شبكة LSTM ومقارنة مع أعمال أخرى. كما حققنا في بحث سابق كشف الواصفات وفق شبكة تعلم عميق عن طريق التعلم المتعدد المهام، وأعطى النظام المقترح دقة أعلى من معظم الأنظمة التي جرت المقارنة معها. يمكن استخدام هذه الواصفات في كشف النطق الخاطئ بعد التعرف بشكل جيد على الصوتيمات، حيث يجري تحديد وجود خطأ النطق وتحديد نوعه عن طريق واصفات الكلام، وهذا يعطي متعلّم اللغة تقييماً أدق يفيد في تصحيح نطقه للأصوات، وهذا ما نسعى لتحقيقه في الأعمال المستقبلية إضافة لاختبار أنواع أخرى من نماذج طرف إلى طرف لتحسين تعرف الصوتيمات.

قائمة المراجع

- 1- Abdel-hamid, O. *et al.* (2014) 'Convolutional Neural Networks for Speech Recognition', 22(10), pp. 1533–1545.
- 2- Abdou, S. M. *et al.* (2006) 'Computer aided pronunciation learning system using speech recognition techniques', *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2(June 2014), pp. 849–852.
- 3- Al-Marri, M. *et al.* (2018) 'Computer Aided Qur'an Pronunciation using DNN', *Journal of Intelligent and Fuzzy Systems*. IOS Press, 34(5), pp. 3257–3271. doi: 10.3233/JIFS-169508.
- 4- Algabri, M. *et al.* (2021) 'Deep learning-based detection of articulatory features in arabic and english speech', *Sensors (Switzerland)*, 21(4), pp. 1–23. doi: 10.3390/s21041205.
- 5- Alghmadi, M., 2003, August. KACST arabic phonetic database. *In the Fifteenth International Congress of Phonetics Science, Barcelona* (pp. 3109-3112).
- 6- Chen, N. F. and Li, H. (2017) 'Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning', in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2016*. IEEE, pp. 1–7. doi: 10.1109/APSIPA.2016.7820782.
- 7- Deng, L. and Yu, D. (2015) *Automatic speech recognition, SpringerBriefs in Computer Science*. doi: 10.1007/978-981-15-0595-9_2.

- 8- Duan, R., Kawahara, T., Dantsuji, M. and Zhang, J. (2017) 'Articulatory modeling for pronunciation error detection without non-native training data based on DNN transfer learning', *IEICE Transactions on Information and Systems*, E100D(9), pp. 2174–2182. doi: 10.1587/transinf.2017EDP7019.
- 9- Duan, R., Kawahara, T., Dantsuji, M. and Nanjo, H. (2017) 'Transfer Learning based Non-native Acoustic Modeling for Pronunciation Error Detection', (August), pp. 42–46.
- 10- E Almfashi; O Dakkak; N Ghneim (2022) 'Classifying Arabic phonemes and extracting their attributes using deep learning in view of mispronunciation detection', *Damascus University Journal for Engineering Sciences*, 3, pp. 1–17.
- 11- *Google Colaboratory* (no date). Available at: <https://colab.research.google.com/>.
- 12- Hagen, A. and Morris, A. (2005) 'Recent advances in the multi-stream HMM/ANN hybrid approach to noise robust ASR', *Computer Speech & Language*, 19(1), pp. 3–30. doi: 10.1016/j.csl.2003.12.002.
- 13- Harrison, A. M. *et al.* (2008) 'Improving mispronunciation detection and diagnosis of learners' speech with context-sensitive phonological rules based on language transfer', in *Ninth Annual Conference of the International Speech Communication Association*.
- 14- Hu, W. *et al.* (2015) 'Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers', *Speech Communication*. Elsevier

- B.V., 67, pp. 154–166. doi: 10.1016/j.specom.2014.12.008.
- 15- Hu, W., Qian, Y. and Soong, F. K. (2013) 'A new DNN-based high quality pronunciation evaluation for computer-aided language learning (call)', *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (August), pp. 1886–1890.
- 16- Hu, W., Qian, Y. and Soong, F. K. (2014) 'A new Neural Network based logistic regression classifier for improving mispronunciation detection of L2 language learners', *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing, ISCSLP 2014*, pp. 245–249. doi: 10.1109/ISCSLP.2014.6936712.
- 17- Hussain, F. *et al.* (2020) 'Improving Mispronunciation Detection of Arabic Words for Non-Native Learners Using Deep Convolutional Neural Network Features', (June). doi: 10.3390/electronics9060963.
- 18- Hussein, A., Watanabe, S. and Ali, A. (2022) 'Arabic speech recognition by end-to-end, modular systems and human', *Computer Speech and Language*, 71, pp. 1–39. doi: 10.1016/j.csl.2021.101272.
- 19- Li, K., Qian, X. and Meng, H. (2017) 'Mispronunciation Detection and Diagnosis in L2 English Speech Using Multidistribution Deep Neural Networks', *IEEE/ACM Transactions on Audio Speech and Language Processing*. Institute of Electrical and Electronics Engineers Inc., 25(1), pp. 193–207. doi: 10.1109/TASLP.2016.2621675.

- 20-Li, W. *et al.* (2016) 'Improving non-native mispronunciation detection and enriching diagnostic feedback with DNN-based speech attribute modeling', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Institute of Electrical and Electronics Engineers Inc., 2016-May, pp. 6135–6139. doi: 10.1109/ICASSP.2016.7472856.
- 21-Lo, W.-K., Zhang, S. and Meng, H. (2010) 'Automatic derivation of phonological rules for mispronunciation detection in a computer-assisted pronunciation training system', in *Eleventh annual conference of the international speech communication association*.
- 22-Lu, L., Zhang, H. J. and Li, S. Z. (2003) 'Content-based audio classification and segmentation by using support vector machines', *Multimedia Systems*, 8(6), pp. 482–492. doi: 10.1007/s00530-002-0065-0.
- 23-Malik, M. *et al.* (2021) 'Automatic speech recognition: a survey', *Multimedia Tools and Applications*. Multimedia Tools and Applications, 80(6), pp. 9411–9457. doi: 10.1007/s11042-020-10073-7.
- 24-Mao, S. *et al.* (2018) 'Integrating Articulatory Features into Acoustic-Phonemic Model for Mispronunciation Detection and Diagnosis in L2 English Speech', *Proceedings - IEEE International Conference on Multimedia and Expo*, 2018-July. doi: 10.1109/ICME.2018.8486462.
- 25-Maqsood, M. *et al.* (2016) 'A Complete Mispronunciation Detection System for Arabic Phonemes using SVM', *IJCSNS International Journal of Computer Science and Network Security*,

- 16(3), p. 30. Available at:
<https://pdfs.semanticscholar.org/2002/54a67fd1b8efa91dda7399095509a6e23b6a.pdf>.
- 26- Meng, H. *et al.* (2007) 'Deriving salient learners' mispronunciations from cross-language phonological comparisons', in *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*, pp. 437–442.
- 27- Morsy, H. *et al.* (2018) 'Automatic Speech Attribute Detection of Arabic Language', *International Journal of Applied Engineering Research*, 13(8), pp. 5633–5639. Available at:
<http://www.ripublication.com>.
- 28- Nazir, F. *et al.* (2019) 'Mispronunciation detection using deep convolutional neural network features and transfer learning-based model for Arabic phonemes', *IEEE Access*. IEEE, 7, pp. 52589–52608. doi: 10.1109/ACCESS.2019.2912648.
- 29- Sak, H., Senior, A. and Beaufays, F. (2014) 'Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition'. Available at:
<https://arxiv.org/abs/1402.1128v1> (Accessed: 21 September 2021).
- 30- Seddiq, Y. *et al.* (2016) 'Reintroducing KAPD as a Dataset for Machine Learning and Data Mining Applications', in *2016 European Modelling Symposium (EMS)*. IEEE, pp. 70–74. doi: 10.1109/EMS.2016.022.
- 31- Shahin, M. *et al.* (2014) 'A comparison of GMM-HMM and DNN-HMM based pronunciation verification techniques for use in the

- assessment of childhood apraxia of speech', in *Fifteenth Annual Conference of the International Speech Communication Association*.
- 32- Thampi, S. M. *et al.* (2014) 'Phoneme-Based Recognizer to Assist Reading the Holy Quran', *Advances in Intelligent Systems and Computing*, 235, pp. V–VI. doi: 10.1007/978-3-319-01778-5.
- 33- Wei, S. *et al.* (2009) 'A new method for mispronunciation detection using Support Vector Machine based on Pronunciation Space Models', *Speech Communication*. Elsevier B.V., 51(10), pp. 896–905. doi: 10.1016/j.specom.2009.03.004.
- 34- Witt, S. M. and Young, S. J. (2000) 'Phone-level pronunciation scoring and assessment for interactive language learning', *Speech communication*. Elsevier, 30(2–3), pp. 95–108.
- 35- Zhou, S., Xu, S. and Xu, B. (2018) 'Multilingual End-to-End Speech Recognition with A Single Transformer on Low-Resource Languages', pp. 2–6. Available at: <http://arxiv.org/abs/1806.05059>.



"إطار عام مرجعي لتوظيف التقنية في تعليم اللغة العربية وتعلمها"

**A General Framework of Reference for
Engaging Technology in Arabic Language
Teaching and Learning**

الدكتور / عيسى صالح الحمادي
مدير المركز التربوي للغة العربية لدول الخليج

ملخص

تُعَدّ التغيرات المتسارعة الناجمة عن التقدم العلمي والتّقنيّ سمةً مميزةً لهذا العصر؛ لذا أصبح من الضروري مواكبة العملية التربويّة بمختلف جوانبها لهذه التغيرات، ومواجهة المشكلات والتحديات التي قد تنجم عنها، متمثلةً في: كثرة المعلومات، وزيادة أعداد الطلاب، ونقص المعلمين المؤهلين، وبُعد المسافات، وازدياد الحاجة إلى التعليم، وقد أدت هذه التغيرات إلى ظهور أنماط عديدة للتعليم الفرديّ، أو الذاتيّ، ولمواجهة هذه التغيرات؛ ظهرت مفاهيم: التعليم المبرمج، والتعليم باستخدام الحاسوب، والتعليم عن بُعد.

إن توظيف التقنيات في التعليم أضحى خيارًا استراتيجيًا لكل دولة تُقدّر -بوعي- ما للقطاع التربويّ من أهميّة بالغة في ازدهار المجتمع وتطوره التنموي، وبدلاً لا يمكن الاستغناء عنه ليس في الظروف الاستثنائية، لكن -أيضاً- لبناء جيل جديد قادر على الاستفادة من معطيات الثورة الصناعيّة الرابعة، يمتلك من المعارف والمهارات التي تُمكنه من الانخراط في وظائف الغد. والتعليم إذا استثمر التقنيات الحديثة، ووظّفها في مختلف جوانب العملية التعليميّة تخطيطاً، وإدارة، وتدريباً، وتقويماً؛ يمكن أن يحقق أهدافه إذا توافرت الشروط لإنجاحه؛ حتى يمكن صياغة سياسات علميّة مناسبة للتعامل مع المستقبل.

والبحث الحالي يُركّز على بناء إطار مرجعيّ لتوظيف التقنيّة في تعليم اللّغة العربيّة وتعلّمها، ويشمل إعداد منظومات معايير: البيئة الرّقمية لتعليم اللّغة العربيّة، ومعلم اللّغة العربيّة الرّقميّ، ومتعلم اللّغة العربيّة الرّقميّ، ومعايير تدريس فنون اللّغة الرّقمية، والدليل المقترح لمُعَلّمي اللّغة العربيّة ومُتعلّمها؛

لتوظيف التقنية في تعليم اللغة العربية وتعلمها، في ضوء الإطار العام المرجعي، ومعايير التقويم الإلكتروني في اللغة العربية.

الكلمات المفتاحية: الإطار المرجعي. توظيف التقنية. اللغة الرقمية. البيئة الرقمية. البيئة الرقمية. فنون اللغة الرقمية. الكفاءة الرقمية. العصر الرقمي.

Abstract

The rapid changes resulting from scientific and technical progress are a distinctive feature of this era. Therefore, it has become necessary to keep pace with the educational process in its various aspects in light of these changes. To face the problems and challenges that may result from it, represented by: the abundance of information, the increase in the number of students, the lack of qualified teachers, the distance, and the increased need for education. These changes have led to the emergence of many patterns of individual or self-education, and to confront these changes, three concepts emerged: programmed education, computer education, and distance education. Employing technology in education has become a strategic choice for every country that consciously appreciates the importance of the educational sector in the prosperity and development of society and an indispensable alternative not only in exceptional circumstances, but also - to build a new generation capable of benefiting from the data of the industrial revolution. Fourth, he possesses the knowledge and skills that will enable him to engage in the jobs of tomorrow. Education, if it invests in modern technologies and employs them in various aspects of the learning process in planning, management, teaching, and evaluation, can achieve its goals if the conditions are met for its success so that appropriate scientific policies can be formulated to deal with the future. The current research project focuses on building a frame of reference for employing technology in teaching and learning the Arabic language, and it includes the preparation of standards systems:

the digital environment for teaching Arabic, the digital Arabic language teacher, the digital Arabic language learner, the grammatical standards for teaching Arabic, and the Linguistic Art Guide, to employ technology in teaching and learning the Arabic language, in light of the general frame of reference and the standards of electronic assessment in the Arabic language.

Keywords: reference frame. Technology employment. The digital language. The digital environment. Numerical Language Arts. Numerical efficiency. The digital age

مقدمة

شهد القرن العشرون ثورة علمية ومعرفية هائلة، شملت تغيرات تكنولوجية، وتحولات رقمية، ونظريات تعليمية، في مختلف ميادين العلوم الإنسانية والطبيعية والتطبيقية، كما شهد مولد ميادين علمية جديدة لم تكن معروفة من قبل، ولم تكن التربية بمنأى عن هذا التطور، بل كانت من أكثر الميادين تأثراً به، وتأثيراً فيه، إذ ظهرت العديد من النظريات والاتجاهات التربوية، التي سعت إلى استيعاب الحجم الهائل من العلوم، والحفاظ على هذا التراث الإنساني من خلال نقله للأجيال المتتالية؛ لتطويره من ناحية، ووضعه موضع التطبيق من ناحية أخرى، وعملت على تحقيق ذلك من خلال الاستفادة من كل ما أنتجه العلم من نظريات وتطبيقات. وهي بدورها أثرت - بشكل واضح - في حياة المتعلمين، واستراتيجيات تعليمهم، وطرق حصولهم على المعلومات داخل المدرسة وخارجها.

وحرص التربويون على توظيف تقنيات الاتصال المختلفة التي شهدها العصر الحاضر، وتسابقت المدارس والجامعات في مختلف الدول من أجل دمج التقنيات في المناهج الدراسية، ومحاولة محو الأمية الرقمية للمعلمين والمتعلمين، وإكسابهم القدرة على التعامل الفعال مع الأجهزة الإلكترونية، والشبكات الدولية، والحصول على المعلومات من مصادر متعددة متنوعة، وبطرق سريعة، والقدرة على إجراء معالجات عقلية معرفية للمعلومات، وفهمها، واستيعابها، وتحليلها، ونقدها، بصورة تعكس ضرورة اكتساب مهارات التقنية اللازمة للتنقل بين المواقع ومصادر المعلومات في عالم رقمي متزايد؛ من أجل القراءة والاطلاع، والدراسة والتحصيل، والبحث والتقصي، والمتعة والتسلية.

إن استخدام التقنيات في التعلم يمكن أن يسهم في تحسين التعلم اللغوي؛ لضمان حصول جميع المتعلمين على خبرات تعليمية عالية الجودة. "لأن التقنيات يُنظر إليها الآن على أنها مكون أساسي من مكونات تعلم المتعلمين. مع التقدم في التقنيات وتطبيقاتها عبر الفصول الدراسية تبين أن المتعلمين:

- يحرزون تقدماً في الاستفادة من التعلم الرقمي بطرق جديدة.
- الاستراتيجيات الرئيسة التي شجعها هذا التعلم الرقمي هي التعلم المخصص الذي يمنحهم مزيداً من الخيارات بشأن، ماذا، وكيف يتعلمون، وبأية وتيرة، وإعدادهم لتنظيم تعلمهم، وممارستهم لمهارات الأداء اللغوي، وتوجيهه لبقية حياتهم.
- التعلم الرقمي يسمح للعديد من المدارس بتحسين قدرات تقييم الأداء اللغوي للمتعلمين، وتكييف التدريس بشكل أفضل مع احتياجات الطلاب اللغوية المتنوعة.

الهدف النهائي للتعلم الرقمي يتمثل في كثير من الأحيان في السماح للمتعلمين بأداء دور نشط في بناء خبراتهم التعليمية؛ لذلك يحتاج المعلمون إلى تشجيع الاستخدام الفعال للتقنيات على مدار اليوم داخل المدرسة، وخارجها.⁽¹⁾

(U.S. Department of Education, 2017, p7)

(1) U.S. Department of Education. (2017). Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update, Office of Educational Technology.

مشكلة البحث:

نادى مؤتمر اللغة العربية الدولي الاستثنائي عن بُعد بالشارقة⁽¹⁾ بعنوان: "التعليم عن بُعد في تدريس اللغة العربية: (الواقع، والمتطلبات، والآفاق)" (25) إلى (29) من أكتوبر (2020م) (الخليج، (2020))، الذي نظّمه المركز التربوي للغة العربية لدول الخليج، بضرورة الاستفادة من التقنيات الذكية في تعليم اللغة العربية للنشء، عبر منصة إلكترونية شاملة تتضمن أنشطة لغوية متنوعة، تُسهم في تنمية المهارات اللغوية ذاتيًا، وضرورة توظيف شبكات التواصل الاجتماعي في تجسير الفجوة بين الفصحى والعامية، ومعالجة الأخطاء اللغوية الشائعة، والارتقاء بلغة الحوار والتواصل الاجتماعي.

وبناءً على ما سبق يمكن القول: إن توظيف التقنيات في تعليم اللغة العربية وتعلّمها عملية تضم أطرافًا متعددة؛ منها: معلم اللغة العربية، والمتعلم، والبيئات الرقمية بما فيها من أجهزة الكمبيوتر، وشبكات الإنترنت، وهواتف ذكية، وكل الإمكانيات المادية والبشرية، والكوادر الفنية المتخصصة، ودعم الإدارة المدرسية، وغير ذلك من العوامل، وينبغي هنا صياغة منظومة معايير خاصة بكلٍّ من مُعلّم اللغة العربية، ومُتعلّمها، والبيئات الرقمية، وفنون اللغة الرقمية، والتقويم الإلكتروني، وبناء دليل يسترشد به المعلمون والمتعلمون في توظيف التقنية في تعليم اللغة العربية وتعلّمها في مراحل التعليم العام.

استنادًا إلى نتائج الدراسات السابقة مثل دراسة الشمراني 2019⁽²⁾، ودراسة

(1) المركز التربوي للغة العربية لدول الخليج، مؤتمر اللغة العربية الدولي الاستثنائي عن بُعد بعنوان: التعليم عن بُعد في تدريس اللغة العربية: (الواقع، والمتطلبات، والآفاق) المنعقد بالمركز التربوي للغة العربية بالشارقة، 2020.

(2) الشمراني، عليه أحمد يحيى آل حمود، أثر توظيف التعلم الرقمي على جودة العملية التعليمية

السيد⁽¹⁾ 2021، السعدون، والبازعي⁽²⁾ 2021، دراسة زيناتي⁽³⁾ 2021 في مجالي معوّقات توظيف التقنية في تعليم اللغة العربية وتعلمها، وما يمكن أن تؤديه بيئات التعلم الرقمية في تعزيز تعليمها وتعلمها، تتصدى الدراسة الحالية للمشكلة من خلال الإجابة عن السؤال الرئيس الآتي:

ما الإطار العام المرجعي لتوظيف التقنية في تعليم اللغة العربية وتعلمها؟ وما يتفرّع عنه من أسئلة:

1. ما المعوّقات التي تحوّل دون توظيف التقنية في تعليم اللغة العربية وتعلمها؟
2. ما معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها بمراحل التعليم العام؟
3. ما معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية في مراحل التعليم العام بدُول الخليج العربي؟
4. ما معايير تعليم فنون اللغة الرقمية (الاستماع الرقمي- السرد الرقمي- القراءة الرقمية - الكتابة الرقمية) في مراحل التعليم العام؟

==

وتحسين مخرجاتها، المجلة العربية للعلوم التربوية والنفسية: المؤسسة العربية للتربية والعلوم والآداب، 2019، ص ص 145-169.

(1) السيد، محمود مصطفى محمود، توظيف الفصول الافتراضية في تدريس اللغة العربية من بعد جوجل كلاس روم نموذجاً، الشارقة، المركز التربوي للغة العربية لدول الخليج مقدم الملتقى أفضل الممارسات والتجارب في مجال توظيف التقنيات في مارس 2021 تدريس اللغة العربية عن بعد، ص1-14

(2) السعدون، نجلاء مهنا، والبازعي حصة حمود، أدوار معلم المرحلة الثانوية في ضوء متطلبات المدرسة الافتراضية بمحافظة حفر الباطن، مجلة الفنون والأدب وعلوم الإنسانيات والاجتماع، (2021).

(3) زيناتي، محمد بشر أيمن (2021). توظيف التقنيات الحديثة في تعليم اللغة العربية، جامعة إسطنبول، مجلة العلوم الإنسانية والطبيعية، 2(2)، www.hnjournal.net

5. ما معايير التقويم الإلكتروني في اللغة العربية بمراحل التعليم العام؟
6. ما معايير الكفاءة الرقمية اللازمة لمتعلمي اللغة العربية في مراحل التعليم العام؟
7. ما الدليل المقترح لمُعَلِّمي اللغة العربية ومُتعلِّمها لتوظيف التقنية في تعليم اللغة العربية وتعلُّمها، في ضوء الإطار العام المرجعي؟

مصطلحات البحث

الإطار المرجعي:

منظومة معايير ومؤشرات خاصة بالكفاءة الرقمية لمُعَلِّمي اللغة العربية ومُتعلِّمها، وفنون اللغة الرقمية: (الاستماع الرقمي- السرد الرقمي- القراءة الرقمية - الكتابة الرقمية) في مراحل التعليم العام، والمواصفات القياسية للبيئة الرقمية، فضلاً عن معايير التقويم الإلكتروني في اللغة العربية بمراحل التعليم العام، ودليل مرجعي لمُعَلِّمي اللغة العربية ومُتعلِّمها يتضمن نماذج واضحة للمعارف والمهارات التي تُمكن مُعَلِّمي اللغة العربية ومُتعلِّمها من التعامل الذي مع التقنيات الحديثة في عمليتي تعليم اللغة العربية وتعلُّمها.

التقنية:

يرتبط مصطلح التقنية بمصطلح آخر؛ وهو التكنولوجيا، ويتكوّن مصطلح التكنولوجيا (Technology) من مقطعين هما: (Techno) والآخر (logy)، والمقطع الأول يوناني؛ يعني حرفه، أو صنعة، أو فناً، أو مهارة، أمّا المقطع الثاني فمعناه علم أو دراسة، وهناك من يرى أن المقطع الأول مأخوذ من الكلمة الإنجليزية (Technique) ومعناه التقنية أو الأداء التطبيقي، والصلة واضحة بين الأصلين؛ فالحرفة أو الصنعة ما هي إلا تقنية أو أداء تطبيقي؛ أي أن التكنولوجيا تعني علم الأداء، أو

علم التطبيق، أو علم التقنية.⁽¹⁾

ويمكن تعريف التقنية إجرائيًا بأنها: جميع التطبيقات الذكية التي تستهدف تيسير عمليّتي التعليم والتعلم؛ بحيث لا تقف عند حدود توظيف الأجهزة والآلات، واستخدام الحاسوب، والإنترنت، والمحمول، وآلات المسح الرقعي، والبريد الإلكتروني في عمليّتي التعليم والتعلم، بل تشمل -أيضًا- استعمال مبادئ علمية في تنظيم المحتوى، وطرائق في التفكير، ومنهج في العمل، وأسلوب في حل المشكلات من أجل تحقيق أهداف تعليمية منشودة.

حدود البحث:

يقتصر البحث على بناء منظومة معايير:

1. البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها بمراحل التعليم العام في الصفوف من (4-1)، ومن (9-5)، ومن (12-10).
2. الكفاءة الرقمية لمُعَلِّمي اللغة العربية في مراحل التعليم العام بدُول الخليج العربي في الصفوف من (4-1)، ومن (9-5)، ومن (12-10).
3. الكفاءة الرقمية اللازمة لمتعلّمي اللغة العربية في مراحل التعليم العام في الصفوف من (4-1)، ومن (9-5)، ومن (12-10).
4. تعليم فنون اللغة الرقمية: (الاستماع الرقعي، والسرد الرقعي، والقراءة الرقمية، والكتابة الرقمية) في مراحل التعليم العام.
5. التقويم الإلكتروني في اللغة العربية بمراحل التعليم العام.

(1) نصر، معاطي، سليمان، محمود جلال، تكنولوجيا تعليم اللغة العربية وتعلمها للعاديين وغير العاديين. دميّاط: مكتبة نانسى، 2019.

أهداف البحث:

1. رصد المعوقات والتحديات التي تواجه استعمال تقنيات العصر الرقمي في تعليم اللغة العربية وتعلمها.
2. تفسير الأسباب التي أدت إلى القصور في توظيف التقنية في تعليم اللغة العربية، وتعلمها.
3. تحليل المعوقات التي تسببت في توظيف التقنية في تعليم اللغة العربية، وتعلمها.
4. تحليل الاحتياجات التقنية للبيئة الرقمية، والمعلم والمتعلم في العصر الرقمي، وتدرّس اللغة العربية، وتقويم تعليمها في العصر الرقمي.
5. بناء دليل مرجعي لتدرّس اللغة العربية في العصر الرقمي.

أهمية البحث:

أ- بالنسبة لميدان تعليم اللغة العربية:

نظرياً:

- رصد التجارب العالمية والعربية في مجال استخدام تقنيات العصر الرقمي، في تعليم اللغات وتعلمها.
- تحليل منطلقات هذه التجارب وأهدافها، وإجراءاتها، وما أسفرت عنه من نتائج.

تطبيقياً:

- بناء منظومة معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية ومُتعلِّميها، والبيئات الرقمية، والتقويم الإلكتروني، وبناء معايير تعليم فنون اللغة الرقمية، وإعداد دليل استراتيجيات التدريس الإلكتروني للغة العربية في العصر الرقمي؛ للاستفادة منها في التطوير الشامل لتعليم اللغة العربية.

ب- بالنسبة لمُعَلِّمي اللغة العربيّة:

نظريًا:

- تزويد المعلم بمعايير الكفاءة الرّقميّة اللّغويّة.

تطبيقًا:

- الاستفادة من تقنيات العصر الرّقميّ في تعليم اللّغة العربيّة، وتعلّمها.
- تمكين المعلم من استراتيجيّات التدريس الإلكترونيّ للّغة العربيّة في العصر الرّقميّ.

ج. بالنسبة لمتعلّمي اللغة العربيّة:

نظريًا:

- إكساب المتعلّم معايير الكفاءة الرّقميّة لمتعلّم اللغة العربيّة.
- تأهيل المتعلّم ليكون مواطنًا رقميًا يمكنه العيش في بيئة رقميّة.
- إعداد المعلم لتعلّم اللّغة العربيّة في بيئة رقميّة، وممارستها رقميًا.

تطبيقًا:

- اكتساب مهارات الأداء اللّغويّ من خلال تطبيقات العصر الرّقميّ.

منهج البحث

المنهج الوصفيّ التحليلي (Descriptive Analytical Method):

- يستخدم البحث المنهج الوصفيّ التحليلي وفقًا لطبيعته، وتحقيقًا لأهدافه:
- للتعرف على الخبرات العالميّة والممارّسات الدوليّة، ودراستها وتحليلها في استخدام تقنيات العصر الرّقميّ في تعليم اللغات وتعلّمها.
- ورصد السياسات والممارّسات التي تُتّبع في عدد من الدول المتقدّمة، والتي

- ثبتت فاعليتها ونجاحها؛ للإفادة منها بما يتلاءم مع سياق البيئة العربية الإسلامية في تطوير منظومة الكفاءة الرقمية لبيئات التعلم، والمعلم، والمتعلم.
- وتحليل البحوث والدراسات الأكاديمية والوثائق ذات العلاقة بتعليم اللغات في العصر الرقمي.
- وتحليل الدراسات والتجارب؛ بهدف الوصول إلى بناء الإطار المرجعي، وإعداد دليل استراتيجيات التدريس، واستراتيجيات التقويم التكويني في اللغة العربية التي تحقق أهداف المشروع البحثي وغاياته.

بغرض:

- تحديد أسس اشتقاق المعايير (البيئات الرقمية - المعلم الرقمي - المتعلم الرقمي - تعليم فنون اللغة الرقمية - التقويم الإلكتروني)، وصياغتها، وما ينبثق عنها من مؤشرات.
- مراجعة الاتجاهات المعاصرة في مجال معايير تعليم اللغات في العصر الرقمي.
- تفسير البيانات، وتحليلها، وتصنيفها.
- تبويب البيانات وفق أطر، وتصورات منطقية:
- الكفاءة الرقمية لمُعَلِّمي اللغة العربية، ومُتعلِّمها، والبيئات الرقمية، والتقويم الإلكتروني بما يتسق مع الاتجاهات العالمية، وواقع المجتمعات العربية، وطبيعة اللغة العربية.
- بناء معايير تعليم فنون اللغة الرقمية بما يتناسب مع مهارات اللغة، ووظائفها، ومجالات التواصل اللغوي.
- استراتيجيات التدريس الإلكتروني للغة العربية في العصر الرقمي بما يحقق متطلبات ممارسة مهارات الأداء اللغوي لمتعلم اللغة العربية.

الإطار النظري

مفهوم التقنية، وأهمية توظيفها في تعليم اللغة العربية وتعلمها في العصر الرقمي

يؤصل الإطار النظري للمفاهيم التي يتضمنها متغيرات البحث من خلال المحورين التاليين:

المحور الأول: التقنية وتعليم اللغة العربية.

أولاً: مفهوم التقنية:

قدم جاك، وهيجنس Jack⁽¹⁾ & Higgins عرضاً لتعريفات تقنيات التعليم على النحو التالي:

- مجال يركز على التطبيق النظري والعملية للتقنيات في البيئات التعليمية.
- تحسين التدريس والتعلم من خلال الاستخدام الأمثل للموارد التكنولوجية داخل الفصل الدراسي.
- دراسة أدوار وتطبيقات التقنيات التناظرية والرقمية فيما يتعلق بالتعلم البشري والممارسات التعليمية.
- تكنولوجيا التعليم هي الاستخدام الفعال للأدوات التكنولوجية في التعلم.
- مدخل متعدد التخصصات مستنير بالنظريات في مجالات التعليم، يدرس أفضل الممارسات لتصميم البيئة التعليمية القائمة على التكنولوجيا والمواد التعليمية؛ وتطويرها، وتنفيذها، وتقييمها بغرض تعزيز التعلم.

(1) Jack, Christine & Higgins, Steven: E What is educational technology and how is it being used to support teaching and learning in the early years?, August 2018, International Journal of Early Years Education 27(4):1-16

- مجموعة معقدة من العمليات المتكاملة التي تشمل الأفراد والإجراءات والاستراتيجيات والأفكار والأجهزة، لتحليل تحديات بيئة التعلم والتعلم.
- ثانيا: مبررات دمج التقنية في تعليم اللغة العربية وتعلمها:
- يصبح التعليم اللغوي عملية إنتاجية، وأكثر واقعية، وأكثر سرعة.
- إتاحة التعلم اللغوي في أي زمان وأي مكان.
- إثراء التعلم اللغوي بمصادر ووسائل متعددة ومتنوعة.
- مراعاة الفروق الفردية، والاحتياجات اللغوية المتنوعة بين الطلاب.
- استيعاب أعداد كبيرة من الطلاب في عملية التعلم اللغوي، وممارسة المهارات اللغوية.⁽¹⁾

ثالثا: تقنيات تعليم اللغة العربية وتعلمها ومهارات القرن الحادي والعشرين.

تتمثل مهارات القرن الحادي والعشرين في القدرة على استخدام التقنيات الرقمية وأدوات الاتصال والشبكات وصولاً إلى المعلومات وإدارتها وتقويمها وإنتاجها، وتشمل⁽²⁾:

- الثقافات: البصرية، والاقتصادية، والعلمية، وإنتاجها، كما تشمل: الثقافة الأساسية والمعلوماتية، والوعي الكوني، وفهم الثقافات المتعددة.
- مهارات التفكير الإبداعي، وتشمل: القدرة على التكيف وإدارة التعقيد،

(1) Hutchison, Amy C. & Woodward, Lindsay: Examining the Technology Integration Planning Cycle Model of Professional Development to Support Teachers' Instructional Practices. Teachers College Record, 120 (10).

(2) شلي، نوال محمد، "إطار مقترح لدمج مهارات القرن الحادي والعشرين في مناهج العلوم بالتعليم الأساسي في مصر"، المجلة التربوية الدولية المتخصصة، 2014.

والتوجيه الذاتي، وحب الاستطلاع، والإبداع، وتحمل المخاطر، ومهارات التفكير العليا.

- مهارات الاتصال الفعال Effective Communication.

- مهارات الإنتاجية العالية High Productivity.

ويجمل تقرير حالة اللغة العربية ومستقبلها،⁽¹⁾ 2020 في الإمارات العربية المتحدة صعوبات توظيف التقنيات، والتحديات التي توجه اللغة العربية على النحو التالي:

صعوبات استخدام التقنيات:

التحديات التي تواجه اللغة العربية في المجالات التقنية والمحتوى الرقمي.

الأسباب المعيقة لتقدم العربية تقنيات عليم اللغة العربية.

رابعاً: معوقات توظيف التقنية في تعليم اللغة العربية في⁽²⁾،⁽³⁾

1. معوقات مادية

2. معوقات بشرية

3. معوقات فنية

4. معوقات نفسية

(1) وزارة الثقافة والشباب، الإمارات العربية المتحدة، تقرير حالة اللغة العربية ومستقبلها، 20 ديسمبر، 2020.

(2) إبراهيم، علي حسن، عبد الله يوسف الفيلكاوي، مدى تحقق كفايات التعليم الإلكتروني لدى أعضاء هيئة تدريس كلية التربية بجامعة الكويت، المجلة التربوية، 2018، ص32-55.

(3) البشير، عادل أكرم، "معوقات توظيف مناهج اللغة العربية الإلكترونية في مراحل التعليم العام بمحافظة الجبراء بدولة الكويت: دراسة تحليلية"، مجلة العلوم التربوية والنفسية، مج20، 2019، ص19.

5. معوقات اجتماعية

6. معوقات علمية

المحور الثاني: تعليم اللغة العربية وتعلمها في العصر الرقمي: الواقع، والتحديات.

أولاً: مفهوم التعليم الرقمي: Digital Learning يشير واطفة⁽¹⁾ 2021 إلى أن مفاهيم: التعليم الإلكتروني، والتعليم الافتراضي، والتعليم الرقمي، تتحرك في فضاء واحد، ونشأت في رحم واحدة هي الثورة الصناعية، فمفهوم الرقمي قد يتطابق مع مفهوم الافتراضي، وكذلك مع مفهوم الإلكتروني، وهي تدل على نمط واحد من التعليم بدلالة التعريفات التي سادت في مجال التخصص.

فالتعليم الرقمي:

- تقديم محتوى تعليمي إلكتروني عبر الوسائط المتعددة على الكمبيوتر، وشبكاته، إلى المتعلم بشكل يتيح له إمكانية التفاعل النشط مع المحتوى، ومع المعلم، ومع أقرانه، سواء كان ذلك بصورة متزامنة أم غير متزامنة، وكذلك إمكانية إتمام هذا التعلم في الوقت والمكان وبالسعة التي تناسب ظروفه وقدراته، فضلاً عن إمكانية إدارة هذا التعلم أيضاً من خلال تلك الوسائط⁽²⁾.

- طفرة في الوسائل والتقنيات، تسخر أحدث ما توصلت إليه التقنية من

(1) وطفة، على أسعد، "إشكاليات التعلم الإلكتروني وتحدياته في ظل جائحة كورونا، قراءة سوسيولوجية في جدليات التفاعل والتأثير"، مجلة مركز دراسات الخليج والجزيرة العربية، جامعة الكويت، 2021، ص 21.

(2) بدرانة، عبد الله، دور التعليم الرقمي في مواجهة الأزمات والتحديات الراهنة "سفريرس"، 2020/6/10. شوهد في 2021/1/7 aYUH/ws.bitly://htztp

أجهزة وبرامج في العملية التعليمية، من استخدام وسائل العرض الإلكترونية في إلقاء الدروس في الفصول التقليدية، وانتهاء بإنشاء الفصول الافتراضية والبيئات التفاعلية المختلفة، والتي أثبتت في معظمها فاعليتها في التعليم.⁽¹⁾

- تعليم يجمع بين أشكال التعليم والتعلم الإلكترونية، باعتماد أجهزة الحاسوب ووسائل التخزين وشبكتها بانفتاح جسورها على كل أشكال التحديث في التكنولوجيا المتغيرة باستمرار.⁽²⁾

- تعليم له أنماط: التعلم المدمج، والتعلم المعكوس، والتعلم المخصص، مع وجود الاتصال المتزامن أو غير المتزامن بين عناصر العملية التعليمية.⁽³⁾

- تعليم لم يعد متقيداً بـ:

• الوقت: أتاح الإنترنت، وانتشار أجهزة الوصول إلى الإنترنت للمتعلمين القدرة على التعلم في أي وقت.

• المكان: أتاح التعليم الرقمي الفرصة للتعلم في أي مكان، وفي كل مكان.

(1) مجاهد، فايزة أحمد الحسيني، "التعليم الإلكتروني في زمن كورونا المأل والأمال"، المجلة الدولية للبحوث في العلوم التربوية، 3-2020-305

(2) الألوسي، تيسر عبد الجبار، التعليم الإلكتروني ما قبل كورونا وما بعده، السومريان،

3/5/2020.com.slates-somerian.www:/http/ 2020/05/03/9819 /

شوهدي في 2020/1/7.

(3) Department of Education and Skills, Digital Learning 2020: Reporting on practice in Early Learning and Care, Primary and Post-Primary Contexts.

متاح من خلال الرابط، شوهدي في 2020/10/1:

file:///C:/Users/TABARK/Downloads/Documents/78007_a37413f9-7423-44bf-86df-01762e04a408.pdf

- المسار: تعددت البرامج التفاعلية والتكيفية التي تتيح للمتعلمين التعلم بأسلوبهم الخاص.
- السرعة: فالبرامج تتيح للمتعلمين التعلم بالسرعة التي تناسبهم، وقضاء وقت أكثر أو أقل في الدروس التكنولوجية: هي الآلية التي توصل المحتوى، بشكل يسهل كيفية تلقيه.
- المحتوى الرقمي: مادة أكاديمية عالية الجودة يتم تسليمها من خلال التكنولوجيا.
- التعليمات: غيرت التكنولوجيا دور المعلم لتقديم التوجيه والمساعدة الشخصية.⁽¹⁾
- مصطلح شامل يشير إلى أي نوع من التعلم المقدم باستخدام التكنولوجيا الرقمية عبر الإنترنت، ويتضمن: المتعلمين الذين يتلقون دروسا عبر الإنترنت، والمعلمين الذين يستخدمون الأدوات الرقمية.⁽²⁾
- ومن أهم مميزات استخدام التقنيات الرقمية في تعليم اللغات:
- تحفّز الطلاب على تعلم اللغة، وتطور كفاءاتهم التواصلية.
- توسّع معرفة الطلاب بالثقافة اللغوية.

(1) Butler, Deirdre (DC), learning Digital Learning Framework for Primary Schools, National Council for Curriculum & Assessment; H2 1-16 ISTE COMMUNITY STANDARDS. ISTE Standards for Students,

متاح من خلال الرابط، شوهد في 2020/10/3:

<https://www.iste.org/standards/iste-standards-for-students>. 2016

(2) كرم، على، ما التعليم الرقمي؟ وأهميته ومميزاته؟ "سواح هوست"، 18 ديسمبر 2020، شوهد في 2021/1/7. متاح من خلال الرابط <http://aYVq/ws.bitly://http>

- تُحَسِّن كفاءة تدريس المفاهيم والمهارات اللغوية.
- تُعزِّز التفاعل بين الطلاب وبين المعلمين والطلاب.
- تخلق بيئة تعليمية لغوية غنية بالمشيرات اللغوية.
- تُوفِّر فرصًا لتدريس اللغة خارج الفصول الدراسية⁽¹⁾.
- تُوفِّر طرقًا متعددة للتعلم.

ثانيا: التحديات التي تواجه تعليم اللغة العربية في العصر الرقمي:

- يمكن أن يؤدي إلى الإفراط فيها إلى الاستخدام اللاوعي من قبل المتعلمين.
- تضع فجوة بين طرق التدريس التقليدية والحديثة.
- تفتقر إلى الواقعية، فوقت التدريس قد يكون في أوقات غير رسمية.
- لا تهنيء للمتعلمين فرصًا لممارسة مهارات التفكير المنطقي.
- إن سرعة تطور تكنولوجيا التدريس جعل من الصعب على العديد من المعلمين معرفة أفضل السبل لاستخدام أجهزة الحاسوب، وأشكال أخرى من التكنولوجيا الرقمية، والتفاعل العالمي الذي تم تمكينه بواسطة الإنترنت في تدريس اللغة.

ثالثا: متطلبات توفير البيئة الرقمية لتعليم اللغة العربية وتعلمها:

- مادية: البنية التحتية، والبرامج، والتقنيات الحديثة.
- بشرية: الفنيون والمهندسون وكوادر صناعة المحتوى العلمي للغة العربية، والمعلمون.

(1) Abbasova, Milana& Mammadova, Narmin "The Role of Digital Technology in English Language Teaching in Azerbaijan", International Journal of English Linguistics 9(2):364. March 2019

- فنية: الصيانة، والشبكات، وتوظيف البرمجيات في تعليم اللغة العربية.
- علمية: مواكبة المناهج مع التطورات والمستجدات التكنولوجية، وضوح الفلسفة والأسس العلمية من خلال الاستخدام.
- تُوفّر أشكالاً من مصادر المعلومات الرقمية: الكتاب الرقمي. الدوريات الرقمية. الموسوعات الرقمية. المعاجم الإلكترونية. الأدلة قواعد البيانات العامة والمتخصصة،
- توافر خصائص البيئة الرقمية: التفاعل، والتكامل، والفردية، والتنوع.
- تنوع أدوات البحث: محركات البحث، والبوابات، والأدلة الموضوعية، وفهارس المكتبات، والوكيل الذكي.
- تتيح البيئة الرقمية فرصاً لتعليم اللغة العربية وتعلمها من خلال وسائط متعددة.
- تقييم المعلومات سلامة، ودقة، وموضوعية، وتنظيمًا، وترتيبًا.
- تلبي البيئة الاحتياجات اللغوية لتعلم اللغة العربية، وتوافق مع ما يريده المتعلم الرقمي. تعلمًا اجتماعيًا، وغير مقيد، وغنيًا بمصادر رقمية، وغنيًا بمصادر تفاعلية.
- تتوافق مع تغير نماذج التعلم والتدريس، المعتمدة على المخرجات، والعمليات، وتوليد المعرفة.
- توافر تقنيات جديدة للتعلم. (التقنيات المتنقلة - مجتمعات التعلم الافتراضية)
- تتوافر فيها عمليات ومفاهيم التكنولوجيا: الاستخدام، والوصول، والإنشاء، والاكتشاف، والاختيار لإكمال مهمة لغوية، أو أداء الواجبات.
- توفر فرصًا واسعة لتحقيق الاتصال والتعاون: في تنفيذ الأنشطة والمشروعات.

- تتيح فرص طلاقة البحث والمعلومات.

رابعاً: الإطار العام لمعايير كفاءة معلم اللغة العربية في العصر الرقمي:

1. أن تتوافر للمعلم الظروف التي تجعله: يمتلك الكفاءات المطلوبة لتسهيل اكتساب الكفاءة الرقمية، ويستخدم الأدوات الرقمية لإنشاء المحتوى الرقمي، ومشاركته للمساعدة في التعلم التفاعلي والتعاوني ولأغراض التقويم، وإنشاء بيئة تعليمية بوساطة التكنولوجيا.
2. تهيئة الظروف لممارسة المعلم لأدواره: ميسر المحتوى، والتقني، والمصمم، والمدير، وميسر التدريس، والمستشار، والمقيم، والباحث.
3. تهيئة الظروف لممارسة الأدوار الجديدة للمعلم في ظل تقنيات التعليم: مشخص، ومصمم، ومخطط برامج، وموجه للعملية التعليمية التعليمية، ومهندس للسلوك، وضابط لبيئة التعلم، ومهندس اجتماعي، وموفر للتسهيلات اللازمة للتعلم، ومتخصص في الوسائل التقنية.

خامساً: الإطار العام لمتطلبات معايير المتعلم الرقمي:

1. مراعاة خصائص المتعلم عند اختيار المحتوى، وتصميم طرق التعلم، والتقويم.
2. مراعاة الاحتياجات الأكاديمية والتقنية.
3. تدريب المتعلم على المهارات الأساسية لاستخدام الأجهزة الرقمية:
4. تنمية مهارات البحث عبر شبكات المعلومات الرقمية (الإنترنت):
5. تنمية مهارات التعلم الذاتي.
6. تنمية مهارات حلّ المشكلات.
7. تنمية مهارات مواجهة الأزمات.
8. تدريب المتعلم على ممارسة كفايات التعاون والتواصل الإلكتروني.
9. تكوين ثقافة المواطنة الرقمية.

إجراءات البحث:

استهدف البحث الحالي إعداد إطار مرجعي لتوظيف التقنية في تعليم اللغة العربية وتعلمها، وتطلب هذا مراجعة التجارب والخبرات العربية والدولية في المجال، وجمع البيانات والمعلومات النظرية وتفسيرها، وتحليلها فيما يتصل بإعداد ما يأتي:

1. قائمة المعوقات التي تحول دون توظيف التقنية في تعليم اللغة العربية وتعلمها.
2. معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها بمراحل التعليم العام في الصفوف من (4-1)، ومن (5-9)، ومن (10-12).
3. معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية في مراحل التعليم العام بدول الخليج العربي في الصفوف من (4-1)، ومن (5-9)، ومن (10-12).
4. معايير الكفاءة الرقمية اللازمة لمتُعَلِّمي اللغة العربية في مراحل التعليم العام في الصفوف من (4-1)، ومن (5-9)، ومن (10-12).
5. معايير تعليم فنون اللغة الرقمية (الاستماع الرقمي - السرد الرقمي - القراءة الرقمية - الكتابة الرقمية) في مراحل التعليم العام.
6. معايير التقويم الإلكتروني في اللغة العربية بمراحل التعليم العام.
7. الدليل المقترح لمُعَلِّمي اللغة العربية، ومُتَعَلِّمها لتوظيف التقنية في تعليم اللغة العربية وتعلمها في ضوء الإطار العام المرجعي.

وسارت إجراءات البحث على النحو الآتي:

1. بناء قائمة المعوقات التي تحول دون توظيف التقنية في تعليم اللغة العربية وتعلمها:

في ضوء المعطيات النظرية في مجال توظيف التقنية في تعليم اللغات، وما أسفرت عنه نتائج الدراسات السابقة والتجارب التي تصدّت لدراسة المعوقات التي

تَحُول دون توظيف التقنية في تعليم اللُّغات، حدّد المشروع الحالي قائمةً للمعوّقات الرئيسة، وما يتفرّع عن كل مُعوّق من معوّقات فرعيّة على النحو الآتي:

أ. المعوّقات الماديّة: وهي التي تتعلّق بتوفير مُقوّمات البيئة التّقنيّة؛ من أجهزة، ووسائل، وتقنيات لكل من المعلم، والمتعلم داخل المدرسة، وخارجها، وتفرّع عن هذا المُعوّق ستّة معوّقات فرعيّة.

ب. المعوّقات الإداريّة: تتصل بجهود الإدارة المدرسيّة من أجل توفير البيئة التكنولوجيّة الملائمة؛ لاستخدام التقنيات في تعليم اللُّغة العربيّة، وتشجيع استخدامها، وتوفير سُبل صيانة الأجهزة والتقنيات، وتفرّع عن المعوّق خمسة معوّقات.

ج. المعوّقات الفنيّة: تتعلّق بتوافر شبكة الإنترنت، والتعامل مع الأعطال الفنيّة، وتوافر المختصين، وتوفير الدعم اللازم لاستخدام التقنيات على أفضل صورة ممكنة، وتفرّع عن هذا المُعوّق ستّة معوّقات فرعيّة.

د. معوّقات تتصل بمعلم اللُّغة العربيّة: وضُمّت المعوّقات الخاصّة بتقدير معلم اللُّغة العربيّة ووعيه لاستخدام التقنيات، وما يتوافر له من دورات تدريبيّة تضمّن له تنمية مهنيّة تقنيّة مستمرّة، وما يَحُول بينهم وبين استخدامهم الكُفء للتقنيات، وتفرّع عن هذا المعوّق عشرة معوّقات فرعيّة.

هـ. معوّقات خاصّة بالمتعلم: وتشمل المعوّقات التي تَحُول بين متعلّم اللُّغة العربيّة وبين استخدامه للتقنيات، واستفادته منها؛ كنقص الكفاءة الرّقميّة، والجوانب النفسيّة المتصلة باستخدامه للتقنيات، وسوء استخدام بعض المتعلّمين لها، وتفرّع عن هذا المعوّق سبعة معوّقات.

وأُعِدَّت المعوّقات في صورة استبانة:

جدول (3)

مكونات استبانة المعوقات

التي تحول دون استخدام التقنية في تعليم اللغة العربية

م	المعوق	عدد المعوقات
1	المعوقات المادية.	6
2	المعوقات الإدارية.	5
3	المعوقات الفنية.	6
4	معوقات تتصل بمعلم اللغة العربية.	10
5	معوقات خاصة بالمتعلم.	7

توصيف العينة:

بلغ عدد أفراد العينة (198).

أ- توزيع أفراد العينة بحسب الدولة:

جدول (4)

توزيع أفراد العينة بحسب الدولة

الدولة	عدد المستجيبين	النسبة المئوية
الكويت	86	43,4%
المملكة العربية السعودية	84	42,9%

ب- توزيع أفراد العينة بحسب النوع:

جدول (5)

توزيع أفراد العينة بحسب النوع

النوع	العدد	النسبة المئوية
ذكر	87	%56,1
أنثى	111	%43,9

ج- توزيع المستجيبين بحسب الوظيفة:

جدول (6)

توزيع أفراد العينة بحسب الوظيفة

الوظيفة	العدد	النسبة المئوية
موجه/ مُشرف تربويّ	158	%79,8
معلم	40	%20,2

د- توزيع أفراد العينة بحسب المؤهل الدراسي:

جدول (7)

توزيع أفراد العينة بحسب المؤهل الدراسي

المؤهل الدراسي	العدد	النسبة المئوية
بكالوريوس	163	%82,3
ماجستير	23	%12,1
دكتوراه	12	%5,6

هـ. توزيع المستجيبين بحسب سنوات الخبرة:

جدول (8)

توزيع أفراد العينة بحسب سنوات الخبرة

عدد سنوات الخبرة	العدد	النسبة المئوية
أقل من 5 سنوات	16	8,1%
من 6-10 سنوات	22	10,6%
أكثر من عشر سنوات	160	81,3%

و. توزيع أفراد العينة بحسب المرحلة التعليمية:

جدول (9)

توزيع أفراد العينة بحسب المرحلة التعليمية

المرحلة التعليمية	العدد	النسبة المئوية
من 1-6	41	20,7%
من 7-9	57	28,8%
من 10-12	100	50,5%

1- بناء منظومة معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلّمها
بمراحل التعليم العام في الصفوف من (1-4)، ومن (5-9)، ومن (10-12):

من منطلق:

- الإطار النظري للبحث، وما تناوله من تأصيل للمفاهيم النظرية المتعلقة بالتعلم الرقمي مفهوماً، وأهدافاً، وأهمية، ومميزات، وفوائد، ومكونات، وخصائص، وأنماطاً، ومهارات العصر الرقمي، ومتطلبات توفير بيئة رقمية لتعليم اللغة العربية، وتعلّمها.

- عَرَضَ التجارب الإقليمية والعالمية في المجال، ورَصَدَ ما أسفرت عنه من معايير البيئات الرقمية.

أُعِدَّت قائمةُ معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها بمراحل التعليم العام في الصفوف من (1-4)، ومن (5-9)، ومن (10-12)، وصيغت القائمة في صورة استبانة.

تكوّنت استبانة معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها بمراحل التعليم العام في الصفوف من ثلاثة أجزاء:

الأول: معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها في الصفوف من (1-4).

الثاني: معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها في الصفوف من (5-9).

الثالث: معايير البيئة الرقمية المناسبة لتعليم اللغة العربية وتعلمها في الصفوف من (10-12).

وعُرضت الاستبانة - في صورتها المبدئية - في ورشة عمل تحكيمية ضمّت اثني عشر أستاذًا من أساتذة الجامعات المصرية (خمسة من المتخصصين في طرائق تعليم اللغة العربية، وخمسة من المتخصصين في تكنولوجيا التعليم، واثنين من المتخصصين في التقويم)، لمناقشة المعايير، والمؤشرات، وإبداء الرأي حولها، وكانت آراؤهم على النحو الآتي:

جدول (10)

يُوضَّح نسب الموافقة على بنود تحكيم معايير البيئة الرقمية المناسبة
لتعليم اللغة العربية وتعلُّمها بمراحل التعليم العام

م	المعايير والمؤشرات	درجة المناسبة		
		كبيرة	متوسطة	صغيرة
1	تشمل الجوانب المختلفة لبناء بيئات التعلم الرقمية: المادية، والبشرية، والاجتماعية، والنفسية، والعلمية.	88,33%	16,67%	-
2	تتسق مع الممارسات العالمية في تخطيط بيئات التعلم الرقمية.	88,33%	16,67%	-
3	تتوافر معايير خصائص البيئة الرقمية التي تستهدف تعليم اللغة العربية في العصر الرقمي.	75%	25%	-
4	تركز على طبيعة المتعلم في العصر الرقمي.	91,66%	8,34%	-
5	تركز على توظيف معلم اللغة العربية لمهارات في العصر الرقمي.	75%	25%	-
6	تُراعي طبيعة تعليم اللغة العربية وتعلُّمها في العصر الرقمي.	75%	25%	-
7	تُتيح فرصاً لتوظيف تطبيقات التعليم الرقمي في تعليم اللغة العربية وتعلُّمها.	75%	25%	-

يتَّضح من الجدول أن درجات الموافقة على البنود التي تم تحكيم المعايير والمؤشرات في ضوءها أن:

- نسب الموافقة بدرجة كبيرة تراوحت بين (75%) إلى (88,33%).

- نِسَب الموافقة بدرجة متوسطة تراوحت بين (8,34%) إلى (25%).
وجاءت الصورة النهائية للمنظومة على النحو الآتي:

جدول (11)

يُوضِّح عددَ معايير البيئة الرِّقْمِيَّة، وما ينبثق عن كل معيار من مؤشِّرات
في كل مرحلة من مراحل التعليم العام

الصفوف من 12-10		الصفوف من 9-5		الصفوف من 4-1	
عدد المؤشِّرات	المعيار	عدد المؤشِّرات	المعيار	عدد المؤشِّرات	المعيار
4	1	5	1	6	1
2	2	3	2	2	2
2	3	2	3	3	3
2	4	4	4	2	4
4	5	2	5	2	5
3	6	3	6	2	6
11	7	11	7	2	7
10	8	7	8	5	8
3	9	3	9	3	9
2	10	3	10	3	10
3	11	3	11	3	11
2	12	4	12	4	12
7	13	5	13	3	13
9	14	2	14	4	14
2	15	4	15	2	15
3	16	2	16	5	16
4	17	4	17	7	17

الصفوف من 12-10		الصفوف من 9-5		الصفوف من 4-1	
عدد المؤشرات	المعيار	عدد المؤشرات	المعيار	عدد المؤشرات	المعيار
4	18	6	18	2	18
2	19	3	19	-	19
2	20	2	20	-	20
-	21	2	21	-	21

2. بناء منظومة معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية في مراحل التعليم العام بدُول الخليج العربي في الصفوف من (4-1)، ومن (9-5)، ومن (12-10). بناءً على المعطيات النظرية المتمثلة في:

- معلم اللغة العربية في العصر الرقبي، ودوره في دعم توظيف التقنيات.
- مهارات معلم اللغة العربية في العصر الرقبي.
- أدوار معلم اللغة العربية في العصر الرقبي (ميسر إلكتروني، ووسيط إلكتروني، وابتكاري، وباحث، ومقوم، وتكنولوجي، ومدير للصف).
- خصائص معلم اللغة العربية في العصر الرقبي.
- الكفايات المهنية لمعلم اللغة العربية في العصر الرقبي.
- التحديات التي يواجهها معلم اللغة العربية في العصر الرقبي.
- متطلبات الكفاءة الرقمية لمعلم اللغة العربية.

أ. عَرَضَ التجارب الإقليمية والعالمية في المجال، ورصد ما أسفرت عنه من معايير الكفاءة الرقمية للمعلم.

أُعِدَّتْ قائمة معايير الكفاءة الرقمية لمعلم اللغة العربية في الصفوف من (1-4)، ومن (9-5)، ومن (12-10)، ووضعت القائمة في صورة استبانة.

تكوّنت استبانة الكفاءة الرقمية لمعلم اللغة العربية في الصفوف من (4-1)، ومن (9-5)، ومن (10-12) من ثلاثة أجزاء:

الأول: معايير الكفاءة الرقمية لمعلم اللغة العربية في الصفوف من (4-1).

الثاني: معايير الكفاءة الرقمية لمعلم اللغة العربية في الصفوف من (5-9).

الثالث: معايير الكفاءة الرقمية لمعلم اللغة العربية في الصفوف من (10-12).

وعُرضت الاستبانة - في صورتها المبدئية- في ورشة عمل تحكيمية ضمت خمسة من المتخصصين في طرائق تعليم اللغة العربية، وخمسة من المتخصصين في تكنولوجيا التعليم من أساتذة الجامعات المصرية، لمناقشة المعايير، والمؤشرات وإبداء الرأي حولها، وكانت آراؤهم على النحو الآتي:

جدول (12)

يُوضّح نسب الموافقة على بنود تحكيم

معايير الكفاءة الرقمية لمعلم اللغة العربية

م	المعايير والمؤشرات	درجة المناسبة		
		كبيرة	متوسطة	صغيرة
1	تُرَكِّز المعايير والمؤشرات على المهارات التكنولوجية لاستخدام الحاسوب وتطبيقاته، والإنترنت.	100%	-	-
2	تُعنى المعايير والمؤشرات بطرائق التواصل الإلكتروني.	100%	-	-
3	تتضمّن المعايير والمؤشرات كفايات استخدام التقنيات الرقمية، وتطبيقها في تدريس اللغة العربية.	80%	20%	-
4	تشمل المعايير والمؤشرات أدوار معلم اللغة	80%	20%	-

م	المعايير والمؤشرات	درجة المناسبة		
		كبيرة	متوسطة	صغيرة
	العربية الجديدة في ظل تقنيات العصر الرقمي.			
5	تشمل المعايير والمؤشرات مهارات معلم اللغة العربية الخاصة بالمعارف في العصر الرقمي، ومصادره، وتقييم المعلومات.	%60	%40	-
6	تضمن المعايير والمؤشرات استجابة المعلم لتغير نماذج التعلم والتدريس في العصر الرقمي.	%60	%40	-
7	تشير المعايير والمؤشرات إلى أهمية التنمية المهنية التكنولوجية لمعلم اللغة العربية في العصر الرقمي.	%80	%20	-

يتضح من الجدول أن درجات الموافقة على البنود التي تم تحكيم المعايير والمؤشرات في ضوءها أن:

- نسب الموافقة بدرجة كبيرة تراوحت بين (60%) إلى (100%).
 - نسب الموافقة بدرجة متوسطة تراوحت بين (20%) إلى (40%).
- وجاءت الصورة النهائية للمنظومة على النحو الآتي:

جدول (13)

يُوضّح عدد معايير الكفاءة الرّقمية لمعلم اللغة العربيّة

في كل مرحلة من مراحل التعليم العامّ، وما ينبثق عن كل معيار من مؤشّرات

الصفوف من 12-10		الصفوف من 9-5		الصفوف من 4-1	
عدد المؤشّرات	المعيار	عدد المؤشّرات	المعيار	عدد المؤشّرات	المعيار
3	1	4	1	8	1
3	2	4	2	22	2
2	3	3	3	2	3
11	4	15	4	2	4
2	5	3	5	3	5
2	6	2	6	3	6
3	7	2	7	3	7
2	8	4	8	2	8
2	9	4	9	4	9
5	10	2	10	4	10
5	11	-	11	-	11
5	12	-	12	-	12

3. بناء منظومة معايير الكفاءة الرّقمية اللازمة لمتعلّمي اللغة العربيّة في

مراحل التعليم العامّ في الصفوف من (4-1)، ومن (9-5)، ومن (12-10).

في ضوء ما تناوَله الإطارُ النظريُّ من تحليل للمفاهيم النظرية السائدة في

مجال المتعلّم الرّقميّ:

- مصطلح المتعلّم الرّقميّ، والمواطن الرّقميّ.

- مزايا نظام التعليم الرقّميّ لمُتعلّمي اللّغة العربيّة.
 - خصائص المتعلّم الرقّميّ.
 - كيفية تلبية الاحتياجات التعليميّة للمُتعلّمين الرقّمين.
 - متطلّبات الكفاءة الرقّميّة لمُتعلّم اللّغة العربيّة.
- أ. ما أسفرت عنه نتائج الدراسات السابقة بشأن معايير المتعلّم الرقّميّ:
- أُعِدّت قائمة معايير الكفاءة الرقّميّة اللازمة لمُتعلّم اللّغة العربيّة في الصفوف من (4-1)، ومن (5-9)، ومن (10-12)، ووُضعت القائمة في صورة استبانة.
- تكوّنت استبانة معايير الكفاءة الرقّميّة اللازمة لمُتعلّم اللّغة العربيّة في الصفوف من (4-1)، ومن (5-9)، ومن (10-12) من ثلاثة أجزاء:
- الأول: معايير الكفاءة الرقّميّة اللازمة لمُتعلّم اللّغة العربيّة في الصفوف من (4-1).
- الثاني: معايير الكفاءة الرقّميّة اللازمة لمُتعلّم اللّغة العربيّة في الصفوف من (5-9).
- الثالث: معايير الكفاءة الرقّميّة اللازمة لمُتعلّم اللّغة العربيّة في الصفوف من (10-12).
- وعُرضت الاستبانة - في صورتها المبدئيّة- في ورشة عمّل تحكيميّة ضمت خمسةً من المتخصصين في طرائق تعليم اللّغة العربيّة، وخمسةً من المتخصصين في تكنولوجيا التعليم من أساتذة الجامعات المصرية، لمناقشة المعايير، والمؤشرات وإبداء الرأي حولها، وكانت آراؤهم على النحو الآتي:

جدول (14)

يُوضَّح نِسَبُ المُوافَاقَةِ على بنود تحكيم معايير الكفاءة الرَّقْمِيَّة لمتعلِّمي اللُّغة العربيَّة

م	المعايير والمؤشرات	درجة المناسبة		
		كبيرة	متوسطة	صغيرة
1	تُقدم المعايير والمؤشرات إطارًا للمهارات الأساسية اللازمة للمتعلم في العصر الرَّقْمِي لاستخدام الأجهزة الرَّقْمِيَّة.	%80	%20	-
2	تغطي المعايير والمؤشرات مهارات البحث عبر شبكات المعلومات الرَّقْمِيَّة.	%100	-	-
3	تؤكد المعايير والمؤشرات على مهارات التعلُّم الذاتي، والتعاوني، والمستمر، وحل المشكلات، والمهارات الحيائية للمتعلم في العصر الرَّقْمِي.	%100	-	-
4	تُركِّز المعايير والمؤشرات على استخدام المتعلم الرَّقْمِي للتقنية في التعلُّم، وأداء المهام اللُّغويَّة، والاختبارات، وحل الواجبات، وممارسة الأنشطة اللُّغويَّة.	%100	-	-
5	تشير المعايير والمؤشرات إلى تحليل المتعلم الرَّقْمِي بالسلوك الآمن والقانوني والأخلاقي عبر الإنترنت.	%70	%30	-
6	تُوضح المعايير والمؤشرات كيفية استخدام المتعلم للتقنيات الرَّقْمِيَّة في الاتصال والتعاون الفعَّال.	%80	%20	-
7	تغطي المعايير والمؤشرات تحقيق متطلَّبات المُواطنِ الرَّقْمِيَّة.	%40	%50	%10

يتَّضح من الجدول أن درجات الموافقة على البنود التي تم تحكيم المعايير والمؤشرات في ضوءها أن:

- نسب الموافقة بدرجة كبيرة تراوحت بين (40%) إلى (100%).
 - نسب الموافقة بدرجة متوسطة تراوحت بين (20%) إلى (50%).
 - نسب الموافقة بدرجة قليلة (10%) على معيار واحد.
- وجاءت الصورة النهائية للمنظومة على النحو الآتي:

جدول (15)

يُوضَّح عدد معايير الكفاءة الرقمية لمتعلِّم اللغة العربية

في كل مرحلة من مراحل التعليم العام، وما ينبثق عن كل معيار من مؤشرات

الصفوف من 10-12		الصفوف من 5-9		الصفوف من 1-4	
عدد المؤشرات	المعيار	عدد المؤشرات	المعيار	عدد المؤشرات	المعيار
3	1	10	1	5	1
2	2	4	2	2	2
3	3	5	3	2	3
2	4	3	4	2	4
3	5	3	5	4	5
4	6	4	6	2	6
8	7	5	7	4	7
4	8	4	8	4	8
6	9	11	9	4	9
2	10	2	10	3	10
3	11	8	11	-	11
4	12	3	12	-	12
11	13	5	13	-	13
2	14	4	14	-	14

4- بناء منظومة معايير تعليم فنون اللغة الرقمية (الاستماع الرقمي- السرد الرقمي- القراءة الرقمية - الكتابة الرقمية) في مراحل التعليم العام.

أعدت قائمة معايير تعليم فنون اللغة الرقمية (الاستماع الرقمي- السرد الرقمي- القراءة الرقمية - الكتابة الرقمية) في مراحل التعليم العام، في ضوء:

أ- ما تضمنه دليل توظيف التقنية في تعليم اللغة العربية وتعلمها في ضوء الإطار العام المرجعي ومعايير ومؤشراته الذي عرض لـ:

- استراتيجيات تدريس مهارات (الاستماع، والتحدث، والقراءة والكتابة) الرقمية.
- استراتيجيات تدريس القواعد.
- استراتيجيات تدريس المفردات.

ب- ما ورد في الدراسات السابقة من معايير لتعليم فنون اللغة.

ووضعت القائمة في صورة استبانة.

وعُرضت الاستبانة - في صورتها المبدئية- في ورشة عمل تحكيمية ضمت خمسة من المتخصصين في طرائق تعليم اللغة العربية، وخمسة من المتخصصين في تكنولوجيا التعليم من أساتذة الجامعات المصرية، لمناقشة المعايير، والمؤشرات وإبداء الرأي حولها، وكانت آراؤهم على النحو الآتي:

جدول (16)

يُوضّح نسب الموافقة على بنود تحكيم معايير تعليم فنون اللغة الرقمية

م	المعايير والمؤشرات	درجة المناسبة		
		كبيرة	متوسطة	صغيرة
1	تُغطي معايير تدريس فنون اللغة العربية الرقمية ومؤشراتها المبادئ الأساسية للتدريس في العصر الرقمي، وخصائصه، وأهدافه.	100%	-	-
2	تُركز معايير ومؤشرات طرق تدريس فنون اللغة الرقمية على إكساب المتعلم مهارات العصر الرقمي، وتنميتها.	80%	20%	-
3	تساعد المعايير على خلق فرص متساوية للتعلم؛ لتحقيق أهداف التعلم اللغوي.	60%	40%	-
4	تتوافق المعايير مع أدوار معلم اللغة العربية، ووظائفه في العصر الرقمي.	100%	-	-
5	تُحقّق المعايير التوازن بين المهارات التقنية، والمهارات اللغوية لمعلم اللغة العربية.	100%	-	-

يتّضح من الجدول أن درجات الموافقة على البنود التي تم تحكيم المعايير والمؤشرات في ضوءها أن:

- نسب الموافقة بدرجة كبيرة تراوحت بين (60%) إلى (100%).
 - نسب الموافقة بدرجة متوسطة تراوحت بين (20%) إلى (40%).
- وجاءت الصورة النهائية للمنظومة على النحو الآتي:

جدول (17)

يُوضَّح عدد معايير تعليم فنون اللغة الرُّقْمِيَّة

وما ينبثق من كل معيار من مؤشِّرات

المعيار	1	2	3	4	5	6	7	8	9	10
عدد المؤشِّرات	15	6	5	5	9	9	8	6	8	8

بناء منظومة معايير التقويم الإلكتروني في اللغة العربيَّة بمراحل التعليم

العام:

أُعدت منظومة معايير التقويم الإلكتروني في اللغة العربيَّة انطلاقًا من:

أ. التأصيل النظريِّ للجوانب الآتية:

- مفهوم التقويم الإلكتروني، وأهم أدواته، ومجالاته.

- مزايا التقويم اللُّغويِّ الإلكترونيِّ.

- مبادئ التقويم اللُّغويِّ الإلكترونيِّ.

- أنواع مهامَّ التقويم اللُّغويِّ الإلكترونيِّ.

- ضوابط إجراء التقويم اللُّغويِّ الإلكترونيِّ.

- أساليب التقويم اللُّغويِّ الإلكترونيِّ.

ب. نتائج الدراسات السابقة في مجال التقويم الإلكتروني:

عُرِضت الاستبانة - في صورتها المبدئيَّة- في ورشة عمَل تحكيميَّة ضمَّت اثني

عشر أستاذًا من أساتذة الجامعات المصرية (خمسة من المتخصصين في طرائق

تعليم اللغة العربيَّة، وخمسة من المتخصصين في تكنولوجيا التعليم، واثنين من

المتخصصين في التقويم)، لمناقشة المعايير، والمؤشِّرات وإبداء الرأي حولها، وكانت

آراؤهم على النحو الآتي:

جدول (18)

يُوضَح نسب الموافقة على بنود تحكيم

معايير التقويم الإلكتروني في اللغة العربية بمراحل التعليم العام

م	المعايير والمؤشرات	درجة المناسبة		
		كبيرة	متوسطة	صغيرة
1	تتضمّن المعايير والمؤشرات التخطيط لاستخدام أدوات التقويم الإلكتروني بما يحقق الدقة، والموثوقية في النتائج.	100%	-	-
2	تُركز المعايير والمؤشرات على استخدام المعلم للتقويم التشخيصي، والتكويني، والختامي إلكترونيًا بما يحقق أهداف التعلم اللغوي.	83,33%	16,67%	-
3	تُتيح المعايير والمؤشرات للمعلم أن يوظف إمكانيات الحاسوب، وبرمجياته، وتطبيقاته، والإنترنت في تصميم التقويم الإلكتروني وتنفيذه، والتعامل مع النتائج.	91,66%	8,34%	-
4	تؤكد المعايير والمؤشرات على قياس المعارف والمفاهيم اللغوية، وممارسة مهارات الأداء اللغوي، وقياس الجوانب الوجدانية في تعلم اللغة العربية إلكترونيًا.	75%	25%	-
5	تُقدم المعايير والمؤشرات لمعلم اللغة العربية إطارًا لمتطلبات العمل مع المتعلمين لاستخدام أساليب التقويم الإلكتروني.	50%	20%	30%

يتَّضح من الجدول أن درجات الموافقة على البنود التي تم تحكيم المعايير والمؤشرات في ضوءها أن:

- نسب الموافقة بدرجة كبيرة تراوحت بين (50%) إلى (100%).
- نسب الموافقة بدرجة متوسطة تراوحت بين (8,33%) إلى (25%).
- نسبة الموافقة بدرجة قليلة على معيار واحد بلغت (30%).

جدول (19)

يُوضِّح عدد معايير التقويم الإلكتروني

في اللغة العربية بمراحل التعليم العام وما ينبثق عن كل معيار من مؤشرات

المعيار	1	2	3	4	5	6	7	8	9	10	11
عدد المؤشرات	9	7	64	9	5	5	7	6	17	7	17

نتائج البحث

أولاً: المعوقات التي تخول دون استخدام التقنية في تعليم اللغة العربية:

1- المعوقات المادية:

ترتيب المعوقات بحسب الأهمية:

جدول (20)

ترتيب المعوقات المادية بحسب الأهمية

م	المعوقات	عدد من يرون أنه الأكثر أهمية
1	عدم تناسب أعداد الحاسوب مع أعداد المتعلمين	103
2	صعوبة توفير بيئة	90
3	صعوبة توافر التقنيات	88
4	قصور الدعم المالي	87
5	قصور في البنية التحتية	77

2- المعوّقات الإداريّة:

ترتيب المعوّقات بحسب الأهمية:

جدول (21)

ترتيب المعوّقات الإداريّة من حيث الأهمية

م	المعوق	عدد من يرون أنّه الأكثر أهميّة
1	ضَعْف الجهود الإداريّة المبذولة من أجل توفير البيئة التكنولوجيّة الملائمة لاستخدام التقنيات في تعليم اللّغة العربيّة وتعلّمها.	89
2	ضَعْف الجهود الإداريّة المبذولة من أجل حل المشكلات التي تواجه مُعلّمي اللّغة العربيّة في استخدام التقنيات في عمليّتيّ التعليم والتعلّم.	81
3	قلة تشجيع إدارة المدرسة لمُعلّمي اللّغة العربيّة على التدريس الفعّال باستخدام التقنيات الحديثة.	75
4	ضَعْف الجهود الإداريّة المبذولة من أجل توفير الصيانة اللازمة للأجهزة والوسائل التعليميّة المستخدمة في تعليم اللّغة العربيّة وتعلّمها.	73
5	قلة اهتمام إدارة المدرسة بتخصيص ساعات في الجداول الدراسيّة؛ للتفاعل الإلكترونيّ بين مُعلّمي اللّغة العربيّة والمتعلّمين.	65

3- المعوّقات الفنيّة:

ترتيب المعوّقات الفنيّة بحسب الأهمية:

جدول (22)

ترتيب المعوّقات الفنيّة بحسب درجة الأهمية

م	المعوق	عدد من يرون أنّه الأكثر أهمية
1	قلة المختصين الفنيين في المدارس يحُول دون توفير الصيانة والمتابعة المستمرة للأجهزة المستخدمة في تعليم اللّغة العربيّة وتعلّمها.	101
2	ضعف شبكة الإنترنت يعوق استخدام التقنيات في تعليم اللّغة العربيّة وتعلّمها.	81
3	نقص الدعم الفنيّ اللازم لاستخدام التقنيات الحديثة في تصميم الاختبارات الإلكترونيّة، ورفعها، وتصحيحها.	76
4	نقص الدعم الفنيّ اللازم لمُعَلِّمي اللّغة العربيّة من أجل متابعة الواجبات، والأنشطة، والمشروعات الإلكترونيّة.	73
5	نقص الأدلة الإرشاديّة اللازمة لمُعَلِّمي اللّغة العربيّة؛ للوعي بكيفية التعامل مع التقنيات الحديثة في عمليّتيّ التعليم والتعلّم.	73
6	كثرة الأعطال الفنيّة للأجهزة المستخدمة في تعليم اللّغة العربيّة وتعلّمها؛ ممّا يُعرّضها للتلف.	68

4- معوّقات تتعلّق بمعلم اللّغة العربيّة:

ترتيب المعوّقات المتصلة بمعلم اللّغة العربيّة حسب درجة الأهمية:

جدول (23)

ترتيب المعوّقات المتصلة باللّغة العربيّة بحسب درجة الأهمية

م	المعوق	عدد من يرون أنّه الأكثر أهميةً
1	قلة إلمام مُعلّمي اللّغة العربيّة بكيفية توظيف استراتيجيّات التعليم الإلكترونيّ في تعليم اللّغة العربيّة وتعلّمها.	106
2	ضيق بعض مُعلّمي اللّغة العربيّة من قلة تلبية الاحتياجات اللازمة لتوظيف التقنيات الحديثة في تعليم اللّغة العربيّة وتعلّمها.	98
3	نقص الكفاءة الرّقميّة لبعض مُعلّمي اللّغة العربيّة؛ ممّا يُقلّل من قدرتهم على إدارة التعليم الإلكترونيّ للّغة العربيّة، والتواصل مع طلابهم.	96
4	كثرة أعباء مُعلّمي اللّغة العربيّة، ممّا يُقلّل من قدرتهم على استخدام التقنيات الحديثة في تعليم اللّغة العربيّة وتعلّمها.	92
5	ضيق وقت بعض مُعلّمي اللّغة العربيّة؛ للتواصل مع الطلاب، ومتابعة مشروعاتهم اللّغويّة الإلكترونيّة، وتصحيح واجباتهم المدرسيّة من خلال البريد الإلكترونيّ، والرد عليها.	92
6	نقص الدورات التدريبية لمُعلّمي اللّغة العربيّة في مجال استخدام التقنيات الحديثة في تعليم اللّغة العربيّة وتعلّمها.	90

م	المعوق	عدد مَنْ يرون أَنَّهُ الأكثر أهميةً
7	قلة وعي مُعلِّمي اللُّغة العربيَّة بكيفية توظيف التقنيات في تعليم اللُّغة العربيَّة وتعلُّمها.	89
8	قلة تقدير مُعلِّمي اللُّغة العربيَّة لأهمية استخدام التقنيات الحديثة في تعليم اللُّغة العربيَّة وتعلُّمها.	82
9	قناعات بعض مُعلِّمي اللُّغة العربيَّة بأن طرق التلقين أكثر فاعليَّةً من استخدام التقنيات الحديثة في تعليم اللُّغة العربيَّة وتعلُّمها.	81
10	خشية بعض مُعلِّمي اللُّغة العربيَّة من صعوبة ضبط الصف؛ في ظل استخدام التقنيات الحديثة في التدريس.	71

5- معوّقات خاصّة بالمتعلم:

ترتيب المعوّقات المتصلة بالمتعلم بحسب درجة الأهمية:

جدول (24)

ترتيب المعوّقات المتصلة بالمتعلم بحسب درجة الأهمية

م	المعوق	عدد مَنْ يرون أَنَّهُ الأكثر أهميةً
1	نقص الكفاءة الرّقميَّة لمتعلِّم اللُّغة العربيَّة تعوق استخدام التقنيات الحديثة في تعليمه وتعلُّمه.	93
2	قلة قناعة بعض المتعلِّمين بجدوى استخدام التقنيات الحديثة في تعليم اللُّغة العربيَّة وتعلُّمها.	89
3	قلّة تفاعل بعض المتعلِّمين مع مُعلِّمي اللُّغة العربيَّة عند استخدام التقنيات الحديثة في تعليم اللُّغة العربيَّة وتعلُّمها.	77

م	المعوق	عدد من يرون أنه الأكثر أهمية
4	افتقاد بعض المتعلمين مهارات التعلم الذاتي اللازمة لاستخدام التقنيات الحديثة في تعليم اللغة العربية وتعلمها.	89
5	تراخي بعض المتعلمين وقلة حماسهم لاستخدام التقنيات الحديثة في البحث، والتواصل، وبناء المعرفة في مجال اللغة العربية.	93
6	تفضيل بعض المتعلمين للطرق التقليدية في تدريس اللغة العربية على حساب التقنيات الحديثة؛ ضماناً لطرح أسئلتهم على المعلمين بشكل مباشر، والفهم والاستيعاب.	68
7	سوء استخدام بعض المتعلمين للإنترنت، وإهدارهم للوقت يعوقان استخدام التقنيات الحديثة في تعليم اللغة العربية وتعلمها.	79

ثانياً: منظومة معايير البيئة الرقمية:

أ. معايير البيئة الرقمية في الصفوف من (4-1).

جدول (25)

معايير البيئة الرقمية في الصفوف من (4-1)

وما ينبثق عنها من مؤشرات

م	المعايير	عدد المؤشرات
1	مادية: تتوافر بالبيئة الرقمية المقومات المادية التي تؤهلها لتعليم اللغة العربية.	6
2	بشرية: تُعنى البيئة الرقمية بالنمو المهني للكفاءات البشرية بصورة تضمن عملها بكفاءة.	2
3	فنية: تتضمن البيئة الرقمية المقومات الفنية لتعمل بفاعلية.	2
4	نفسية: تحقق عناصر البيئة الرقمية الجوانب النفسية الإيجابية اللازمة لتعليم اللغة العربية، وتعلمها.	2
5	اجتماعية: يحدث التعلم اللغوي في البيئة الرقمية في سياق اجتماعي.	3
6	علمية: ترى البيئة الرقمية مناخاً تعليمياً إيجابياً لاكتساب مهارات الأداء اللغوي.	2
7	فلسفية: تنطلق بيئة تعليم اللغة العربية وتعلمها من فلسفة بيئات التعلم الرقمي وأهدافها.	2
8	توفر بيئة تعلم اللغة العربية وتعلمها أشكالاً من مصادر المعلومات الرقمية.	2
9	تتوافر بيئة تعلم اللغة العربية وتعلمها خصائص البيئة الرقمية.	5
10	تتيح البيئة الرقمية فرصاً متنوعة لتعليم اللغة العربية وتعلمها.	3

المؤتمر الدولي الأول للسانيات الحاسوبية واللغة العربية التصورات والتطبيقات والتطلعات

م	المعايير	عدد المؤشرات
11	تحقق معايير البيئات الرقمية فيما يتصل بتقييم المعلومات اللغوية.	3
12	تلي البيئة متطلبات التعلم اللغوي، وتتوافق مع ما يريده المتعلم الرقمي.	3
13	تتوافق مقومات بيئة تعلم اللغة العربية وتعلمها مع تغير نماذج التعلم والتدريس.	4
14	تتوافر بيئة تعليم اللغة العربية وتعلمها تقنيات متنوعة للتعلم.	3
15	تراعى بيئة تعليم اللغة العربية وتعلمها طبيعة المتعلمين وخصائصهم.	4
16	تراعى بيئة تعلم اللغة العربية وتعلمها احتياجات المتعلمين اللغوية، وقدراتهم.	5
17	توظف تطبيقات التعليم الرقمي في بيئة تعليم اللغة العربية وتعلمها.	7
18	تتوافر في البيئة الرقمية مقومات السلامة.	2

ب. معايير الكفاءة الرقمية لبيئات تعليم اللغة العربية وتعلّمها في الصفوف من (5-9)

جدول (26)

معايير البيئة الرقمية في الصفوف من (5-9)
وما ينبثق عنها من مؤشرات

م	المعايير	عدد المؤشرات
1	مادية: تتوافر بالبيئة الرقمية المقومات المادية الرقمية التي تؤهلها لتعليم اللغة العربية.	5
2	بشرية: تضم البيئة الرقمية الكفاءات البشرية التي تضمن عملها بكفاءة.	3
3	فنية: تتضمن البيئة الرقمية المقومات الفنية لتعمل بفاعلية.	2
4	نفسية: تحقّق عناصر البيئة الرقمية الجوانب النفسية الإيجابية اللازمة لتعليم اللغة العربية وتعلّمها.	2
5	اجتماعية: يحدث التعلّم اللغوي في البيئة الرقمية في سياق اجتماعي يتّسق مع طبيعة اللغة، وخاصيتها الاجتماعية.	2
6	علمية: تهيئ البيئة الرقمية مناخًا تعليميًا إيجابيًا للتعلّم اللغوي، وممارسة مهاراته.	3
7	فلسفية: تنطلق بيئة تعليم اللغة العربية وتعلّمها من فلسفة بيئات التعلّم الرقمي وأهدافه.	2
8	توفر بيئة تعلّم اللغة العربية وتعلّمها أشكالًا من مصادر المعلومات الرقمية.	11
9	توفر بيئة تعلّم اللغة العربية وتعلّمها خصائص البيئة الرقمية.	7
10	تنوّع في البيئة الرقمية أدوات البحث.	3

م	المعايير	عدد المؤشرات
11	تُتيح البيئة الرقمية فُرصًا لتعليم اللغة العربية وتعلّمها من خلال.	3
12	تُحقّق معايير البيئات الرقمية فيما يتصل بتقييم المعلومات اللغوية.	3
13	تُلبيّ متطلّبات تعلّم اللغة العربية، وتوافق مع ما يريده المتعلّم الرقّي.	4
14	تتوافق مقومات بيئة تعلّم اللغة العربية وتعلّمها مع تغيّر نماذج التعلّم والتدريس.	5
15	تتوافر ببيئة تعليم اللغة العربية وتعلّمها تقنيات جديدة للتعلّم.	2
16	تراعي بيئة تعليم اللغة العربية وتعلّمها طبيعة المتعلّمين وخصائصهم.	4
17	تعمل بيئة تعليم اللغة العربية وتعلّمها على تحقيق أهداف التعلّم الرقّي.	4
18	تراعي بيئة تعلّم اللغة العربية وتعلّمها احتياجات المتعلّمين اللغوية، وقدراتهم.	6
19	تُوظّف تطبيقات التعليم الرقّي في تعليم اللغة العربية وتعلّمها.	3
20	تتوافر في البيئة الرقمية مقومات الحفاظ على الخصوصية.	2

ج. معايير الكفاءة الرقمية لبيئات تعليم اللغة العربية وتعلّمها في الصفوف من (10-12)

جدول (27)

معايير البيئة الرقمية في الصفوف من (10-12)
وما ينبثق عنها من مؤشرات

م	المعايير	عدد المؤشرات
1	مادية: تتوافر بالبيئة الرقمية المقومات المادية التي تُوفّر مناخاً يُسهّم بفاعلية في تحقيق أهداف تعليم اللغة العربية.	4
2	بشرية: تضم البيئة الرقمية الكفاءات البشرية المؤهلة التي تضمّن عمل البيئات الرقمية بكفاءة واستمرارية.	2
3	فنية: تتضمن البيئة الرقمية المقومات الفنية لتعمل بفاعلية.	2
4	نفسية: تُحقّق عناصر البيئة الرقمية الجوانب النفسية الإيجابية اللازمة لتعليم اللغة العربية، وتعلّمها.	4
5	اجتماعية: يحدث التعلّم اللغوي في البيئة الرقمية في سياق اجتماعي يتّسق مع طبيعة اللغة، وخاصيتها الاجتماعية.	2
6	علمية: تُبني البيئة الرقمية مناخاً تعليمياً إيجابياً للتعلّم اللغوي، وممارسة مهاراته.	3
7	فلسفية: تنطلق بيئة تعليم اللغة العربية وتعلّمها من فلسفة بيئات التعلّم الرقمي وأهدافه.	2
8	توفر بيئة تعلّم اللغة العربية وتعلّمها أشكالاً من مصادر المعلومات الرقمية.	11
9	تتوافر في بيئة تعلّم اللغة العربية وتعلّمها خصائص البيئة الرقمية.	10
10	تتنوع أدوات البحث عن المصادر اللغوية في البيئة الرقمية.	3

م	المعايير	عدد المؤشرات
11	تُتيح البيئة الرقمية فرصًا لتعليم اللغة العربية.	2
12	تُحقق معايير البيئات الرقمية ما يتصل بتقييم المعلومات اللغوية.	3
13	تتوافق مقومات بيئة تعلم اللغة العربية وتعلمها مع تغيّر نماذج التعلم والتدريس.	2
14	تتوافر بيئة تعليم اللغة العربية وتعلمها تقنيات جديدة للتعلم.	7
15	تُراعى بيئة تعليم اللغة العربية وتعلمها طبيعة المتعلمين وخصائصهم.	9
16	تعمل بيئة تعليم اللغة العربية وتعلمها على تحقيق أهداف التعلم الرقمي.	3
17	تُراعى بيئة تعلم اللغة العربية وتعلمها احتياجات المتعلمين اللغوية، وقدراتهم.	4
18	توظف تطبيقات التعليم الرقمي في تعليم اللغة العربية وتعلمها.	4
19	تتوافر في البيئة الرقمية مقومات الحفاظ على الخصوصية.	2
20	تُتيح البيئة الرقمية فرصًا للتواصل.	2

ثالثاً: معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية:

أ- معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية في الصفوف من (1-4)

جدول (28)

معايير الكفاءة الرقمية لمُعَلِّمِي اللُّغة العربيَّة في الصفوف من (1-4)

م	المعايير	عدد المؤشرات
1	يُلمّ المعلم بالمهارات التكنولوجية الأساسية للتعامل مع الشبكة الدولية للإنترنت.	8
2	يُلمّ المعلم بالمهارات التكنولوجية الأساسية للتعامل مع الحاسب الآلي، وبرامجه، وتطبيقاته في مجال تعليم اللُّغة العربيَّة وتعلمها.	22
3	يملك المعلم معرفةً جيدةً بطرائق التواصل الإلكتروني، وخصائصها الأساسية من إرسال واستقبال، وخيارات، ونقل، وتخزين، وغيرها.	2
4	يُلمّ المعلم بكفايات استخدام التقنيات الرقمية وتطبيقها في تدريس اللُّغة العربيَّة.	2
5	يُمارس معلم اللُّغة العربيَّة أدواره الجديدة بوصفه مشخّصًا في ظل تقنيات العصر الرقفي.	3
6	يُمارس معلم اللُّغة العربيَّة أدواره الجديدة بوصفه مصمّم برامج في ظل تقنيات العصر الرقفي.	3
7	يُمارس معلم اللُّغة العربيَّة أدواره الجديدة بوصفه مخطّطًا وموجِّهًا للعملية التعليمية التعلمية في ظل تقنيات العصر الرقفي.	3
8	يُمارس معلم اللُّغة العربيَّة أدواره الجديدة بوصفه مهندسًا للسلوك، وضابطًا لبيئة التعلم في ظل تقنيات العصر الرقفي.	2
9	يُقيّم معلم اللُّغة العربيَّة المحتوى الرقفي المتاح لتعليم اللُّغة العربيَّة.	4
10	يُلمّ معلم اللُّغة العربيَّة بالمعارف المطلوبة في عصر التقنية.	4

ب- معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية في الصفوف من (5-9)

جدول (29)

معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربية في الصفوف من (5-9)

م	المعايير	عدد المؤشرات
1	يُلمّ المعلم بالمهارات التكنولوجية الأساسية للتعامل مع الشبكة الدولية للإنترنت.	4
2	يُلمّ المعلم بالمهارات التكنولوجية الأساسية للتعامل مع الحاسب الآلي، وبرامجه، وتطبيقاته في مجال تعليم اللغة العربية وتعلمها.	2
3	يملك المعلم معرفةً جيدةً بطرائق التواصل الإلكتروني وخصائصها الأساسية من إرسال واستقبال، وخيارات، ونقل وتخزين.	3
4	يُتقن المعلم استخدام التقنيات الرقمية وتطبيقها في تدريس اللغة العربية.	15
5	يُمارس معلم اللغة العربية أدواره الجديدة بوصفه مشخصاً تعليمياً في ظل تقنيات العصر الرقمي.	3
6	يُمارس معلم اللغة العربية أدواره الجديدة بوصفه مصمّم برامج تعليمية في ظل تقنيات العصر الرقمي.	2
7	يُمارس معلم اللغة العربية أدواره الجديدة بوصفه مخططاً وموجهاً للعملية التعليمية التعليمية في ظل تقنيات العصر الرقمي.	2
8	يُمارس معلم اللغة العربية أدواره الجديدة بوصفه مهندساً للسلوك وضابطاً لبيئة التعلم في ظل تقنيات العصر الرقمي.	2
9	يُقيّم معلم اللغة العربية المحتوى الرقمي المتاح لتعليم اللغة العربية.	4
10	يُمارس معلم اللغة العربية أدواراً تتناسب مع متغيرات العصر الرقمي.	4
11	يُلمّ معلم اللغة العربية بالمعارف المطلوبة للمعلم في عصر التقنية.	2

ج. معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربيّة في الصفوف من (10-12)

جدول (30)

معايير الكفاءة الرقمية لمُعَلِّمي اللغة العربيّة في الصفوف من (10-12)

م	المعايير	عدد المؤشرات
1	يُلمّ معلم اللغة العربيّة بالمهارات التكنولوجيّة الأساسيّة للتعامل مع الشبكة الدوليّة للإنترنت	3
2	يُتقن معلم اللغة العربيّة بالمهارات التكنولوجيّة الأساسيّة للتعامل مع الحاسب الآلي، وبرامجه، وتطبيقاته في مجال تعليم اللغة العربيّة وتعلّمها	3
3	يمتلك معلم اللغة العربيّة معرفةً جيدةً بطرائق التواصل الإلكترونيّ وخصائصها الأساسيّة، من إرسال، واستقبال، وخيارات، ونقل، وتخزين، وغيرها.	2
4	يُتقن معلم اللغة العربيّة كفايات استخدام التقنيات الرقمية وتطبيقها في تدريس اللغة العربيّة.	11
5	يُمارس معلم اللغة العربيّة أدواره الجديدة بوصفه مشخّصًا في ظل تقنيات العصر الرقفيّ.	2
6	يُمارس معلم اللغة العربيّة أدواره الجديدة بوصفه مصمّم برامج في ظل تقنيات العصر الرقفيّ.	2
7	يُمارس معلم اللغة العربيّة أدواره الجديدة بوصفه مخطّطًا وموجّهًا للعملية التعليمية التعلّميّة في ظل تقنيات العصر الرقفيّ.	3
8	يُتقن معلم اللغة العربيّة أدواره الجديدة بوصفه مهندسًا للسلوك، وضابطًا لبيئة التعلّم في ظل تقنيات العصر الرقفيّ.	2
9	يُمارس معلم اللغة العربيّة أدواره الجديدة بوصفه مخطّطًا وموجّهًا للعملية التعليمية التعلّميّة في ظل تقنيات العصر الرقفيّ.	2

م	المعايير	عدد المؤشرات
10	يُقيّم معلّم اللغة العربيّة المحتوى الرّقميّ المتاح لتعليم اللغة العربيّة.	5
11	يُمارِس معلّم اللغة العربيّة أدوارًا تتناسب مع متغيّرات العصر الرّقميّ.	5
12	يُلمّ معلّم اللغة العربيّة بالمعارف المطلوبة للمعلّم في عصر التقنية.	5

رابعاً: معايير الكفاءة الرّقمية لمُتعلّمي اللغة العربيّة:

أ- معايير الكفاءة الرّقمية لمُتعلّمي اللغة العربيّة في الصفوف من (4-1)

جدول (31)

معايير الكفاءة الرّقمية لمُتعلّمي اللغة العربيّة في الصفوف من (4-1)

م	المعايير	عدد المؤشرات
1	يكتسب التلميذ المهارات الأساسيّة لاستخدام الأجهزة الرّقمية.	5
2	يُمارِس التلميذ مهارات البحث عبر شبكات المعلومات الرّقمية (الإنترنت).	2
3	يُمارِس التلميذ مهارات التعلّم الذاتي الرّقميّ.	2
4	يوظّف التلميذ مهارات حل المشكلات.	2
5	يستخدم التلميذ التقنية في إنجاز المهام اللّغويّة.	4
6	يلتزم التلميذ السلوك الآمن والقانوني والأخلاقيّ عبر الإنترنت.	2
7	يكتسب التلميذ مهارات التعلّم التعاونيّ الإلكترونيّ عند تعلّمه اللغة.	4
8	يُشارك التلميذ في الأنشطة اللّغويّة الإلكترونيّة.	4
9	يتعرّف التلميذ بعض مفاهيم تقنيات التعليم.	4
10	يكتسب مهارات البحث عن المعلومات.	3

ب- معايير الكفاءة الرقمية لمتعلمي اللغة العربية في الصفوف من (5-9)

جدول (32)

معايير الكفاءة الرقمية لمتعلمي اللغة العربية في الصفوف من (5-9)

م	المعايير	عدد المؤشرات
1	يمارس المتعلم المهارات الأساسية لاستخدام الأجهزة الرقمية.	10
2	يُمارِس المتعلِّم مهارات البحث عبر شبكات المعلومات الرقمية (الإنترنت).	4
3	يُمارِس المتعلِّم مهارات التعلُّم الذاتي الرقمي.	5
4	يستخدم المتعلم مهارات مواجهة الأزمات.	3
5	يستخدم المتعلم التقنية في إنجاز المهام اللغوية.	4
6	يلتزم المتعلم السلوك الآمن والقانوني والأخلاقي عبر الإنترنت.	5
7	يُوظَّف المتعلِّم مهارات التعلُّم التعاوني الإلكتروني عند تعلُّمه اللغة.	4
8	يُشارِك المتعلِّم في الأنشطة اللغوية الإلكترونية.	11
9	يُوظَّف المتعلِّم المهارات الحياتية في تعامله مع التقنيات الرقمية.	2
10	يُطبِّق مفاهيم تكنولوجيا التعليم وعملياتها في تعلُّمه للغة.	5
11	يُحقِّق المتعلِّم من خلال التقنيات الرقمية مهارات الاتصال والتعاون الفعال.	3
12	يتعامل المتعلم مع التقنيات الرقمية؛ ليحقق مبدأ المواطنة الرقمية.	5
13	يبحث المتعلم عن المعلومات في الإنترنت بقدر مناسب من الدقة والطلاقة.	4

ج-معايير الكفاءة الرقمية لتعليم اللغة العربية في الصفوف من (10-12)

جدول (33)

معايير الكفاءة الرقمية لتعليم اللغة العربية في الصفوف من (10-12)

م	المعايير	عدد المؤشرات
1	يُتقن المتعلم المهارات الأساسية لاستخدام الأجهزة الرقمية.	3
2	يُتقن المتعلم مهارات البحث عبر شبكات المعلومات الرقمية (الإنترنت).	2
3	يُتقن المتعلم مهارات التعلم الذاتي الرقمي.	3
4	يوظف المتعلم مهارات حل المشكلات في مواقف التعلم اللغوي.	2
5	يستخدم المتعلم مهارات مواجهة الأزمات.	3
6	يستخدم المتعلم التقنية في إنجاز المهام اللغوية.	4
7	يلتزم المتعلم السلوك الآمن والقانوني والأخلاقي عبر الإنترنت.	8
8	يطبق المتعلم مهارات التعلم التعاوني الإلكتروني عند تعلم اللغة.	4
9	يشارك المتعلم في الأنشطة اللغوية الرقمية.	6
10	يوظف المتعلم المهارات الحياتية في تعامله مع التقنيات الرقمية.	1
11	يُتقن المتعلم عمليات التقنيات ومفاهيم العصر الرقمي.	3
12	يُحقق المتعلم -بفاعلية- متطلبات الاتصال والتعاون في العصر الرقمي.	4
13	يتعامل المتعلم كمواطن رقمي في مجتمع رقمي يفي بمتطلبات المواطنة الرقمية.	11
14	يبحث بطلاقة عن المعلومات والروابط.	2

خامساً: معايير تعليم فنون اللغة الرقمية:

جدول (34)

معايير تعليم فنون اللغة الرقمية وما ينبثق عنها من مؤشرات

م	المعايير	عدد المؤشرات
1	تُرَاعَى معاييرُ تدريس فنون اللغة العربية الرقمية المبادئ الأساسية للتدريس في العصر الرقمي.	15
2	تُحَقِّق معايير تدريس فنون اللغة العربية الخصائص الرئيسة للتعليم في العصر الرقمي.	6
3	تُحَقِّق معايير تدريس فنون اللغة العربية الخصائص الرئيسة للتعليم في العصر، الرقمي وأهداف التعلم الرقمي.	5
4	توظف معايير تدريس فنون اللغة العربية الرقمية الكفايات المنهجية لاستخدام التعلم الرقمي.	5
5	تعمل معايير تدريس فنون اللغة الرقمية على إكساب المتعلم مهارات العصر الرقمي، وتنميتها.	9
6	تُتِيح معايير تدريس فنون اللغة العربية فُرَصًا متساوية للتعلم؛ لتحقيق أهداف التعلم اللغوي.	9
7	تُمْكِّن طرق تدريس فنون اللغة العربية المعلم من أداء وظائفه في العصر الرقمي.	8
8	تُتِيح معايير تدريس فنون اللغة العربية الرقمية للمعلم أداء أدواره في العصر التكنولوجي، وما يترتب عليها من مسؤوليات.	6
9	تتفاعل معايير تدريس فنون اللغة العربية الرقمية مع ما يقدمه التعليم الرقمي للمتعلم.	8
10	تُتِيح معايير تدريس فنون اللغة العربية الرقمية للمتعلم فُرَصًا لتطوير مهارات الأداء اللغوي.	8

سادساً: معايير التقويم الإلكتروني في اللغة العربية:

جدول (35)

معايير التقويم الإلكتروني في اللغة العربية وما ينبثق عنها من مؤشرات

م	المعايير	عدد المؤشرات
1	يُخَطِّط المعلمُ عبرَ الإنترنت للتقويم اللُّغويّ، وينفذه في بيئات التعلم الرّقميّة بطُرُق تضمّن دقّة الأدوات والإجراءات وموثوقيتهما.	9
2	يُوظّف المعلمُ التقويمَ اللُّغويّ لأغراض التعرّف على الاحتياجات الأكاديميّة للمُتعلّمين، ويعمل على الوفاء بها.	7
3	يُصمّم المعلمُ أساليبَ التقويم الإلكترونيّة التكامليّة لتعزيز التنظيم الذاتي للمُتعلّمين وقياسه، والقدرات المرتبطة بالتعلم مدى الحياة.	6
4	يَستخدِم المعلمُ مهامَّ التقويم المختلفة لقياس مهارات الأداء اللُّغويّ.	4
5	يُوظّف المعلمُ إمكاناتِ الحاسوب، وبرمجياته، وتطبيقاته، والإنترنت في تصميم التقويم الإلكترونيّ وتنفيذه، والتعامل مع النتائج.	5
6	يُوظّف المعلمُ أدوات التقويم الإلكترونيّ لقياس المعارف والمفاهيم والمهارات اللُّغويّة.	5
7	يُطبّق المعلمُ المبادئ التي تقوم عليها الأساليب الجديدة في التقويم الإلكترونيّ.	7
8	يقيس المعلم -من خلال تطبيق أساليب التقويم الإلكترونيّ- ما يقوم به المتعلّم من أنشطة، ويُرَاقِب تقدّمهم في أداء المهام اللُّغويّة، ويُقدِّم تغذية راجعةً تساعد على تلافي أوجه القصور في التعلّم مرحلياً قبل الانتقال إلى مهمات تعليميّة جديدة.	6

م	المعايير	عدد المؤشرات
9	يُحقّق المعلمُ متطلّباتِ العمل مع المتعلّمين لاستخدام أساليب التقويم الإلكترونيّ.	17
10	يقيس المعلمُ -من خلال التقويم الإلكترونيّ- المعرفةَ والمفاهيمَ والمهاراتِ اللُّغويّةَ، ويساعد المتعلّمين على إدارة تقدّمهم عبر الإنترنت.	7
11	يُخطّط المعلمُ للتقويم الإلكترونيّ، محدّدًا مدى ملاءمة تقنيات التقييم الإلكترونيّ للتقييم، ومحدّدًا متطلّباته.	17

سابعاً: الدليل المقترح لمعلّمي اللغة العربيّة، ومتعلّميها لتوظيف التقنية في تعليم اللغة العربيّة وتعلّمها في ضوء الإطار العام المرجعي:
فلسفة الدليل:

تُوطّر الجوانب التالية لفلسفة إعداد الدليل:

- تحقيق أخلاقيات استخدام التقنيات بالتوازي مع اكتساب مهارات استخدامه في التعلّم اللُّغويّ.
- مسؤولية المعلم والمتعلم عن صحة استخدام التقنيات، ومراعاة قواعده.
- تدريب المعلم والمتعلم على تقدير المخاطر (Risk Assessment)، استخدام التقنيات، وتقييمها (Risk Evaluation)، وإدارتها (Risk Management).
- ربط التكنولوجيا الحديثة كالوسائط المتعدّدة والإنترنت وبرامج الحاسب المختلفة في المحتوى الدراسي لمنح عمليّتي التربية والتعليم والتعلّم بُعداً إضافيّاً؛ بهدف رفع مستوى التحصيل الدراسي من خلال استغلال تقنية المعلومات بما توفره من أدوات جديدة للتعلّم والتعليم.

- توفير التقنيات واستخدامها في المدارس يُسهم في جعل المتعلم متعلماً نشطاً وفعالاً وباحثاً عن المعلومة، ومُدرِّكاً لأهمية التعلم، وقادراً على اكتساب مهارات حل المشكلات، والاستقصاء، والتواصل، والتنوع.
- الإسهام في التحول الجذري إلى الخدمات الإلكترونية بغية الاستغناء عن استخدام الورق والكتب الدراسية.
- تحقيق جودة التعليم، وتوسيع مدارك المتعلم، والتواصل المباشر والنشط بين الطالب والمعلم.
- تنمية مهارة الإبداع والاستكشاف لدى المتعلم.
- تحويل الفصول التقليدية إلى عالم افتراضي يحاكي الواقع.
- خلق فضاء تفاعليّ بإمكانيات مثيرة ومثيرة لتفكير المتعلم ومعرفته في آنٍ واحدٍ، وفي ظل بيئة صحيّة وأمنة.
- تيسير استخدام المنصّات التعلّميّة، والأجهزة المحمولة، وبرامج التواصل الاجتماعيّ التي تُوفّر بيئةً آمنةً للاتصال والتعاون، وتبادل المحتوى التعليمي بين المتعلم والمعلم.
- مساهمة التعليم للتغيرات المتسارعة الناجمة عن التقدّم العلميّ والتّقنيّ، وتقنيّة المعلومات؛ لذا أصبح من الضروريّ مواكبة العمليّة التربويّة لهذه التغيرات لمواجهة المشكلات التي قد تنجم عنها؛ مثل: كثرة المعلومات، وزيادة أعداد المتعلّمين، ونقص المتعلّمين المؤهلين، وبُعد المسافات، وازدياد الحاجة إلى التعليم، وقد أدّت هذه التغيرات إلى ظهور أنماط وطرائق عديدة للتعليم الفرديّ، أو الذاتيّ الذي يسير فيه المتعلّم حسب طاقاته وقدراته، وسرعة تعلمه، ووفقاً لما لديه من خبرات ومهارات سابقة، لمواجهة هذه التغيرات، ظهرت مفاهيم: التعليم المبرمج، والتعليم باستخدام الحاسوب، والتعليم عن بُعد.

- مع ظهور تقنية المعلومات زادت الحاجة إلى تبادل الخبرات مع الآخرين، وحاجة المتعلم إلى بيئات غنية متعددة المصادر للبحث والتطوير الذاتي، فظهر التعليم الإلكتروني، الذي يعتمد على التقنيات الحديثة للحاسوب، والشبكة العالمية والوسائط المتعددة.

التوصيات:

- توظيف منظومات المعايير التي نتجت عن الدراسة في تعليم اللغة العربية.
- إعادة النظر في برامج إعداد معلم اللغة العربية بما يتواءم مع معايير المتعلم الرقمي.
- توفير البنى التحتية بما يتيح فرصاً لإعداد بيئات التعلم في العصر الرقمي.
- تطبيق معايير تدريس فنون اللغة العربية الرقمية، وما ينبثق عنها من مؤشرات
- تطبيق معايير التقويم الإلكتروني للغة العربية، وما ينبثق عنها من مؤشرات.

قائمة المراجع

- 1- إبراهيم، على حسن، عبد الله يوسف الفيلكاوي، "مدى تحقق كفايات التعليم الإلكتروني لدى أعضاء هيئة تدريس كلية التربية بجامعة الكويت"، المجلة التربوية، 13، 2018، 55-32
- 2- الألوسي، تيسر عبد الجبار، التعليم الإلكتروني ما قبل كورونا وما بعده، "السومريان، متاح من خلال الرابط، شوهدي في 2020/1/7.
3/5/2020.com.slates-somerian.www://http/ 2020/05/03/9819 /
- 3- بدرانة، عبد الله، دور التعليم الرقمي في مواجهة الأزمات والتحديات الراهنة، "سفيريس"، متاح عبر الرابط، 2020، شوهدي في 2021/1/7
aYUH/ws.bitly://htztp
- 4- زيناتي، محمد بشر أيمن، توظيف التقنيات الحديثة في تعليم اللغة العربية، جامعة إسطنبول، مجلة العلوم الإنسانية والطبيعية، 2021،
www.hnjournal.net
- 5- البشير، عادل أكرم، "معيقات توظيف مناهج اللغة العربية الإلكترونية في مراحل التعليم العام"، بمحافظة الجبراء بدولة الكويت: دراسة تحليلية، مجلة العلوم التربوية والنفسية، مج2019.
- 6- السعدون، نجلاء مهنا، والبازعي حصة حمود، "أدوار معلم المرحلة الثانوية في ضوء متطلبات المدرسة الافتراضية بمحافظة حفر الباطن"، مجلة الفنون والأدب وعلوم الإنسانيات والاجتماع، 64، 2021.
- 7- السيد، محمود مصطفى محمود، توظيف الفصول الافتراضية في تدريس اللغة العربية من بعد جوجل كلاس روم نموذجاً، الشارقة، المركز التربوي

- لغة العربية لدول الخليج مقدم لملتقى أفضل الممارسات والتجارب في مجال توظيف التقنيات في مارس 2021 تدريس اللغة العربية عن بعد، 1-14، 2021.
- 8- شلبي، نوال محمد، "إطار مقترح لدمج مهارات القرن الحادي والعشرين في مناهج العلوم بالتعليم الأساسي في مصر"، المجلة التربوية الدولية المتخصصة، 2014.
- 9- الشمراني، عليه أحمد يحيى آل حمود، "أثر توظيف التعلم الرقمي على جودة العملية التعليمية وتحسين مخرجاتها"، المجلة العربية للعلوم التربوية والنفسية: المؤسسة العربية للتربية والعلوم والآداب، (8)، 2019.
- 10- كرم، على (2021). ما التعليم الرقمي؟ وأهميته ومميزاته؟ "سواح هوست"، 18 ديسمبر 2020، 2021. متاح من خلال الرابط شوهد في 1/7
aYVq/ws.bitly://http
- 11- مجاهد، فائزة أحمد الحسيني، "التعليم الإلكتروني في زمن كورونا المآل والآمال"، المجلة الدولية للبحوث في العلوم التربوية، 2020.
- 12- المركز التربوي للغة العربية لدول الخليج، مؤتمر اللغة العربية الدولي الاستثنائي عن بُعد التعليم عن بُعد في تدريس اللغة العربية: (الواقع، والمتطلبات، والآفاق) المنعقد بالمركز التربوي للغة العربية بالشارقة، 2020.
- 13- نصر، معاطي، سليمان، محمود جلال، تكنولوجيا تعليم اللغة العربية وتعلمها للعاديين وغير العاديين. دمياط: مكتبة نانسي، 2019.
- 14- وزارة الثقافة والشباب، الإمارات العربية المتحدة، تقرير حالة اللغة العربية ومستقبلها، 20 ديسمبر، 2020.
- 15- وطفة، على أسعد، "إشكاليات التعلم الإلكتروني وتحدياته في ظل جائحة كورونا، قراءة سوسيولوجية في جدليات التفاعل والتأثير"، مجلة مركز دراسات الخليج والجزية العربية، جامعة الكويت، 2021.

المراجع الأجنبية:

- 1- Abbasova, Milana& Mammadova, Narmin "The Role of Digital Technology in English Language Teaching in Azerbaijan", *International Journal of English Linguistics* 9(2):364. March 2019.
- 2- Butler, Deirdre (DC), learning Digital Learning Framework for Primary Schools, National Council for Curriculum & Assessment; *H2 1-16 ISTE COMMUNITY STANDARDS*. ISTE Standards for Students, متاح من خلال الرابط، شوهد في 2020/10/3: <https://www.iste.org/standards/iste-standards-for-students>. 2016
- 3- Department of Education and Skills, Digital Learning 2020: Reporting on practice in Early Learning and Care, Primary and Post-Primary Contexts. متاح من خلال الرابط، شوهد في 2020/10/1: file:///C:/Users/TABARK/Downloads/Documents/78007_a37413f9-7423-44bf-86df-01762e04a408.pdf
- 4- Hutchison, Amy C. & Woodward, Lindsay: *Examining the Technology Integration Planning Cycle Model of Professional Development to Support Teachers' Instructional Practices*. *Teachers College Record*, 20(10).
- 5- Jack, Christine & Higgins, Steven: E *What is educational technology and how is it being used to support teaching and learning in the early years?*, August 2018, *International Journal of Early Years Education* 27(4):1-16
- 6- 1 U.S. Department of Education. (2017). *Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update*, Office of Educational Technology.

(1) ينظر: عبيزة، د. عائشة، البرامج التفاعلية الإلكترونية ودورها في تعليم اللغة العربية للناطقين بغيرها، (بحث مقدم للمؤتمر الدولي الأول، تركيا، 2019م، ص36.



توصيف العربية حاسوبياً لغير الناطقين بها **- إشكالات وتحديات -**

Computerized description of Arabic for non-native speakers

إعداد

أ.د. محمد بن عبد العزيز العميري

أستاذ اللغة العربية في جامعة محمد بن زايد للعلوم الإنسانية

الملخص

يُعد التعليم الإلكتروني مفهومًا حديث الظهور على سطح عالم التعليم الحديث، مواكبًا لظهور شبكات الإنترنت وانتشارها المذهل في التطبيقات التعليمية المختلفة، وتبرز في الساحة المعلوماتية مصطلحات عدة تحاول وصف هذه الظاهرة التكنولوجية التعليمية حديثة الظهور، منها: التعليم الإلكتروني، والتعليم المحوسب، والتعليم الافتراضي، والتعليم عبر الإنترنت، ونحو ذلك من المصطلحات التي تتباين صياغتها، لكنها تشير في الوقت ذاته إلى جميع أنواع الأنشطة التعليمية التي يتم تصميمها وتنظيمها وإيصالها من خلال شبكة الإنترنت

وأشارت الدراسات المختلفة إلى أهمية توظيف التكنولوجيا في تعليم العربية لغير الناطقين بها، إذ يحقق ذلك مبدأ التفاعل الجيد بين الأستاذ والمتعلم، وتجاوز مرحلة التلقين والطرح الجاف للمعلومة، كما أنه يحدث تحسنًا كبيرًا في اتجاهات التلاميذ نحو التعلم

هذا الاتجاه التقني في التعليم يُمكن من إدارة تعليم اللغة وتعلمها في ضوء برمجيات تعليمية، ومقررات إلكترونية نشطة من أجل إكساب المتعلمين مهارات اللغة، لتحقيق التواصل اللغوي البناء، والتعامل الدقيق مع كل مقوماته

الكلمات المفتاحية: توصيف، الحاسوب، إشكالات، تحديات، توظيف

Abstract

Electronic learning (e-learning) has spread widely and internationally, side by side with traditional learning. Among these e-learning methods are computational learning, virtual learning, and distance learning. Many educational applications have been developed to be used to teach various types of educational activities that were developed through networks. Many studies show the importance of technology in learning and teaching Arabic to non-native Arabic speakers. This contributes to creating a positive interaction between teachers and learners. It also helps to transfer teaching Arabic from the repetitive classroom to the learner-centered classroom. This new technical direction in education enables Arabic language institutions to produce many electronic programs that help learners develop new skills and acquire Arabic language easily.

Keywords: Description-Computer-Problems-Challenges-Function.

مدخل

يُعد التعليم الإلكتروني مفهومًا حديث الظهور على سطح عالم التعليم الحديث، مواكبًا لظهور شبكات الإنترنت وانتشارها المذهل في التطبيقات التعليمية المختلفة، وتبرز في الساحة المعلوماتية مصطلحات عدة تحاول وصف هذه الظاهرة التكنولوجية التعليمية حديثة الظهور، منها: التعليم الإلكتروني، والتعليم المحوسب، والتعليم الافتراضي، والتعليم عبر الإنترنت، ونحو ذلك من المصطلحات التي تتباين صياغتها، لكنها تشير في الوقت ذاته إلى جميع أنواع الأنشطة التعليمية التي يتم تصميمها وتنظيمها وإيصالها من خلال شبكة الإنترنت⁽¹⁾.

وأشارت الدراسات المختلفة إلى أهمية توظيف التكنولوجيا في تعليم العربية غير الناطقين بها، إذ يحقق ذلك مبدأ التفاعل الجيد بين الأستاذ والمتعلم، وتجاوز مرحلة التلقين والطرح الجاف للمعلومة، كما أنه يحدث تحسنًا كبيرًا في اتجاهات التلاميذ نحو التعلم⁽²⁾.

هذا الاتجاه التقني في التعليم يُمكن من إدارة تعليم اللغة وتعلمها في ضوء برمجيات تعليمية، ومقررات إلكترونية نشطة؛ من أجل إكساب المتعلمين مهارات اللغة؛ لتحقيق التواصل اللغوي البناء، والتعامل الدقيق مع كل مقوماته⁽³⁾.

(1) ينظر: شحادة، أمل عايد، التكنولوجيا التعليمية، دار كنوز المعرفة للنشر والتوزيع، عمان، ط2، 2010م، ص18.

(2) ينظر: الزهراني، مرضي، المدخل التقني في تعليم اللغة العربية، مفهومه وأسس ومطالبه وتطبيقاته، ورقة عمل مقدمة إلى المؤتمر العالمي الأول للغة العربية وآدابها، ماليزيا، ص5.

(3) صالح، عبد الرحمن الحاج، بحوث ودراسات في اللسانيات العربية، موفم للنشر، الجزائر، 1966م، ص62.

في ظل هذا الاتجاه التعليمي العام الذي يقصد إلى استعمال (الشبكة العنكبوتية) وتيسيرها للمتعلم، أصبح من المهم لدى علماء العربية الإفادة من تلك التقنية الحاسوبية في الدرس اللغوي والنحوي، وتمكين التعليم الإلكتروني ليصل إلى أعداد كبيرة من الدارسين، وتوظيف التقنيات الحديثة في إنشاء برامج متنوعة، ذات صياغات محددة ومناسبة، يُقصد منها تعليم العربية بأسر الطرق، وأكثرها جدية خاصة للناطقين بغيرها، لذا بات مجال الحوسبة اللغوية أو اللسانيات الحاسوبية ميداناً علمياً وتطبيقياً واسعاً يشمل التطبيقات المتعددة كتعليم اللغات بالحاسوب، وإصلاح الأخطاء المطبعية آلياً، والترجمة الآلية⁽¹⁾.

من هذا المنطلق فإنه من المؤكد أن الدراسات النحوية يمكنها أن تُفيد من هذا الجانب التقني إفادة واضحة، بعد تطويره وصياغته ليقدم الجوانب التطبيقية، والقواعد المختلفة في النحو العربي، عن طريق تمكين المعرفة، والإلهام الذهني الذي يتماشى مع إمكانات الحاسوب وقدرته على تفهم تلك القواعد، وسلكها في نظام التطبيق العلمي القائم على الدقة والضبط، خاصة إذا قام بتلك العملية الفنية والعلمية أساتذة متخصصون في هذا الجانب المهم، كما يستوجب ذلك "الاستيعاب الكامل لإمكانية إدخال اللغة العربية إلى الحاسوب، وإمكانية معالجتها داخل الحاسوب، إضافة لإمكانية طباعتها بشكل عادي"⁽²⁾.

(1) خضر، محمد، الحرف العربي والحوسبة، الموسم الثقافي لمجمع اللغة العربية، عمان - الأردن، 2001م، ص2.

(2) ينظر: شريف، سمير، اللسانيات، المجال والوظيفة والمنهج، استيتة، عالم الكتب الحديث، الأردن، ط1، (2005م)، 107، وعلي، نبيل، اللغة والحاسوب، مؤسسة تعريب الكويتية، الكويت، 1988م، 111.

وقد ألمح عدد من الباحثين إلى أنه بمقدور اللغويين والحاسوبيين تأمين مركز مرموق في حقل اللسانيات الحاسوبية، وأن يمارسوا دوراً مهماً في تمكينه وإثرائه عندما يتركز اهتمامهم بالموضوعات والجوانب ذات الصلة باللغة العربية، وإدخال التقنيات المعلوماتية في البلدان العربية⁽¹⁾.

وقد دعا عدد من الدارسين إلى وجوب معالجة اللغة العربية حاسوبياً، وأنه يُعد أمراً مهماً لا يمكن تجاهله، لا سيما أن الدراسات الحاسوبية والمعلوماتية حققت نتائج كثيرة للعربية في مجالات متعددة كالتعلم، والتعريب والإحصاء اللغوي والترجمة الآلية، والمعالجة الآلية، وغيرها من المجالات. وأشاروا إلى أن العربية تواجه اليوم تأخراً وضعفاً تقنياً بسبب قلة الاهتمام بمثل تلك الجوانب⁽²⁾.

ومما ينبغي الاهتمام به هذا الجانب هو المحاولة الجادة في صياغة قوانين النحو العربي وقواعده صياغة مناسبة يمكن تنفيذها عن طريق برامج الحاسوب، تكون ملائمة للمحتوى العربي، وقابلة للتطبيق المرن من خلال برامج تقنية تصاغ من قبل المختصين والمهتمين الذي يتمتعون بالدربة والدراية في هذا المجال (العلمي والتقني في آن واحد) من أجل تجنب الوقوع في الخطأ أو الخلط، ومما قد يلاحظ في هذا الاتجاه أن "التردد الملحوظ من قبل نحاة عصرنا في حوسبة النحو العربي، حمل الحاسوبيين الذي لا تتجاوز خبرتهم التقانة الفنية على الاستئثار بحوسبة اللغة العربية، فتجروؤوا على قواعد النحو وطوّعوها لأنظمة الحاسوب بعيداً عن أعين النحويين، فأنجزوا في هذا الجانب أعمالاً

(1) ينظر: الحاج، وليد، اللغة العربية ووسائل الاتصال الحديثة، دار البداية، عمان، (2007م)، 22.

(2) الأنصاري، عبد الله، علاقة النحو العربي بالبرمجة الآلية للغة، جامعة الإمام محمد بن سعود الإسلامية، (1435هـ)، 6.

مركوزة على الجانب الفني التقني المجرد، وغاب الجانب اللغوي الذي يلامس اللغة مباشرة، فجاءت نتائج الأبحاث اللغوية التي أنجزوها متأثرة بالمنطق الرياضي أكثر من تأثرها بقوانين العربية⁽¹⁾، من أجل هذا فإن: "حوسبة النحو ومعالجته آلياً تتطلب تداخل الفكر اللساني مع العقل الحاسوبي من أجل تحقيق الغاية"⁽²⁾، غير أن "حوسبة اللغة تستدعي الدراسة المنطقية لها، لأن متطلبات الحاسوب تقتضي تحديد أشكال الوحدات اللغوية بواسطة رموز رياضية لكي يمكن تمريرها إلى الحاسوب"⁽³⁾.

يتضح مما سبق أن حوسبة اللغة تحتاج إلى إعادة صياغة لكثير من الضوابط والقواعد المستمدة من كتب النحو العربي، من غير إخلال بالعربية ولا تقويض لنظامها⁽⁴⁾.

مفهوم اللسانيات الحاسوبية:

لا يزال تعريف اللسانيات الحاسوبية أو حوسبة اللغة غير مستقر، وما ذاك إلا لتعدد ترجمات المصطلح في الأصل، إضافة إلى التشتت في تحديد المجالات التي تشكل محور الدرس فيه⁽⁵⁾.

غير أن بعض الباحثين حاول تعريفها انطلاقاً من مجالها العام الذي تنطلق منه، أو من خلال التطبيقات المختلفة التي تمثلها، فذكروا أنها: "المجال الذي

(1) الكمار، رأفت، الحاسوب وميكنة اللغة العربية، دار الكتب العلمية، القاهرة، ط1، (2006م)، 440.

(2) فريحي، مليكة، "النحو الحاسوبي"، مجلة عود الند، ع82، الجزائر.

(3) ينظر: علاقة النحو العربي بالبرمجة الآلية للغة، 12.

(4) ينظر: أحمد، رضا بابا، اللسانيات الحاسوبية مشكل المصطلح والترجمة، جامعة تلمسان، الجزائر، 4-5.

(5) السابق، 2.

ترتبط فيه اللسانيات أو علوم اللغة العربية كلغة من اللغات الطبيعية، تدخل في علم مخصص وليد التطورات التكنولوجية المتقدمة...، ومجالها البحثي دقيق وجديد يعرض آخر النظريات والتطبيقات الحاسوبية المجربة على جميع اللغات"⁽¹⁾.

وارتأى د. العارف تعريف اللسانيات الحاسوبية من خلال تنوع مستوياتها واختلافها من حيث التنظير والتطبيق، فيشير إلى أنه إذا أردنا تعريف هذا العمل بشكل مختصر قلنا: إنه العمل الذي يبحث في اللغة البشرية كأداة طيعة لمعالجتها في الحاسبات الإلكترونية، وهي بهذا الشكل تتألف من مبادئ اللسانيات العامة بجميع مستوياتها التحليلية الصوتية النحوية، والدلالية، ومن علم الحاسبات الإلكترونية، ومن علم الذكاء الاصطناعي وعلم المنطق والرياضيات⁽²⁾. وهي بهذا التعريف تحقق غرض الدراسات البينية التي تجمع بين العلم باللغة، وآلية عمل الحاسوب وتطبيقاته المختلفة، وتتألف بهذا الشكل من مكونين رئيسين: الأول: تطبيقي يهتم بالنتائج العملية، ويهدف إلى إنتاج برامج مزودة بكثير من المعارف اللغوية.

والآخر: نظري يتناول قضايا في اللسانيات النظرية، أبرزها إقامة نظريات صورية للمعرفة اللغوية التي يحتاج إليها الإنسان لتوليد اللغة وفهمها⁽³⁾.

(1) ينظر: العارف، عبد الرحمن، "توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، جهود ونتائج"، مجلة مجمع اللغة العربية الأردني، الأردن، عدد (73)، (2007م)، ص 18.

(2) ينظر: الموسى، نهاد، حصاد القرن في اللسانيات، مؤسسة عبد الحميد شومان، الأردن، (2000م)، 47/2.

(3) ينظر: قنديل، أحمد، التدريس بالتكنولوجيا الحديثة، عالم الكتب، القاهرة، 2006م، ص 6.

ثم اتسع مجالها بعد ذلك ليشمل تخطيط وتنفيذ عناصر عملية التدريس على أسس علمية⁽¹⁾.

وتتعدد مجالات استعمال اللسانيات الحاسوبية، ويمكن أن تتوسع حسب الحاجة، وحسب تنوع المجالات، ومن أبرز مجالاتها اليوم:

- 1- الإحصاء اللغوي والذي يشمل التصريف والجذور والمشتقات.
- 2- التحليل الصرفي الآلي، وذلك في عمليات الإحصاء للجذور ومشتقاتها اللغوية.
- 3- الترجمة الآلية.
- 4- الدراسات المقارنة والتقابلية.
- 5- التدقيق الإملائي والنحوي، وهنا- لا يمكن الاعتماد على الحاسوب في التدريس اللغوي، إذ لا بد أن يراجع مختص بعد ذلك.
- 6- تعليم العربية لغير الناطقين بها، ويدخل ذلك ضمن الدراسات اللسانية التقابلية التي تسرع عملية التعليم، وتقدمه بصورة مبسطة.
- 7- تحويل النص إلى كلام والعكس، وتعد هذه العملية من أكثر العمليات نفعا من تطبيقات اللسانيات الحاسوبية؛ لأن فائدها عامة ومشتركة بين جميع الفئات⁽²⁾.

(1) ينظر: أمهان، طارق عبد الحكيم، اللسانيات الحاسوبية ومشكلة حوسبة اللغة العربية، خطوة باتجاه الحل، بحث منشور في شبكة الألوكة، ص 10-11.

(2) يراد منها تجديد المناهج النحوية في كتابه (الرد على النحاة) حيث يهاجم نظرية العامل وما تصوره النحاة لعواملهم وما تؤدي إليه من تقديرات وعلل وأقيسة ملأت صفحات النحو بمسائل معقد لا تجدي في تقويم اللسان. ينظر: فكرة تجديد النحو عند شوقي ضيف بين النظرية والتجديد.

مشكلات تعليم العربية لغير الناطقين بها

المطلب الأول: مشكلات ذاتية:

أول ما يعيق العملية التعليمية للغة العربية للناطقين بغيرها؛ ما يعود إلى اللغة العربية نفسها، فالمشتغلون باللسانيات العربية خاصة اللغويات التطبيقية، طرّقوا أبواباً عديدة من الصعوبات التي وصفوها بالكؤود، فمنذ قرون طويلة وهم يحاولون التيسير والتسهيل، منذ محاولات ابن مضاء القرطبي⁽¹⁾ مروراً باجتهادات الدكتور نهاد الموسى، إلى محاولات الدكتور شوقي ضيف وغيرهم، ويعترف علماء اللسانيات العربية في شتى البلاد العربية بأن عملية تعليم العربية يواجه مشكلات كثيرة وصعوبات جمّة؛ لذلك نجدهم عاكفين على محاولات تيسير عملية تعليم العربية، فإذا كان هذا حالهم مع أبناء العربية، فمن الطبيعي أن يواجه مشكلات أكبر أو أكثر تعقيداً عند متعلمي العربية من الناطقين بغيرها، وإن اختلفت في طبيعتها ودرجة صعوبتها.

إن اللغة العربية تعدّ من أغزر اللغات مادة وأطوعها في تأليف الجمل وصياغة العبارات، فهي لغة مليئة بالألفاظ والكلمات التي تناسب مدارك بنائها وتتفرع في المرحلة الأساسية إلى أنماط لغوية.

ولما كان لهذه اللغة تلك الأهمية البالغة على النطاق الديني والفكري والعلمي؛ أصبح لزاماً على روادها العمل على تيسير تعلمها وكسر جدر الصعوبة التي تمخضت في نفوس الناشئة وهي في حقيقتها ليست صعبة كما يعتقد البعض لكن صعوبتها تكمن في الطرائق والأساليب المتبعة في تدريسها⁽²⁾.

(1) ينظر: الحديدي، علي، مشكلة تعليم اللغة العربية لغير العرب، طبع دار الكتاب العربي للطباعة والنشر، القاهرة (1967م، ص135).

(2) ينظر: ماسيري، ديكوري، المشكلات الصوتية في تعلم اللغة العربية للناطقين بغيرها، جامعة المدينة العالمية أنموذجاً، (عضو هيئة التدريس في كلية اللغات - ماليزيا)، مقدمة الدراسة.

وهنا نلقي الضوء على بعض الجوانب التي لمسها المعنيون والقائمون على تذليل صعوباتها، ووضع التدابير الملائمة لحلها والتي تلمس اللغة العربية ذاتها ومن ذلك:

1-المشكلات الصوتية:

يهدف هذه الجانب إلى التعرف على المشكلات الصوتية التي تواجه الدارس الناطق بغير العربية عند تعلّمه للعربية، والتعرف على التحديات التي تواجه المعلم القائم بتدريس هذه اللغة لهذا الصنف من الطلاب، وقد أدرك الباحثون خلال دراساتهم التطبيقية القائمة على تحليل عيّنات واقعية لمتعلمين للغة العربية من غير أبنائها واستخدام المنهج المسحي الكمي⁽¹⁾ أن المشكلات المتعلقة بالنظام الصوتي من أعمق وأكبر المشكلات في تعلّم اللغة العربية للناطقين بغيرها؛ وعليه ظهرت الحاجة إلى تحليل هذا المستوى اللغوي (المستوى الصوتي)، لدى دارسي العربية من غير أبنائها؛ من أجل الوقوف على أهمّ مشكلات النظام الصوتي، ثم اقتراح بعض الحلول والآليات المناسبة لها.

إن من أهم مشكلات تعلم اللغة العربية لغير الناطق بها؛ تأثره بلغته الأم "ونقل بعض الجوانب اللغوية إلى اللغة العربية فنجدّه -مثلاً- يحاول أن ينقل عادات النطق من لغته الأم إلى اللغة العربية، فنشأ على أثره مشكلات في النظام الصوتي التي قد تعوق مسيرة التعلم الصحيح لديه"⁽²⁾.

من هنا يمكن أن نقرر أن أهم ما يشكل الصعوبات اللغوية هو التداخل

(1) الباكير، مجد محمد، مشكلات اللغة العربية المعاصرة، طبع مكتبة الرسالة الحديثة عمان - الأردن (1989م)، ص 11.

(2) ينظر: الراجحي، عبده، النحو العربي والدرس الحديث، طبع دار النهضة العربية، بيروت- لبنان، ط6، (1406هـ- /1986م)، ص 116.

(interference) بين اللغة العربية، ولغاتهم الأصلية في الجوانب الصوتية والنحوية والدلالية والكتابية.

كما ترجع صعوبة نطق الأصوات العربية إلى أنها لا توجد في كثير من لغات العالم؛ لذلك يواجه معظم متعلمي اللغة العربية صعوبة في نطق أصواتها، والملاحظ أن هذه الصعوبات التي تمثل عوائق لغوية أمام متعلميها قد تتفاوت درجتها من شخص إلى آخر؛ بناءً على العوامل اللغوية، والشخصية، والتعليمية، ويرى علماء اللغة التطبيقيون⁽¹⁾ أن وقوع المتعلم الأجنبي بهذه الأخطاء يعود إلى أربعة أسباب:

- اختلاف اللغتين في مخارج الأصوات.
- اختلاف اللغتين في التجمعات الصوتية.
- اختلاف اللغتين في مواضع النبر والتنغيم والإيقاع.
- اختلاف اللغتين في العادات النطقية.

وهذه الصعوبات قد تختلف من شخص لآخر تبعاً لطبيعة اللغة الأم التي يتحدث بها المتعلم، ولن يواجه المتعلم أية صعوبة في تعلم ونطق أحرف مشابهة لأحرف اللغة الأم؛ فمثلاً: لا ترى صعوبة في نطق المتحدث باللغة الإنجليزية لأحرف: (الباء التاء الثاء الجيم الدال الذال الراء السين الشين الزاي الكاف الميم النون اللام)، لكن من الصعب تعلم بعض الأصوات المطبقة؛ مثل: (الصاد الضاد الطاء الظاء)؛ لأن هذه الأصوات غير موجودة في لغته الأم⁽²⁾.

(1) ينظر: المشكلات الصوتية في تعلم اللغة العربية للناطقين بغيرها - جامعة المدينة العالمية أنموذجاً، ص 24.

(2) ينظر: زقوت، محمد شحادة، المرشد في تدريس اللغة العربية، طبع مكتبة الأمل للطباعة والنشر، غزة، ط 2، (1999م)، ص 90.

إن اللغة العربية متميزة من الناحية الصوتية، حيث إن العربية أكثر أخواتها احتفاظاً بالأصوات السامية، فلقد اشتملت على جميع الأصوات التي اشتملت عليها أخواتها السامية، ما عدا (الباء والفاء)، بل وزادت عليها بأصوات كثيرة لا وجود لها في واحدة منها مثل: (الثاء والذال والغين والضاد)، كما أن للحروف العربية مخارجها الدقيقة، التي قد يقع الخلط في نطقها نتيجة لتقارب المخارج، كما أن تميزها من الناحية الصوتية يتمثل -أيضاً- في ثبات أصوات الحروف فيها لدقة مخارجها، الأمر الذي جعل طريقة نطق الأصوات ثابتة عبر العصور⁽¹⁾.

لكن ينبغي التنبيه إلى أن المتعلم لا يجد صعوبة في نطق حروف لغة ما، إذا كان تعلمه لها في وقت الصغر وكانت ممارسته لها من خلال التواصل المباشر مع المجتمع على عكس المتعلم كبير السن؛ لأنه لم يمارس اللغة تواصلياً فصعب عليه نطق حروف لم توجد في لغته الأم وقد تكون موجودة فيها لكنها تنطق نطقاً مختلفاً عما هو عليه في اللغة الثانية⁽²⁾.

إن معالجة مشكلة الأصوات ليست من السهولة بمكان للناطقين بغير العربية في مرحلة متأخرة من العمر؛ لأن الجهاز الصوتي قد تشكل وأخذ قالباً مناسباً لأصوات اللغة الأم وأصوات الطفولة المكتسبة مثل اللهجات أو اللغة الثانية الشبيهة باللغة الأم⁽³⁾.

(1) ينظر: بوقرة، نعمان، النظرية اللسانية عند ابن حزم الأندلسي قراءة نقدية في مرجعيات اللسان الخطابي وأبعاده المعرفية، منشورات اتحاد الكتاب العربي، دمشق، (2004م)، 48/1.

(2) ينظر: مذكور، علي، تعليم اللغة العربية لغير الناطقين بها، النظرية والتطبيق، طبع دار الفكر العربي، القاهرة، ط1، (1427 هـ-2006م)، ص 7.

(3) ينظر: المرشد في تدريس اللغة العربية، ص 81-84.

2- جوانب لغوية أخرى:

بجانب تلك المشكلة الصوتية؛ فإن ثمة مشكلات أخرى تعود إلى ما تتفرد به اللغة العربية من سمات وخصائص لا تكاد توجد في غيرها من اللغات، فيجد المتعلم صعوبة في تفهمها وإدراكها؛ مثل: إشكالية الإعراب وتفهم الدلالة والمعنى عن طريقه، وافتقاد قيمته في تنسيق الجمل وترتيبها وفق نظام خاص، وازدواجية اللغة في البلاد العربية التي يواجهها الدارس، وكثرة المترادفات في العربية التي تؤدي إلى الاضطراب التعليمي لدى الدارسين، حيث تتمتع بثناء عز نظيره في معظم لغات العالم، وظاهرة الترادف وما نتج عنها من ثراء في مفردات اللغة والقدرة على التوليد، إذ هي لغة اشتقاق ومع وجود هذه الظاهرة في بعض اللغات الأخرى إلا أنها في العربية أوسع وأغنى⁽¹⁾، ومرونتها وطواعية ألفاظها للدلالة على المعاني، وهي أوسع وأدق في قواعد النحو والصرف من أخواتها السامية⁽²⁾، إضافة إلى ذلك فلكل صوت في العربية صفة ومخرج، وإيحاء، ودلالة، ومعنى، وظل، وإشعاع، وصدى، وإيقاع، لا أنه صوت فقط، فلقد أثبت علماءنا القدامى القيمة التعبيرية للصوت البسيط وصعوبة فهم القضايا البلاغية، وتشابه بعضها البعض مع بعض الظواهر اللغوية الأخرى؛ مثل ظاهرة (المثنيات التي لا تُفرد) مثل: الثقلان (الجن والإنس)، والجديدان: (الليل والنهار)، الأبوان، القمران، العمران، المشرقان...⁽³⁾، وكذلك ظاهرة ثلاثية الألفاظ

(1) ينظر: المدهون، نور، ملف إنجاز متدربة اللغة العربية - مدرسة المأمونية الإعدادية للبنات، (الملف الإلكتروني)، كلية التربية بالجامعة الإسلامية، بغزة، (2015م).

(2) ينظر: العصيلي، عبد العزيز، أساسيات تعليم اللغة العربية للناطقين بلغات أخرى، جامعة أم القرى، مكة المكرمة (1422هـ)، ص54.

(3) ينظر: السابق، ص55.

اللغوية، فقد أقام العرب معظم ألفاظهم على حروف ثلاثة لخفتها وإيجازها وسهولة النطق بها.

كل ذلك يضع المتعلم أمام لغة يصعب إدراكها والإحاطة بسماتها وخصائصها، ولذا وجب على المعنيين بتعليمها للناطقين بغيرها تذليلها وذلك عن طريق اختيار المناهج المناسبة لدراستها، وكذلك المواد المدروسة والوسائل المعينة على ذلك كله.

المطلب الثاني: مشكلات منهجية:

1- مشكلات طرق التدريس:

تعد طرق التعليم الوسيط بين المعلم والطالب داخل الصف، طرقاً ينقل من خلالها المعلم المعلومات للطالب بغرض فهمها، ويقصد بطرائق تدريس اللغات بما فيها اللغة العربية تدريس المهارات، والتي غالباً ما تصنف في أربع مهارات؛ هي: الاستماع، والحديث، والقراءة، والكتابة؛ لأن التمكن من هذه المهارات يؤدي إلى تحقق الكفاية اللغوية التي هي الهدف الرئيس من تعلم اللغة، وليس المقصود بطرائق التدريس أساليب تدريس مقررات المنهج منفصلاً بعضها عن بعض، كطريقة تدريس القواعد أو الإملاء، أو الخط أو التاريخ، أو غير ذلك⁽¹⁾، فهذه المقررات إن وجدت في المنهج في هيئة مقررات، فإنما تقدم للمتعلمين بوصفها محتويات تخدم مهارة اللغة، وتقود إلى تحقيق الهدف الرئيس.

ويمكننا بادئ الأمر تلخيص أبرز الأسباب في تلك المشكلة من واقع الدراسات العديدة التي قامت على عرضها ومحاولة إيجاد طرق حلها فيما يلي:

(1) الحديدي، علي، مشكلة تعليم اللغة العربية لغير العرب، ص 47.

- الاعتماد على طريقة الترجمة.
- الاعتماد على التلقين وذاكرة الطالب.
- الاعتماد على دور المعلم وتجاهل المشاركة الفعالة.
- عدم مراعاة الفروق الفردية بين الطلاب.
- الحرية المطلقة لدى المعلم في الاعتماد على نفسه واختيار ما يشاء من المواد والطرائق لتأدية عمله، وهذا ربما يعني قلة الوعي بما ينطوي عليه تعليم اللغة من تعقيد، وأنه سيجد طريقه في متاهة من النظريات والتطبيقات والأفكار المتعلقة بالموضوع.

إن المتتبع لسير حركة تطوّر طرائق تدريس اللغات الأجنبية، يلحظ هيمنة طريقة القواعد والترجمة، فيما يمكن أن يندرج تحت الطرق التقليدية أو الكلاسيكية، فترات طويلة من الزمن لدرجة جعلت بعض المشتغلين يعتقدون بصعوبة الانعتاق منها، تلك الطريقة التي تقوم على ترجمة النصوص ودراستها لغايات حفظ الكلمات، والقوالب اللغوية والنحوية، ومن المنصف القول إنّ ظهورها في ذلك الوقت كان له مسوغاته، خاصة إذا علمنا ارتباط نشأتها بدراسة اللغة اللاتينية، وفهم الكتب الموضوعية بها، تلك اللغة الحية بثقافتها وحضارتها ورموزها، الميته في التواصل بها، ولكنّ استخدام تلك الطريقة الذي ما زال شائعاً في تدريس اللغة العربية للناطقين بغيرها يواجه مشاكل جمة صعبة على الحصر، فمن المفيد معرفة نتائج هذه الطريقة وذلك بالاستماع لرأي أحد الدارسين وفقها، وهو المستشرق صمويل زويمر، إذ يقول: " إنّ طريقة تعليم العربية للأجانب قاصرة على تربية الذاكرة في حفظ كلمات وجملٍ وصيغٍ وأنماط لغوية، دون الالتفات إلى درجة استعمالها في الغرض من التعليم... وتهمل قوى العقل، تتركها في خمولها التام، ولا يجد العقل مجالاً لفهم المعاني التي تمسّ الحياة، ويتطلبها المتعلّم، ولا

تستغل المعاني، التي تختزنها ذاكرة الطالب من ثقافته ولغته، فيعطى له مقابلها.. تضيع السنون، ويخرج المتعلم بثروة قيّمة غير أنّها لا تصلح إلا للعرض في واجهة بائعي الجواهر الكلاسيكيين⁽¹⁾.

والحقيقة أن إيجاد الطريقة المناسبة ليس أمراً سهلاً، ونظرة إلى الطرق المتبعة في التدريس نجدها تقليدية تعتمد على التلقين، كما يعتمد تدريسها في البلاد الأجنبية على الترجمة، فينال هذا الجانب الاهتمام أكثر من التطبيق لفهم اللغة واستعمالها "فالابتعاد عن الترجمة هو خير طريقة لتعليم اللغة العربية للأجانب حيث إن الاقتصار على اللغة العربية يجبر الطالب على استعمالها ويعود حسن استعمالها والتعرف على ألفاظها وأصواتها"⁽²⁾.

ومن المعلمين كذلك من يعتمد على التلقين - المعتمد على ذاكرة الطالب - لتوصيل معلومات الدرس إليه، وبالرغم من نفع التلقين في العملية التعليمية إلا أن هذا النفع يتسم بالآنية، أما على المدى الطويل فلا يبقى من تلك المعلومات إلا القليل، الذي ربما لا يسمن ولا يغني لتحقيق الاتصالات الحياتية التي تخزن في ذاكرة الطالب، فيضعف تلقيه للعلوم داخل الصف لانشغال الذاكرة بأمور يومية وحياتية تفرض نفسها عليها، وعلى هذا الأساس فإن اعتماد العملية التعليمية على التلقين تجعل قياس استيعاب الطالب للمواد المدروسة مبنياً على قوة ذاكرته دون مهارته في ممارسة اللغة.

وتستتبع طريقة التلقين قيام العملية التعليمية داخل الصف على المعلم مما

(1) بدران، محمد ممدوح، اللغة العربية وتدريسها لغير الناطقين بها، نحو منهج لإعداد مدرس اللغة العربية لغير الناطقين بها، من منشورات المنظمة العربية للتربية والثقافة والعلوم، تونس (1992م)، ص 42.

(2) ينظر: العناتي، وليد، اللسانيات التطبيقية وتعليم اللغة العربية لغير الناطقين بها، طبع دار الجوهرة، عمان (2003م)، ص 76.

يستدعي انعدام مشاركة الطلاب للمعلم بالصورة المطلوبة في الصف، وهذا على أقل تقدير يصيب الطلاب بالملل ولا يضمن بقاء أذهانهم مع المعلم طول مدة تلقي الدرس من ناحية، ومن أخرى فإن فرصة التطبيق وممارسة ما درسه الطلاب من القواعد قد انعدمت تمامًا داخل الصف فيكون حال الطالب كحال من يقرأ عن كيفية قيادة السيارة فيجمع المعلومات الكافية دون أن يتدرب على القيادة، وإذا استمرت المعلومات النظرية لفترة طويلة في ذهنه دون التطبيق بممارسة القيادة فسوف تتلاشي تلك المعلومات شيئًا فشيئًا من ذهنه.

ولما اختلفت أهداف تعلم اللغات الأجنبية وغاياتها مع بدايات القرن الماضي، استتبع ذلك ميلاد طرائق لغوية جديدة تلبي الحاجات الإنسانية الحديثة، ولعل أهم مسوغات هذا التطور ما شهده العالم من تقدم تقني وعلمي، إذ شملت أرجاءه شتى مناحي الحياة⁽¹⁾، وبصرف النظر عن اختلاف العلماء في منطلقاتهم الفكرية وتوجهاتهم المختلفة واتباعهم طرقًا لا يحيد بعضهم عنها؛ فإنهم يجتمعون على وجود فروق فردية بين الناس من حيث قدراتهم وطبائعهم ودوافعهم وأمزجتهم واستعداداتهم وانفعالاتهم⁽²⁾.

من - هنا - يمكن القول بلزوم تعدد المناهج والطرق الدراسية لمواءمة العقول المختلفة، ومواكبة الظروف المتغيرة؛ لما لذلك من دور فاعل في عملية تعلم العربية وتعليمها، وتحقيقًا للفائدة نعرض طرق تعليم اللغات الحية، التي قام على حصرها (مكاي)، في كتابه (تحليل تعليم اللغة)⁽³⁾، حيث حصرها في خمس

(1) ينظر: عامود، بدر الدين، علم النفس في القرن العشرين، من منشورات اتحاد الكتاب العرب، دمشق، (2001م)، ص 432.

(2) Macky W. F. Language Teaching Analysis. London: Long man Group 6th Edition 1976.

(3) ينظر: أبو عمشة، خالد، والفاعوري، عوني، تعليم العربية للناطقين بغيرها: مُشكلات وحلول،

عشرة طريقة، هي: الطريقة المباشرة، والطريقة الطبيعية، والطريقة السيكلوجية، والطريقة الصوتية، وطريقة القراءة، وطريقة القواعد، وطريقة الترجمة، وطريقة القواعد والترجمة، والطريقة التوليفية، وطريقة الوحدة، وطريقة ضبط اللغة، وطريقة التقليد والحفظ، وطريقة المران، وطريقة المفردات المتشابهة، وطريقة اللغة المزدوجة.

ولذا فإن اقتصار بعض المدرسين على طرق ومسالك محددة لا يحيدون عنها في تعليم شتى المهارات، يعد من الإشكاليات التي ينبغي إيجاد بديل عنها، إذ إن واقع تعليم العربية الحي أثبت أن بعض الطرق أفضل من بعضها الآخر في تعليم مهارة أو عنصر لغوي؛ لذلك أضحي التوجه إلى رؤية المهارة اللغوية أو العنصر لكي تتحدد أفضل وسيلة أو طريقه لتقديمه للدارسين، أي لا بدّ من الإفادة من شتى الطرائق اللغوية، من ترجمة وسمعية وتواصلية وكلامية، خاصّة تلك الجوانب التي أثبتت جدواها، فضلاً عن مراعاة مبادئ علم النفس من تعلم واكتساب، عدا الجوانب الثقافية، التي لا بد أن تكون نصب الأعين في عملية تدريس العربية للناطقين بغيرها بذلك يمكن التغلب على كثير من المشكلات، من تركيبية، ودلالية، وصوتية⁽¹⁾.

2- مشكلات المناهج التعليمية:

إنّ عمليّة التدريس أيّاً كان نوعها أو نمطها أو مادتها ومحتواها؛ فإنها تعتمد اعتماداً كبيراً على الكتاب المنهجي، وهو بهذا المفهوم يعدّ ركناً من أركان عمليّة التعليم وعنصرًا من عناصرها، وركيزة من ركائزها، ولذلك تعدّ نوعيّة الكتاب

==

الجامعة الأردنية نموذجاً، منشور في مجلة دراسات الجامعة الأردنية، العلوم الإنسانية والاجتماعية، مج32، ع3، سنة (2005م)، ص 487-497.

(1) ينظر: الناقه، محمود كامل، خطة مقترحة لتأليف كتاب أساسي لتعليم العربية للناطقين بغيرها، وقائع ندوات تعليم العربية للناطقين بغيرها، مكتب التربية العربي لدول الخليج، 239/2.

وجودته أبرز الأمور التي تشغل بال المهتمين بحقل تعليم العربية للناطقين بغيرها⁽¹⁾، ولعل الحاجة لمثل هذا الكتاب الممنهج ضرورة في ظل إقرار معظم المراكز المشتغلة بتعليم العربية للناطقين بغيرها بعدم توفر المدرس الكفاء المتخصص في تعليم العربية للناطقين بغيرها، الذي قد يغطي وجوده جوانب النقص التي تعجّ في كتب تعليم العربية للناطقين بغيرها.

وعلى هذا الأساس ينبغي مراعاة الأسس التالية في منهجية الكتاب المطروح للعملية التعليمية:

- طبيعة المادة المقدمة وطرائق إخراجها.
 - اختيار المفردات والتراكيب اللغوية وأساليب معالجتها.
 - محتويات الكتب اللغوية والثقافية والدينية.
 - الوسائل التعليمية المرافقة للكتاب.
 - الأهداف المرجوة من الكتاب، ومدى مناسبة التدريبات لتحقيقها.
 - ملائمة الكتاب لمستويات الطلبة.
- لذا، لا بدّ أن يؤلف هذا الكتاب في ضوء خطة تعليمية محكمة، تحدد أهدافها، ويربط محتواها بتلك الأهداف، ولا بدّ أن يكون الكتاب انعكاساً لهذه الأهداف وساعياً إلى تحقيقها، ولا شكّ في أنّ أيّ عمل جاد يبدأ بتحديد الأهداف بوضوح تامّ، ثمّ اختيار الوسائل التي تحقق هذه الأهداف⁽²⁾.

(1) ينظر: الغالي، ناصر عبد الله، أسس إعداد الكتب التعليمية لغير الناطقين بالعربية، طبع دار الغالي، الرياض، (1991م)، ص 56.

(2) بدوي، السعيد محمد، مقتضيات الكفاءة في تعليم العربية كلغة إضافية، منشورات المنظمة العربية للتربية والثقافة والعلوم، تونس، (1992م)، ص 48.

طبيعة الكتب الدراسية:

إن عملية اختيار الكتاب الرسمي الذي يدرس للطلاب داخل الصف، من أصعب الخطوات في تعليم اللغة العربية للناطقين بغيرها، إذ يمثل الكتاب صلة تربط الطالب باللغة العربية. وتعد مشكلة إعداد الكتاب من المشكلات التي تداولها المختصون بهذا العلم؛ للوصول إلى وضع كتاب يعمل على رفع مستوى الطالب إلى درجة الكفاءة في إتقان اللغة، وهو ما حدا بأحد المختصين بهذا العلم إلى الاعتراف بالقصور في وضع الكتاب الذي يصل بالطالب إلى تلك الدرجة من الكفاءة، ف" قضية الكفاءة، بمعنى المستوى النهائي في إتقان اللغة العربية بمهاراتها المختلفة من: حديث، واستماع، وقراءة، وكتابة، والذي نسعى بالمتعلم الأجنبي الجاد الذي يسلم نفسه كلية لنا إلى الوصول إليه، هذه القضية لم نتناولها بالفحص لا عند كتابة الكتب التي ألفناها ولا عند إقامة المناهج التي وضعناها لتعليم هذا الطالب"⁽¹⁾.

من - هنا - فإنه يلزم أن تتحقق في الكتاب الرسمي الشروط التالية:

- الاعتماد على اللغة المتداولة بين الناطقين بالعربية في الماضي والحاضر على السواء، وذلك عن طريق توزيع النصوص المدروسة بين القديم والحديث، وليس الحديث وحده، مع مراعاة أن تكون المفردات والتركيب موزعة بين القديم والحديث -أيضاً-، وذلك لسد الفجوة بين ما يدرسه الطالب في العام التحضيري وبقية الأعوام التالية، وقد وضحت تلك الفجوة من خلال صدمة طلاب الصف الأول عند قراءة متون التفسير والحديث وغيرها⁽²⁾.

(1) ينظر: الخولي، كريم فاروق، مشكلات تعليم العربية لغير الناطقين بها وطرق حلها، بحث منشور في مجلة كلية اللاهوت، جامعة سلجوق، ع (32)، سنة (2011م)، ص 195.

(2) السابق، ص 195.

- وضع القواعد الموضحة لكيفية فهم الطالب للتراكيب المجازية، ويمكن ذلك من خلال توزيع التراكيب المجازية على الدروس، ووضع التراكيب الحقيقية المقابلة لها مع تكرارها في سياقات مختلفة، فيعلم الطالب المقابل الحقيقي للتراكيب المجازية، عندما يتلقاها كما يمكن اختيار منهج مفصل لبعض دروس البلاغة المبسطة للمرحلة التمهيديّة.
- الاعتماد على الموضوعات المنفصلة والمتصلة معاً "فأهمية النص المتصل قد تفوق تلك النصوص المنفصلة في تعليم العربية، إذ يتميز النص المتصل ببقاء مفرداته وتراكيبه الحقيقية والمجازية ودلالاتها في ذهن الطالب، يسترجعها خلال دراسته لأجزاء النص المتصل، ويسترجعها عند الحاجة لبقائها في ذهنه لفترات طويلة، كما يمكنه من تلخيص أجزاء النص ووضع العناوين الفرعية وربط الدلالات الفرعية بالدلالات الكلية للنص المتصل، مع مراعاة أن تشمل النصوص المنفصلة موضوعات خاصة بالتفسير والحديث والفقه والسير والتصوف وغيرها"⁽¹⁾.
- تنوع التراكيب المستخدمة في النصوص المدروسة، مع تسليط الضوء على التراكيب بتكرارها وتوضيح دلالاتها في سياقات مختلفة، ووضع ما يدعو الطالب إلى استخدامها داخل الصف وخارجه وذلك عن طريق الأنشطة والتدريبات.
- توظيف القواعد النحوية والصرفية في الفهم والتحدث، وذلك عن طريق توضيح الدلالات الصرفية للصيغ والفروق الدلالية بينها، ومثل ذلك مع القواعد النحوية؛ لتحديد من اعتماد الطالب على الترجمة أو النظر في المعجم.
- توظيف الفروق اللغوية بين الكلمات في سياقات مختلفة يفهمها الطالب من

(1) ينظر: الخولي، كريم فاروق، مشكلات تعليم العربية لغير الناطقين بها وطرق حلها، ص 196.

خلال الدلالة الكلية للتركيب أو السياق، وكذلك الاعتماد على أنواع الدلالات المختلفة من نحوية، وصرفية، وتركيبية ومجازية، وسياقية، وعرفية، وشرعية، وغيرها من الدلالات.

- اتخاذ معاجم لغوية يراعى فيها تدرج منهجها ليناسب الفروق في القدرات أو الكفاءات العلمية والعمرية.

- ومن اللازم ألا تنقطع صلة الطلاب بدروس العربية وعلومها وآدابها في الصفوف التالية للصف التحضيري، بل من الأفضل إفساح المجال لها في تلك الصفوف لتحقيق الكفاءة المطلوبة في إتقان اللغة⁽¹⁾.

3- مشكلات أفراد

أولاً: ما يرجع إلى المعلم:

إن المعلم هو حجر الزاوية في العملية التعليمية، لذا يقع عليه العبء الأكبر في تحقيق أهدافها عند تعليم العربية للناطقين بغيرها، وعلى المعلم بدايةً فهم تلك الأهداف والبحث عن طرق تحقيقها والهدف الأسى لذلك وهو فهم الطلاب لغة القرآن الكريم، وفهم كتب التراث، من تفسير، وفقه، وسير، وغيرها.

ولأهمية دور المعلم في تحقيق تلك الأهداف، يتوجب عليه أن يتسلح بأدوات تفوق في فاعليتها تلك الأدوات التي يتسلح بها معلم العربية للناطقين بها، وأن يتميز بخصائص تقوده إلى تحقيق ذلك، ولذا يلزم أن "يتميز تكوين أستاذ اللغة العربية لغة ثانية، بخصائص تضاف إلى التكوين العام لأستاذ اللغة الأم، وذلك بالتركيز على كسب معرفة لغوية، ومهارات عملية، وقدرات تقنية معينة"⁽²⁾.

(1) الغالي، ناصر عبد الله، أسس إعداد الكتب التعليمية لغير الناطقين بالعربية، ص 76.

(2) ينظر: الخولي، كريم فاروق، مشكلات تعليم اللغة العربية لغير الناطقين بها وطرق حلها، ص 189.

- وتتمثل طرق تحقيق تلك الأهداف فيما يخص المعلم في:
- إدراك الدور المنوط به من توسيع رقعة الناطقين بالعربية وقيمة تعليمها وتعلمها، وأثرها على الأصعدة المختلفة.
 - الإجابة التامة للعربية وقواعدها وأساليبها التي تميزها عن غيرها من اللغات.
 - العمل على ترقى مستواه، وزيادة التثقيف والتأهيل، بالخضوع للدورات الخاصة بتأهيل معلمي العربية لتعليم غير الناطقين بها، ومطالعة ما يستجد ويُصنف من الكتب الخاصة بذلك.
 - محاولة الاستفادة من الخبرات التي حققت تقدماً ملموساً في تعليم العربية لغير الناطقين بها، وذلك عن طريق عقد المؤتمرات والندوات والدورات التدريبية أو المقابلات الفردية إن أمكن، وإذا تحقق للمعلم ذلك فسوف تكون لديه القدرة على توصيل الدلالات الصرفية، والنحوية، والمجازية، والتركيبية والسياقية للطالب، دون الحاجة إلى اللغة الوسيطة.
 - الاستفادة من الخبرات التربوية وطرق التدريس الحديثة عمومًا مثل طرق التدريس التي تعتمد على الطالب وتلك التي تراعي الفروق الفردية بين الطلاب⁽¹⁾.
 - الإمام بعلوم اللغة الحديثة: كعلم اللغة النفسي، وعلم اللغة الاجتماعي، وعلم اللغة التطبيقي، وتطبيقاتها في تعليم العربية للناطقين بغيرها، يعدّ من أبجديات تعليم اللغات الحية، وقد بيّن الأستاذ محمد صالح بن عمر أهميّة ذلك في كتابه (كيف نعلم العربية لغة حيّة) بقوله: "إنّ إمام مدرّس اللغة الحية

(1) بن عمر، محمد صالح، كيف نعلم العربية لغة حيّة، سلسلة لسانيات عربية، معهد بورقيبة للغات الحية، تونس، ص 44.

بأسس علم النفس اللغوي أبعد ما يكون عن الترف أو الثقافة التكميلية، إنّما هو ضرورة ملحة اعتباراً لما يقدمه من حلول عمليّة للكثير من المشكلات التي يواجهها المعلم⁽¹⁾.

ثانياً: ما يرجع إلى المتعلم:

إن الطالب هو أساس العملية التعليمية، وهو المستهدف الذي وضعت من أجله كل الإمكانيات المتاحة؛ ليفهم اللغة العربية باعتبارها الأداة التي تعبر به الطريق إلى فهم كتاب الله وسنة رسوله الكريم، لينقلها إلى طلابه في المستقبل إذا أصبح معلماً، أو إلى عامة المسلمين إذا أصبح إماماً أو مفتياً، وهذا جزء من الرسالة التي يحملها على عاتقه لنشر دين الله تعالى في أرضه وتصحيح المفاهيم الخاطئة، وقد رصد بعض الدارسين⁽²⁾ بعض المشكلات التي تتعلق بالطالب باعتباره هدفاً لهذه العملية التعليمية ذات الأبعاد الخاصة، وتتلخص فيما يلي:

- صعوبة التأقلم مع المجتمع العربي من حيث اللغة، ومن حيث العادات والتقاليد، فلا بد له أن يتعلم اللغة العربية بشكل سليم حتى يستطيع أن يندمج مع المجتمع ويتعايش معهم.
- مشكلة الدمج مع أبناء اللغة الثانية وخوفهم من ردة الفعل، فعندما يخطئ المتعلم في لفظة معينة فإنه قد يتعرض لسخرية البعض مما يؤدي بالضرورة إلى إرباكه.
- محدودية إدراكه للغة عمقاً ودلالة ووظيفة وأهدافاً وتعليمياً.
- عدم توافر الدوافع المختلفة لتعلم اللغة العربية.

(1) ينظر: الخولي، كريم فاروق، مشكلات تعليم اللغة العربية لغير الناطقين بها وطرق حلها، ص 187.

(2) ابن عمر، محمد صالح، كيف نعلم العربية لغة حيّة، ص 67.

- اكتفاؤه بما يتعلمه في قاعات الدراسة والتقيد بالمناهج المطروحة، وعدم الانطلاق إلى أبعاد رحبة في ساحات اللغة العربية وفي ميادين غاياتها المختلفة.
- عدم اجتهاده في فهم القواعد وتحصيلها، ومن ثمّ تطبيقها.
- استخدام بعض العبارات والمصطلحات في غير سياقها، لذلك عليه أن يتعلم المواقف الفعلية لتلك التعابير.
- وتفاديا لذلك وعملاً على الاستفادة من العملية التعليمية بشتى جوانبها؛
يجب على الطالب أن يتبع الآتي:
- تحديد أهدافه وإدراك أهمية دراسة العربية؛ من واقع الغايات المختلفة التي تتصل بأنماط حياته المختلفة، ويدرك أنها لغة الدين ولسان الدراسة في الأقاليم العربية، ولغة التجارة على مختلف أصعدها، ووسيلة التلاقي الدبلوماسي، وغير ذلك مما تمثل اللغة العربية فيه المحرك الأول لشتى مسارات الحياة، وعلى المعلم أن يستغل في الطالب رؤيته وتطلعه وارتباطه بالعربية في خلق دوافعه، والاجتهاد في إتقان ذلك فمن المسلم به عند علماء التربية أننا " نتعلم بشكل أحسن إذا كان الدافع أقوى"⁽¹⁾.
- المداومة والتعلق بالكتب الحرة والمؤلفات المدونة باللغة العربية في مختلف المجالات، ولا يحدد دائرة معارفه أو يقتصر على ما يتعلمه في قاعات الدراسة.
- العمل على التدرج في مستويات القراءة، وذلك بأن يبدأ بالكتب ذات الموضوعات البسيطة ثم ما سهل عليه ثم ما صعب، حتى تستوعب عقله الترقى في مدارج اللغة، فلا يصطدم بمعوقات القواعد والتراكيب والدلالات، كما أن عليه أن يتخذ من كتب المعاجم رفيقاً في هذا المنحى مما ييسر له العمل والدرس المنظم.

(1) ينظر: الخولي، كريم فاروق، مشكلات تعليم اللغة العربية لغير الناطقين بها وطرق حلها، ص 190.

- معرفة الدلالات الكلية للقوالب التركيبية العربية مثل: الجملة الفعلية، والجملة الاسمية، وقوالب الشرط، والوصف، والإضافة، والحال، وغيرها...، ثم استخدام تلك القوالب في التحدث بالرجوع إلى المعلم ليدربه على استخدام تلك القوالب، ومن ثَمَّ استخدامها في التحدث مع معلمه وزملائه في الصف أو غيرهم إن أمكن.

الممارسة العملية أو المهنية للغة ومحاولة المداومة على التحدث بها داخل وخارج قاعات الدراسة، وتطبيق القواعد التي تعلمها؛ يرسخ فهم الطالب لها فلا يكون في حاجة إلى مراجعتها باستمرار، كما يكسبه توسعاً وثروة للتراكيب والكلمات الجديدة التي لم يقف عليها خلال تعلمه، وكذلك فهم بعض القواعد التي لم يعمها في دراسته النظرية، كما أن الممارسة العملية تكسر تهيُّب الطالب من العربية والحاجز النفسي تجاه تعلمها، ويتهيأ لفهم المزيد والتحصيل والاطلاع، يضاف إلى ذلك أن الحديث الشفهي مهم للغاية ولن يتم التعلم بدونه، والتشجيع عليه واجب والصبر إزاءه أشد وجوباً؛ إذ على المدرس أن يقدر ظروف تعلم العربية للناطقين بغيرها، ويدرك أنها ليست الظروف الطبيعية المتسعة الميدان التي تلقى فيها اللغة الأم، كما لا يجوز للمدرس أن يطارده أخطاء الطلاب في الحديث الشفهي، وعليه -هنا- أن يكتفي بتصحيح الفاحش منها، الذي قد يسيء عن غير قصد أو يقرب المعنى⁽¹⁾.

(1) الأنصاري، عبد الله، تعرُّر التوصيف النحوي الحاسوبي، جامعة الإمام محمد بن سعود الإسلامية، (1435هـ)، ص 201.

طرق التوصيف النحوي حاسوبياً:

يُقصد بذلك "تقديم النحو العربي للحاسوب بطريقة شبيهة أو مقارب للوصف المقدم للعقل البشري، وذلك يعني أن الباحث -هنا- يريد أن يتقدم بالذكاء الاصطناعي تقدماً يضاهي أو يزيد على قمة العمليات الرياضية الدقيقة المعقدة، لما يتصف به النحو العربي من حساسية بالغة"⁽¹⁾.

وتعتمد هذه العملية على قضايا التحليل حاسوبياً ليتم بعد ذلك الإفادة منها وفق القواعد النحوية واللغوية المقررة.

وتخضع عملية التحليل هذه لاكتمال أركان التحليل اللغوي عامة، أبرزها:

1- التحليل الصوتي الدقيق للمواد اللغوية بغية أمن اللبس فيها، فالتاء -مثلاً- تختلف دلالاتها في آخر الفعل الماضي باختلاف سكونها وما يتلوها من حركات (قرأت - قرأت - قرأت).

2- التحليل الصرفي الكامل والدقيق لجميع مفردات اللغة، لا سيما الأوزان والصيغ المحصورة.

3- وضع المعاني المعجمية المحتملة لكل لفظة؛ ليختار منها المحلل ما يناسب السياق.

4- التحليل التركيبي الشامل للسياقات المحتملة في العربية لكل لفظة.

5- التوصيف الكتابي (الإملائي) الدقيق لكل كلمة، لإضافته إلى التركيب⁽²⁾.

(1) ينظر: الأنصاري، عبد الله، الدرس النحوي في ضوء الحاسب الآلي، جامعة الإمام محمد بن سعود الإسلامية، (1438هـ)، ص 12.

(2) ينظر: كنالي، وجدان، اللسانيات الحاسوبية العربية، الإطار والمنهج، المعهد الإسلامي للبحوث والتدريب، البنك الإسلامي للتنمية، (2011م)، ص 12.

وقدّم بعض الباحثين تصوراً تأسيسياً لحل مشكلة حوسبة وتوصيف العربية، لتقدم بصورة واضحة للحاسوب، وتكون صورتها مناسبة لدارسي العربية من العرب، ومن غير الناطقين بالعربية، وهذا التصور (العلمي والتقني)، تتعدد فيه المقاربات تبعاً لتباين الأنظمة اللغوية الجزئية التي تتكامل فيما بينها، وتشكل بذلك النظام الكلي للغة العربية⁽¹⁾.

ويمكن توصيف تلك الأنظمة اللغوية حسب الآتي⁽²⁾:

1- توصيف النظام الفوناتيكي:

تؤوّل العربية في نظامها الصوتي إلى أربعة وثلاثين فونيمًا، ثمانية وعشرون صامتًا، تتمثل في حروف الأبجدية العربية، وستة صوائت تمثلها الحركات الثلاث القصيرة: الفتحة، والكسرة، والضمّة. والحركات الثلاث الطويلة حروف المد، ويتم توصيف تلك الأصوات بحيث يكون لكل صوت صورة طيفية مرئية ذات ثلاثة أبعاد: أفقي (يمثل الوقف)، وعمودي (يمثل التردد)، وبعد ثالث يمثل درجة الشدة.

2- توصيف النظام الفونولوجي:

ويعني هذا أن التوصيف الفوناتيكي لأصوات العربية، لا يكفي وحده ليتمكن الحاسوب من تمييز الفونيمات، وأنه يحتاج إلى توصيف آخر لما يعتري فونيمات العربية في سياقها الوظيفي، نحو: التضعيف، و(أل الشمسية)، والمد وغيرها.

(1) ينظر: السابق، 10-17، والمعالجة الآلية للكلام المنطوق، التعرف والآلية من كتاب: الفزالي، سالم، استخدام اللغة العربية في المعلوماتية، المنظمة العربية للتربية والثقافة والعلوم، تونس، (1996م)، 134.

(2) ينظر: حركات، مصطفى، الصوتيات والفونولوجيا، المكتبة العصرية، بيروت، ط1، (1411هـ-). 144.

3- توصيف النظام المورفولوجي:

يهدف إلى تمكين الحاسوب من التعامل مع الوحدات الصرفية (المورفيمات) التي تشكّل التراكيب اللغوية المتفوقة، وفق مطلبين رئيسين:

- 1- القدرة على التوليد، وصياغة الأبنية اللغوية السليمة.
- 2- القدرة على التحليل عن طريق تعيين المعاني الصرفية وتفكيك التراكيب اللغوية إلى وحداتها الصرفية الصغرى.

4- توصيف النظام الإعرابي:

يتم هذا التوصيف عن طريق تمكين الحاسوب من تعيين حركات أواخر الكلم، وفق إجراءات محددة تبدأ من تصنيف كلم العربية إلى المبني والمعرب، ذلك أن المبني سيدخل الحاسوب على هيئة لا تتغير، فلا يحتمل غير وجه واحد لقراءته أو لضبط آخره، وهكذا يستمر العمل وفق تلك الطريقة في أبواب النحو المختلفة، مع استثناء المعرب الذي لا تتبدل حركة آخره، كالمثنى، وجمع المذكر السالم، والأسماء الخمسة، والأفعال الخمسة، والمضارع المجزوم المعتل الآخر.

5- توصيف النظام الدلالي:

هذا الإجراء يعتمد على أنواع الدلالة المختلفة: المعجمية، والصرفية، والنحوية، والمجازية، وغيرها، ولا يتم فيه رسم قواعد شاملة، وإنما يلتزم فيه دعم كل مفردة معجمية بتفاصيل ما يرد إلى الذهن من علائق دلالية تتعلق باللفظ أو تشتق منه كالعلاقات الصرفية، أو النحوية أو السياقية، أو الاشتقاقية، كما هو الحال - مثلاً - في اشتقاق اسم الفاعل، والمفعول، والفعل، وغيرها من المشتقات، تشكل بعد ذلك وحدة محورية محددة.

والواقع أن الحاسوب لا يمكنه التعامل مع القوانين الصوتية (الفونولوجية) العربية، إلا بعد تحويلها إلى نماذج رياضية، يتعرف عليها الجهاز على أنها رموز،

ثم يحللها ليخرجها بعد ذلك أصواتاً حقيقية، وواقع الترميز يعتمد على محاولة تخصيص كل حالة نطقية وكل وصف برمزي خاص⁽¹⁾.

وينبغي الإشارة -هنا- إلى أنه يجب أن يراعي توصيف العربية (لغةً ونحوًا) حاسوبياً، حال ومستوى المتعلم، وثقافته، وبيئته التي نشأ فيها خاصة غير الناطقين بالعربية ويرغبون في تعلمها، والإفادة منها، ويشكل (التعلم التفاعلي الإلكتروني) محوراً مهماً في هذا المجال؛ لأنه يعتمد على أسس علمية وتقنية "تمزج الصوت والصورة والكتابة، وهدفها تجاوز الطرق التعليمية التقليدية القائمة على التلقين والتحفيز والتسميع، في نقل المعرفة واقتراح طرق بديلة من خلال استغلال قدرات الحاسوب، الشيء الذي أدى إلى إعداد برامج حاسوبية تعليمية، تتماشى مع النظريات البيداغوجية والتعليمية الراهنة"⁽²⁾.

وتظهر أهمية توظيف التكنولوجيا في تعليم اللغات من خلال الآتي⁽³⁾:

- 1- أنها تتيح الفرصة الجادة أمام الطلاب لتجريب اللغة بأساليب لها معنى، وتعرض عليهم مواقف اجتماعية ليتفاعلوا معها، ويكتسبوا منها المهارات اللازمة.
- 2- أنها تمنح المتعلم دوراً إيجابياً في العملية التعليمية.
- 3- أنها تتيح للمعلم استعمال وسائل تعليمية متعددة ومشوقة، وطرائق تدريس مختلفة.

(1) اللسانيات الحاسوبية، رقمنة اللغة العربية ورهان مجتمع المعرفة، موقع شبكة الألوكة.

(2) ينظر: معزوي، منصور عبد الرحمن، توظيف التكنولوجيا في تعليم اللغة العربية للناطقين بغيرها، الجامعة القاسمية نموذجاً، (بحث مقدم للمؤتمر الدولي الأول، تركيا 2019م)، ص 359.

(3) طعيمة، رشدي، الأسس العامة لمناهج تعليم اللغة العربية، إعدادها، تطويرها، تقويمها، دار الفكر العربي، القاهرة، (2008م)، ص 17.

- 4- أنها توفر بيئة تعليمية تفاعلية، تتيح ممارسة اللغة الأجنبية بصورة طبيعية.
 - 5- أنها تراعي العوامل النفسية للمتعلمين، كالرهبة، والخوف، والخل.
 - 6- أنها ترفع درجة الانتباه والتركيز لدى متعلمي اللغة.
 - 7- أنها تخلق جوًّا من المتعة والتشويق عند ممارسة اللغة.
 - 8- أنها تراعي الفروق الفردية بين المتعلمين-خصوصًا- في تعلم اللغات.
- وتُعد البرامج الإلكترونية داعماً رئيساً لتعليم العربية لغير الناطقين بها، خاصة في عصرنا الحاضر، إذ يمثل (التعلم التفاعلي) نسبة كبيرة من البرامج المطروحة لتعلم اللغات المختلفة؛ لذا فإنه ينبغي الإفادة من إمكانيات الحاسوب في هذا المجال، وتقديم البرامج التي تشكّل قواعد النحو والعربية بأسلوب واضح وميسر، بالطريقة التي يرغبها المتعلم، ووفق قدراته وإمكاناته؛ ولهذا فإنه من المؤكد أن "المتعلم يصبح أكثر انخراطاً واندماجاً في تعلّم اللغة عندما تقدم له البرامج المبنية على اهتماماته وأغراضه وحاجاته، كما أن ذلك يجعل عملية تعلم اللغة لها معنى، كما تخلق لدى المتعلم الرغبة في السيطرة عليها"⁽¹⁾.

إشكالات توصيف العربية حاسوبياً وتحدياته:

يتميز التعلم القائم على استعمال تطبيقات الحاسوب التعليمية بطريقة تقديمه التراكيب النحوية عن طريق ربطها بالمعنى والموقف التواصلية، فالطلاب يكتسبون المهارات من خلال المواقف التواصلية الحقيقية، كما أن التراكيب والوظائف النحوية التي يكتسبونها من خلال التفاعل يكون لها معنى وقيمة، وبالرغم من هذا فإن استعمال تطبيقات الحاسوب التعليمية في هذا المجال، واكتساب

(1) ينظر: رمضان، محمد، اكتساب الوظائف النحوية باستخدام تطبيقات الحاسوب التعليمية لدى الناطقين بغير اللغة العربية في المستوى المتقدم، دراسة تطبيقية إجرائية، (بحث مقدم للمؤتمر الدولي الأول، تركيا 2019م)، ص 889.

الوظائف النحوية في ضوءه، لم يلق حظه من الاهتمام حتى اليوم في فصول تعليم العربية لغة ثانية⁽¹⁾.

غير أن هذا التقديم الذي يحتاجه متعلم اللغة خاصة غير الناطقين بها، القائم على التوصيف المنهجي والتقني للقواعد والقوانين النحوية المختلفة، ربما يواجه عدداً من الإشكالات والتحديات التي تحتاج إلى معالجة وحل جذري سريع يتوافق مع الإمكانيات والتقنيات الحديثة المتاحة، وهذه المشكلات المتعددة "يمكن أن تصنف في أكثر من مجموعة واحدة من هذه المشكلات، فالمشكلات التي تعود إلى اللغة العربية -مثلاً- منها ما هو لغوي، كالمشكلات الصوتية، والمشكلات الكتابية، والمشكلات الصرفية، والمشكلات النحوية، والمشكلات الدلالية، ومنها ما هو غير لغوي"⁽²⁾.

ومن الواضح أن الإشكالات والتحديات ذات شقين، منها ما يخص المتعلم نفسه، ومنها ما يتجه إلى اللغة العربية ذاتها، ومناهجها، وقواعدها، وطرق تقديمها وتدريسها.

ومن أجل تلك التحديات بات لزاماً على علماء العربية وأهلها، معالجة عدد من الأمور المهمة التي تسهل عملية التعليم المقدمة لغير الناطقين بالعربية عن طريق الحاسوب، أو البرامج التفاعلية التقنية، أو حتى تلك البرامج والتطبيقات المقدمة عن طريق الهواتف المحمولة بأنواعها المختلفة، ولهذا فإنه يعد من الأولويات في هذا المجال محاولة تحليل العربية تحليلًا آلياً في مجالاتها المختلفة: قواعدياً، وصرفياً، وصوتياً، ودلالياً. وإنشاء معاجم متخصصة لمختلف المراحل

(1) العصيلي، عبد العزيز، أساسيات تعليم اللغة العربية للناطقين بلغات أخرى، ص. 192-193.

(2) عبد الهادي، حسني، الاتجاهات الحديثة لتدريس اللغة العربية في المرحلتين الإعدادية والثانوية، المكتب العربي الحديث، الإسكندرية، ص 220.

التعليمية تقدم بصورة واضحة، بالإضافة إلى تمكين التدقيق الإملائي والكتابي حاسوبياً تمكيناً جيداً يخدم المتعلمين خاصة في مراحلهم الأولى.

يُضاف إلى هذا أنه لا بُدَّ من ابتكار طرائق جديدة لتقديم قواعد النحو واللغة بصورة أكثر ملاءمة، تتناسب مع إمكانات الحاسوب، وقدرات الطلاب التفاعلية، والتواصلية، خاصة حين ندرك الضعف العام لدى متعلمي اللغة العربية غير الناطقين بها في هذا المجال، وربما يعود ذلك لأمرين: الأول: هو تدريس القواعد باعتبارها مادة مستقلة منفصلة. والثاني: هو التركيز في تعليم النحو على القواعد، وإغفال الجانب المهم في تعلم اللغة وتذوقها، وهو جانب التذوق اللغوي، والإحساس باللغة، وطرق استعمالها

كما أنه مما يسهل تعليم القواعد النحوية، أن يتم ربط نماذجها وأمثلتها، وصيغها بالمواقف الحياتية الواقعية التي تجعل ذهن المتعلم متحفزاً لربط التصورات الذهنية بالوقائع الماثلة، ولهذا فإنه ينبغي على المعلم "أن يرشد الطلاب حتى يكونوا واعين أن التعليم والنشاطات المدرسية جزء من الحياة"⁽¹⁾.

من هذا المنطلق باتت دعوات تسهيل القواعد النحوية، وتقديمها لغير الناطقين بالعربية، محل اهتمام لدى المنشغلين بهذا المجال، ونادى عدد من المهتمين بأن يكون الشعار في تعليم القواعد "البساطة والوضوح، فقليل يُفيد ويُستوعب ويُفهم ويُستخدم غير من كثير يُحفظ ويردد بدون فهم، ثم يتلاشى كضباب كثيف خانق"⁽²⁾.

(1) ديب، إلياس، مناهج وأساليب التربية والتعليم لتراكيب اللغة العربية، دراسة لغوية، دار الكتب اللبنانية، بيروت، (1981م)، ص 278.

(2) الأنصاري، عبد الله، تعثر التوصيف النحوي الحاسوبي، ص 6.

هذه الدعوات تزايدت في الآونة الأخيرة، خاصة مع الثورة التقنية والحاسوبية، ومع التحديات التي تواجه اللغة العربية وتعرثر توصيف قواعدها للحاسوب توصيفاً دقيقاً وواضحاً، يعتمد على تحديد نوع أنظمة البرمجة التي يمكن استعمالها في التحليل الصرفي، والدلالي، والصوتي، والاشتقائي، والمعجمي، والتحكم بها، ومتابعة تطورها.

إن المعالجة الآلية لقواعد العربية وأنظمتها، ومحاولة توصيفها يهدف إلى وضع "قوانين نظرية تناسب تقنية الحاسب لتأليف اللغة، وتنظيم أنساقها، وفهمها واستعمالها...، لأجل أن يخدم ذلك الدراسات اللغوية بعامة، فيتمكن الدارس من تخزين المعلومات في الحاسب بسهولة، فإنه إذا أودعها بطريقة سليمة في الحاسب، أمكنه الاستفادة منها لأغراضه اللغوية بسرعة وسهولة دون عناء"⁽¹⁾.

وقد أشار أحد الباحثين إلى أهم التحديات التي تواجه توصيف العربية حالياً، وأرجع ذلك إلى أمرين⁽²⁾:

الأول: ضعف المحتوى الرقمي العربي الآلي، وعلاج تلك المشكلة يكمن في تطويع التقانة الحديثة بما يناسب اللغة العربية، وبناء قواعد بيانية حاسوبية ملائمة للنحو العربي في أنظمتها وقوانينها، مع التركيز على المعالجة الآلية المعمقة للغة العربية.

والآخر: قلة الذخائر اللغوية المجهزة آلياً، ويُقصد بذلك قواعد البيانات

(1) ينظر: السابق، 10-11.

(2) O'Dell, F. and McCarthy, M. English collocations in use, Cambridge University Press. 2011, p:4

المستوعبة للمعجم العربي الواسع، و-هنا- نحتاج لعلاج مثل تلك المشكلة إلى عدة أمور منها: الإحصائيات الشاملة للصيغ العربية وأوزان الكلمات، وأنواع الجمل والتراكيب وأحوالها، وأشكال التقديم والتأخير والحذف، والضبط الإعرابي.

ومثل تلك التحديات والحلول التي قدّمها بعض الباحثين، إنما هي في الواقع محاولات توصيفية للظواهر التي تعاني منها العربية عند محاولة توصيفها للحاسوب، غير أنها لا تقدّم حلولاً سريعة وناجعة، إذ مثل تلك العوائق تحتاج إلى تضافر جهود المنظمات التي تهتم باللغة العربية، ومحاولة تقسيم الأعمال ذات الأهمية مع البدء بما يعالج أساس التوصيف، والمتركز في تقديم المعاجم العربية بألفاظها، واشتقاقاتها، وصيغها إلى الحاسوب بصورة واضحة وبمبسطة، مع إعادة النظر في سبك القواعد النحوية وصياغتها بالطرق المناسبة التي تتوافق مع إمكانيات الحاسوب، ومتابعة تطويرها مع مراعاة أن تخضع تلك القوانين لصياغة علمية نظرية ومنطقية تنطلق من الحقائق التي تصف اللغة وصفاً واضحاً، مع ربط أمثلتها ونماذجها المختلفة بواقع الحياة لتكون أقرب للتحقيق، وأحكم في الإفهام والتناول، مع عدم الإخلال بجوهرها ومدلولاتها.

الخاتمة

في نهاية الحديث حول هذا الموضوع الذي عنوانته بـ(توصيف العربية حاسوبياً لغير الناطقين بها) الذي أشرت فيه إلى أهمية التعليم الإلكتروني، وتوظيف الاستراتيجيات الحديثة في تعليم العربية لغير الناطقين بها، لما يحققه هذا النوع من التعليم من التفاعل المثمر بين المعلم والمتعلم، ويتجاوز بمفهومه هذا مرحلة التلقين والطرح الجاف للمعلومات، ويحفّز المتعلمين على المشاركة المستمرة، والحوار البناء الذي يرسّخ المعلومات، ويرتكز على طرح المهم المفيد ومناقشته.

بعد هذه الجولة يحسن إبراز نتائج هذه الدراسة، وهي نتائج اجتهادية تخضع للحوار والتحليل وأهمها الآتي:

- 1- ينبغي لعلماء العربية والمهتمين بها، والمشتغلين بعلومها المختلفة الاستفادة قدر الإمكان من التقنية الحاسوبية في تعليم العربية والدرس النحوي لغير الناطقين بها، ليصل ذلك إلى أعداد وافرة من الدارسين والمهتمين.
- 2- تتنوع طرائق الاستفادة من الحاسوب في تعليم العربية لغير الناطقين بها، والمهم هو تمكين المعرفة والإلهام الذهني الذي يتوافق مع إمكانيات الحاسوب، وقدرته على تفهم تلك القواعد، وتقديمها بصورة، واضحة، وميسرة.
- 3- لا بُدَّ أن يدرك القائمون على تلك العملية، أنه ينبغي الاستيعاب الكامل لإمكانية إدخال اللغة العربية إلى الحاسوب بكل دقة وضبط، وكذا إمكانية معالجتها، وطباعتها بشكل عادي ومقبول.
- 4- من المهم في مثل تلك العمليات التعليمية القدرة على صياغة قوانين النحو العربي وقواعده صياغة مناسبة يمكن تنفيذها عن طريق الحاسوب، تتميز بملاءمة المحتوى للمنطوق العربي، وقابلة للتطبيق المرن.

5- من أبرز المشكلات التي تواجه القائمين على محاولات برمجة اللغة العربية حاسوبيًا، وتجعل الأمر صعبًا هو أن المباشرين لعملية الإدخال والبرمجة-في الغالب- لا يتمتعون بحصيلة لغوية وقواعدية مناسبة، تمكنهم من تجنب الوقوع في الخطأ أثناء عملية التوصيف، وهذا الجانب ينبغي الاهتمام به وتطويره.

تعد هذه النتائج من أبرز ما ركزت عليه في هذه الدراسة، التي أرجو أن تكون لبنة مهمة في مشروعات مستقبلية تُعنى بحوسبة اللغة العربية، وتقديمها إلى الحاسوب بصورة دقيقة وواضحة، والله الموفق.

قائمة المراجع

- 1- أحمد، رضا بابا، اللسانيات الحاسوبية مشكل المصطلح والترجمة، جامعة تلمسان، الجزائر.
- 2- أمهان، طارق عبد الحكيم، اللسانيات الحاسوبية ومشكلة حوسبة اللغة العربية، خطوة باتجاه الحل، بحث منشور في شبكة الألوكة.
- 3- الأنصاري، عبد الله، الدرس النحوي في ضوء الحاسب الآلي، جامعة الإمام محمد بن سعود الإسلامية، (1438هـ).
- 4- الأنصاري، عبد الله، تعثر التوصيف النحوي الحاسوبي، جامعة الإمام محمد بن سعود الإسلامية، (1435هـ).
- 5- الأنصاري، عبد الله، علاقة النحو العربي بالبرمجة الآلية للغة، جامعة الإمام محمد بن سعود الإسلامية، (1435هـ).
- 6- الباكير، مجد محمد، مشكلات اللغة العربية المعاصرة، طبع مكتبة الرسالة الحديثة عمان - الأردن، (1989م).
- 7- بدران، محمد ممدوح، اللغة العربية وتدريسها لغير الناطقين بها، نحو منهج لإعداد مدرس اللغة العربية لغير الناطقين بها، من منشورات المنظمة العربية للتربية والثقافة والعلوم، تونس، (1992م).
- 8- بدوي، السعيد محمد، مقتضيات الكفاءة في تعليم العربية كلغة إضافية، منشورات المنظمة العربية للتربية والثقافة والعلوم، تونس، (1992م).
- 9- بوقرة، نعمان، النظرية اللسانية عند ابن حزم الأندلسي قراءة نقدية في مرجعيات اللسان الخطابي وأبعاده المعرفية، منشورات اتحاد الكتاب العربي، دمشق، (2004م).

- 10- الحاج، وليد، اللغة العربية ووسائل الاتصال الحديثة، دار البداية، عمان، (2007م).
- 11- الحديدي، علي، مشكلة تعليم اللغة العربية لغير العرب، طبع دار الكتاب العربي للطباعة والنشر، القاهرة (1967م).
- 12- حركات، مصطفى، الصوتيات والفونولوجيا، المكتبة العصرية، بيروت، (1411هـ).
- 13- خضر، محمد، الحرف العربي والحوسبة، الموسم الثقافي لمجمع اللغة العربية، عمان - الأردن، 2001م.
- 14- الخولي، كريم فاروق، مشكلات تعليم العربية لغير الناطقين بها وطرق حلها، بحث منشور في مجلة كلية اللاهوت، جامعة سلجوق، ع32، سنة (2011م).
- 15- ديب، إلياس، مناهج وأساليب التربية والتعليم لتراكيب اللغة العربية، دراسة لغوية، دار الكتب اللبنانية، بيروت، (1981م).
- 16- الراجحي، عبده، النحو العربي والدرس الحديث، طبع دار النهضة العربية، بيروت- لبنان، ط6، (1406هـ -/ 1986م).
- 17- رمضان، محمد، اكتساب الوظائف النحوية باستخدام تطبيقات الحاسوب التعليمية لدى الناطقين بغير اللغة العربية في المستوى المتقدم، دراسة تطبيقية إجرائية، (بحث مقدم للمؤتمر الدولي الأول، تركيا، 2019م).
- 18- زقوت، محمد شحادة، المرشد في تدريس اللغة العربية، طبع مكتبة الأمل للطباعة والنشر، غزة، ط2، (1999م).
- 19- الزهراني، مرضي، المدخل التقني في تعليم اللغة العربية، مفهومه وأساسه ومطالبه وتطبيقاته، ورقة عمل مقدمة إلى المؤتمر العالمي الأول للغة العربية وآدابها، ماليزيا.

- 20- شحادة، أمل عايد، التكنولوجيا التعليمية، دار كنوز المعرفة للنشر والتوزيع، عمان، ط2، 2010م.
- 21- شريف، سمير، اللسانيات، المجال والوظيفة والمنهج، استيتة، عالم الكتب الحديث، الأردن، (2005م).
- 22- صالح، عبد الرحمن الحاج، بحوث ودراسات في اللسانيات العربية، موفم للنشر، الجزائر، 1966م.
- 23- طعيمة، رشدي، الأسس العامة لمناهج تعليم اللغة العربية، إعدادها، تطويرها، تقويمها، دار الفكر العربي، القاهرة، (2008م).
- 24- العارف، عبد الرحمن، "توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، جهود ونتائج"، مجلة مجمع اللغة العربية الأردني، الأردن، ع 73، (2007م).
- 25- عامود، بدر الدين، علم النفس في القرن العشرين، من منشورات اتحاد الكتاب العرب، دمشق، (2001م).
- 26- عبد الهادي، حسني، الاتجاهات الحديثة لتدريس اللغة العربية في المرحلتين الإعدادية والثانوية، المكتب العربي الحديث، الإسكندرية.
- 27- عبيزة، عائشة، البرامج التفاعلية الإلكترونية ودورها في تعليم اللغة العربية للناطقين بغيرها، (بحث مقدم للمؤتمر الدولي الأول، تركيا، 2019م).
- 28- العصيلي، عبد العزيز، أساسيات تعليم اللغة العربية للناطقين بلغات أخرى، جامعة أم القرى، مكة المكرمة، (1422هـ).
- 29- علي، نبيل، اللغة والحاسوب، مؤسسة تعريب الكويتية، الكويت، (1988م).
- 30- ابن عمر، محمد صالح، كيف نعلم العربية لغة حيّة، سلسلة لسانيات عربية، معهد بورقيبة للغات الحية، تونس.

- 31- أبو عمشة، خالد، والفاعوري، عوني، تعليم العربية للناطقين بغيرها: مُشكلات وحلول، الجامعة الأردنية نموذجاً، منشور في مجلة دراسات الجامعة الأردنية، العلوم الإنسانية والاجتماعية، مج(32)، ع3، (2005م).
- 32- العناتي، وليد، اللسانيات التطبيقية وتعليم اللغة العربية لغير الناطقين بها، طبع دار الجوهرة، عمان (2003م).
- 33- الغالي، ناصر عبد الله، أسس إعداد الكتب التعليمية لغير الناطقين بالعربية، طبع دار الغالي، الرياض، (1991م).
- 34- فريحي، مليكة، النحو الحاسوبي، مج: عود الند، ع82، الجزائر.
- 35- الفزالي، سالم، المعالجة الآلية للكلام المنطوق، التعرف والآلية من كتاب: استخدام اللغة العربية في المعلوماتية، المنظمة العربية للتربية والثقافة والعلوم، تونس، (1996م).
- 36- قنديل، أحمد، التدريس بالتكنولوجيا الحديثة، عالم الكتب، القاهرة، (2006م).
- 37- الكمار، رأفت، الحاسوب وميكنة اللغة العربية، دار الكتب العلمية، القاهرة، (2006م).
- 38- كنالي، وجدان، اللسانيات الحاسوبية العربية، الإطار والمنهج، المعهد الإسلامي للبحوث والتدريب، البنك الإسلامي للتنمية، (2011م).
- 39- اللسانيات الحاسوبية، رقمنة اللغة العربية ورهان مجتمع المعرفة، موقع شبكة الألوكة.
- 40- ماسيري، ديكوري، المشكلات الصوتية في تعلم اللغة العربية للناطقين بغيرها، جامعة المدينة العالمية أنموذجاً، (عضو هيئة التدريس في كلية اللغات - ماليزيا)، مقدمة الدراسة.

- 41- مذكور، علي، تعليم اللغة العربية لغير الناطقين بها، النظرية والتطبيق، طبع دار الفكر العربي، القاهرة، (1427 هـ -/2006م).
- 42- المدهون، نور، ملف إنجاز متدربة اللغة العربية - مدرسة المأمونية الإعدادية للبنات، (الملف الإلكتروني)، كلية التربية بالجامعة الإسلامية، غزة، (2015م).
- 43- معذوي، منصور عبد الرحمن، توظيف التكنولوجيا في تعليم اللغة العربية للناطقين بغيرها، الجامعة القاسمية نموذجاً، بحث مقدم للمؤتمر الدولي الأول، تركيا، 2019م.
- 44- الموسى، نهاد، حصاد القرن في اللسانيات، مؤسسة عبد الحميد شومان، الأردن، (2000م).
- 45- الناقة، محمود كامل، خطة مقترحة لتأليف كتاب أساسي لتعليم العربية للناطقين بغيرها، وقائع ندوات تعليم العربية للناطقين بغيرها، مكتب التربية العربي لدول الخليج.
- 46- Macky W. F. *Language Teaching Analysis*. London: Long man Group 6th Edition 1976



تصور مقترح لبناء معجم رقمي للمتصاحبات اللفظية فكرة تطبيق ذكي

**A Proposed Vision to Design a Digital Lexicon
of Collocations
Smart Appl Proposal**

د. بدرية براك بنيه العنزي

**الأستاذ المشارك في اللغويات التطبيقية - المدونات المحوسبة
جامعة الأميرة نورة بنت عبد الرحمن**

ملخص

يهدف هذا البحث إلى تقديم تصور مقترح لتطبيق ذكي لمعجم رقمي لغوي للمتصاحبات اللفظية في اللغة العربية المعاصرة، يستهدف أبناء العربية ومتعلميها لغة ثانية؛ لتحقيق مقاصد منها: تعزيز هوية اللغة العربية، وتنمية الكفاية اللغوية، بالإضافة إلى زيادة الثروة اللفظية، ومساعدة الفئة المستهدفة للارتقاء بمستواها اللغوي وإثراء رصيدها المعرفي، وتمكينها من فهم اللغة العربية وممارستها وفق المواقف اللغوية المختلفة.

وكذلك الإسهام في إثراء المحتوى العربي على الإنترنت بمشروع نوعي يخدم اللغة العربية ومتعلميها، ويضاهي المعاجم اللغوية في اللغات الأخرى، إذ لا يوجد حتى اليوم معجم تكاملي بهذه الكيفية، مما يؤكد حاجة الميدان اللغوي لمثل هذه المشروعات النوعية، كون تعليم اللغة العربية أصبحت حالة ملحة تفرضها طبيعة التحولات التي نشهدها داخلياً وعربياً ودولياً، فيما يسمى باقتصاديات المعرفة، والتنمية المعرفية المستدامة، والهوية اللغوية، والأمن اللغوي وغيرها. وتعتمد فكرة التطبيق على برامج اللغويات الحاسوبية وأدواتها التحليلية، ويعتمد بناؤه على مراحل علمية منهجية تتمثل في الجمع والوضع المعجمي والإخراج النهائي (تطبيق ذكي).

كلمات مفتاحية: تطبيق ذكي - معجم رقمي - المتصاحبات اللفظية - المدونات اللغوية - الكلمة الأساس - المدخل المعجمي

Abstract

This paper aims to present a proposed vision for designing a smart app for a digital lexicon of collocations in contemporary Arabic language. The smart app targets Arabic native speakers and learners of Arabic as a second language to achieve several purposes, including: enhancing the identity of the Arabic language and developing linguistic competence. In addition to increasing verbal strength. Besides helping the target group to improve their linguistic level, enrich their knowledge, and enable them to understand and practice the Arabic language according to different linguistic situations. Also contributing to the enrichment of the Arabic content on the internet with a qualitative project that serves the Arabic language and its learners, and is comparable to linguistic lexicons in other languages, as there is no such integrated lexicon to date. This confirms the need of the linguistic field for such qualitative projects because teaching the Arabic language has become a crucial situation imposed by the nature of the transformations that we are witnessing internally, regionally, and globally. By which it is called knowledge economies, sustainable cognitive development, linguistic identity, linguistic security, and others. The idea of the app depends on computer linguistics programs and its analytical tools. The design of the app depends on systematic scientific stages represented in plurality, lexical status, and final output (smart app).

Keywords: Smart app, digital lexicon, collocations, linguistic corpuses, core word, lexical entry.

المقدمة

للغة العربية دور ريادي في إسهامها في حفظ ونشر الحضارة الإنسانية، بوصفها تمثل لغة 22 دولة من أعضاء منظمة اليونسكو، ولكونه يتحدث بها أكثر من 422 مليون عربي، وعطفاً إلى كونها لغة عالمية يحتاج إلى استعمالها أكثر من مليار ونصف المليار مسلم في شتى أنحاء العالم، جاء اعتماد اللغة العربية ضمن اللغات الست الرسمية للأمم المتحدة بمختلف منظماتها، منذ 18 ديسمبر 1973م.

وقد زاد الإقبال على تعلم العربية مؤخراً من كافة أقطار العالم، ولرؤية 2030 دور بارز في تعزيز الهوية العربية، فقد نص الهدف التفصيلي العاشر من أهداف رؤية 2030 على العناية باللغة العربية، والهدف التفصيلي الأول على تعزيز قيم الوسطية والتسامح مما يستوجب العناية بالعربية والاهتمام بها ومسايرتها للركب التقني والحضاري.

وتتقدم الثورة الرقمية بصورة مذهلة حتى طالت جميع الجوانب الحياتية، وتعد الهواتف الذكية نقطة التحول إلى المعرفة الرقمية بما فيها العملية التعليمية بكافة إداراتها، وحتماً ستكون الثقافة الرقمية حاسمة في المستقبل فهي تفرض التعلم الرقمي لتحقيق أقصى فوائد التحول الرقمي.

ومن فوائد هذا التحول الرقمي صناعة المعاجم الرقمية التي أصبحت ضرورة ملحة لمواكبة التطور التقني المتسارع؛ وبما أن سمة اللغة التطور والتجدد، فإنه لا يزال النشاط المعجمي قاصراً على مستوى المعاجم العربية المحوسبة؛ إذ لم تعد المعاجم التقليدية ذات فائدة في ظل الثورة الرقمية الحديثة، وأن تعلم المعاني اللغوية لم يعد معتمداً على قوالب المفردات، كما لو كان سابقاً، بل لا بُدَّ

من تعليمها في سياقات طبيعية من خلال مواقف، ولعل من أفضل المواقف هو سياق الوحدات المعجمية المتعددة (المتصاحبات اللفظية)؛ لذا برزت الحاجة الماسة إلى بناء معجم رقمي يتضمن المتصاحبات اللفظية المستعملة في المجالات الحديثة، ولعل بناء هذا النوع من المعاجم قد يغطي جانبًا من هذا القصور المعجمي الرقمي. ويمكن تعريف المتصاحبات اللفظية بإيجاز على أنها ارتباط كلمتين معاً⁽¹⁾ بتوليفة لغوية ذهنية بحيث إذا ذكرت المفردة يتبادر إلى الذهن مباشرة المفردة التي تتصاحب معها دائماً.

أهداف البحث

- 1- الإسهام في خدمة اللغة العربية من خلال الريادة في بناء معاجم لغوية عربية تقنية محوسبة مختصة بالتصاحب اللفظي موجهة لأبناء العربية ومتعلميها من غير الناطقين بها.
- 2- تعزيز الهوية العربية من خلال العناية باللغة العربية وبناء معاجم لغوية متنوعة للفئات للشرائح العمرية كافة.
- 3- الارتقاء بجودة التعليم ومخرجاته لإثراء التراث اللغوي وفق المستجدات العصرية الحديثة.
- 4- بناء معجم ورقي وحاسوبي لتنمية كفاءة الطالب اللغوية، والارتقاء بمستواه اللغوي والمعرفي، وزيادة ثروته اللفظية من خلال التعرض لقوائم من المتصاحبات اللفظية، في سياقات وأمثلة وشواهد إثرائية متنوعة.

(1) الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) بالتعاون مع مجمع الملك سلمان العالمي للغة العربية، معجم البيانات والذكاء الاصطناعي (انجليزي - عربي)، المملكة العربية السعودية، 2022،

- 5- تنمية ميدان تعليم العربية، وإثراؤه من خلال بناء معجم لغوي تقني حاسوبي.
- 6- بناء معاجم لغوية عربية تقنية حديثة في ضوء مستجدات علم صناعة المعاجم وعلم لسانيات المدونات وأدوات تحليلها.
- 7- الربط بين المدونات اللغوية والرصيد المعجمي الضخم؛ لبناء معجم تقني ممنهج قائم على أسس الصناعة المعجمية.
- 8- خدمة اللغة العربية من خلال إنتاج معاجم تقنية محوسبة مختصة بالمتصاحبات اللفظية، قائمة على مبدأ الشيوخ والتدرج عبر قوائم مدعمة بأمثلة وسياقات توضح معنى التصاحب اللفظي.

أهمية البحث

- 1- أهمية المعجمات اللغوية الرقمية في تنمية الثروة اللغوية لدى أبناء العربية ومتعلميها.
- 2- سد الفجوة المتمثلة في عدم وجود معجم لغوي عربي رقمي مختص بالمتصاحبات اللفظية لأبناء العربية ومتعلميها من غير الناطقين بها.
- 3- الحاجة الماسة لأبناء العربية ومتعلميها لمعجم ذكي يمكنهم من فهم اللغة ويثري الرصيد المعرفي.
- 4- الريادة في بناء معاجم المتصاحبات اللفظية الرقمية؛ حيث سيكون الأول من نوعه في اللغة العربية (عربي - عربي) يخدم أبناء العربية ومتعلميها لغة ثانية؛ (حاسوبي، تطبيقات على الأجهزة الذكية).
- 5- إثراء المحتوى العربي على الإنترنت وتطبيقات الأجهزة الذكية بمشاريع نوعية تخدم أبناء العربية ومتعلميها من غير الناطقين بها.

- 6- اعتماد منهجية المعجم على برامج اللغويات الحاسوبية وأدواتها التحليلية، وهذا من شأنه إنتاج معجم قائم على مبدأ الشيوخ.
- 7- الإسهام في تقدم الصناعة المعجمية من خلال معالجة اللغة العربية حاسوبياً.

مصطلحات البحث

1. التطبيق الذكي: منتج بقدرات ذكية يتطور باستمرار لتلبية احتياجات العملاء.
2. المتصاحبات اللفظية: هي وحدة معجمية متواترة في الاستعمال ومؤلفة في المعنى، وقائمة على الانسجام والاستخدام.
3. الكلمة الأساس: يطلق عليها العقدة أو الكلمة المفتاحية أو الكلمة المحورية.
4. الكلمة المصاحبة أو المحورية: وهي الكلمة التي تتصاحب دائماً أو غالباً مع الكلمة الأساس⁽¹⁾.

ووضعت الباحثة عدة معايير للتمييز بين المتصاحبات اللفظية وما يلتبس بها من التراكيب الأخرى، وهي كالآتي:

- التركيب. (مقياس تركيبى)⁽²⁾
- احتفاظها بمعناها الفردي، كلياً أو جزئياً، وبذلك تختلف عن الأمثال والتعبيرات الاصطلاحية. (مقياس دلالي)
- تفاوت مكوني التصاحب، بحيث يكون الأساس مستقلاً ويحتفظ بمعناه المؤلف؛ ويكون الملازم تابعاً. (مقياس دلالي)

(1) انظر: التعبيرات الاصطلاحية في لغة نجيب محفوظ، مرجع سابق، ص 48.

(2) انظر: ابن عمر، عبدالرزاق، المتلازمات اللفظية في اللغة والقواميس العربية، مجمع الأطرش للنشر، تونس، 2007، ص 38.

- التواتر والشيوع، ويعود إلى السليقة اللغوية للغويين العرب⁽¹⁾. (مقياس تداولي)

وبهذه المعايير يمكن أن يتفرع التصاحب اللفظي إلى نوعين، هما:

- التصاحب المقيد، والمتصاحبات اللفظية لطلبة المرحلة الابتدائية، يقتضي ظهور الكلمة الأساس ظهور الكلمة المصاحبة لها، أي أن الأساس يفرض انتقاء متصاحب بعينه يغدو تابعاً.
- المتتابعات المعكوسة وغير المعكوسة، ومنها المترادفات والمتضادات والمتكاملات، ك (ليل، نهار- عيش، وملح- السراء، والضراء).

منهج البحث

اعتمد البحث على المنهج الوصفي التحليلي؛ الذي يهدف إلى تحديد المشكلة، ووصفها، ودراستها ومعالجتها، ومن ذلك بناء تطبيق ذكي بنسخته الأولية وتجريبه وسيتم العمل لاحقاً على تطويره.

مشكلة البحث:

مما دفع الباحثة لتقديم هذا المقترح (تصور مقترح لبناء معجم رقمي للمتصاحبات اللفظية، فكرة تطبيق ذكي) هو عدم وجود تطبيقات معجمية عامة ذكية وقائمة على أسس الصناعة المعجمية الحديثة، فضلاً عن عدم وجود تطبيقات ذكية معجمية مختصة بالتصاحب اللفظي، بالإضافة إلى ندرة المعاجم اللغوية الحديثة وفق المحتوى العربي التقني، ولحاجة الميدان التعليمي الماسة للتطبيقات المعجمية الذكية في تنمية الإجابة اللغوية لأبناء العربية ومتعلميها،

(1) انظر: المتلازمات اللفظية في اللغة والقواميس العربية، ص 177.

وتعزيز مهارات اللغة وعناصرها لدى دارس اللغة ومتعلمها، وللتصدي لهذه المشكلة سعت الباحثة للإجابة عن الأسئلة التالية:

1. ما المنهجية التي قام عليها بناء المعجم الذكي؟
2. ما المدونات التي سيقوم عليها؟ وما مراحل بنائه؟
3. ما التصور المقترح لبناء معجم رقمي للمتصاحبات اللفظية، ويكون متاحًا على شكل تطبيق ذكي؟

مخرجات البحث:

• معجم تقني رقمي عربي مختص بالمتصاحبات اللفظية، موجه لأبناء العربية، وآخر لمتعلميها من غير الناطقين بها. متاح على الانترنت وعلى الأجهزة الذكية، يهدف إلى استخراج المتصاحبات اللفظية من خلال البحث عن الكلمة الأساس (النواة/العقدة) ثم ظهور المفردات التي تتصاحب وتتلازم معها، بحسب الأكثر ورودًا تبعًا، ثم ظهور شرح للمعنى التام للمتصاحبتين ثم السياقات التي تظهر فيها معًا، وهو على عدة أقسام، وتفصيلها على الشكل الآتي، وعلى مستوى البحث:

- تطبيق ذكي لمعجم المتصاحبات اللفظية لطلاب المرحلة الابتدائية 300 مدخل معجمي.
- تطبيق ذكي لمعجم المتصاحبات اللفظية لطلاب المرحلة المتوسطة (الإعدادية) والثانوية 500 مدخل معجمي.
- تطبيق ذكي لمعجم المتصاحبات اللفظية لطلاب المرحلة الجامعية 1000 مدخل معجمي.
- تطبيق ذكي لمعجم متصاحبات لغير الناطقين بالعربية 1000 مدخل معجمي.

مميزات المعجم الرقمي

- 1- حاجة الميدان اللغوي لمعاجم محوسبة مختصة تعكس التطورات التقنية الحديثة وتسخيرها لخدمة اللغة.
- 2- مساهمة العصر الرقمي في توفير القوالب التقنية للوحدات المعجمية المتعددة وتقديمها بشكل أسهل لكل مهتم بالعربية.
- 3- يعد هذا البحث نواة لمشروع معجمي رقمي، يسعى إلى تنمية الثروة اللغوية وتصحيحها لدى أبناء العربية ومتعلميها من غير الناطقين بها.
- 4- فتح الآفاق لدى الباحثين والمهتمين بالصناعة المعجمية، بإقامة دراسات وأبحاث تساهم في تطوير مخرجات هذا المشروع لاحقاً.
- 5- قيام فكرة البحث على أسس ومعايير علمية رقمية في الصناعة المعجمية.
- 6- إمكانية إفادة كل مهتم ومحب للغة العربية سواء كان داخل المملكة أو خارجها.

الفئات المستهدفة

- المرحلة الجامعية.
- المرحلة المتوسطة (الإعدادية) والثانوية.
- المرحلة الابتدائية.
- معجم المتصاحب اللفظية لغير الناطقين بالعربية يخدم المستويات المبتدئ والمتوسط والمتقدم.

الأدبيات السابقة

من خلال مسح قامت به الباحثة للمعاجم الخاصة بالمتصاحبات اللفظية في العربية المعاصرة، لم تجد إلا معجم الحافظ للمتصاحبات اللفظية (عربي-إنجليزي) الذي صدر سنة 2004م، وقاموس دار العلم للمتلازمات اللفظية

(إنجليزي- عربي) الذي صدر سنة 2007م. كما تبين أن المعاجم لم تول ظاهرة التصاحب العناية أو الاهتمام الكافي، إذ ترد في بعضها، وتُهمَل في بعضها الآخر، ويُكتفى بتفسير المعنى دون عناية بوضع الكلمة في سياقاتها اللغوية، فليس هناك خطة محددة للتعامل مع هذه القضية، إذ تلتبس بين الأمثلة التوضيحية، والشواهد، والتلازم المنتظم، والتعبيرات الاصطلاحية.

ومن جانب آخر حظيت المتصاحبات اللفظية بمعاجم ودراسات واهتمام ملحوظ في اللغة الإنجليزية، كونها من العوامل التي تُسهم في إبراز المعنى. ويُلاحظ وجود معاجم في هذا المجال فضلاً عن المعاجم غير المتخصصة التي أدخلت المتصاحبات اللفظية، من أجل إعطاء قيمة أكبر لهذه المعاجم من حيث المعنى، ولعل أهم هذه المعاجم الخاصة بالتلازم الآتي:

1- The BBI Combinatory Dictionary of English

وهو من تأليف Morton Benson, Evelyn Bensin, Robert Ilson، وBBI يرمز إلى اسم العائلة لكل مؤلفي المعجم الثلاثة، وهو معجم متخصص بالمتلازمات الإنجليزية، وموجه لتعليمي اللغة الإنجليزية لغة أجنبية؛ إذ اعتمد في جمع مادته على المدونة الوطنية البريطانية. ويوضح المعجم الخلافات النحوية بين الإنجليزية الأمريكية والإنجليزية البريطانية، ويقدم الكثير من الأمثلة التي تعد نماذج للاستخدام الصحيح في اللغة الإنجليزية. ويساعد أيضاً في تعرف الصعوبات التي تكتنف العبارات الإنجليزية الشائعة. وصدرت منه ثلاث طبعات (1986م - 1993م - 1997م)، وبحكم التطور المتلاحق في كافة مجالات الحياة، أُضيفت متلازمات جديدة في الطبعة الأخيرة لم تكن موجودة في الطبعتين السابقتين؛ إذ ضمَّ المعجم ما يزيد عن 90 ألف متلازم لفظي ونحوي.

2- Oxford Collocations Dictionary for Students of English

وهو معجم موجه لمتعلمي اللغة الإنجليزية فوق المتوسط إلى المتقدمين في مستويات عليا، ويساعد المتعلمين على الكتابة والتحدث بطلاقة؛ إذ اعتمد على المدونة الوطنية البريطانية، وقد ضم ما يزيد عن 150 ألف متلازم لفظي إنجليزي، فضلاً عن السياقات التي ترد فيها المتلازمات اللفظية والتعبيرات الاصطلاحية.

3- Macmillan Collocations Dictionary

Editor- in-chief, Michael Rundell, 2014

وهو معجم خاص بالمتلازمات الإنجليزية، وموجه لمتعلمي اللغة الإنجليزية فوق المتوسط إلى المتقدمين بمستويات عليا، ويضم 4500 متلازمة؛ لمساعدة الطلبة على الكتابة الأكاديمية الصحيحة؛ إذ اعتمد على مدونة حجمها (2 بليون كلمة) من النصوص الأكاديمية، والصحافة ولغة الإعلام، والقصص، والروايات، (لغة الأدب)؛ بالإضافة إلى المكالمات الهاتفية والمحادثات المسجلة، ومقابلات لرجال الأعمال⁽¹⁾.

يتضح مما سبق أن هذه الظاهرة قد سلط الضوء عليها في الدراسات الغربية (الإنجليزية خاصة)، ونالت نصيباً وافراً من البحث بوصفها ظاهرة دلالية تسهم في إيضاح المعنى، في حين نجد اللغويين العرب القدامى لم يفرّدوا دراسة خاصة للتلازم في الخطاب اللغوي القديم، غير أنه من الواضح وعي العرب القدامى بها لاسيما المعجميين، ولا يزال الجهد العربي المعجمي مقصوراً على مستوى المعاجم الخاصة والمعاجم العامة الأحادية والثنائية.

(1) Rundell, M. Macmillan Collocations Dictionary, Bridgwater, UK, 2014, p: vii

مراحل بناء المعجم الرقمي:

أولاً: مرحلة الجمع: وستكون مادة الجمع مقتصرة على المدونات الحاسوبية الموجودة على برنامج SKETCH ENGEN (AR TENTEN)، وهي مجموعة من النصوص العربية تم جمعها من شبكة الإنترنت منذ عام 2012، وعدد كلماتها حوالي 6 مليار كلمة، وتم توسيمها (POS) باستخدام أداة (MADA)، ثم تم تحميله على SKETCH ENGEN حتى يكون متاحاً للاستخدام في البرنامج الذي يصف عملية بناء ومعالجة النصوص، ويعد مورداً خصباً للأبحاث اللغوية بكافة مستوياتها النحوية والصرفية والمعجمية والدلالية، بالإضافة إلى المدونة العربية الصادرة عن مدينة الملك عبد العزيز للعلوم والتقنية، وهي مدونة عربية تحوي مليار كلمة، بدءاً من العصر الجاهلي حتى العصر الحديث، شاملة لكافة أوعية النشر من (مخطوطات وصحف وكتب ومجلات علمية، مع الأخذ في الحسبان حجم النشاط اللغوي والفكري لكل حقبة زمنية معينة، وللإستزادة بمعلومات أكثر عن المدونة العربية يمكن زيارة الموقع التالي (<http://www.kacstac.org.sa>) وصولاً إلى المتصاحبات الأكثر وروداً وفق ما يتيسر من أدوات إحصائية لكل مدونة.

ثانياً: مرحلة الوضع: وسيتم فيها تصنيف الوحدات المعجمية المتواردة معاً تقنياً؛ بحيث يتم تحديد الكلمة الأساس ويوضح معناها، ثم تُورد الكلمات المصاحبة لها الأكثر وروداً، ثم يُوضح معناها معاً، ثم يستشهد بأمثلة توضيحية للمتصاحبتين معاً.

ثالثاً: مرحلة تصميم التطبيق وفق المحددات الآتية:

- الإخراج الرقمي (التطبيقات الذكية)، وتشمل: الصفحة الرئيسية التي تضم شعار المعجم الحاسوبي، ومقدمة موجزة عن المعجم وكيفية البحث في التطبيق من خلال استخدام أيقونات محددة، وفي التطبيق الحاسوبي وظيفتان

رئيستان، هما: البحث في المعجم، وتصفح المعجم، وتفصيلهما على النحو التالي:

(أ) تصفح المعجم:

من خلال الضغط على زر "تصفح المعجم" ينتقل التطبيق إلى صفحة التصفح، كما هو في الصورة رقم (1) في الشكل التالي:



ثم تظهر أربع أيقونات رئيسية للمتصاحبات اللفظية، الأيقونة الأولى من الأعلى (قائمة المتصاحبات اللفظية بشكل عام، وتليها قائمة المتصاحبات اللفظية ذات التركيب الوصفي، ثم قائمة المتصاحبات ذات التركيب الإضافي، وأخيراً قائمة المتصاحبات ذات التركيب العاطفي). وفي كل أيقونة من هذه الأيقونات تنقل المستخدم إلى صفحة أخرى فيها قائمة المتصاحبات المطروحة، كما هو موضح في الصورة (2).

وفي حالة اختيار قائمة المتصاحبات اللفظية بشكل عام كما في الصورة (3) تظهر قائمة منسدلة بكافة المتصاحبات العامة في المعجم مرتبة أبجدياً بحسب حروف المعجم.

وفي حال اختيار قائمة المتصاحبات اللفظية ذات التركيب الوصفي ستظهر قائمة منسدلة بكافة المتصاحبات المركبة تركيباً وصفياً في المعجم، منها على سبيل المثال: (عدو لدود - عدو متريص - عدو غاشم) كما في الصورة رقم (4) و(5).



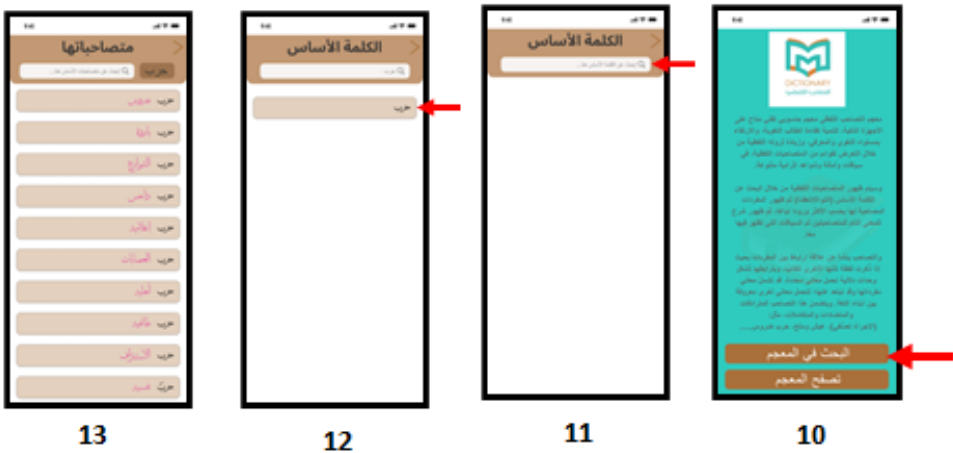
وفي حال اختيار قائمة المتصاحبات اللفظية ذات التركيب الإضافي، ستظهر قائمة منسدلة بكافة المتصاحبات المركبة تركيباً إضافياً في المعجم، منها على سبيل المثال: (حرب باردة - حرب الشوارع - حرب العصابات ...) كما في الصورة رقم (6) و(7).

في حال اختيار قائمة المتصاحبات اللفظية ذات التركيب العاطفي، ستظهر قائمة منسدلة بكافة المتصاحبات المركبة تركيباً عاطفياً في المعجم، على سبيل المثال: (عدو وصديق - عدو وحاقد) كما في الصورة رقم (8) و(9)



(ب) البحث في المعجم:

عند اختيار أيقونة (البحث في المعجم)، ستظهر أيقونة بحث الكلمة الأساس كما في الصورة (10) و (11) وعند إدخالها ستظهر الكلمات المصاحبة لها كما في كلمة (حرب) التي تصاحب معها (حرب ضروس - حرب باردة - حرب أهلية...) كما في الصورة رقم (12) و (13).



وفي حال اختيار إحدى المتصاحبات، كما في "حرب باردة"، ستظهر الأيقونات التالية:

1. (معنى التصاحب) الذي يوضح معنى المتصاحبتين معًا (حرب باردة: حرب سلاحها الدعاية والكلام واختلاق الإشاعات والتصريحات الاستفزازية) كما في الصورة (14)

2. (نوع التصاحب) الذي يبين نوع التصاحب (وصفي - إضافي - عطف). كما في الصورة (15)

3. (أمثلة) الذي يذكر مثال توضيحي للتصاحب من خلال السياق الاستعمالي للغة، (هناك قوة عظمى لاندلاع حرب باردة) كما في الصورة (16)



16



15



14

النتائج

- ترتيب منطقي وفق الأهمية، وتشتق من الأهداف والأهمية.
- حاجة الميدان اللغوي إلى البرامج النوعية والتطبيقات الرقمية المعينة على اكتساب اللغة وتعلمها.
- فاعلية مثل هذه البرامج في تعزيز دراسة اللغة العربية واكتسابها.
- توفير الوقت والجهد على دارس اللغة ومتعلمها.
- قدرة المعاجم اللغوية المتخصصة في إثراء الرصيد المعرفي والثقافي لمستخدميها.
- إسهام المتصاحبات اللغوية في تعزيز المهارات الإنتاجية والاستقبالية لمتعلم اللغة.
- تمكين دارس اللغة ومتعلمها من عناصر اللغة واستخدامها في المواقف اللغوية الحية.
- الحاجة الماسة إلى بناء معاجم رقمية.
- حاجة الميدان التعليمي لمعاجم متخصصة تسهم في تنمية الثروة اللغوية للمتعلمين.
- دور المعجمات اللغوية في تنمية الثروة اللغوية.
- فقر الميدان التعليمي لتطبيقات المعاجم الرقمية.
- ما أنتجته الدراسات السابقة من أهمية المتصاحبات في تنمية الإجابة اللغوية لأبناء العربية ومتعلميها.
- ثراء المتصاحبات بالمعاني الظاهرة والمعاني العميقة التي لديها القدرة على تنمية التفكير الناقد للطلاب.

الخاتمة

ستنفذ فكرة هذا البحث في ضوء مستجدات علم صناعة المعاجم، وعلم لسانيات المدونات، وأدوات تحليلها، وهذا بدوره سيسهم في تقدم الصناعة المعجمية من خلال معالجة اللغة العربية حاسوبياً. وسيتحقق لهذا المشروع ما لم يتحقق لغيره، وهو قيامه على مدونة عربية حاسوبية ضخمة، من شأنها إنتاج معجم قائم على مبدأ الشيوخ، لا سيما أن المدونة حديثة معتمدة على العربية المعاصرة، مع الإفادة من التراث العربي، كما أنها قائمة على أسس ومعايير علمية في الصناعة المعجمية، وبإذن الله سيرى النور قريباً.

قائمة المراجع

- 1- بعلبكي، رمزي، معجم المصطلحات اللغوية، إنجليزي-عربي، دار العلم للملايين، بيروت، 1990م.
- 2- بعلبكي، رمزي، المورد، إنجليزي-عربي، دار العلم للملايين، بيروت، 1997م.
- 3- دلول، إيمان، معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللغة العربية، الجامعة الإسلامية، كلية الآداب، غزة، 2014م.
- 4- الطاهر، هاشم حافظ، معجم الحافظ للمتصاحبات العربية، مكتبة لبنان ناشرون، بيروت، 2004م.
- 5- عمر، أحمد مختار، معجم اللغة العربية المعاصرة، عالم الكتب، القاهرة، 2008م.
- 6- ابن عمر، عبدالرزاق، المتلازمات اللفظية في اللغة والقواميس العربية، مجمع الأطرش للنشر، تونس، 2007م.
- 7- العناتي، وليد "اللسانيات الحاسوبية العربية. المفهوم، التطبيقات، الجدوى" مجلة الزرقاء للبحوث والدراسات، عمادة البحث العلمي بجامعة الزرقاء الأهلية. مج 7، ع 2، مج 7، 2005م.
- 8- غزالة، حسن، قاموس دار العلم للمتلازمات اللفظية، دار العلم للملايين، لبنان، 2007م.
- 9- القاسمي، علي، ترتيب المداخل في المعجم العربي، مجلة اللسان العربي، مكتب تنسيق التعريب، الرباط، مج 19، 1982م.
- 10- كنالي، وجدان، اللسانيات الحاسوبية العربية (الإطار والمنهج)، المؤتمر الدولي الثاني للغة العربية، دبي، 1013، ص 9-39

- 11- الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) بالتعاون مع مجمع الملك سلمان العالمي للغة العربية، معجم البيانات والذكاء الاصطناعي (انجليزي - عربي)، المملكة العربية السعودية، 2022، ص78
- 12- Baisa, V., Suchomel, V., Kilgarrieff, A. and Jakubíček, M., 2015. Sketch Engine for English Language Learning .*Corpus Linguistics* 2015، p.33.
- 13- Davies, A. and Elder, C. eds., 2008 .*The handbook of applied linguistics* .John Wiley & Sons.
- 14- Lantolf, J.P., 2006. Alan Davies and Catherine Elder (eds): The Handbook of Applied Linguistics. Blackwell, 2004 .*Applied Linguistics* ،(1)27، pp.147-152.
- 15- McCarthy, M. and O'Dell, F., 2005. English Collocations in English.
- 16- Nurmukhamedov, U., 2011. Macmillan Collocations Dictionary for Learners of English.
- 17- O'Dell, F. and McCarthy, M., 2011. *English collocations in use: advanced; how words work together for fluent and natural English; self-study and classroom use*. Cambridge University Press.
- 18- Sinclair, J., 2004 .*Trust the text: Language, corpus and discourse* . Routledge.
- 19- Stubbs, M., 1995. Collocations and semantic profiles: On the cause of the trouble with quantitative studies .*Functions of language*، (1)2، pp.23-55



رقمنة اللغة العربية وتحديات الانخراط في عالم الذكاء الاصطناعي: توظيف الأسناد الرقمية في تعليم العربية للناطقين بغيرها

**Digitizing The Arabic Language and the
Challenges of Engaging in the World of
Artificial Intelligence:
For a Digital Revolution Whose Language
Is Arabic**

أ.د. بلقاسم الجطاري

أستاذ اللغة العربية في جامعة محمد بن زايد للعلوم الإنسانية

الملخص

تستعرض هذه الورقة ما يتيح الإعلام الرقمي من فرص لتطوير تعليم اللغة العربية للناطقين بغيرها، وهي فرص كثيرة واعدة يمكن ملامستها من خلال تأمل تجارب دولية ومؤسسية مختلفة، نجحت في خلق جيل جديد من التعليمات الخاصة باللغات، بل وتغلبت على كثير من العراقيل التي كانت تقف أمام بلوغ غاياتها بهذا الخصوص، كما يمكن الوقوف عليها من خلال تأمل ما يزخر به الفضاء الرقمي من وسائط وإمكانات تفاعلية، تسرع التعلم وتختصر المسافات وتقلص الميزانيات.

هذا وتناقش الورقة أيضاً، من خلال تتبع منهجي موصول إلى حقلي: تعليمية اللغات واللسانيات الاجتماعية، بعض التحديات التي تفرضها الغاية المذكورة، إذ تضع الأصبع على بعض الصعاب التي تعيق بلوغ الأهداف المرسومة، منطلقة في ذلك من خلفية إبستمية تنظر إلى ثورة الرقمية بعين ناقدة، توازن بين مقادير الكلفة من جهة، وبين مبالغ العائدات المادية والمعنوية والقيمية من جهة ثانية، وذلك بالنظر إلى المتاح الاقتصادي والحضاري للأمة العربية، كما تضع الانفجار الرقمي الحاصل راهناً في مشرح التحري الموضوعي، على نحو تبحث فيه عن السبل الكفيلة بجعل هذا الانفجار أداة للرفع من تنافسية اللغة العربية في سوق اللغات المطلوب تعلمها عبر العالم.

كلمات مفتاحية: اللغة العربية، الحوسبة، تعليم، لسانيات تطبيقية، ديداكتيك.

Abstract

This paper is premised on the theories of artificial intelligence, basically the achievements of computational linguistics that are cognitively integrated with natural language engineering and cognitive science. The latter is based, in essence, on the findings of research in linguistics, mathematics, psychology, and physiology, which are sciences that have known great breakthroughs in recent years. Although we are still in the early stages of big data theories, it has become possible to program the computer with algorithms derived from the images of digitized linguistic blogs to perform many linguistic operations and provide practical solutions to various issues. Indeed, digitization has made it possible to analyze, deduce, classify, and evaluate language statistical data. In line with this, and in order to benefit from experiences in this field, this presentation will be focused on the digitization of the language of Dhad, given the importance it enjoys in the field of language applications via the computer.

Keywords: Arabic language, computing, education, applied linguistics, didactic.

المقدمة

لا خلاف في القول إن الثورة الرقمية الجارية قد أحدثت تحولات كبرى في مضمار تعليميات (Didactics) العلوم واللغات المختلفة، إذ سمحت بتجاوز كثير من الطرائق والوسائل والوسائط التعليمية، كما أتاحت أسنادًا (Supports) جديدة لتعليم المعارف والمهارات والقيم المتنوعة، علاوة على فتحها آفاقًا بحثية واعدة أمام المشتغلين في مضمار البيداغوجيا وما يتصل بها من علوم، وهي آفاق أخذت في بعض الأحيان أوجهًا من التنافس الحامي، بسبب ارتباط الثورة الرقمية المذكورة بمصالح مادية، وكذا بسبب علاقتها القوية بمستويات التنمية التي تسجلها البلدان الرائدة في هذا المجال.

وهكذا فقد سارعت المؤسسات المسؤولة عن صياغة السياسات التربوية والثقافية لمختلف بلدان العالم، ونظيراتها المشتغلة بقضايا اللغويات، واللسانيات التطبيقية، إلى البحث في مضمار الرقميات، وتوظيف التقنيات المستجدة، لا لتحسين اكتساب اللغات وتسريع وتيرتها لدى الصغار والكبار، بل لتشبيك العلاقات بين اللغات عبر جسور الترجمة والتواصل الرقمي، مع ما ينتج عن ذلك من عائدات حضارية تعود على متكلمي لغات العالم المختلفة، ولا بأس من التذكير هنا بما يتيح التلاقح اللغوي من فرص لتطوير اللغات والذهنيات.

نروم من خلال هذه المداخلة مقارنة موضوع تعليم العربية للناطقين بغيرها من منظور مزدوج، يقابل بين الحوافز والإكراهات، فيما يمكن وصفه مرافعة غايتها قياس للجدوى، ودعوة إلى الاستثمار في هذا المضمار؛ بشكل موضوعي ومريح، لا سيما أن الرفع من جاذبية اللغة العربية لا يمكن بلوغه ما لم يتم ابتكار تعليميات رقمية مستجدة، وليس من النافل القول: إن الإقبال على تعلم اللغات يستتبع بالضرورة إقبالاً على استهلاك منتجات ثقافية، مادية ورمزية.

أما عن العتاد المنهجي الذي نستند إليه في هذه الورقة فمصدره الفضاء الاستدلالي الذي أنتجته تعليمية اللغات، باعتبارها فرعاً وسطاً تنتهي إلى البيداغوجيا واللسانيات التطبيقية، علاوة على ما تتيحه اللسانيات الاجتماعية من مقولات ومفاهيم مناسبة، تصف السياق الاجتماعي والسياسي الذي يساعد على شرح الفرص والتحديات المرتبطة بتوظيف العنصر الرقمي سنداً في تعليم العربية للناطقين بغيرها.

1- الإقبال على تعلم العربية؛ لمحة عن الدوافع والحركات

تحتل العربية مرتبة متقدمة ضمن قائمة اللغات العالمية ذات الجاذبية، كما تزداد درجات الاهتمام بها سنة تلو أخرى، وذلك رغم تواضع الترتيب العربي في مضمار الإنتاج العلمي، وعدم انخراط المجتمعات العربية في دائرة مجتمعات المعرفة. ولئن حازت العربية هذا الترتيب المتقدم، فإن ذلك لا يعود بالضرورة إلى التعداد الديموغرافي العربي الذي يكاد يصل ربع مليار نسمة، ولكنه يعود إلى وجود إقبال متسارع على تعلم العربية من قبل غير الناطقين بها، لاعتبارات متعددة نعرض منها:

(1) تزايد عدد المسلمين بالعالم، وحرص أبناء المسلمين غير الناطقين بالعربية، وكذا الداخلين في دين الإسلام، على تعلم العربية رغبة منهم في فهم معاني القرآن والحديث، لغايات تعبدية أو معرفية ثقافية؛ وهي زيادة عددية تمس جهات العالم الأربع، من أسبابها سماحة الدين الإسلامي، واستجابته للحاجات الروحية والمادية الإنسانية المختلفة، علاوة على ارتفاع معدلات الخصوبة لدى المرأة المسلمة وارتباط ذلك بقدسية الرابطة الزوجية لدى المسلمين، خلافاً لما يقضي به الأمر في العديد من المجتمعات الحديثة.

(2) الإحاطة بالوعاء اللغوي للحضارة الإسلامية، وذلك في سياق متطلبات نديّة

الجوار، الممتدة إلى عصور قديمة، كما هو الشأن مثلاً مع الأمة الهندية، إذ "يجري اليوم تدريس العربية في العديد من الجامعات والكليات الهندية، في جميع أنحاء البلاد. هذه الجامعات تلعب دوراً في تعزيز العلاقات الثقافية القديمة في الأزمة المعاصرة. هذا وتدرس العربية حالياً في أكثر من ثلاثة آلاف مدرسة إسلامية ومدرسة حكومية على مستوى الشهادة الثانوية، وفي أكثر من سبعين كلية و35 جامعة على مستوى البكالوريوس، والماجستير، والشهادات العليا، والدكتوراه.⁽¹⁾

كما يجري ذلك في سياق الاستشراق الذي توجهه البواعث العلمية، ظاهرياً على الأقل، أو في سياق الحرص على الاستخبار الاستراتيجي المباشر، والذي تحض عليه المؤسسات السياسية المدفوعة بمقصدية مصلحة معروفة، والجدير بالذكر هنا أن هذا الاهتمام قد أسهم، رغم دوافعه المشار إليها، في إحياء التراث العربي، كما لفت الأنظار إليه⁽²⁾، بحثاً عن خصائصه اللغوية والفكرية المتنوعة.

(3) توفير السند اللغوي لمساعي الولوج إلى السوق الاقتصادية العربية، وهي سوق واعدة أمام العديد من المنتجات الاقتصادية العالمية، ذات المحتوى المادي أو الرمزي، وخصوصاً السوق الخليجية التي تعرف ارتفاعاً على مستوى معدل الدخل الفردي لمواطنيها.

(1) محمد أيوب الندوي، «المركز الثقافي العربي الهندي بالجامعة المالية الإسلامية نيودلهي ودوره في نشر الثقافة العربية»، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020، ص76.

(2) هاني، إسماعيل رمضان، تعليم العربية لغير الناطقين بها؛ رؤية استشرافية، منشورات المنتدى العربي التركي، 2017، ص12.

(4) امتلاك العربية لخصائص لسانية- تاريخية لا تتوفر لأي لغة عالمية حية، فهي لغة قديمة يفوق عمرها عمر مجمل اللغات العالمية الحية، إذ تماثل في عراقتها اللغة اللاتينية، التي تعتبر أم كثير من اللغات الأوروبية الجارية اليوم على الألسن (الفرنسية، والإيطالية، والبرتغالية..)، ونظرًا لحيازة العربية هذه الخاصية الفريدة، فقد جرت العادة أن تستنفر جهود الدارسين اللسانيين الباحثين في شؤون التطورات الدياكرونية للغات من خلال متون لغوية شفوية (لا وجود لتسجيلات شفوية تخص اللغات القديمة)، فضلًا عن الباحثين في موضوع التقابلات اللسانية الجارية بين فروع اللغات المختلفة، المنتمية إلى العائلات اللغوية المتقاربة والمتباعدة.

(5) وقوع البلاد العربية في مجال جغرافي، عرف -وما يزال- صراعات جيوعسكرية مع خصوم مختلفين، وهو ما استوجب حوافز عسكرية واستخباراتية، جعلت من العربية لغة مطلوبة في سوق اللغات ذات الجاذبية من الزاوية البرغماتية المباشرة.

(6) تطور الحوامل الرقمية المستعملة في تعليم اللغة العربية، أسوة بكثير من اللغات العالمية، وهو ما سهل مهمة تعلمها أمام الراغبين في ذلك باعتماد الصيغة التفاعلية (عن بعد).

هذا، وتنضاف إلى هذه الاعتبارات شروط وظروف أخرى عديدة مرتبطة بمحركات شخصية أو بحثية جامعية، كما هو الشأن مع المراكز والجامعات الغربية التي تخصص وحدات بحثية ودراسية عليا في تخصص اللغة العربية وتراثها اللغوي والأدبي والفكري، وإن كان هذا الباعث، في الغالب الأعم، موصولًا إلى مشاريع سيادية توجه السياسات الخارجية لعدد من دول الشمال المتقدم، والساعية إلى المشاركة الفاعلة في صياغة المخططات الجيواستراتيجية، والاستفادة من ثروات

المجال العربي الحيوي. وهي ثروات كبرى، طاقة (الطاقة الأحفورية، الطاقة الشمسية، الطاقة الريحية..)، وبشرية، وبيئية مختلفة.

2- المتاح والفرص الرقمية في تعليم العربية للناطقين بغيرها

يعيش العالم طورًا عاليًا من أطوار التطور التقني، يعرف بطور الهيمنة الرقمية، استطاع أن يخلق واقعًا تواصلًا جديدًا، سهل على الأفراد إنتاج الخطابات والمشاركة في التواصل العمومي⁽¹⁾، وهو "يشمل كل مناحي حياتنا الاجتماعية والمعرفية والتربوية، خاصة البحث في اللغات الطبيعية، تعليمًا وتعلمًا وتقييمًا، بما في ذلك تعليمها للناطقين بغيرها، فقد أصبح المتعلم في نفس مستوى معلمه في استهلاك التقنية، ولذلك بات لزامًا على المختصين تقديم المادة التعليمية على شكل موارد رقمية، إذ لم يعد في الإمكان اعتماد الموارد الورقية لتعليم اللغة الثانية، فهذه أصبحت من الماضي، لذا كان واجبًا على مصممي مناهج تعليم هذه اللغة الخضوع الحتمي لدكتاتورية التقنية، وذلك برقمتها وفق ما تفرضه مقتضيات الواقع الرقمي الجديد.⁽²⁾

نحن، إذًا، إزاء وضعية ليس لنا فيها خيار، كما قد توحى بذلك بعض الخطابات التربوية التقليدية، بل بوضعية فيها كثير من القسر والإجبار، بمعنى أن تعليم اللغة العربية، للناطقين بها وبغيرها، لن يكون ممكنًا في المستقبل القريب خارج مدار الأسناد والبرامج الرقمية. ومعنى ذلك أن مجازاة روح العصر، تقضي جعل

(1) ماهر، عودة الشمايلة، ومحمود عزت اللحام، ومصطفى يوسف كافي، الإعلام الرقمي الجديد، دار الإعصار للنشر والتوزيع، عمان، 2015، ص152.

(2) محمد الحناش، المتطلبات الرقمية لتعليم اللغة العربية لغة ثانية، ضمن التقرير الشامل لأشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها بعد كوفيد 19، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020، ص16.

الثورة الرقمية مطية لتحسين النتائج التعليمية في هذا المضمار.

لا بد أن نثير ما تتيحه الثورة المذكورة من فرص كثيرة ذات أثر على واقع تعليم العربية موضوع الحديث، إذ أصبح متاحًا أمام الباحثين المنتمين إلى أطراف الدول والقارات، إقامة شبكات متعددة للتعاون البحثي في مجال تعليمات العربية، وتقاسم الخبرات والتجارب، وتأسيس الهيئات الساهرة على تقييم النتائج وتحسين الأداءات، وتوحيد النظم الشكلية المتعلقة بتعليم العربية، من قبيل قواعد النحو والإملاء، ونوع العتاد المعجمي والمصطلحي.

توفر الأسناد الرقمية فرصًا واعدة أمام تعليمية اللغة العربية الخاصة بالناطقين بغيرها، وينبغي أن نستحضر هنا أن كثيرًا من المشتغلين بالمجال، عبر ربوع العالم، هم من فئة الباحثين ذوي التكوين المرموق في مجال تعليمية اللغات المختلفة، وعلى رأسها اللغات التي راكمت تعليميتها إنتاجًا علميًا غزيرًا، قام على أساس اشتغال نظري وتطبيقي طويل، تطلب الكثير من التحسين والتشذيب والتجاوز، فهؤلاء لا بد أن ينقلوا خبرتهم ومهارتهم إلى مضمار تعليمية اللغة العربية، وأن يقلصوا زمن تطورها إلى مسافة أقل مما سيجري لو تركت لمصيرها الطبيعي الذي يحكمه إيقاع التطور الحضاري للمجتمعات العربية.

تتيح الأسناد الرقمية موارد أخرى عديدة تشكل فرصًا أمام المشتغلين في مضمار تعليمية اللغات، إذ تسمح باعتماد التسجيلات الكتابية والسمعية والمرئية، كما تفتح الباب أمام أصناف من التفاعل والتواصل عن بعد بين مختلف الأطراف، مما يسهل المهمة التعليمية، ويقلص زمن التعلم، ويخفف من الاعتمادات المالية التي يمكن رصدها لهذه المهام. يقول أنور غيتسيتش، عميد كلية الدراسات الإسلامية في نوفي بازار، بصربيا: "إن لهذا الأسلوب [التعليم عن بعد] فوائد وإيجابيات جمة، فهو يختصر ويدلل العديد من الصعوبات، ومنها انتقال الطلاب

وسفرهم إلى مكان الدراسة، حيث يوفر مصاريف السفر والإقامة وغيرها، بالإضافة إلى تذليل عقبة تأشيرات الدخول من بلد إلى آخر، كما يمكن الكلية من التعاقد مع عدد أكبر من المدرسين المؤهلين دون أن تلزمهم بالسفر والحضور شخصيًا إلى مبنى الكلية لإلقاء المحاضرات، كما يوفر هذا النظام التواصل مع عدد أكبر من طلاب الدراسات الجامعية في مناطق عديدة من العالم⁽¹⁾.

وفي السياق نفسه، يميز الباحث محمد الحناش، في مضممار تعليم اللغات الثانية، بين عدد من المبادئ التعليمية، نعرض منها:

■ **التعلم الآلي Machine learning**، وهو على نوعين: إشرافي وغير إشرافي، والمقصود به تدريب الحاسوب برمجياً على تعرف البيانات الصحيحة، مما يمكنها من رفع قدرتها على التوقع، واتخاذ القرارات المناسبة. فكلما تحسنت قدرة الآلة على التوقع، قللنا من اعتمادها على البيانات، إلى أن نتخلص منها نهائياً، فتصبح قادرة على اتخاذ القرارات المناسبة في أي موضوع اعتماداً على القواعد التي تعلمتها بالتجربة والمراس. وهذا هو الأسلوب الذي يتبع في تعليم اللغة الثانية، حيث نبدأ بتدريب المتعلمين على تعرف البيانات التي تبنى عليها القواعد، إلى أن يتخلص المتعلم من البيانات ليصبح قادراً بنفسه على بناء بيانات جديدة وتعرف ما لم يسبق له أن خبره أثناء عملية التعلم.

■ **التعلم العميق Deep learning**، ويستخدم البيانات الالتفافية على شكل عصبونات الدماغ البشري، وهو وإن كان يطبق جزئياً بيانات التعلم الآلي، إلا أنه يفوقه من حيث استخدام البيانات غير الخطية، مثل النقاط التي تتكون

(1) أنور، غيتسيتش، التحديات التي تواجه التعليم الإسلامي العالي في أوروبا بعد جائحة كورونا، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020، ص71.

منها الصور، وهي تستخدم في تعرف الوجوه البشرية، والخرائط، وكل ما هو مبني على بيانات غير خطية، دون إغفال البيانات الخطية. وقد حقق التعلم العميق نجاحًا كبيرًا في التعليم بالألعاب والصور والفيديوهات، فأصبح بذلك، تقانة معتمدة في تعليم اللغة الثانية في أكبر مراكز التعليم في العالم.⁽¹⁾

لكن هذه المبادئ وغيرها لن تكون قادرة لوحدها على تجويد مخرجات التعلم موضوع الحديث، ذلك أن نجاحها، حسب الباحث نفسه، يبقى رهينًا بتطوير منصة اختبار لقياس المهارات عن بعد، لأن الاختصار على تصميم مناهج رقمية تفاعلية للتعليم والتعلم فقط، قد لا يمكننا من معرفة أثر كل ذلك في المتلقي، على أن تكون هذه المنصة جزءًا من منصة شاملة تقوم على الأركان الثلاثة: تعليم، وتعلم، وتقييم، كما يجب أن تشكل هذه المنصة العمود الفقري لإطار مرجعي مؤسس على أحدث المعايير العالمية التي تأخذ في الحسبان معطيات التقنيات الجديدة.⁽²⁾

3- تعليم العربية للناطقين بغيرها في عصر الثورة الرقمية؛ عراقيل وتحديات

عرضنا بضع فرص يتيحها التقدم الجاري في العالم الرقمي، وتحديدًا في مجال تعليمية العربية للناطقين بغيرها، وهي فرص ذات طبيعة نظرية أو افتراضية، لا يمكن الحكم على مدى واقعيتها إلا بعد وضعها في محك الميدان، وقياس مدى تلاؤمها مع حدود الإمكانيات الحضارية التي تحوزها المجتمعات العربية.

ولعل الجدير بالذكر في هذا المقام أن عقبات كأداء كثيرة تقف أمام إنجاح ورش رقمنة الأسناد موضوع الحديث، سواء أعلق الأمر بتعليم العربية للناطقين

(1) محمد، الحناش، المتطلبات الرقمية لتعليم اللغة العربية لغة ثانية، م. س، ص 17.

(2) نفسه، ص 20.

بها أو بغيرها، فالملاحظ أن تعليم العربية وفق الصيغة الكلاسيكية غارق في مشاكل لا حصر لها، بسبب أعطاب بنيوية يتداخل فيها العامل الاقتصادي بالعامل الثقافي والسياسي، فهل يمكن مباشرة هذا الورش التعليمي في ظل البنيات القائمة، وبالاستناد إلى موارد بشرية غير مؤهلة من الزاوية العلمية، وغير قادرة على الانخراط في ذلك بسبب ضعف الحافزية؟

أما "التحدي الآخر فينطلق من أن اللغة العربية بصفة عامة تعيش أزمة، والأزمة متنوعة تمس السياسات والتشريعات، ذلك أن لغات أجنبية دخلت على سوق اللغة العربية وأصبحت تنافسها في مجالها، فقد أصبحت الكثير من الجامعات العربية والإسلامية تدرس باللغات الأجنبية حتى في مجال تعليم اللغة العربية، فعندما ننظر إلى تعليم العربية في بعض الدول في أوروبا وأمريكا نجد أن العربية تدرس بواسطة اللغات الأجنبية⁽¹⁾.

ينضاف إلى ذلك غياب النموذج الذي يمكن الاحتذاء به، علاوة على اختلاف أهداف مؤسسات تعليم اللغة العربية للناطقين بغيرها، فكل مؤسسة، وكل دولة إلا ولها أهداف مختلفة في تعليم العربية للناطقين بغيرها، إذ تختلف المصالح وتتضارب، بالإضافة إلى وجود تضارب في المصالح بين المؤسسات، سواء أكانت خاصة أو حكومية (نفسه)، وهذا الأمر مرتبط بواقع الدولة القطرية، الذي يفرض خيارات تدبيرية وسياسية لا تقع بالضرورة على أثر خيارات غيرها من الدول العربية. كما يرتبط بغياب رؤية دولية أو قومية استراتيجية، واضحة ومتفاوض بشأنها من قبل المؤسسات العربية المختلفة.

(1) علي، عبد الله موسى، تحديات تعليم اللغة العربية للناطقين بها وبغيرها، تجديد الاستراتيجيات وتطوير الوسائل والمنهجيات، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020، ص 27.

هذا وينتصب مأزق العاميات عائقًا في هذا الباب أيضًا، إذ تتباين المواقف المؤسسية من مسألة استدعاء العاميات في تعليم اللغة العربية للناطقين بغيرها، بين رافض ومرحب ومنفتح على العاميات على سبيل الاستئناس، ولعل ما يزيد من تلبيس الوضع وجود فئات- زبناء يرغبون في تعلم العاميات العربية المختلفة، لاعتبارات اقتصادية أو سياسية.

أما العائق الأكبر فهو قلة الاعتمادات المالية المخصصة في هذا المضمار، وذلك بسبب ضعف الميزانيات التي تخصصها الدول العربية لقطاعي التعليم والثقافة، وهكذا يستمر الطابع الإرادوي (غير موصول إلى سياسات مؤسسية) في تعليم العربية للناطقين بغيرها، إذ يعود الفضل في إنجاز هذه المهام، وخصوصًا بالدول الغربية، إلى "عدد من المراكز الإسلامية والجمعيات الثقافية والخيرية، وللعلماء والمشايخ والدعاة والبعثات الدينية، وبعض المتطوعين والغيورين من أبناء الجاليات"،⁽¹⁾ وهو ما يتضح من إفادة محمد محسن الغزاوي، مدير أحد المراكز الإسلامية بالبرازيل، إذ يقول: "لقد فتحت بعض المدارس العربية، وإن بشكل متواضع وخجول، واستطاعت أن تواكب المناهج الأجنبية (البرتغالية) إلى حد ما، إلا أن هذه المدارس تعثرت بعض الشيء؛ لأنها لم تحظ بأي دعم من أية جهة خارجية، عربية كانت أم إسلامية. فهي تارة تضعف وتارة تعود من جديد، والسبب قلة الموارد بالإضافة إلى الأزمات الاقتصادية التي يمر بها العالم منذ حوالي عقدين من الزمن، وحتى ظهور هذه الجائحة التي أوقفت البلاد والعباد."⁽²⁾

(1) الغزاوي، محمد حسين، جهود المراكز الإسلامية في البرازيل لخدمة اللغة العربية، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020، ص83.

(2) نفسه.

تركيب:

عرضنا، في هذه المداخلة، بعضاً مما يتيح السند الرقمي في تعليم اللغة العربية للناطقين بغيرها، ولمحة عن أهم التحديات التي تعيق هذا المشروع اللساني-البيداغوجي، من خلال تأمل دراسي ينطلق من معطيات الواقع، ويتخذ تجارب محددة سنداً في النظر والحكم.

لقد تأكد لدينا أن الأسناد الرقمية تشكل فرصاً واعدة أمام المشتغلين في مضمار تعليمية اللغات على نحو عام، وفي مضمار تعليم العربية لغير الناطقين بها على نحو خاص، إذ تسمح باعتماد التسجيلات الكتابية والسمعية والمرئية، كما تفتح الباب أمام أصناف من التفاعل والتواصل عن بعد بين مختلف الأطراف، مما يسهل المهمة التعليمية في عموميتها؛ حتى وإن ظل ذلك رهيناً بتوفر شرط رئيس، هو تطوير منصة اختبار لقياس المهارات عن بعد، لأن الاقتصار على تصميم مناهج رقمية تفاعلية للتعليم والتعلم فقط، قد لا يمكننا من معرفة أثر كل ذلك في المتلقي، على أن تكون هذه المنصة جزءاً من منصة شاملة تقوم على الأركان الثلاثة: تعليم، وتعلم، وتقييم.

أما عن العوائق التي تحول دون بلوغ الغاية المرجوة في هذا الصدد فكثيرة ومختلفة، لكن أكبرها، وأشدّها وطأة، قلة الاعتمادات المالية، وذلك بسبب ضعف الميزانيات التي تخصصها معظم الدول العربية لقطاعي التعليم والثقافة، وهكذا نجد أن الفضل في بلورة مشاريع تعليم العربية لغير الناطقين بها، وخصوصاً بالدول الغربية، يعود غالباً إلى عدد من المراكز الإسلامية والجمعيات الثقافية والخيرية، وللعلماء والمشايخ والدعاة والبعثات الدينية، وبعض المتطوعين والغيورين من أبناء الجاليات.

لا بأس من التأكيد-ختاماً- على الأهمية الشديدة التي يكتسبها العامل الثقافي-

السياسي في تصميم مشاريع في هذا الصدد. إذ لا مناص من صياغة مشروع عربي كبير، شريطة أن يكون نتاجاً لاجتهاد جماعي تتم صياغته على نحو مؤسسي، تساهم فيه كل الدول العربية بمواردها المادية والبشرية الكفؤة، وهو ما يجب الانخراط فيه بشكل مستعجل؛ لأن من شأن هذا المشروع إكساب الدول العربية عائدات اقتصادية وحضارية كبرى، لا سيما وأن كثيراً من المنتجات والتعبير الفنية العربية تحوز، في ذاتها، إمكانات كبرى للانتشار، ولا يقف أمام تنافسيتها سوى عائق اللغة، وما يتعلق به من الشروط التواصلية المختلفة.

قائمة المراجع

- 1- أنور، غيتسيتش، التحديات التي تواجه التعليم الإسلامي العالي في أوروبا بعد جائحة كورونا، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020.
- 2- علي، عبد الله موسى، تحديات تعليم اللغة العربية للناطقين بها وبغيرها، تجديد الاستراتيجيات وتطوير الوسائل والمنهجيات، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020.
- 3- الغزاوي، محمد حسين، جهود المراكز الإسلامية في البرازيل لخدمة اللغة العربية، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020.
- 4- ماهر، عودة الشمايلة، ومحمود عزت اللحام، ومصطفى يوسف كافي، الإعلام الرقمي الجديد، دار الإعصار للنشر والتوزيع، عمان، 2015.
- 5- محمد، أيوب الندوي، المركز الثقافي العربي الهندي بالجامعة المليية الإسلامية نيودلهي ودوره في نشر الثقافة العربية، تقرير عن أشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020.
- 6- محمد، الحناش، المتطلبات الرقمية لتعليم اللغة العربية لغة ثانية، ضمن التقرير الشامل لأشغال المنتدى الدولي الافتراضي للغة العربية للناطقين بغيرها بعد كوفيد 19، منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة، 2020.
- 7- هاني، إسماعيل رمضان، تعليم العربية لغير الناطقين بها؛ رؤية استشرافية، منشورات المنتدى العربي التركي، 2017.



التعابير المسكوكة في اللغة العربية من التوصيف اللساني إلى المعالجة الحاسوبية

**Idioms in Arabic Language
From Linguistic characterization to
Computer Processing**

إعداد

د. شائق عبد الرحمان

كلية الآداب والعلوم الإنسانية – مراكش

ملخص البحث

يندرج هذا البحث ضمن حقل اللسانيات الحاسوبية، ويتوخى وضع تصور لساني-حاسوبي؛ لبناء قاعدة بيانات لتعابير اللغة العربية المسكوكة، وتحويلها إلى نموذج صوري يتم ضبطه والتحكم فيه بواسطة الحاسوب. وتتجلى أهمية البحث في هذه الظاهرة اللغوية، في النقاط التالية:

- وجود مواقع إلكترونية غربية اهتمت بحوسبة هذه الظاهرة اللغوية في لغات متعددة، كاللغة الإنجليزية، واللغة الفرنسية، واللغة الألمانية، في غياب تام لموقع عربي يهتم بجمع ونشر هذه التعابير التي هي جزء لا يتجزأ من متن اللغة العربية.

- بناء قاعدة بيانات تضم هذه التعابير قديمها وحديثها، ونشرها على الشبكة العنكبوتية ليعرفها ويتعلمها الناطقون باللغة العربية والناطقون بغيرها؛ فهي جزء مهم من تعلم لغة ما ومعرفة استعاراتها.

- الإسهام في إثراء المحتوى الرقمي العربي.

ننتقل في معالجة هذه الظاهرة اللسانية آلياً من الإشكالية التالية: كيف يمكن للحاسوب أن يفهم التعابير المسكوكة وينتجها؟. تفرعت عن هذه الإشكالية مجموعة من الأسئلة التي سنجيب عنها، وهي:

ما المقصود بالتعابير المسكوكة؟ ما سماتها المميزة؟ ما التوصيف اللساني المقترح لمعالجتها حاسوبياً؟ وما الصعوبات التي ستواجه الحاسوب عند معالجتها؟ وما الحلول المقترحة لحل هذه الصعوبات؟ وكيف سنستثمر نظرية الإطارات في بناء قاعدة بيانات آلية للتعابير المسكوكة، تستطيع بواسطتها توليد معطياتها اللغوية على مستويي الفهم والإنتاج؟.

وللإجابة عن هذه الأسئلة المتعلقة بالمستويين النظري والتطبيقي للبحث، سنتبع المنهجية التالية:

- الجمع: جمع متن من التعابير المسكوكة.
- الوصف: وصف الخصائص اللسانية للتعابير المسكوكة.
- التطبيق: إعداد نظام معلوماتي يستطيع فك رموز التعابير المسكوكة عند معالجتها آلياً.

الكلمات المفتاحية: التعابير المسكوكة، التوصيف اللساني، اللسانيات الحاسوبية، التمثيل الآلي، قاعدة البيانات، نظرية الإطارات.

Abstract

This study falls within the field of computational linguistics. It intends to develop a computational linguistic visualization to build a database of Arabic idioms and convert them into a graphic model that is computer-tuned and controlled. The importance of this study is reflected in the following points:

- The existence of Western websites that are interested in computerizing this linguistic phenomenon in multiple languages, such as English, French and German, while the absence of an Arabic website that is interested in collecting or publishing these idioms.
- Build a database of old and modern idioms and publish it on the Web so that Arabic and non-Arabic speakers know and learn them, as they are an important part of language learning.
- Contribute to enriching Arabic digital content.

The study's main problem is: How can a computer understand and produce idioms? To answer this question, we first have to answer some other sub-questions, including: What is meant by idioms? What are their distinctive features? What is the suggested linguistic description for its computational processing? What are the difficulties that the computer will face when dealing with idioms? What are the proposed solutions to solve these difficulties? How will we invest frame theory in building an automated database of idioms through which they can generate their linguistic data at the understanding and production levels?

To answer these questions, we will follow the following methodology:

- Data collection: collect a corpus of idioms.
- Description: describe the linguistic characteristics of idioms.
- Application: develop an information system that can decode collocations when processed automatically.

Keywords: Idioms, linguistic characterization, computational linguistics, automated representation, database, frame theory.

المقدمة

أثار انتباهنا عند تصفح المواقع الإلكترونية المتخصصة في المعاجم الحاسوبية الموجهة للطلبة الناطقين بلغاتها والناطقين بغيرها، اهتمامها الكبير بصناعة معاجم حاسوبية للتعابير المسكوكة، تضم متوناً ضخمة من هذه التعابير قديمها وحديثها، إضافة إلى خاصية التحيين (Update) التي يُضاف بها كل تعبير جديد ظهر استعماله في لغاتهم ونشرها إلكترونياً، كما هو حال اللغات الإنجليزية، والألمانية والفرنسية، وهذا على سبيل المثال لا على سبيل الحصر. في غياب- على حد علمنا- لموقع عربي يهتم بجمع هذه التعابير ونشرها، التي تمثل متناً ضخماً وجزءاً مهماً من معجم لغتنا العربية. كما نذكر أنّ أخطر مجال أظهره اندماج اللغة والحاسوب، هو الغزو الثقافي للغات المنتجة للمعرفة كما هو الحال بالنسبة للغة الإنجليزية، التي غزت بقوة جميع جوانب المعرفة الإنسانية، مما أدى إلى عزوف الكثير عن لغتهم لضعف محتواها الرقمي إنتاجاً معرفياً وتحييناً.

دفعنا ذلك، إلى البحث عن طريقة فعالة تسير عصر السرعة والانتشار السريع للمعلومات، فكانت الطريقة هي التفكير في معالجة التعابير المسكوكة من منظور لساني-حاسوبي. لكن السؤال المطروح: كيف يمكن القيام بهذا التحليل اللساني وهذه الحوسبة؟ ولمعرفة ذلك، يتطلب منا الوقوف على مستويين اثنين وهما: المستوى النظري، والمستوى التطبيقي، فمن الناحية النظرية: سنعرف بهذه الظاهرة اللسانية وسماتها المميزة وإعطاء أمثلة منها، ومن خلال ذلك سنقترح توصيفاً دقيقاً للخصائص اللسانية للتعابير المسكوكة في أبسط جزئياتها، إذ يعد عملية مهمة وضرورية؛ لأنها تعابير يصعب فهمها عند الإنسان

نفسه حتى باستعمال حواسّه التي تساعد على تفسير ما يسمع من أقوال تطابق الواقع أو لا تطابقه، ناهيك عن فهم الحاسوب لها. أما من الناحية التطبيقية: سننطلق من الفرضية التالية: من الممكن أن يتعامل الحاسوب فهمًا وإنتاجًا مع التعابير المسكوكة إذا تم إعداد نواة لغوية صلبة تحلل هذه التعابير، وتفكّكها إلى خصائصها اللسانية، على مستوى الصوت، والتركيب، والدلالة، والتداول، في إطار فهم شامل لخصائص الحاسوب وكيف تمثل المعرفة اللغوية فيه، مستثمرين في ذلك نظرية الإطارات (Frames) التي أثبتت فاعليتها ونجاحتها في معالجة التعابير المسكوكة آليًا، وذلك ما سنعرضه من خلال تطبيق حاسوبي يربط بين المستويين النظري والتطبيقي نبين فيه باللموس، تطبيق الجانب النظري على الحاسوب باعتماد قاعدة بيانات آلية تم تصميمها وصياغتها بشكل يطابق عدد الأطر التي حددها التوصيف اللساني للتعابير المسكوكة.

- ثم نختم هذا البحث بأهم الخلاصات والنتائج التي تم التوصل إليها. ولربط بين المستويين النظري والتطبيقي والإجابة عن أسئلتهم، سنتبع المنهجية التالية:
- **الجمع:** نجتمع فيها متنا من التعابير المسكوكة بمراعاة السمات المميزة لها.
 - **الوصف:** نصف فيها الخصائص اللسانية للتعابير المسكوكة على مستويات الصوت، والتركيب، والدلالة، والتداول، ووضعها في قاعدة بيانات.
 - **التطبيق:** يتم فيها تمثيل مستويات الوصف اللساني وترجمته إلى لغة الآلة، بتتبع نموذج آلي لتمثيل المعرفة اللغوية؛ لإعداد نظام معلوماتي يستطيع فك رموز التعابير المسكوكة عند معالجتها آليًا.

أولاً: المستوى النظري

1. التعابير المسكوكة: ماهيتها وسماتها

عرف كريم زكي التعابير المسكوكة بقوله: "نمط تعبير خاص بلغة ما، يتميز بالثبات ويتكون من كلمة أو أكثر، تحولت عن معناها الحرفي إلى معنى يغايره اصطلحت عليه الجماعة اللغوية".⁽¹⁾

نستنتج من هذا التعريف، ثلاث سمات تحدد مفهوم التعبير المسكوك وهي:

أ- ثبات التعبير المسكوك.

ب- إمكانية اقتصار التعبير على كلمة أو كلمتين أو أكثر.

ت- تحول كلمة التعبير من معناها الحرفي، إلى معنى متفق عليه من الجماعة اللغوية.

وعرفها أندرياس لانكلوتز (Andreas Langlotz) التعابير المسكوكة⁽²⁾ (Idioms) أو (Idiomatic expression)، بقوله: "هي وحدات لغوية عرفية، متعددة الكلمات، معناها خفي ولها تركيب ثابت، تتميز بعدم الشفافية الدلالية".⁽³⁾

يحدد هذا التعريف مجموعة سمات للتعبير المسكوك وهي:

أ- وحدات لغوية متعددة الكلمات.

(1) . زكي كريم، حسام الدين، التعبير الاصطلاحي: دراسة في تأصيل المصطلح ومفهومه ومجالاته الدلالية وأنماطه التركيبية، المكتبة الأنجلو المصرية، القاهرة، 1985، ص. 34-35.

(2) Idiom

مصطلح مشتق من الكلمة اليونانية Idiomatikos التي تحمل معنى التفرد والخصوصية، كما تعني كلمة Idiom أيضاً لغة أو لهجة أو منطقة أو طبقة أو مجموعة من الناس، كما تعني "تعبير مميز في لغة ما أو نمط ثابت من التعبير".

(3) Langlotz, Andreas, Idiomatic Creativity: A cognitive-Linguistic model of idiom-representation and idiom-variation in English, John Benjamins Publishing Company, 2006, p. 2.

ب- تتعارفها جماعة لغوية.

ت- تركيبها ثابت.

ث - خفاء معناها وعدم شفافيتها الدلالية.

يلاحظ من هذين التعريفين أنهما لم يذكرنا سمات أخرى للتعابير المسكوكة، مثل كونها تعابير استعارية لا يمكن ترجمتها حرفياً، وكونها محدودة الانتشار، وخالية من الحكمة والموعظة والإرشاد والتذكرة.

1. أمثلة من التعابير المسكوكة

- رِيحٌ يُوسَفُ: يقال فيما يُحسَّ به من أثر الشَّيء السَّار. (1)
- رَشْحُ الْحَجَرِ: يُقال في البَخِيلِ يَجُودُ بِالشَّيْءِ القليل على عُسْرَةٍ وَنَكْدٍ. (2)
- دُودَةُ الْقَرِّ: يقال فيمن يَضُرُّ نَفْسَهُ وَيَنْفَعُ غَيْرَهُ. (3)
- إِحْرَاقُ الْبُخُورِ: الْمُبَالِغَةُ فِي الْمَدِيحِ تَزْلُفًا وَنِفَاقًا. (4)
- ثُلَّ عَرَشُهُ: أَي زَالَ قَوَامُ أَمْرِهِ، وَذَهَبَ عِزُّهُ وَسَاءَتْ حَالُهُ. (5)
- مَلَحَهُ عَلَى رُكْبَتَيْهِ: يُقال في سَيِّءِ الْخُلُقِ يَغْضَبُ مِنْ كُلِّ شَيْءٍ وَيَصِيحُ مِنْ أَدْنَى شَيْءٍ. (6)

(1) الثعالبي، أبو منصور عبد الملك بن محمد بن إسماعيل (350-429هـ)، ثمار القلوب في المضاف والمنسوب، تحقيق: محمد أبو الفضل، دار المعارف، القاهرة، 1965، ص 50.

(2) السابق، ص 558.

(3) الثعالبي، ثمار القلوب، ص 434.

(4) داود، محمد محمد، معجم التعبير الاصطلاحي في العربية المعاصرة، دار غريب، القاهرة، د. ط، 2003، ص 25.

(5) العبدري، أبو الحسن بن محمد الشيبني (837/1433م)، تمثال الأمثال، تحقيق: أسعد ذبيان، دار المسيرة، بيروت، 1982، 2/ 403.

(6) الأنباري، أبو بكر محمد بن القاسم (328هـ)، الزاهر في معاني كلمات الناس، تحقيق: حاتم صالح الضامن، دار الشؤون الثقافية العامة-آفاق عربية، بغداد، ط 2، 1987، 1/ 325.

نلاحظ من خلال هذه الأمثلة أن متكلمي اللغة العربية، يستعملون ألفاظاً مركبة تتسم بالتحجر أو التكلس (figement) ثابتة بالتواضع، تأخذ معناها من الإطار الثقافي والبيئي للمجتمع، للتواصل بأوجز تعبير عن أمور يمكن أن تأخذ وقتاً كثيراً من الشرح والتفصيل، وهذا التواضع يجعل سامع هذه التعبيرات على علم بالمعنى الجديد للفظ وتحولّه من معناه الحرفي إلى معنى استعاري.

هذا ما ذهب بمحليي التعابير المسكوكة إلى القول، بأنه: "من المستحيل أن نحيط بما يستخلص من التركيب اللغوي للتعابير المسكوكة دون الاعتماد على حزمة من الأبعاد التعريفية".⁽¹⁾

نُجمل معرفة هذه الأبعاد التعريفية للتعابير المسكوكة، في الآتي:

- أ- الخصائص الدلالية، التي تخرج باللفظ من معناه الحرفي إلى معنى مجازي.
- ب- التراكيب الغريبة والشاذة، التي تظهر في تركيبها استعمال أسماء الحيوانات، الأطعمة، الصحة... إلخ.
- ت- القيود المفروضة على وظيفتها المعجمية والنحوية، التي لا يمكن تفسيرها من خلال القواعد النحوية للغتها.

2. ترجمة التعابير المسكوكة من لغة إلى أخرى

ذهب المتخصصون في دراسة هذه الظاهرة اللغوية إلى صعوبة ترجمة التعابير المسكوكة⁽²⁾. لأن هذا أحد المجالات الذي يمكن أن تكون فيه اللغات مختلفة جداً. ففي بعض الأحيان يمكن ترجمة التعابير المسكوكة، ولكن ذلك غير ممكن في أغلب الأحيان. على سبيل المثال: التعبير المسكوك الإنجليزي التالي:

(1) Langlotz, Idiomatic Creativity, pp. 2-3.

(2) Wright, Jhon, Idioms organizer by metaphor, topic and key word, Zrinski d.d, 2002, P. 10.

Break the ice⁽¹⁾ إذا ترجمناه إلى اللغة العربية حرفيًا فسنحصل على الترجمة التالية: كَسَّرَ الجَلِيد وهذا ليس المعنى المراد منه، وإنما معناه عند متكلمي اللغة الإنجليزية مَهَّد الطريق لعمل ما.

وفي اللغة الفرنسية نورد التعبير المسكوك التالي: De bouche à oreille⁽²⁾ فترجمته إلى اللغة العربية حرفيًا تعطينا الترجمة التالية: من الفم للأذن وهذا ليس المقصود منه بين متكلمي اللغة الفرنسية، وإنما يراد به في ثقافتهم من نشر الخبر بسرعة كبيرة.

يجب أن يكون المترجم حذرًا عند ترجمة التعابير المسكوكة، حيث لا يمكن ترجمتها كلمة بكلمة، بل يجب ترجمة التعبير ككل. في بعض الأحيان يمكن أن يترجم التعبير المسكوك إلى لغة ما بوجود مكافئ له في اللغة الهدف، ولكن في الغالب الأعم قد لا يكون هناك تعبير مسكوك يشابهه في اللغة الهدف، فيقوم المترجم فقط بشرح المعنى، وترجع أسباب صعوبة ترجمة التعبيرات إلى أنها:

- مجال يعكس ثقافة متكلمي لغة ما.
- الاستعارات في الثقافة الواحدة تختلف عن تلك التي في ثقافة أخرى.

3. التعابير المسكوكة وتعلم اللغة الثانية

تعد التعابير المسكوكة واستعاراتها مهمة جدًا في تعلم اللغة الثانية ومعرفة ثقافتها للأسباب التالية⁽³⁾:

أ- لأنها شائعة جدًا، فمن المستحيل التحدث، والقراءة، والاستماع، بتعلم لغة

(1) Spears, A, Richard, NTC's American Idioms Dictionary, McGraw-Hill, 2000, p. 49.

(2) Bordon, Jean-Luc, Best of Idioms, Ellipses Edition, 2008, p. 22.

(3) Wright, Idioms organizer by metaphor, P. 9.

ثانية دون معرفة تعابيرها المسكوكة، فمثلاً التعبير المسكوك الإنجليزي التالي: It rains cats and dogs⁽¹⁾ الذي يعني عند متكلمي الإنجليزية هطل الأمطار بغزارة، فمتكلم العربية حينما يسمع أو يقرأ هذا التعبير رغم قدرته على التحدث بلغتهم وأغفل هذا الجانب من لسانهم تعد معرفته بمعجمهم اللغوي ناقصة. وهذا الأمر لا يمكن تركه حتى الوصول إلى مرحلة متقدمة من تعلم اللغة. فكل المتحدثين باللغة الإنجليزية يستعملون التعابير المسكوكة، كما أن الصحف مليئة بهذه اللغة المجازية.

ب- لأن الاستخدام المجازي للكلمة في كثير من الأحيان يكون أكثر شيوعاً - اليوم - من الاستخدام الحرفي لها.

ت- لأن هذا النوع أو الجانب من اللغة ممتع في تعلم اللغة الثانية وعلى تذكر الأشياء والقدرة على إبداع استعارات جديدة.

إذن توصف التعابير المسكوكة على أنها رموز مركبة ومعقدة، تتضمن خصائص شكلية ودلالية وتداولية وسوسiolسانية عرفية تدخل في نطاق القواعد النحوية للغة ما، تقوم بوظائف خطابية وتواصلية ذات معنى، لا يمكن أن يستمد من مجرد فهم معاني كلماتها منفصلة؛ أي لا يمكن الاعتماد على معناها الحرفي. فرغم معرفتنا وإلمامنا بمعاني كل كلماتها وتركيبها وجوانبها النحوية، يظل المعنى الإجمالي للتعبير مُعْتَمَداً. فما الصعوبات التي ستواجه الحاسوب في معالجته للتعابير المسكوكة؟ وما التوصيف اللساني المقترح لحل هذه الصعوبات وتجاوزها؟ وكيف سنستثمر نظرية الإطارات في تصميم قاعدة بيانات آلية للتعابير المسكوكة في اللغة العربية؟ هذا ما سنتطرق إليه على المستوى التطبيقي.

(1) Spears, NTC's American Idioms Dictionary, p. 335.

ثانياً: المستوى التطبيقي

1. الصعوبات التي ستواجه الحاسوب في معالجته للتعابير المسكوكة

يمكن أن تعترض الحاسوب في معالجته للتعابير المسكوكة صعوبتان اثنتان تأويليتان وهما: الصعوبة التأليفية، والصعوبة التداولية.

- الصعوبة التأليفية⁽¹⁾: هي الصعوبة الناتجة عن اختفاء بعض مظاهر الدلالة وراء تركيب الكلمات في العبارة، ويتجلى دور التأليف في دراسة التعابير المسكوكة على مستويين اثنين هما: المستوى النحوي: احترام التعابير المسكوكة لقواعد اللغة وهذا هو الغالب في جلها. المستوى الدلالي: كشف اللبس أو الغموض في معاني التعابير المسكوكة، وهذا أصعب تحد ستواجهه الآلة؛ لأنها ستعمل على تحديد معنى عبارة تتألف من كلمات مترابطة لا تقبل التجزيء.

- الصعوبة التداولية: وهي الناتجة عن اختفاء بعض مظاهر المعنى المتعلقة بالأبعاد الخارجة عن منطوق الكلام، كالسياق والنية وغيرهما. ومفاد هذه الصعوبة أن العقل الإنساني يستثمر "التنافر" الدلالي⁽²⁾ لإعطاء التعابير المسكوكة بعداً بلاغياً وجمالياً وفنياً، معتمداً في ذلك على لغته وحده وما تتيحه الطبيعة والبيئة التي يعيش فيها. أما عقل الحاسوب فلا يستطيع أن يؤول "التنافر" الدلالي لعدم توفره على الحدس والسياق والقدرة على معرفة الانسجام الخطابي لهذه التعابير ومطابقتها للواقع (المتكلم، المخاطب، السياق "الإشاري").

(1) جرمان، كلود؛ ولوبلون ريمون، علم الدلالة، ترجمة: نور الهدى لوشن، جامعة قار يونس، بنغازي، 1997، ص 33.

(2) استخدام لفظ واحد بمعنيين متضادين، مثل كلمة "جون" التي تستعمل في اللغة العربية بمعنى أبيض وأسود كليهما.

2. المنهج المقترح

نقترح لحل هاتين الصعوبتين أن نزود الحاسوب بقاعدة بيانات آلية، التي تعتبر مرحلة مهمة في أي تمثيل لغوي أريدت حوسبته آلياً؛ لأن الحاسوب في معالجته للظواهر اللغوية لا ينطلق من فراغ، بحيث تعطيه مُدخلًا وتنتظر منه مُخرجًا دون أن توفر له قاعدة معطيات محددة بمعلومات لغوية منتظمة واضحة، تم توصيفها بشكل دقيق وبسيط تراعى فيه خصائصه⁽¹⁾. تتضمن قاعدة البيانات هاته تحليل التعابير المسكوكة بدلالة خصائصها اللسانية على مستويات: الصوت، والتركيب، والدلالة، والتداول، أخذاً الترتيب التالي:

- التمثيل الصوتي.
- التمثيل التركيبي.
- التمثيل الدلالي.
- التمثيل التداولي.

أ- التمثيل الصوتي:

الكتابة الصوتية⁽²⁾ عبارة عن رموز موحدة تستعمل لتمثيل الأصوات في جميع اللغات الطبيعية⁽³⁾، حيث تسهل لمن يعرف رموزها قراءة كلمات اللغات التي لا يتقنها⁽⁴⁾. تمثيل التعابير المسكوكة بالكتابة الصوتية بتحويلها من حروف

(1) Moreau, R, Introduction à la théorie des langages, Librairie Hachette, 1975, pp. 30-31.

(2) انظر: موقع الجمعية الصوتية الدولية: <http://ipa.typeit.org/full/>.

(3) الألفبائية الصوتية الدولية (إنجليزية: IPA، فرنسية: API) نظام كتابة شامل لجميع الوحدات الصوتية الموجودة في أي لغة، من ابتكار الجمعية الصوتية الدولية، وذلك كتمثيل موحد للأصوات في اللغات المتكلم بها.

(4) حنا، سامي عياد؛ وكريم زكي حام الدين، نجيب جريس، معجم اللسانيات الحديثة: انجليزي-عربي، مكتبة لبنان، لبنان، 1997، ص.ص 67-68.

إلى رموز، سيساعد الحاسوب على ضبط وفهم صوائها وصوامتها؛ لأن الآلة تتعامل بالرموز. كما أن التمثيل الصوتي تمثيل صوري يفهمه الحاسوب، وسنعزز الكتابة الصوتية بالصوت المسموع لكيفية قراءة كل تعبير مسكوك قراءة صحيحة.

ب- التمثيل التركيبي:

تتوزع التعابير المسكوكة في بنيتها التركيبية بين: المركبات الاسمية (م س)، والمركبات الإضافية (م إض) والمركبات الفعلية (م ف)، والمركبات الاستفهامية (م اس)، والمركبات الحرفية (م ح)، كما تندمج ضمنها المركبات الحدية (م حد). تشجير⁽¹⁾ التعابير المسكوكة على شكل رسم بياني، سيظهر ترتيب العناصر والعلاقات البنيوية الرابطة بينها وتحليل قواعد تركيبها لتبيين كيفية تقسيمها وتحديد أجزائها. وهو تمثيل صوري لا يشكل عائقاً للآلة.

ت- التمثيل الدلالي:

لا يستطيع الحاسوب فهم مجازية التعابير المسكوكة لأسباب ذكرناها آنفاً (المجازية، التنافر الدلالي، وحدة دلالية متكاملة)، ولتبسيط معناها للحاسوب سنوصف المعاني الحقيقة الأصلية المتواضع عليها⁽²⁾ لهذه التعابير بين الجماعة اللغوية، قبل نقلها واستعمالها في غير ما وُضعت له لغاية بلاغية وجمالية⁽³⁾ وتأثيرية في نفس المتلقي، وبذلك نوفر للحاسوب معرفتين: معرفة بالمعنى الحقيقي،

(1) التشجير: رسم بياني ابتكره تشومسكي يوضح فيه طريقة تطبيق القواعد وكيفية وضع المادة المعجمية لتناسب جملة معينة ومعرفة بنيتها العميقة.

(2) رشوان، محسن وآخرون، مقدمة في حوسبة اللغة العربية، دار وجوه للنشر، المملكة العربية السعودية، 2019، ص 220.

(3) السابق، ص 220.

لأن الوضوح شرط من شروط عمل الحاسوب⁽¹⁾، ومعرفة بالمعنى المجازي، الذي يظهر فيه انتقال الكلمة من معناها الأصلي إلى معنى آخر. والذي يُشكل لبساً أو غموضاً دلاليّاً للحاسوب.

ث- التمثيل التداولي:

تتميز التعابير المسكوكة بخروجها عن المؤلف تركيبياً ودلاليّاً؛ لتحمل ذهن سامعها على فهم معين وإدراك مخصوص لمضمونها في سياق خطابي، يتميز بالتعارض بين المعنى الظاهر للمفوضها والعالم الحقيقي (الواقع). هنا تطرح الصعوبة في كيفية تمثيل السياق بالنسبة للحاسوب.

نقترح لذلك توصيف الانزياحات الداخلية والخارجية للتعابير المسكوكة؛ لأن ألفاظها انزاحت من المؤلف إلى غير المؤلف في معانيها. فالانزياح معيار أسلوبِي يتحدد به السياق الذي يؤدي بدوره إلى تحديد المعنى.

يعرف ريفاتير (Riffaterre)⁽²⁾ الانزياح بقوله: "هو انزياح عن النمط التعبيري المتواضع عليه، بخرق القواعد حيناً ولجوءاً إلى ما نذر من الصيغ حيناً آخر"⁽³⁾. فالانزياح عند " ريفاتير " مبني على التضاد⁽⁴⁾ (مخالفة)، وطرّاه السياق والمخالفة، وقاعدته هي:

سياق أصغر + تضاد (مخالفة) = انزياح⁽⁵⁾

(1) Moreau, Introduction à la théorie des langages, p. 26.

(2) باحث ألسني، وناقد أدبي بنيوي أمريكي، وأستاذ في جامعة كولمبيا.

(3) المسدي، عبد السلام، الأسلوبية والأسلوب، الدار العربية للكتاب، القاهرة، ط 2، 1982، ص 103.

(4) جيرو، بيير، الأسلوبية، ترجمة: منذر عياش، دار الحاسوب للطباعة، حلب، ط 2، 1994، ص.ص 123-124-125.

(5) فضل، صلاح، علم الأسلوب: مبادئه وإجراءاته، دار الآفاق الجديدة، بيروت، 1985، ص.ص 194-206.

ونمثل لذلك بالتعبير المسكوك التالي: أبو البيضاء

أبو ← سياق أصغر؛ البيضاء ← تضاد (مخالفة) = انزياح

فهي استعارة تقوم على وصف الشيء بما لا يعد من صفاته. حيث انتقل المعنى ممن له بنت بيضاء، إلى كنية لكل من هو أسود البشرة. مما خلق تعارضاً بين التعبير والشخص الذي أشير إليه به في الواقع. وبذلك يعتبر الرجل الأسود عنصراً غير متوقع في البنية التأليفية للتعبير المسكوك، مما شكل انزياحاً داخلياً حدده الانزياح الخارجي الذي أشار إلى أن " أبو البيضاء " هو الرجل الأسود، فشكل عنصراً غير متوقع في الواقع. فالانزياح الداخلي يكون في التعابير المسكوكة التي لا يكون فيها تعارض بين معناها والعالم الخارجي. والانزياح الخارجي يكون في التعابير المسكوكة التي في معناها تعارض مع العالم الخارجي بعدم موافقة العبارة للإشارة. وبذلك تكون القاعدة كالتالي:

- الانزياح الداخلي لا يحتاج الانزياح الخارجي لتحديد معناه.
- الانزياح الخارجي يفترض الانزياح الداخلي؛ لإخراج بنية التعبير المسكوك من تعارض معناه مع الواقع وتحديد سياق قوله.

إذن، الانزياح بنوعيه الداخلي والخارجي، إجراء سيساعد الحاسوب على تحديد السياق لمعرفة معنى التعابير المسكوكة، ومعرفة مطابقة أو تعارض ملفوظها مع الواقع.

من خلال ما سبق ذكره، سيأخذ التمثيل اللغوي للتعابير المسكوكة في قاعدة البيانات الترتيب التالي:



الشكل (1): تمثيل التعابير المسكوكة في قاعدة البيانات

هذا الترتيب هرمي لا يمكن تقديم أو تأخير مستوى عن آخر، فهو محفوظ الرتبة سيتم تطبيقه على التعابير المسكوكة؛ لأن من شروط إعداد قاعدة البيانات التسلسل والانتظام في وضع المعطيات هذا من جهة، ومن جهة أخرى سنقوم بتفريغ خصائص التعابير المسكوكة في الإطارات بنفس الترتيب الهرمي فهي طريقة

ميكانيكية⁽¹⁾ ستمكن الحاسوب من استيعاب وفهم وإنتاج هذه التعابير.
1. التوصيف اللساني للتعابير المسكوكة: نموذج التعبير المسكوك "أبو البيضاء"

حرف الألف

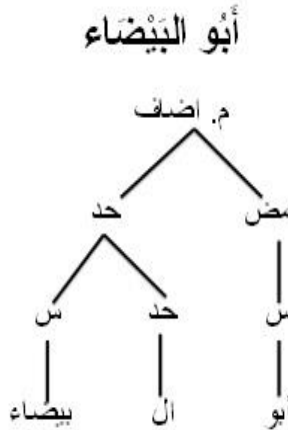
أَبُو الْبَيْضَاءِ⁽²⁾

التمثيل الصوتي:

أبو البيضاء: |ʔaβu: ʔa|βajda:ʔ|

التمثيل التركيبي:

مركب إضافي: أبو، مضاف ؛ البيضاء، ال: أداة تحديد، بيضاء: مضاف إليه.



(1) Moreau, Introduction à la théorie des langages, p. 27.

(2) أبو أسعد، أحمد، التراكيب والعبارات الاصطلاحية العربية القديم منها والمولد، دار العلم للملايين، لبنان، 1987، ص24.

التمثيل الدلالي:

- ❖ المعنى الحقيقي: ▪ أبُو: الأب، أصله " أَبُو "، بالتحريك، لأن جمعه " آبَاء " ⁽¹⁾.
- البَيْضَاء: البياض ضد السواد، يكون ذلك في الحيوان، والنبات، وغير ذلك مما يقبله غيره. والبياض: لون أبيض. " البَيْضَاء ": تطلق على السمراء أيضًا ⁽²⁾.
- ❖ المعنى المجازي: كنية الحبشي الأسود أو الزنجي.

التمثيل التداولي:

- الانزياح الداخلي: انزياح دلالي من معنى من له " بنت بيضاء " إلى معنى " الرجل الأسود ".
- الانزياح الخارجي: انزياح إحالي على رجل أسود.
- التجربة الوجودية: السخرية.

3. نظرية الإطارات: مفهوما ومكوناتها

أ- مفهوما:

يقول مارفن منسكي ⁽³⁾ (Marvin minsky) في نظرية الإطار التي وضعها: "إليك جوهر نظرية الإطار. حين يواجه المرء موقفًا جديدًا (أو يدخل تغييرًا محسوسًا على رأيه في إحدى المسائل)، يختار من ذاكرته بنية تدعى إطارًا، وهذا

(1) ابن منظور، جمال الدين بن محمد، لسان العرب، دار إحياء التراث العربي، بيروت، ط 3، 1419هـ/ 1999م، 5756/1، (أب).

(2) السابق 1/ 550-551، (بيض).

(3) مارفن منسكي عالم أمريكي مختص بالعلوم الإدراكية والمعرفية في مجال الذكاء الاصطناعي، وتعتبر نظريته نموذج من نماذج تمثيل المعرفة الآلية للغات الطبيعية.

الإطار شبكة من الذاكرة تستدعي لتعديلها لكي تتلاءم والواقع بتغيير التفصيلات حسب الواقع".⁽¹⁾

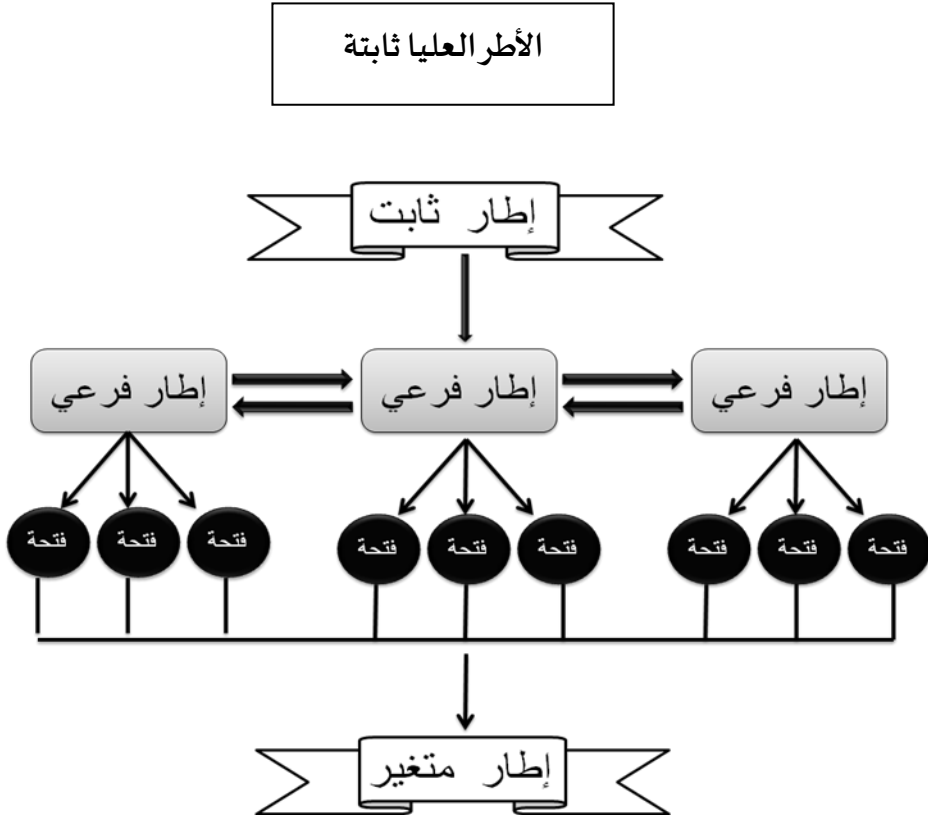
ب- مكوناتها

يتصور منسكي الإطار شبكة من العقد والعلاقات⁽²⁾، يتكون من مستويين اثنين وهما:

- مستويات الإطار العليا: تتسم بالثبات؛ لكونها تمثل الأشياء الصادقة دائماً عن الموقف المفترض.
 - مستويات الإطار الدنيا: تتسم بالتغير، وتتكون من فتحات يتم ملؤها بمعلومات، وأمثلة معينة افتراضية، بحيث إن هذه الأطر الفرعية يمكن حذفها أو زيادة المعلومات فيها حسب مطابقة افتراضاتها للواقع.
 - أنظمة -أطر: أطر فرعية ذات علاقات فيما بينها، تخضع لعملية التحويل حسب المعطيات والافتراضات الموجودة في كل واحدة منها.
- ونوضح في الشكل التالي تصميم الإطارات، بحيث كل سهم من الأسهم يمثل التحويلات الممكنة:

(1) Johnson-Laird et al, Thinking: Reading in cognitive Science, Cambridge University Press, 2010, P. 355.

(2) Johnson-Laird, Thinking: Reading in cognitive Science, PP. 355-356.



الشكل (2): تصميم الإطارات

وتملأ الإطارات بالطريقة التالية⁽¹⁾:

(Frame-name	اسم-الإطار
<slot-name1 Filler1>	الفتحة- اسم 1 المائلات 1
<slot-name2 Filler2>	الفتحة- اسم 2 المائلات 2

(1) Brachman, Ronald J; Hector J, Levesque, Knowledge Representation and Reasoning, Elsevier, 2003. p. 137.

4. تفرغ التعابير المسكوكة في الإطارات وخطوات معالجتها آلياً: نموذج التعبير المسكوك "أبو البيضاء"

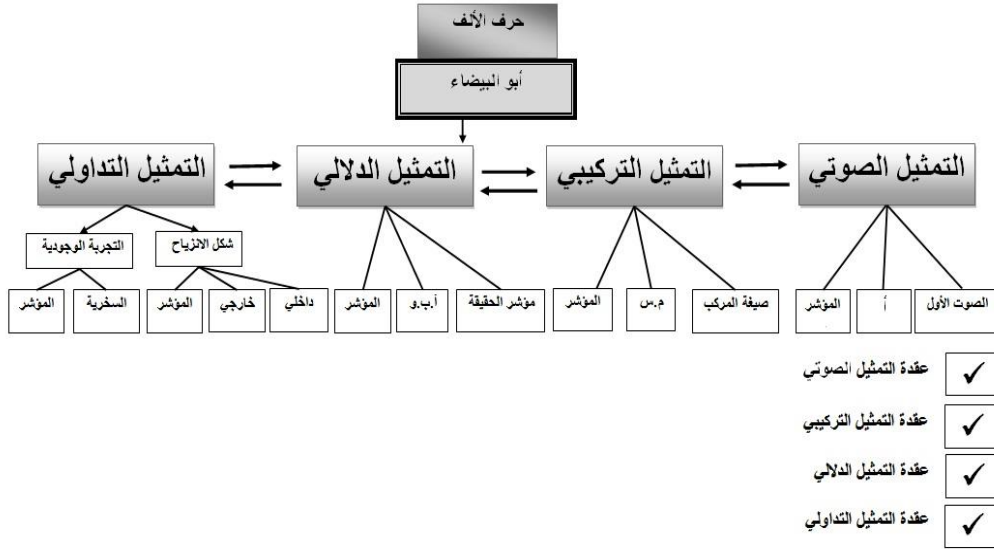
نشير بهذا الصدد، أن شرح كيفية تفرغ التعابير المسكوكة في الإطارات ومراحل معالجتها آلياً، يوازيه في الواقع تصميم قاعدة بيانات آلية لهذه التعابير، وكيفية تمثيلها حاسوبياً، ورابطها على شبكة الإنترنت هو التالي:

<https://www.ichraf-data.com/almaskoka/index1.php>

أ- التمثيل الآلي بالإطارات للتعبير المسكوك "أبو البيضاء"

سنطبق نظرية الإطارات على التعبير المسكوك "أبو البيضاء"⁽¹⁾ مع مراحل تمثيله الآلي وطريقة عد الحاسوب وبثه في خصائصه اللسانية التي حددناها في المستويات التالية: المستوى الصوتي، والمستوى التركيبي، والمستوى الدلالي، والمستوى التداولي، عن طريق المسح الكلي أو الجزئي في معطياتها اللغوية، مع إمكانية التعديل والتبسيط والحذف أو الزيادة في أي معطى لغوي يخص المستويات اللسانية السالفة الذكر. لنبين باللموس قابلية التعابير المسكوكة للتمثيل؛ فكل تعبير من التعابير المسكوكة يتميز بخاصية واحدة على الأقل تميزه عن باقي التعابير المسكوكة الأخرى، رغم أن كل التعابير تفرغ في نفس البنية الهرمية، فهي تضمن تمثيلاً هرمياً موحداً ومنسجماً لجميع التعابير المسكوكة من أجل تسهيل تخزينها وتمثيلها وقابليتها لامتصاص تعابير مسكوكة جديدة، ولكنها في الوقت نفسه تضمن أن يتميز كل تعبير مسكوك عن غيره من التعابير المسكوكة بحد أدنى (والحد الأدنى هو واحد) من السمات المميزة، للتفسير لنلاحظ الشكل التالي:

(1) هذا التعبير المسكوك هو بمثابة مثال تُطبق مراحل تمثيله ومعالجته آلياً على التعابير المسكوكة التي ستوضع في قاعدة البيانات بنفس الطريقة.



الشكل (3): تمثيل "أبو البيضاء" بالإطارات

- ب - أعطى تمثيل التعبير المسكوك "أبو البيضاء" بالإطارات، الأطر الآتية:
- إطار "عال" وضعنا فيه الترتيب الألفبائي للتعبير المسكوك وهو حرف "الألف".
 - إطار "ثابت" وضعنا فيه التعبير "أبو البيضاء"، وقد تفرع عن هذا الإطار الثابت أربعة أطر فرعية وهي كالاتي:
 - الإطار الفرعي الأول: يحمل معلومات عن التمثيل الصوتي. وتتفرع عنه ثلاث فئات، تضم الفتحة الأولى "الصوت الأول"، وتضم الفتحة الثانية "الحرف الأول"، وتضم الفتحة الثالثة "المؤشر"، وهو رمز أو رقم عددي، مثلاً الرقم (10.1)⁽¹⁾ يميز التعبير المسكوك "أبو البيضاء" عن باقي التعابير المسكوكة الأخرى التي تبتدئ بحرف الألف.

(1) هذه الأرقام افتراضية للتمثيل فقط.

■ الإطار الفرعي الثاني: يحتوي على معلومات عن التمثيل التركيبي. وتتفرع عنه ثلاث فتحات، الفتحة الأولى "لصيغة المركب"، والفتحة الثانية "بنية المركب": وهو مركب اسمي في التعبير المسكوك "أبو البيضاء"، والفتحة الثالثة "المؤشر" وذلك بتمييزه برمز أو رقم عددي، مثلاً الرقم (10.2) عن باقي المركبات الاسمية الموجودة في قاعدة البيانات.

■ الإطار الفرعي الثالث: يتضمن معطيات عن التمثيل الدلالي. وتتفرع عنه ثلاث فتحات، تحتوي الفتحة الأولى على "مؤشر الحقيقة"، والفتحة الثانية على الكلمة التي يبتدئ بها التعبير المسكوك "أبو البيضاء": وهي "أ.ب.و" تكون مفتاحاً للتعبير المسكوك بأكمله، لتعرض بعد ذلك، جميع المعلومات المتعلقة بتمثيله الدلالي على مستويي "المعنى الحقيقي" بتفكيك ألفاظه ومعرفة معانيها الحقيقية و"المعنى المجازي"، بالنظر إلى التعبير المسكوك مترابطاً بين جميع ألفاظه، ويوضع في الفتحة الثالثة "المؤشر"، وهو عبارة عن رمز أو رقم عددي، مثلاً الرقم (10.3) يميز معطياته الدلالية عن ما هو موجود في قاعدة البيانات.

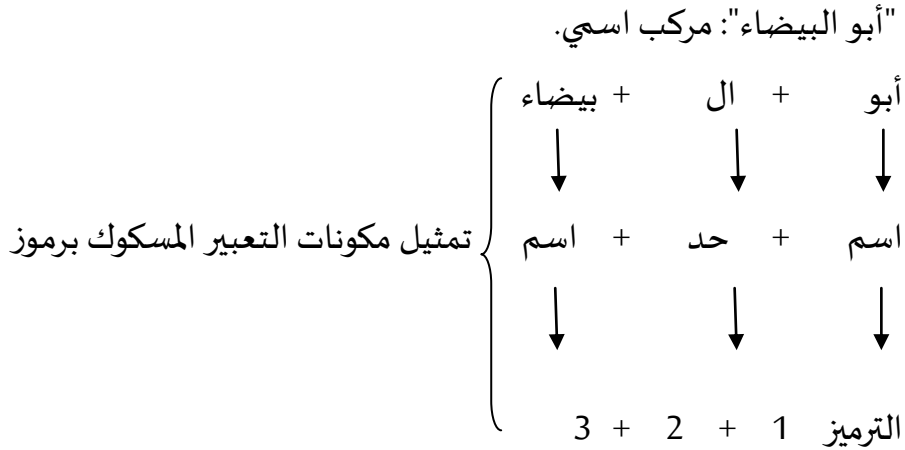
■ الإطار الفرعي الرابع: يتضمن معطيات عن التمثيل التداولي. وتتفرع عنه فتحتان: الفتحة الأولى: خاصة "بشكل الانزياح" وتتفرع عنه ثلاث فتحات صغرى: تضم الأولى "الانزياح الداخلي"، والثانية "الانزياح الخارجي"، وهو خارجي في التعبير المسكوك "أبو البيضاء"، والفتحة الثالثة: "مؤشرهما" عبارة عن رقم أو رمز عددي، مثلاً الرقم (10.4) لتمييزهما عن باقي الانزياحات الواردة في قاعدة البيانات. أما الفتحة الثانية: خاصة "بالتجربة الوجودية" التي أظهرت التعبير المسكوك وأدت إلى تداوله بين الناس. وتضم فتحتين صغيرتين: يوضع في الفتحة الأولى "نوع التجربة الوجودية" ففي حالة التعبير المسكوك "أبو البيضاء" فهو في تجربته الوجودية يحيل على "السخرية" من الرجل الأسود،

ووصف بشرته بالبياض. وتضم الفتحة الثانية: "المؤشر" عبارة عن رقم أو رمز عددي، مثلاً الرقم (10.5) لتمييز تجربته الوجودية عن باقي التجارب الوجودية للتعبير المسكوكة الموجودة في قاعدة البيانات.

بذلك تتوزع المعلومات التي وُصِّفت للتعبير المسكوك "أبو البيضاء" إلى أربعة عقد، كل عقدة تحوي معطيات عن أحد تمثيلاته الأربعة (الصوتية، التركيبية، الدلالية، التداولية).

6- مراحل التمثيل الآلي للتعبير المسكوك "أبو البيضاء"

أ- تمثيل مكونات التعبير المسكوك "أبو البيضاء" برموز (أرقام)
التعبير المسكوك "أبو البيضاء" عبارة عن حزمة⁽¹⁾، يمكن تمثيله آلياً بالشكل التالي:



الشكل (4): تمثيل التعبير المسكوك برموز

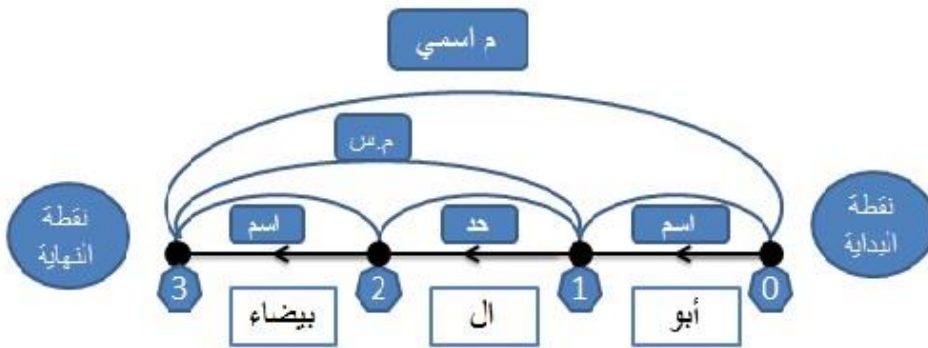
(1) هي التي تمثل جزءاً من المعرفة وتحتوي على المعلومات الخاصة بكل تعبير مسكوك.

ب- تقييد الرموز

التعبير المسكوك "أبو البيضاء" يكون ترتيبه بالشكل التالي:

يسبق الاسم، ثم يليه الحد، ثم يليه الاسم، ويتحكم في هذا التسلسل خاصية N-Gram⁽¹⁾ التي تجعل التعبير المسكوك "أبو البيضاء" يختلف عن تعابير مسكوكة أخرى؛ لفك الغموض الذي يمكن أن يحدث عند معالجة عدد ضخم من هذه التعابير في قاعدة البيانات الآلية، وبذلك يأخذ ترميز التعبير المسكوك بالطريقة الآتية:

أبو البيضاء



الشكل (5): الحزمة الفرعية سليمة الصياغة للتعبير المسكوك "أبو البيضاء"

(1) المبرمج العربي، شرح مفصل لنموذج N-Gram في معالجة اللغات الطبيعية، متاح على الرابط (شاهد يوم 2022/08/15): <https://arabicprogrammer.com/article/3237261129>

- يمثل هذا الشكل ما يسمى بالحزمة الفرعية سليمة الصياغة،⁽¹⁾ وهو رسم موجه غير دائري تم فيه تمثيل التعابير المسكوكة، يتكون من:
- نقطة بداية: تمثل قيمة التعبير المسكوك وقيمتها صفر.
 - نقطة نهاية: تمثل نهاية التعبير المسكوك المدخل وقيمتها 3.
 - نقاط وسطى: تمثل المسافة بين كلمات التعبير المسكوك مرقمة من اليمين إلى اليسار.
 - أسهم: تصل كل نقطة بالنقطة التي تليها.

ت- الإجراءات

- تبتدئ المعالجة من اليمين إلى اليسار. إذن: لا يمكن أن تبتدئ المعالجة من اليسار إلى اليمين.
- لا يمكن تصريف أو حذف أو زيادة كلمة من كلمات التعبير المسكوك " أبو البيضاء ". كما لا يمكن تشكيل حركات حروفه، بل تبقى ثابتة.
- لا يمكن القلب المكاني بين كلمات التعبير المسكوك.

ث- معنى الرموز

مضمون الرمز: 1	←	مقولة اسمية
مضمون الرمز: 2	←	مقولة حدية
مضمون الرمز: 3	←	مقولة اسمية

(1) الناجم، صلاح الدين، علم اللغة الحاسوبي، متاح على الرابط (شاهد يوم 2014/04/16):

7- العملية العددية والبثية للحاسوب في معالجة التعبير المسكوك "أبو البيضاء"

أ- العملية العددية

يقدر الحاسوب على عد⁽¹⁾ الكلمات المكونة للتعبير المسكوك "أبو البيضاء"؛ لأن له نقطة بداية ونقطة نهاية.

ب- العملية البثية

يمكن للحاسوب أن يَبْثَ⁽²⁾ في التعبير المسكوك "أبو البيضاء" من خلال قاعدة البيانات بالمسح في جميع مستوياته التمثيلية، أو في أحد تمثيلاته حسب اختيار المستعمل. وكمثال على ذلك: التعبير المسكوك "أبو البيضاء" "أورد جميع مستوياته التمثيلية".

بذلك يقوم الحاسوب بالمسح الكلي في جميع تمثيلاته، وإعطاء جميع المعلومات التي وُصِفَتْ له في قاعدة المعطيات، أو تطلب من الحاسوب أن يعطيك معلومات حول أحد تمثيلاته، مثال ذلك: التعبير المسكوك "أبو البيضاء" "ما خصائص تمثيله الدلالي" أو "أعط معلومات عن تمثيله الدلالي". وبذلك يبث الحاسوب في معناه الحقيقي ومعناه المجازي. هذه العملية تسري على تمثيلات التعبير المسكوك الأخرى وذلك باختيار أي واحد منها.

تبين عند تفريغ مستويات التمثيل اللساني للتعابير المسكوكة في الإطارات قابليتها لمعالجة هذه الظاهرة اللغوية بالتحليل، والتصنيف، والترتيب، والتنظيم، وانفراد كل تعبير مسكوك بما يميزه عن التعابير المسكوكة داخل قاعدة البيانات، مما يسهل على الحاسوب البث في كل ما يطلب منه دون أي خلل ميكانيكي عند المعالجة.

(1) Moreau, Introduction à la théorie des langages, pp. 30-31.

(2) Moreau, Introduction à la théorie des langages, pp. 30-31.

النتائج والتوصيات:

النتائج:

النتائج التي استخلصناها في هذا البحث من خلال الربط بين المستويين النظري والتطبيقي، مايلي:

1- أن ظاهرة التعابير المسكوكة في اللغة العربية تمثل جزءاً مهماً من معجمها اللغوي، فهي كثيرة الاستعمال بين المتكلمين، ولا غنى عنها في تعلم اللغة الأم وفي تعلم أي لغة ثانية، وفي فهم مضمون النصوص المترجمة من لغة إلى أخرى.

2- أنه من الممكن تحليل التعابير المسكوك بدلالة خصائصها اللسانية على مستويات: الصوت، والتركيب، والدلالة، والتداول. تميز فيه كل تعبير مسكوك بصوته الأول، ونوع مركبه، ودلالته الحقيقة، والمجازية، ونوع انزياحه، وتجربته الوجودية، إذ إنها تعابير مُغلقة لا يستطيع الحاسوب أن يعطي أي معلومة إذا لم تتوفر له المعلومات الكافية عنها.

3- قابلية التعابير المسكوكة للتمثيل الحوسبي: كل تعبير من التعابير المسكوكة يتميز بخاصية واحدة على الأقل تميزه عن باقي التعابير المسكوكة الأخرى، رغم أن كل التعابير تفرغ في نفس البنية الهرمية.

4- أن التمثيل الآلي للتعابير المسكوكة ممكن باستعمال نظرية الإطارات، التي ستحلل المستويات اللسانية لهذه التعابير وتمثيلها في نموذج صوري يعالجه الحاسوب فهماً وإنتاجاً، لما يتيح في أن كل تعبير مسكوك يُفَرِّغ فيه بشكل هرمي وفي نفس عدد الإطارات، ومضمون كل تعبير مسكوك يتميز عن مضمون الآخر محفوظ الرتبة ورقم يميزه عن الأطر الأخرى، مع إمكانية الحذف والزيادة في المعطيات اللغوية المُدخلة. مما سيمكن الحاسوب من بناء

قاعدة بيانات منظمة بشكل جيد ومفتوحة لضم عدد كبير من التعابير المسكوكة قديمها وحديثها.

5- تحليل الحاسوب للتعابير المسكوكة من خلال الإطارات التي وُزعت فيها الخصائص اللسانية لكل تعبير مسكوك، يتميز بالفاعلية، والسرعة، والمرونة، في معالجة التعبير المطلوب. هذا فعلاً ما تم تحقيقه وإنجازه في البرنامج الحاسوبي الذي يضم نموذجاً من قاعدة بيانات للتعابير المسكوكة، الذي تبين في مرحلة تطبيقه تجاوب الحاسوب فيما يطلب منه فهماً وإنتاجاً.

6- يحتاج هذا العمل الذي نريد به خدمة اللغة العربية والإسهام في إثراء محتواها الرقمي، إلى فريق عمل من اللغويين ينكبّ على جمع التعابير المسكوكة من الكتب العربية قديمها وحديثها، بكل أصنافها، ومحاولة تصنيفها، بالاعتماد على الحقب التي ظهرت فيها هذه الكتب، وتبويب هذه التعابير حسب تجاربها الوجودية؛ لأن العمل الفردي في مثل هذه المشاريع سيتسم بالبطء وضعف النتائج المحصل عليها.

التوصيات:

1. عقد مؤتمرات وندوات وأيام دراسية بشكل دوري؛ للاطلاع على ما تم تقديمه من أبحاث سابقة للتعرف على المُنجز منها، واستقبال أبحاث جديدة تناولت قضايا لغوية تهتم مجال اللسانيات الحاسوبية العربية.
2. عقد شراكات بين الجامعات العربية والجامعات الغربية التي تضم مختبرات ومراكز بحثية تعنى بدراسة اللسانيات الحاسوبية العربية، وذلك لتبادل الأفكار والعمل على مشاريع علمية ضخمة تهتم مُتون اللغة العربية، بحيث يكلف كل فريق بحثي بإنجاز الجزء الذي وُكِّل إليه وفي الآجال المحددة.
3. إتاحة البرامج الحاسوبية والمواقع الإلكترونية، التي عالجت اللغة العربية

لجمهور الأساتذة والباحثين والمهتمين؛ لأن الكثير منها يطلب مقابلًا ماديًا للولوج إلى الموقع، أو شراء قرص مدمج لتثبيته على الحاسوب، وفي حالة قرصنته يتابع قانونيًا.

4. الاهتمام بهذا المجال الخصب والواعد؛ إذ يعتبر فرصة حقيقية لسد الفجوة الرقمية للغة العربية الذي تعرفه بين اللغات المنتجة للمعرفة.

5. توفير الدعم العلمي والمادي والبشري للأبحاث التي فيها جدّة وتكون مردوديتها العلمية جيدة في خدمة اللغة العربية.

ملحق

فهرس الأشكال		
الصفحة	عنوان الشكل	رقم الشكل
14	تمثيل التعابير المسكوكة في قاعدة البيانات	الشكل (1)
17	تصميم الإطارات	الشكل (2)
19	تمثيل "أبو البيضاء" بالإطارات	الشكل (3)
21	تمثيل التعبير المسكوك "أبو البيضاء" برموز	الشكل (4)
22	الحزمة الفرعية سليمة الصياغة للتعبير المسكوك "أبو البيضاء"	الشكل (5)

قائمة المراجع

قائمة المراجع العربية

- 1- أبو أسعد، أحمد، التراكيب والعبارات الاصطلاحية العربية القديم منها والمولد، دار العلم للملايين، لبنان، 1987.
- 2- الأنباري، أبو بكر محمد بن القاسم (328هـ)، الزاهر في معاني كلمات الناس، تحقيق: حاتم صالح الضامن، دار الشؤون الثقافية العامة-آفاق عربية، بغداد، ط 2، 1987، 2 مج.
- 3- الثعالبي، أبو منصور عبد الملك بن محمد بن إسماعيل (350-429هـ)، ثمار القلوب في المضاف والمنسوب، تحقيق: محمد أبو الفضل، دار المعارف، القاهرة، 1965.
- 4- جرمان، كلود؛ ولوبلون ريمون، علم الدلالة، ترجمة: نور الهدى لوشن، جامعة قار يونس، بنغازي، 1997.
- 5- جيرو، بيير، الأسلوبية، ترجمة، منذر عياش، دار الحاسوب للطباعة، حلب، ط 2، 1994.
- 6- حنا، سامي عياد؛ وكريم زكي حسام الدين، نجيب جريس، معجم اللسانيات الحديثة: انجليزي-عربي، مكتبة لبنان، 1997.
- 7- داود، محمد محمد، معجم التعبير الاصطلاحي في العربية المعاصرة، دار غريب، القاهرة، د. ط، 2003.
- 8- رشوان، محسن؛ والمعتز بالله السعيد، أحمد راغب، محمد عطية، سامح الأنصاري، مقدمة في حوسبة اللغة العربية، دار وجوه، المملكة العربية السعودية، 2019.

- 9- زكي كريم، حسام الدين، التعبير الاصطلاحي: دراسة في تأصيل المصطلح ومفهومه ومجالاته الدلالية وأنماطه التركيبية، المكتبة الأنجلو المصرية، القاهرة، 1985.
- 10- العبدري، أبو الحسن بن محمد الشيبني (837هـ/1433م)، تمثال الأمثال، تحقيق: أسعد ذبيان، دار المسيرة، بيروت، 1982، 2 مج.
- 11- فضل، صلاح، علم الأسلوب: مبادئه وإجراءاته، دار الآفاق الجديدة، بيروت، 1985.
- 12- المسدي، عبد السلام، الأسلوبية والأسلوب، الدار العربية للكتاب، القاهرة، ط 2، 1982.
- 13- ابن منظور، جمال الدين بن محمد، لسان العرب، دار إحياء التراث العربي، بيروت، ط 3، 1419هـ/1999م، 19 مج.

قائمة المراجع الأجنبية

- 1- Bordron, J.-L. (2008) *Best of idioms*. Paris: Ellipses.
- 2- Brachman, R.J. and Levesque, H.J. (2009) *Knowledge representation and reasoning*. Amsterdam: Elsevier.
- 3- Johnson-Laird, P. N. and Wason, P. C. (Eds) (1977). *Thinking: Readings in cognitive science*. New York: Cambridge U Press.
- 4- Langlotz, A. (2006) *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: Benjamins.
- 5- Moreau, René (1975) *Introduction à La théorie des langages*. Paris: Hachette.
- 6- Spears, R.A. (2000) *NTC's American idioms dictionary the most*

practical reference for the everyday expressions of contemporary American English. Chicago, IL: NTC Pub. Group.

7- Wright, J. (2002) *Idioms organizer by metaphor, topic and key word*. Croatia: Zrinski d.d.

8- قائمة المواقع الإلكترونية

9- الناجم، صلاح الدين، علم اللغة الحاسوبي، متاح على الرابط (شاهد يوم 2014/04/16): www.alnajem.com.

10- الألفبائية الصوتية الدولية، إنجليزية: IPA، فرنسية: API، ابتكار الجمعية الصوتية الدولية، متاح على الرابط (شاهد 2022/09/2): <http://ipa.typeit.org/full/>

11- المبرمج العربي، شرح مفصل لنموذج N-Gram في معالجة اللغات الطبيعية، متاح على الرابط (شاهد يوم 2022/08/15): <https://arabicprogrammer.com/article/3237261129/>



"الأصوات العربية"
وتعليمها للناطقين بغيرها
تطبيق Arabic Phonetics نموذجاً
على الهواتف المحمولة والحواسيب اللوحية
IPhone/iPad
(عرض وتحليل)

**The Arabic Phonetics and teaching them to
non-native speakers - the Arabic Phonetics
application as a model on IPhone/iPad Phones
(View and Analyze)**

إعداد
باحث دكتوراه / محمد الغليظ
كلية الألسن - قسم اللغة العربية
جامعة عين شمس

ملخص البحث

غالبًا ما يواجه الطلاب الناطقين بغير اللغة العربية الفصحى المعاصرة تحديات في اكتساب النظام الصوتي للغة العربية؛ مما قد يشكل غالبًا تحديًا للمعلمين أثناء عمليات انتقال هؤلاء الطلبة من اللغة الأم إلى اللغة المستهدفة أثناء التعلم. على المستوى الصوتي، هناك عدد محدود من أبحاث تحليل التباين بين اللغتين العربية والإنجليزية، وندرة مثل هذه الأبحاث بين العربية والصينية، والفرنسية، والألمانية، والهوسا، والإندونيسية، والروسية، والإسبانية، والتركية، والأردية. يحتاج المتحدثون بهذه اللغات إلى تطبيق يساعد في تعلم الصوتيات العربية وعلى وجه الخصوص الأصوات التي ليس لها مكافئ في لغاتهم.

من ناحية أخرى، نتيجة لوباء Covid-19، هناك رغبة في تطوير أدوات التعلم التكنولوجي؛ لاستخدامها في أي مكان بدلاً من الفصول الدراسية العادية. نتيجة للعمل المكثف لإيجاد حل لهذه الحاجة، وبلاستعانة بأدوات المنهج الوصفي، قمنا بتطوير تطبيق جوال يملأ هذه الثغرة. في هذه الورقة، سيتم تقديم الحلول لهذه التحديات من خلال تطبيقنا للأصوات العربي ACOCO. هذا التطبيق تم تصميمه لتحسين نطق اللغة العربية لغير الناطقين بها. يركز ACOCO على "العربية الدولية"، بما في ذلك العربية من الشرق الأوسط وشمال أفريقيا. التطبيق متوفر حاليًا على متجر آبل: <https://apps.apple.com/app/id1482546188>.

ختامًا، توصي الورقة بالعمل على تقييم التطبيق وتقويمه؛ للوقوف على جوانب القصور، ومحاولة معالجتها بطريقة منهجية تتفق وأساليب صناعة التطبيقات والأرصدة اللغوية، وطبيعة اللغة العربية.

الكلمات الافتتاحية: تعليم اللغة العربية بوصفها لغة ثانية، اللسانيات الحاسوبية، التطبيقات اللغوية، علم الأصوات العربية.

Abstract

Non-native students of Modern Standard Arabic often face challenges in acquiring the phonology system of Arabic. Processes of transfer from the mother tongue to the target language often take place in learning, which could constitute a challenge for educators. At the phonological level, there is a limited number of contrastive analysis research studies between Arabic and English and a scarcity of such research between Arabic and Chinese, French, German, Hausa, Indonesian, Russian, Spanish, Turkish, and Urdu. Speakers of these languages need a program that helps them learn Arabic phonemes and, in particular, the sounds they do not have in their languages.

On the other hand, as a result of the COVID-19 pandemic, there is an urge to develop technological learning tools that can be used anywhere rather than in a regular classroom setting. As a result of extensive work on finding a solution for this need and by using a descriptive-analytical method, we have developed a mobile app that fills this lacuna.

In this paper, solutions to these challenges will be presented, and our Arabic phonetic app will be introduced. The ACOCO App is designed to improve the Arabic pronunciation of non-native speakers of Arabic. ACOCO focuses on 'international Arabic', including Arabic from the Middle East and North Africa. It is available on the Apple Store: <https://apps.apple.com/app/id1482546188>.

Recommendations of the study: further work on the evaluation and

refinement of ACOCO to identify any shortcomings and address them in a systematic manner consistent with the methods of app processing and corpus linguistics. This would include further observing the modern structure and use of the Arabic language.

Keywords: Teaching Arabic as a second language, computational linguistics, linguistic applications, Arabic phonology.

المقدمة

إن الاهتمام العالمي بمجال "اللسانيات الحاسوبية" في حاجة ماسة لأن يواجه باهتمام من قبل باحثينا، لاسيما مع التقدم في تقنيات الآلة (فمن منا الآن لا يحمل هاتفًا، أو جهاز حاسوب محمول؟)؛ لخدمة لغتنا العربية الشريفة، لغة القرآن الكريم؛ لاسيما مع توجه أنشطة بعض أقسام اللغات في المؤسسات الأكاديمية نحو استثمار المخزون اللغوي العربي، والمستمد من النصوص المتنوعة مستخدمة الأنظمة الحاسوبية في معالجة تلك النصوص؛ بهدف استخراج رصيد إلكتروني من أبنيتها وتراكيبها وأساليبها، ومن ثمّ توظيفها في إنتاج مشاريع علمية وتطبيقات لغوية مما يحتم على الباحث العربي ضرورة تناول تلك المشاريع بالتحليل والدراسة؛ للوقوف على أهمّ الإيجابيات والسلبيات ومحاولة الاستفادة منها، أو الوقوف على خطورتها ومواضع قصورها والاحتياط منها.

إن العمل الحاسوبي على اللغة هو جزء من السعي وراء السيطرة عليها، وذلك في ظل ظهور سياسات عالمية جديدة منها مبادرة الرئيس الأمريكي جورج بوش الابن، الذي دعا إليها في أحد خطاباته بعد أحداث سبتمبر 2001م،⁽¹⁾ التي كانت تمثل علامة فارقة ومهمة في تدريس اللغة العربية في الولايات المتحدة؛ حيث ربط تعلمها هي واللغة الصينية بالأمن القومي الأمريكي، وبمستقبل المصالح الأمريكية في الشرق الأوسط. مما كان له كبير الأثر في ازدياد نسبة دارسي اللغة العربية فبلغت 86%، على حين وصلت نسبة دارسي اللغة الصينية 31%،

(1) الخطيب، محمد عبد الفتاح، مؤتمر "اللغة العربية وتحديات العصر" بالتعاون بين كلية دار العلوم جامعة القاهرة ومركز الدراسات المعرفية بالقاهرة، كلية دار العلوم، القاهرة، 2008م.

والإسبانية 28%⁽¹⁾.

ولن يتأتى النهوض باللغة العربية إلا عن طريق الخوض في الميادين الثلاثة؛ البحث اللساني الحاسوبي، والصناعة، والتدريس عمومًا، ومجال صناعة التطبيقات/ البرمجيات اللغوية على وجه الخصوص. إن مجال "صناعة اللغة العربية الرقمية" من شأنه مساعدتنا في تقديم منتجنا الثقافي للمهتمين باللغة العربية وثقافتها وآدابها، من العرب، وغيرهم في شتى أنحاء العالم.

مشكلة البحث:

تتحدد مشكلة هذا البحث في ندرة أبحاث مجال "اللسانيات الحاسوبية العربية"، وقصور الاعتماد على المدونات اللغوية في صناعة تطبيقات تعليمية لخدمة اللغة العربية، ومهاراتها.

أهداف البحث:

استعراض أحد الحلول المقترحة والمستخدم بالفعل للتدريس عن بُعد، وهو تطبيق إلكتروني "الأصوات العربية" للهواتف المحمولة والحواسيب اللوحية iPhone/iPad، المستمد من مدونة الصوامت العربية Arabic consonant corpus (ACOCO)، والذي يساهم في حل مشكلات النظام الصوتي التي تواجه الدارس الناطق بغير العربية.

(1) نابي عبدالله، رحاب، تعليم الكتابة للناطقين بغير العربية بين النظرية والتطبيق، الملتقى الأول/ الأبحاث. azhar- al. com /go// نقلا عن كحولي، (زهرة سعدلاوي حرم). "البعد الاستثماري في تدريس اللغة العربية لغير الناطقين بها، التجربة التونسية أنموذجاً"، ص1.

منهج البحث:

تطلبت طبيعة هذا البحث الاستعانة بأدوات المنهج الوصفي (التحليلي).

أهمية البحث:

في حدود ما انتهى إليه علمُ الباحث، فإنه لم يقف إلا على بعض دراسات أكاديمية باللغة العربية في هذا الموضوع نظراً لحدثة هذا المجال الذي يخدم اللغة العربية في آفاق جديدة؛ لهذا السبب يقف الباحث على تجربة عملية، قام فيها ببناء تطبيق لغوي، يراعي التركيز على الوحدات الصوتية العربية التي لا يوجد لها نظير في اللغات المختلفة، صاحبة الأعداد الكبيرة من البشر محل الدراسة: "الصينية، والإسبانية، والإنجليزية، والروسية، والألمانية، والإندونيسية، والفرنسية، والأردية، والتركية، والهوسا"، التي تمثل صعوبات لدارسي العربية من أصحاب تلك الألسن؛ ومن ثم تذليل تلك المشاكل من شأنه نشر اللغة العربية لدى الأجانب.

أكثر الألسنة غير العربية من حيث عدد الدارسين لها:

البيان التالي يقف على أكثر الألسنة غير العربية من حيث عدد الناطقين بها، والبلاد التي تحدث تلك اللغات غالباً، وطبيعة تلك اللغات، من حيث الفروق الصوتية بينها وبين اللغة العربية.

م	اللغة	عدد الناطقين (بالمليون)	النسبة	البلاد التي تتحدث اللغة العربية
1	اللغة الصينية	897	38.8	الصين، وتايوان، وسنغافورة.
2	اللغة الإسبانية	477	20.6	أمريكا اللاتينية، وإسبانيا، والولايات المتحدة الأمريكية، وغينيا الاستوائية، والصحراء الغربية، وجزر المحيط الهادئ.
3	اللغة الإنجليزية	371	16	أستراليا، وكندا، والهند، وأيرلندا، ونيوزيلندا، والمملكة المتحدة، والولايات المتحدة، وجنوب أفريقيا، وسنغافورة، والفلبين.
4	اللغة الروسية	171	7.4	روسيا والجمهوريات السابقة للاتحاد السوفيتي، ومنغوليا، والأراضي الفلسطينية المحتلة.
5	اللغة الفرنسية	80	3.5	بلجيكا (والونيا وبروكسل)، كندا (خاصة كيبيك، ونيو برونزويك وشرقها أونتاريو)، فرنسا، سويسرا، إفريقيا الناطقة بالفرنسية (الفرانكفونية)، منطقة البحر الكاريبي الفرنسية، وبولينيزيا الفرنسية، والجزر المختلفة في المحيطين الهندي والهادئ.
6	اللغة الإندونيسية	77	3.33	إندونيسيا، وماليزيا، وبروناي، وسنغافورة.

"الأصوات العربية" وتعليمها للناطقين بغيرها

م	اللغة	عدد الناطقين (بالمليون)	النسبة	البلاد التي تتحدث اللغة العربية
7	اللغة الألمانية	76	3.3	النمسا، بلجيكا (أوبين- مالميدي)، وألمانيا، ولوكسمبورغ وليختنشتاين، وسويسرا، وجنوب تيرول (في إيطاليا)
8	اللغة الأردنية	66	2.85	الهند وباكستان.
9	اللغة التركية	63	2.7	تركيا، وقبرص، وبلغاريا.
10	لغة الهوسا	34	1.5	نيجيريا.
	المجموع	2,312		

من خلال الجدول السابق يتبين أن اللغة الصينية هي أكثر اللغات، من حيث عدد الدارسين للعربية بنسبة 38.8% من حجم الألسنة الأكثر تحدثًا على مستوى العالم، وذلك طبقًا لموقع الألفبائية الدولية IPA⁽¹⁾ إحصاء عن عام 2010م، وتقرير للمجلس البريطاني⁽²⁾ إحصاء عن عام 2017م، يأتي بعدها في المرتبة الثانية والثالثة كلٌّ من اللغات الإسبانية والإنجليزية بنسب 20.6% و 16% على التوالي، وتأتي اللغة الروسية في المرتبة الرابعة بنسبة 7.4% تقريبًا، يليها اللغة الفرنسية بنسبة 3.5% من حجم العينة، في حين جاءت النسبة الأصغر من

(1) <http://www.internationalphoneticalphabet.org/>

(2) <https://www.britishcouncil.org/sites/default/files/languages-for-the-future-report-v3.pdf>

نصيب اللغات الإندونيسية، والألمانية، والأردية، والتركية، وأخيرًا لغة الهوسا بنسب تراوحت بين 3.3% و 1.5%؛ وهو مما يدل على ضرورة العمل على الأبحاث اللغوية لنشر اللغة العربية بين الناطقين بغيرها، لا سيما التي تتضمن تطبيقات لغوية لتلك الألسنة بأفضلية للغة الصينية.

كما يتضح مدى انتشار اللغات المختارة، ومتحدثيها في أنحاء بلاد العالم. من الشرق الأقصى نجد اللغة الصينية المنتشرة في بلاد شرق وجنوب شرق آسيا، ذات الأطراف المترامية الشاسعة والقوميات الوفيرة صاحبة اللهجات المختلفة. يليها اللغة الإسبانية، وهي واحدة من اللغات الرومانسية التي تطورت عن اللاتينية، وتستعمل الآن في الأماكن التي كانت من قبل جزءًا من الإمبراطورية الرومانية. وتتضمن تلك العائلة اللغات الفرنسية، والإيطالية، والبرتغالية والرومانية⁽¹⁾. أما اللغة الإنجليزية والألمانية فأنتهيا تنتميان إلى عائلة اللغات الجرمانية⁽²⁾، على خلاف اللغة الروسية التي تنتمي إلى العائلة السلافية الشرقية⁽³⁾. أما اللغة الإندونيسية التي تنتمي إلى عائلة اللغات الأسترونيزية، فهي تمتد عبر الجزر في جنوب شرق آسيا والمحيط الهادئ⁽⁴⁾. اللغة الأردنية لغة هندية آرية من فرع اللغات الهندية الإيرانية، وهي اللغة الرسمية في باكستان، تنتمي إلى عائلة اللغات الهندية الأوروبية، كما أنها تعد واحدة من (22) لغة في الهند⁽⁵⁾. تنتهي اللغة التركية إلى مجموعة الأسرات

(1) <https://www.theguardian.com/education/gallery/2015/jan/23/a-language-family-tree-in-pictures#img-1>

(2) المصدر السابق.

(3) المصدر السابق.

(4) George Quinn, The Learner's Dictionary of Today's Indonesian. Sydney: Allen & Unwin 2001 ISBN 1864485434

(5) <https://www.theguardian.com/education/gallery/2015/jan/23/a-language-family-tree-in-pictures#img-1>

اللغوية أورال- آلتاي Ural- Altay، وهي اللغة الرسمية في الجمهورية "التركية" كما تحتل اللغات التركية مكانة متميزة من قبل الدول الممتدة بين تركيا والصين (أذربيجان، وأوزبكستان، وتركمانستان)⁽¹⁾. من العائلة الأفروآسيوية تنتمي لغة الهوسا التي تعد من أكبر اللغات الإفريقية، وأوسعها انتشارًا بعد اللغة العربية⁽²⁾.

طبيعة اللغات المستهدفة للتطبيق:

التعرف على خصائص اللغات وترجمة ذلك إلى دراسات وأبحاث تطبيقية ذات جدوى؛ لخدمة نشر اللغة العربية وتعليمها من قبل الباحثين المتخصصين- من شأنه أن يقف بنا على الصعوبات المختلفة التي يواجهها الطالب الأعجمي، ومن ثمَّ يكون الأثر البناء في التيسير والتسهيل على الناطقين بغير العربية من الجنسيات المختلفة في تعلم ونشر لغة الضاد.

مشكلات الطلاب الناطقين بغير العربية من أهل اللغات محل الدراسة في نطق الأصوات الصعبة في العربية:

نخلص من الجدول التالي أن الطبيعة الصوتية للعربية (الأصوات المطبقة، والحلقية، والاحتكاكية، والمستعلية، والمفخمة)، تمثل مشكلة كبرى للطلاب الناطقين بغير العربية، أصحاب اللغات محل الدراسة من حيث نطقها، أو الاستماع إليها بالشكل الصحيح وفق قواعد اللغة العربية، وضوابطها. أضف إلى ذلك أهم المعوقات التي تواجه الطالب الناطق بغير العربية عند دراسته للغة الضاد، وهي تأثره بالعوامل الآتية:

(1) Ammon, U. (2006). Sociolinguistics: An international handbook of the science of language and society = Soziolinguistik. Berlin: W. de Gruyter.

(2) حجازي (مصطفى حجازي السيد): العربية والهوسا، نظرات تقابلية، معهد اللغة العربية وحده البحوث والمناهج- جامعة أم القرى، مكة المكرمة، 1985م.

- لغته الأم (بلهجاتها المختلفة)، من حيث التأثيرات السلبية في تعلم نطق الأحرف العربية؛ إذ يحاول- على سبيل المثال- نقل عادات النطق الخاصة بلغته الأم إلى اللغة العربية، كذلك تغيير وضعية اللسان، وما يمثله من مشقة أثناء نطق بعض أصوات اللغة العربية غير الموجودة في لغته الأم، بالإضافة إلى ملاحظة تلك الحروف عند الاستماع إلى الكلام العربي؛ وهو ما يدفع الطلبة إلى الاهتمام بمهارة الكتابة على حساب مهارتي الاستماع والقراءة، وذلك ليس مستحَبًّا⁽¹⁾.

- عدم قدرة الطلبة على التمييز عند النطق أو الاستماع بين صوتين مُتقاربي المخرج، وتختلف درجات الصعوبة من أصوات تُمثِّل صعوبة بالغة، إلى أصوات متوسطة الصعوبة⁽²⁾، من هذه الأصوات- مثلاً- الضاد وتداخلها مع الدال، والسين وتداخلها مع الصاد، والهاء وتداخلها مع الحاء، والحاء نفسها وتداخلها مع الخاء، وكذلك الطاء والتاء والقاف والكاف، والزاي والظاء، والعين، والغين⁽³⁾.

(1) انظر: تشانغ جينغشو، (آسيا). 2011م.

(2) انظر: روزير نبوت، (نيقولاس). 2015م.

(3) انظر: ماسيري، (كوري)، ودفع الله أحمد الأمي، (سميه). 2012م، المشكلات الصوتية في تعلم اللغة العربية للناطقين بغيرها، جامعة المدينة العالمية أنموذجًا، جامعة المدينة العالمية، وكالة البحوث والتطوير، دولة ماليزيا.

"الأصوات العربية" وتعليمها للناطقين بغيرها

[illegible]

الحلول:

إن كل العوامل السابق ذكرها وغيرها تلعب دورًا كبيرًا في تعسير تعليم العربية، إن لم يكن في إفشال تعليمها بوصفها لغة ثانية على وجه خاص، ومن كل ما سبق يبدو واضحًا أهمية اقتراح حلول لمشكلة نطق الأصوات العربية بالشكل الصحيح وفق ضوابط اللغة العربية وقواعدها؛ وذلك عن طريق استحداث وسائل جديدة في مجال تعليم اللغة العربية للناطقين بغيرها، تراعي - التركيز على- الوحدات الصوتية العربية التي لا يوجد لها مكافئ في اللغات المختلفة محل الدراسة: "الصينية، الإسبانية، الإنجليزية، الروسية، الألمانية، الإندونيسية، الفرنسية، الأردنية، التركية والهوسا" من حيث نطقها؛ لإكساب الجودة والفاعلية والمرونة وسرعة التكيف مع المتغيرات الحضارية، فإن من شأن هذه الوسائل أن تسهم في تفعيل تعليم اللغة العربية وتطويرها، والسبيل إلى ذلك من وجهة نظر الباحث يتطلب الآتي:

العمل على تطبيق يراعي أساسيات البرمجة الصحيحة والتصميم الجيد، يستمد مادته من مدونة محدودة - مخطط لها - في حدود (مليون) كلمة تقتصر على جمع الكلمات التي تحتوي على الوحدات الصوتية التي تمثل صعوبات لدارسي العربية من غير الناطقين، هذه الكلمات تتميز بكونها كلمات واقعية حية دالة على حيوية اللغة العربية وتفاعلها مع الواقع الحضاري المتغير.

ما قبل بناء مدونة صوامت اللغة العربية الفصيحة:

تضم مدونة صوامت اللغة العربية الفصيحة التعليمية موضوع الدراسة "الأصوات العربية" قائمة بالمفردات شائعة الاستخدام في مجالات مختلفة، تلي حاجات المبتدئين- مبدئيًا (3000) كلمة- من الدارسين الناطقين بغير العربية؛

وهو مما يمكنهم من الاتصال بمتحدثي العربية في أقطارها المختلفة؛ ومن ثمّ يتمكن الناطقون بغير العربية من إشباع رغباتهم في الاتصال، وكذلك مساعدة المعلم في استخدام طرق ووسائل، تؤهل الدارس لتنمية المعارف والمهارات خاصة، للأسباب الآتية:

لكي يمارسوا عملية التعليم في حياتهم اليومية في بلد ناطق باللغة العربية، وليساعدوا كذلك الناطقين بالعربية المقيمين في خارج بلادهم على التعامل مع مواقف الحياة اليومية.

تحسين فرص تبادل الأفكار والمهارات والمعلومات ذات الطابع الاجتماعي مع الشباب البالغين ممن يتحدثون اللغة العربية.

ليصلوا إلى فهم أفضل وأعمق لمختلف مظاهر أسلوب حياة الناطقين باللغة العربية أولاً، وثانياً لإثراء موروّثهم الثقافي والأدبي⁽¹⁾.

مصادر كلمات المدونة:

استعان الباحث بعددٍ من القوائم في جهوده لجمع كلمات المدونة، وهذه القوائم هي:

استبيان حول المفردات، والمواقف، والملاحق الثقافية في تعليم العربية للناطقين بغيرها، من عمل الدكتورين: محمود فهمي حجازي، ورشدي أحمد طعيمة⁽²⁾.

قائمة مكة للمفردات الشائعة والقائم عليها نخبة من الباحثين بوحدة البحوث والمناهج بمعهد اللغة العربية بجامعة أم القرى، وذلك بالاستعانة بعدة

(1) الإطار المرجعي الأوروبي المشترك للغات، الطبعة العربية، ص 15.

(2) انظر: طعيمة، رشدي أحمد، الأسس المعجمية والثقافية لتعليم اللغة العربية لغير الناطقين بها، جامعة أم القرى، معهد اللغة العربية، وحدة البحوث والمناهج، مكة المكرمة، 1982م.

قوائم أخرى وهي: قائمة الرياض، وقائمة معهد الخرطوم، والرصيد اللغوي الوظيفي للمرحلة الأولى من التعليم الابتدائي، التي وضعتها الهيئة الاستشارية للمغرب العربي، وقائمة جامعتي متشجان، والقاهرة الأمريكيتين، بالإضافة إلى الاستبيان السابق ذكره للدكتورين محمود فهمي حجازي ورشدي أحمد طعيمة⁽¹⁾.

مراحل بناء مدونة صوامت اللغة العربية الفصحى:

وفقاً لمبادئ الصناعة المعجمية، ولطبيعة المدونة المنشودة؛ فإن بناء المدونة تضمن المراحل الآتية:

المرحلة الأولى: مرحلة جمع الكلمات:

وتضمنت جمع كلمات فصيحة معاصرة حية، تتسم بالشيوخ والتكرار، يتعامل الناس بها في كل الأقطار العربية، سواء أكان على مستوى الحديث أم الكتابة، التي تعكس بدورها واقع اللغة العربية المعاصر، وتتميز- كذلك- بكونها تتضمن الوحدات الصوتية الصعبة بالنسبة لدارسي العربية من الناطقين بغيرها.

المرحلة الثانية: مرحلة التسجيل الصوتي:

- نظراً للطبيعة الصوتية للمدونة سُجلت مخارج حروف صوامت اللغة العربية وعددها (28) صوتاً من واقع رموز الألفبائية الدولية IPA، مع شرح تلك المخارج بأسلوب مبتكر يقرؤها إلى الناطقين بها وبغيرها.

- تسجيل أمثلة من الكلمات تشمل تشكيلات الحرف المختلفة؛ من حيث:

- الشدة، والفتحة، والضمة، والكسرة.
- الحرف في بداية الكلمة، وفي أثنائها، ونهايتها.

(1) جامعة أم القرى، معهد اللغة العربية، وحدة البحوث والمناهج، مكة المكرمة.

ووفقاً لهدف مدونة صوامت اللغة العربية- التي يرتجى أن تكون قبلة المعنيين بتدريس اللغة العربية للناطقين بغيرها-، فإن التركيز كان على اختيار الكلمات التي تتضمن (الأصوات المطبقة والحلقية والاحتكاكية والمستعلية والمفخمة) ثم تسجيلها صوتياً⁽¹⁾، وهذه الكلمات هي التي تمثل مشكلة كبرى للطلاب الناطقين بغير العربية أصحاب اللغات محل الدراسة؛ من حيث نطقها أو الاستماع إليها بالشكل الصحيح وفق قواعد اللغة العربية وضوابطها.

المرحلة الثالثة: مرحلة تصنيف الكلمات:

صُنِفَت الكلمات طبقاً لمجالات الاستبيان حول المفردات، والمواقف، والملامح الثقافية في تعليم العربية للناطقين بغيرها من عمل الدكتورين: محمود فهمي حجازي، ورشدي أحمد طعيمة، الذي يضم عددًا من المواقف التفصيلية الموزعة على 20 مجالاً⁽²⁾؛ هي:

- 1- بيانات شخصية.
- 2- السكن.
- 3- العمل.
- 4- وقت الفراغ.
- 5- السفر.
- 6- العلاقات مع الآخرين.
- 7- المناسبات العامة والخاصة.

(1) سُجِّلَت الأصوات بواسطة باحث الدكتوراه الزميل / محمد سلام الحاصل على إجازة القرآن الكريم من الأزهر الشريف، 2017م.

(2) طعيمة، رشدي أحمد، الأسس المعجمية والثقافية لتعليم اللغة العربية لغير الناطقين بها، جامعة أم القرى، معهد اللغة العربية، وحدة البحوث والمنهج، مكة المكرمة، 1982م.

8- الصحة والمرض.

9- التربية والتعليم.

10- في السوق.

11- في المطعم.

12- الخدمات.

13- البلدان والأماكن.

14- اللغة الأجنبية.

15- الجو (المناخ).

16- المعالم الحضارية.

17- الحياة الاقتصادية.

18- الدين والقيم الروحية.

19- الاتجاهات السياسية والعلاقات الدولية.

20- العلاقات الزمانية والمكانية.

المرحلة الرابعة: مرحلة البرمجة

- صُمِّمَت واجهة للتطبيق بنظام iOS على الهواتف المحمولة والحواسيب اللوحية iPhone/iPad.

- يراعي التطبيق أساسيات البرمجة والتصميم، من حيث الشكل والمضمون (الألوان، والصوتيات، ومقاطع الفيديو، والاستخدام الميسر-لوحة مفاتيح عربي للبحث عن المفردات).

- رفع التطبيق على App store <https://apps.apple.com/app/id1482546188>

بالإضافة إلى العناصر الأساسية التي ينبغي أن تتوافر في المدونة؛ وهي:

ا. الشمولية: بأن تتسم كلمات المدونة بالواقعية والحيوية، الدالة على اللغة

- العربية المعاصرة (الحدود الزمانية والمكانية).
- II. الكفاية: بأن تفي كلمات المدونة بالغرض الذي وُضعت لأجله (تذليل الصعوبات الصوتية التي يواجهها الطلاب الناطقون بغير العربية عند نطق الحروف العربية الساكنة، أو الاستماع إليها).
- III. التعبير عن واقع اللغة: من خلال استهداف المفردات، والتعبيرات الشائعة اليومية في البلاد الناطقة بالعربية، التي تُمكن متعلم اللغة من التواصل بشكل جيد مع الناطقين بهذه اللغة، مستخدماً اللغة العربية الفصحى Modern Standard Arabic المستخدمة في أجهزة الإعلام، والأحاديث الرسمية، والكتابات الصحفية والأدبية، وتيسيراً لصعوبات النطق والاستماع إلى الأصوات العربية غير الموجودة في لغته الأم.

مرحلة ما بعد بناء مدونة صوامت اللغة العربية الفصحى:

قبل استعراض متطلبات مرحلة ما بعد بناء المدونة، يؤكد الباحث على أن المدونة اقتصر على الصوامت دون الصوائت؛ لأسباب تتعلق بسهولة وصف الصوامت. الجدير بالذكر أن وصف الصوائت يكون أصعب من وصف الصوامت، لأن نطق الصوامت يستلزم احتكاكاً أو التقاءً محكماً بين أعضاء معينة أثناء النطق.⁽¹⁾ وهو ما أورده د. بشر من خلال قصتين مهمتين أوردهما عالمان جليان من علماء العربية، ولسوف يتبين لنا مما قرره هذان العالمان أننا عند وصفنا للصوامت يتوفر لنا عناصر مادية محسوسة، يمكن أن نلاحظها بسهولة، بعكس الصوائت/الحركات التي تتميز بالانفتاح وعدم الالتقاء المحكم

(1) الشايب، فوزي، محاضرات في اللسانيات، عالم الكتب الحديث-الأردن، الطبعة الثانية، (مزيدة ومنقحة)، 2016م، ص236.

بين الأعضاء، كذلك مجرى الهواء الحر المطلق ومن ثم يصعب وصفها، إضافة إلى التشابه الواضح بين الحركات/الصوائت لدرجة يصعب الفصل بينها بحدود واضحة وجلية، خلافاً للصوائت التي تكون إما انفجارية أو احتكاكية.⁽¹⁾

- خلاصة القصة الأولى أن الخليل بن أحمد، عند الكلام على حروف العربية، نراه يوزع هذه الحروف على مخارجها، وينسب كل واحد (أو مجموعة) منها إلى مدرجة أو حيز معين من أحياز النطق المعروفة، كالحلق واللسان والشفة إلخ.. ولكنه في الوقت نفسه لا يسلك هذه المسلك مع الألف والياء والواو (والهمزة كذلك)، فلا يربطها بمخرج من هذه المخارج، ولا ينسبها إلى أي واحد منها، بل ينسبها إلى الهواء.⁽²⁾

- خلاصة القول في القصة الثانية أن ابن جني قد أدرك خاصية حروف المد بوصفها حركات، وهي أن هواءها يمر حرّاً طليقاً دون مانع يمنعه، على حين يحس إحساساً صادقاً بخاصية النوع الآخر من الحروف وهي الأصوات الصامتة، فيلاحظ أن هواءها قد يقف وقوفاً تاماً. فلا "تجد للصوت منفذاً هناك" أو لا يقف ولكنه ينسل من خلال طريق ضيق.⁽³⁾

ومن هذا الذي قرره لغويو العرب نستطيع القول بأن "مدونة الصوائت العربية" اقتصرَت على الصوائت العربية فقط دون الصوائت/الحركات؛ لما تنماز به الصوائت من خصائص وعلامات مستقلة يستطيع الباحث الوقوف عليها بشكل يسهل وصفه وتحديده، ومن ثم فإن الأصوات الصامتة في هذه

(1) المصدر السابق.

(2) بشر، كمال، علم الأصوات، دار غريب للطباعة والنشر، القاهرة، مصر، 2000م، ص 156.

(3) المصدر السابق، ص: 159.

المدونة هي:

همزة القطع ب ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ ع غ ف ق ك ل م ن
هو (في نحو ولد، يوم) ي (في نحو ولد).
تتطلب مرحلة ما بعد بناء المدونة:

- تجريب الباحث لمخرجات المدونة على أكثر من طالب غير ناطق بالعربية؛
للوقوف على مدى فاعلية استخدام المدونة في تذليل صعوبات النطق
والاستماع إلى الأصوات العربية غير الموجودة في لغته الأم.
- فتح المدونة لإسهام المتطوعين من الباحثين والطلبة الناطقين بغير العربية
(باقتراح كلمات تمثل صعوبة في نطقها/ الاستماع إليها، أو بتسجيل تلك
الكلمات الصعبة بواسطة ناطقي اللغة العربية وفق قواعد اللغة العربية
الفصحى وضوابطها)؛ وذلك تحت إشراف لُغوي متمرس يعي طبيعة عمل
المدونة، عن طريق توظيف التقنيات الحديثة Online أسوة بما يتم في اللغات
الأخرى كاللغة الإنجليزية مثلاً.

الخاتمة

أولاً- النتائج:

- تفاعل المستخدمين مع التطبيق -Online- أثمر النتائج الآتية:
- (1) استخدام التطبيق من قبل مدرسي صعوبات النطق والكلام.
 - (2) تعامل الأطفال مع التطبيق بوصفه لعبة تعليمية لمخارج الحروف.
 - (3) وُجد أن التطبيق مفيد للطلبة الكوريين بالإضافة إلى اللغات العشر محل الدراسة.
 - (4) عُرِزَت المدونة بعدد (2000) كلمة، هي الأكثر تَكَرُّراً من المدونة العربية الدولية International Corpus of Arabic التابعة لمكتبة الإسكندرية، وهي مدونة للعربية المعاصرة، تطمح إلى بناء مدونة قوامها مائة مليون كلمة.⁽¹⁾
 - (5) عُرِزَت المدونة ببيان الكلمات الأكثر تَكَرُّراً من مدونة مدينة الملك عبد العزيز للعلوم والتقنية (المدونة العربية الدولية) KACST، وهي مدونة تطمح إلى بناء مدونة لغوية عربية تحوي مليار - بليون - كلمة مما دُوِّن بالعربية ابتداءً من العصر الجاهلي حتى العصر الحديث.⁽²⁾

جوانب القصور:

اعتور البحث بعض جوانب القصور الناتجة عن طبيعة العمل الفردي ومحدودية الإمكانيات المتاحة للباحث، وتمثلت فيما يلي:

(1) الأنصاري وآخرون، 2007 Alansari et al.

(2) Al-Thubaity, A. O. (2015). A 700M+ Arabic corpus: KACST Arabic corpus design and construction. Language Resources and Evaluation, 49(3), 721-751.

- (1) عدم القدرة على الاطلاع على بعض الأرصدة اللغوية، لعدة أسباب منها:
 - التكلفة العالية التي تفرضها الجهات القائمة على تلك الأرصدة.
 - عدم إتاحتها إلا من خلال قنوات فريق مشروع تلك الأرصدة، والباحث قد يكون على اتصال ببعضهم دون بعضهم الآخر من خلال البريد الإلكتروني.
- (2) قلة الدراسات التقابلية بين اللغة العربية وبعض اللغات الأخرى مثل اللغة الروسية والألمانية- إن التقدم الكبير في تعليم اللغة الإنجليزية بوصفها لغة أجنبية يعود إلى المقابلة اللغوية بين الإنجليزية ولغة الدارس الأصلي، وقد واجه الباحث تلك الصعوبة بالاستعانة بالزملاء الأفاضل في أقسام اللغات بكليات الآداب جامعة القاهرة، وجامعة عين شمس، بالإضافة إلى بعض الباحثين العرب الناطقين بتلك اللغات، بالإضافة إلى الباحثين من غير العرب والناطقين باللغة العربية وهم:
 - أ.د. فرانتشيك أوندراش، مدير برنامج اللغة العربية بكلية الآداب جامعة تشارلز، براغ، جمهورية التشيك.
 - أ.د. رانيا محمد فوزي، رئيس قسم اللغات الشرقية بكلية الآداب جامعة عين شمس، القاهرة، مصر.
 - د. إسلام صالح تهايمي، أ. مساعد اللغة التركية وآدابها، كلية الآداب جامعة عين شمس، القاهرة، مصر.
 - د. ندى عيد، مدرس مساعد اللغة الإسبانية وآدابها، كلية الآداب جامعة القاهرة، القاهرة، مصر.

ثانيا: التوصيات:

رغبة في معالجة جوانب القصور لمدونة صوامت اللغة العربية... يوصي
البحث بما يلي:

- العمل على تقييم المدونة وتقويمها؛ للوقوف على جوانب القصور، ومحاولة معالجتها بطريقة منهجية تتفق وأساليب صناعة المدونات والأرصدة اللغوية، وطبيعة اللغة العربية.

ملحق (1) روابط ونماذج للتطبيق

لقاء تلفزيوني يشرح الهدف من التطبيق:

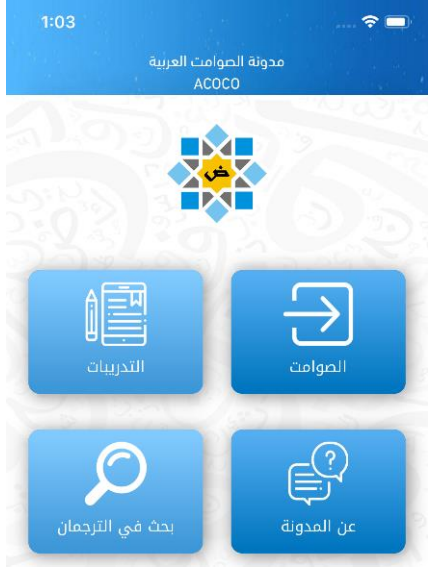
<https://youtu.be/J0M9iUi7-qw>

نموذج مجاني للتطبيق على الموقع الإلكتروني:

<https://code-grow.com/acoco>

كامل التطبيق على متجر التطبيقات لأجهزة الأيفون والأيباد:

<https://apps.apple.com/app/id1482546188>

	
شكل (1) واجهة التطبيق	شكل (2) أقسام التطبيق

صورة اليوم

نِزاع (ج) أزرع / اسم

صوامت الألفبائية الصوتية العربية الدولية IPA

	Labial شفتي	Labio-dental شفتي أسناني	Interdental فج أسناني	Alveolar سني	Palatal قشري	Velar لحوي	Uvular فج اللسان
osive النفجار	b ب			t ت ṭ ط			k ك q ق
ricate وقفي احد				d د ḏ ذ			
cative المنكبة		f ف	θ ث ð ذ	s س z ز	ʃ ش ʒ ج	x خ g غ	
asal الغمر	m م		n ن				
lteral جانب				l ل			
lapp تكرار				r ر			
i vowel * نصف هـ	w و				y ي j ج		

* شفتي بمعنى خلف "الفأرو طيفي والباء غاري"

اضغط على الحرف للاستماع

شكل (3) قسم صورة اليوم

شكل (4) قسم جدول الصوامت

ملحق (2)
عينة من قائمة كلمات مدونة الصوامت العربية

م	الصوت	الكلمة	م	الصوت	الكلمة
1.	ب	بَطَّل	210.	ن	نَظَّفَ
2.		بَاعَ	211.		نَظَرَ
3.		بِطَاقَة	212.		نَاطِقِينَ
4.		بِئْرٌ	213.		صُنُبُور
5.		بَا حِثْ	214.		إِنْ سَحَبَ
6.		يَتَسَابَقُ	215.		إِنْضِمَام
7.		لَيْثٌ	216.		قَنْطَرَة
8.		حَبَسَ	217.		غَنِيٌّ
9.		أَرْبَعَاء	218.		غَنِيٌّ
10.		مَغْرِب	219.		مُنْظَمَة
11.		ضَرَبَ	220.		تَدْخِين
12.		وَهَبَ	221.		طَيَّرَان
13.		أَحَبَّ	222.		قَرَأَ
14.		أَعَزَبَ	223.		غَضَبَان
15.	م	مَطْبَخ	224.	ر	رَابِطَة
16.		مُجْتَمَعٌ	225.		رَعْدٌ
17.		مَشْغُول	226.		رِيَا ضِيٌّ
18.		مُصْحَف	227.		رِيَا ضَة
19.		رَمَضَانُ	228.		رَأَى

م	الصوت	الكلمة	م	الصوت	الكلمة
.20		عَاصِمَة	.229		قَرِيب
.21		مُمَطِّرٌ	.230		مَرَضٌ
.22		تَصْمِيم	.231		سَرِيع
.23		صَامَ	.232		نَظَرِيَّة
.24		يُقَدِّمُ	.233		مَعْرِض
.25		رَقْمٌ	.234		مَعْرِضٌ
.26		أَقْدَمَ	.235		كَثِير
.27	و	وَثِيقَة	.236		مَخَاطِر
.28		وَجْهَة	.237		قَرَار
.29		وُزَرَاء	.238		أَسْعَار
.30		أَوْضَاع	.239		حَارٌّ
.31		أَقْوَال	.240		مَنَاطِرُ
.32		تَوَضَّأَ	.241		مَنَاطِر
.33		ثَوْرَة	.242	ز	زَرْقَاء
.34		تَصْوِير	.243		زَوْرَق
.35		تَشْوِيش	.244		زَائِر
.36		خَوْخ	.245		زِرَاعَة
.37		خُرْطُوم	.246		غَزَل
.38		طَاوُوس	.247		فِيْزِيَاءُ
.39		صَوَّرَ	.248		فِيْزِيَاء
.40		عُضْوِيَّة	.249		عَزِيْزٌ
.41		مَوَاضِيع	.250		عَزِيز

م	الصوت	الكلمة	م	الصوت	الكلمة
.42		أَدْوِيَّة	.251		حَجَز
.43		بَدُو	.252		حَوَافِز
.44		سُمُو	.253		غَاز
.45		جَوّ	.254		عَجُوز
.46		حَشَو	.255		جَوَائِز
.47		عُلُوّ	.256	ل	لَا حَظَّ
.48		عَدُوّ	.257		أَلْفَاظ
.49		عُضُوّ	.258		أَلْفَاظُ
.50	ف	فَجَاءَ	.259		لُغَة
.51		فَجَرَّ	.260		لُغَة
.52		فِضَّة	.261		أَصْلَحَ
.53		فَقِير	.262		مُغْلَق
.54		فُرْصَة	.263		مُغْلَقُ
.55		رَفُضَ	.264		ذَلِيل
.56		مُقَاوَضَ	.265		أَفْلَعَ
.57		تَصَقَّحَ	.266		شُغِلَ
.58		إِتِفَاقِيَّة	.267		شُغِلَ
.59		تَصْفِيْقُ	.268		قَابَلَ
.60		خَفِيف	.269		رَحَلَ
.61		طَوَافُ	.270		مُعْطَل
.62		مَتَحَف	.271		مُعْطَلُ
.63		تَنْظِيف	.272	ش	شَاطِئُ

م	الصوت	الكلمة	م	الصوت	الكلمة
.64		رَفّ	.273		شَاطِئُ
.65	ث	ثَلَاجَة	.274		شَقَّة
.66		تُعْبَان	.275		شَرِيحَة
.67		ثَانَوِيّ	.276		شُرْطَة
.68		تَوْب	.277		شِرَاء
.69		ثَرِيّ	.278		شِعْر
.70		ثَوْرَة	.279		شِعْر
.71		تَوَانٍ	.280		عَطْشَانُ
.72		ثَقِيل	.281		تَأْشِيرَة
.73		تُوم	.282		مِشْرَط
.74		مُتَلَجَّات	.283		مِشْرَطُ
.75		أَثْمَرَ	.284		أَنْشُودَة
.76		أُنْقَى	.285		نَاقِشَ
.77		مُمَثِّل	.286	ج	جُغْرَافِيَا
.78		تِمثال	.287		جِرَاحَة
.79		أَثْنَاء	.288		جَرِيدَة
.80		مُتَقَف	.289		جَائِع
.81		مُتَلَّث	.290		جَائِعُ
.82		حَدِيث	.291		عَجَز
.83		بَحْث	.292		أَجْنَبِيّ
.84		حَادِث	.293		تَأْجِيل
.85	ذ	ذُرَة	.294		رُجَاج

"الأصوات العربية" وتعليمها للناطقين بغيرها

م	الصوت	الكلمة	م	الصوت	الكلمة
.86		ذِرَاع	.295		زُجَاجٌ
.87		ذَكَرَ	.296		خَرَجَ
.88		ذَكَاءَ	.297		أَزْعَجَ
.89		مِثْدَنَة	.298		ثَلَجٌ
.90		عَذَبٌ	.299		حَجَّ
.91		أَذَان	.300	ي	يُصَرِّحُ
.92		مُؤَدِّن	.301		يَسْتَيْقِظُ
.93		مَذْهَب	.302		يَوْمِئِذٍ
.94		بُودِيَّة	.303		يَقْتَرِضُ
.95		بُدُور	.304		قيد
.96		احذر	.305		صيد
.97		إِذَاعَة	.306		لِيلٌ
.98		قَذِر	.307		حياة
.99		قَذِرٌ	.308		جُنَيْه
.100		لَذِيذ	.309		صِنَاعِيٌّ
.101	ظ	ظُهُور	.310		رَضِيَ
.102		ظَلَام	.311		كَهْرَبَائِيٌّ
.103		يَظْهَرُ	.312		كَهْرَبَائِيٌّ
.104		ظالم	.313		عادي
.105		وَضِيفَة	.314		عادي
.106		يُنْظَفُ	.315	ك	كَارِثَة
.107		مُوظَّف	.316		كَهْرَبَاء

م	الصوت	الكلمة	م	الصوت	الكلمة
.108		مَنْظَر	.317		رَكَعَ
.109		نَظَّارَة	.318		تَذَكَّرَه
.110		مُنَاطِر	.319		اِخْتِكَاك
.111		مُنَظَّف	.320		مُحَرِّك
.112		نِظَام	.321		مُحَرِّك
.113		نُظْم	.322	غ	غِنَاء
.114		يُنَظِّمُ	.323		غَادَرَ
.115		ظَفُرٌ	.324		غِذَاءٌ
.116		تَحَقُّظَ	.325		إِلْغَاء
.117		إِخْتَفَظَ	.326		إِلْغَاءٌ
.118		غَلِيظٌ	.327		فَرَاغ
.119	ت	تَعَاقَدَ	.328	خ	خَطَأً
.120		تَوْحِيد	.329		خَاصَّةً
.121		تَوْضُحًا	.330		خَلَعَ
.122		تُرَاب	.331		مُنْخَفِضٌ
.123		إِسْتَيْقَظَ	.332		أَخْضَرَ
.124		إِنْتَظَرَ	.333		تَأَخَّرَ
.125		أُسْتَاذ	.334		أَخِيرَ
.126		إِتَّضَحَ	.335		رَخِيسَ
.127		وَقْتُ	.336		أُخُوَّة
.128		وَقْتُ	.337		بِطِّيخٍ
.129		صَوْتُ	.338		بِطِّيخ

"الأصوات العربية" وتعليمها للناطقين بغيرها

م	الصوت	الكلمة	م	الصوت	الكلمة
.130		هَاتِف	.339		تَارِيخ
.131		أَنْتِ	.340	ق	قَائِد
.132	ط	طَرِيق	.341		قُنْصُل
.133		طَائِرَة	.342		قَمْح
.134		طِبَاعَة	.343		قَمْح
.135		مُطَلَّق	.344		دَقِيقَة
.136		مَنَاطِقُ	.345		إِنْقِلَابٌ
.137		نَاطِقٌ	.346		إِنْقِلَاب
.138		نُقْطَة	.347		عَلَاَقَات
.139		مَحْطَة	.348		بَرْقٌ
.140		قِطَار	.349		إِسْتَحَقَّ
.141		حَائِط	.350		أَزْرَق
.142		حَائِطٌ	.351		رَقِيق
.143		ضَغْطٌ	.352	ح	حَادِث
.144		ضَوَابِطُ	.353		حَادِثٌ
.145	د	دَوْرَة	.354		حَوْضٌ
.146		دَلِيل	.355		حِذَاءٌ
.147		دَفْتَر	.356		صَحَافَة
.148		دُكَّان	.357		صَحَافَة
.149		دِرَاسَة	.358		يَحْفَظُ
.150		دَقِيقَة	.359		ضَحِكَ
.151		بَدَأَ	.360		سَائِح

الكلمة	الصوت	م	الكلمة	الصوت	م
وُضُوْحٌ		.361	عَقِيْدَة		.152
صَحِيْح		.362	تَصْدِيْر		.153
عَاصِيفَةٌ	ع	.363	صُنْدُوق		.154
عَاصِيفَةٌ		.364	ضِدُّ		.155
عَرَضَ		.365	عَهْد		.156
عَطَشَ		.366	عَهْدٌ		.157
عَطَشُ		.367	طَرَدَ		.158
مَطْعَمٌ		.368	نُقُود		.159
إِعْتَذَرَ		.369	ضَبَقَ	ض	.160
صَعِدَ		.370	ضَخَمَ		.161
صَغِبَ		.371	ضَوء		.162
صَغِبَ		.372	وُضُوءٌ		.163
مُعَقَّد		.373	حَضَانَة		.164
مُعَقَّدٌ		.374	رَوْضَة		.165
مُذِيع		.375	غَضِبَ		.166
مَمْنُوع		.376	مَوْضُوع		.167
إِيْدَاع		.377	مُمْرِضَة		.168
يَدْفَعُ		.378	أَرْضِيَّ		.169
أَخَذَ	ء	.379	حَاضِر		.170
أَنْبِيَاءُ		.380	أَرْضٌ		.171
أَنْبِيَاءُ		.381	تَخْفِيضٌ		.172
أَقْرَاصُ		.382	قَرَضٌ		.173

"الأصوات العربية" وتعليمها للناطقين بغيرها

م	الصوت	الكلمة	م	الصوت	الكلمة
174.		أَبْيَضُ	383.		أَبْيَضُ
175.	س	سَائِقُ	384.		ثَائِر
176.		سَدَدٌ	385.		قِرَاءَةٌ
177.		سُحُورٌ	386.		تَأْخِير
178.		سُقُوط	387.		عُلَمَاءُ
179.		إِسْتَعْرَقَ	388.		عُلَمَاءُ
180.		إِسْتِقَالَةٌ	389.		تَوْضُحًا
181.		بَسِيط	390.		بُطْءٌ
182.		بَسِيطٌ	391.	ه	هَنَأٌ
183.		إِسْتِئْذَان	392.		هَبَطَ
184.		مُسْتَشَار	393.		هَضَبَةٌ
185.		دُسْتُور	394.		هَضَبَةٌ
186.		سِيَّاسِيٌّ	395.		مَعْمَدٌ
187.		فَاسِدٌ	396.		مَعْمَدٌ
188.		كُرْسِيٌّ	397.		زُهُورٌ
189.		حَسَّاس	398.		أَسْهُم
190.		جَلَسَ	399.		وَاللَّهِ
191.		دَرَسَ	400.		هَذِهِ
192.		رَّئِيس			
193.		بَطَاطِيس			
194.		قَاسٍ			
195.		قَامُوس			

م	الصوت	الكلمة	م	الصوت	الكلمة
196.	ص	صَدَقَة			
197.		صُعُوبَة			
198.		صَحَّ			
199.		أَصْغَر			
200.		قَصِير			
201.		إِقْتِصَاد			
202.		مَصْرِف			
203.		رَصِيف			
204.		مَصْنَع			
205.		تَخَصَّصَ			
206.		قَمِيص			
207.		مِقْصَّ			
208.		نُصُوص			
209.		خَاصَّ			

قائمة المراجع

أولا- المراجع العربية.

- 1- الإطار المرجعي الأوروبي المشترك للغات، الطبعة العربية "ترجمة إلى العربية عن د.علا عادل عبد الجواد - د. ضياء الدين زاهر - ماجدة مذكور - نهلة توفيق"، دار إلياس العصرية للطباعة والنشر بالتعاون مع معهد جوته، القاهرة، مصر، 2008م.
- 2- بشر، كمال علم الأصوات، دار غريب للطباعة والنشر، القاهرة، مصر، 2000م.
- 3- حجازي، مصطفى حجازي السيد، العربية والهوسا، نظرات تقابلية، معهد اللغة العربية وحده البحوث والمناهج- جامعة أم القرى، مكة المكرمة، 1985م.
- 4- جينغشو، تشانغ، المؤتمر الدولي لتعليم اللغة العربية، آفاق وتحديات ماليزيا والصين 5- 6 ديسمبر 2011م / 9- 10 محرم 1432هـ كلية اللغة العربية، جامعة الدراسات الأجنبية ببيكين، الصين، 2011م.
- 5- الخطيب، محمد عبد الفتاح، مؤتمر "اللغة العربية وتحديات العصر" بالتعاون بين كلية دار العلوم جامعة القاهرة ومركز الدراسات المعرفية بالقاهرة في الفترة من 4-5 نوفمبر 2008م، كلية دار العلوم، القاهرة، 2008م.
- 6- روزير نبوت، نيقولاس، المؤتمر الدولي الأول لتعليم العربية للناطقين بغيرها الرؤى والتجارب، إسطنبول 25- 26 إبريل 2015م.
- 7- الشايب، فوزي، محاضرات في اللسانيات، عالم الكتب الحديث-الأردن، الطبعة الثانية، (مزيدة ومنقحة)، 2016م.

- 8- طعيمة، رشدي أحمد، الأسس المعجمية والثقافية لتعليم اللغة العربية لغير الناطقين بها، جامعة أم القرى، معهد اللغة العربية، وحدة البحوث والمناهج، مكة المكرمة، 1982م.
- 9- ماسيري، كوري، ودفع الله أحمد الأمي، سمي، المشكلات الصوتية في تعلم اللغة العربية للناطقين بغيرها، جامعة المدينة العالمية أنموذجاً، جامعة المدينة العالمية، وكالة البحوث والتطوير، دولة ماليزيا، 2012م.
- 10- نابي عبدالله، رحاب، تعليم الكتابة للناطقين بغير العربية بين النظرية والتطبيق، الملتقى الأول/الأبحاث . azhar- al. ;com /go// نقلا عن كحولي، (زهرة سعدلاوي حرم). "البعد الاستثماري في تدريس اللغة العربية لغير الناطقين بها، التجربة التونسية أنموذجاً"، ص1.
- 11- الأنصاري، سامح، المؤتمر الثامن للجمعية المصرية لهندسة اللغة، بناء المدونة العربية العالمية: سير مرحلة جمع البيانات، القاهرة، 2007م.
- 12- الأنصاري، سامح، المؤتمر الثامن للجمعية المصرية لهندسة اللغة، نحو تحليل المدونة العربية العالمية: سير مرحلة التحليل الصرفي، القاهرة، 2008م.

ثانياً. المراجع الأجنبية:

- 1- Al- Sulaiti, L., Atwell, E., (2004). *Designing and Developing a Corpus of Contemporary Arabic*. In Proceedings of the sixth TALC conference. Granada, Spain.
- 2- Ali, A. R., & Ijaz, M. (2009). *English to Urdu transliteration system*. In Proceedings of the Conference of Language and Technology (CLT'09).

- 3- Alwabari, S. (2013). *Non-Native Production of Arabic Pharyngeal and Pharyngealized Consonants*. Unpublished master's thesis, Carleton University, Ottawa.
- 4- Allen, Roger. (1992). "11. *Teaching Arabic in the United States: Past, Present, and Future*." *The Arabic language in America*.
- 5- Al-Thubaity, A. O. (2015). A 700M+ Arabic corpus: *KACST Arabic corpus design and construction*. *Language Resources and Evaluation*, 49(3), 721-751.
- 6- Ammon, U. (2006). *Sociolinguistics: An international handbook of the science of language and society = Soziolinguistik*. Berlin: W. de Gruyter.
- 7- Amr Gomaa. (2015). صناعة المناهج وتطويرها باستخدام التقنيات الحديثة... نحو منهج معاصر للغة العربية. QScience Proceedings: Vol. 2015, Conference on Education 2015. Partners in Excellence, 40. <https://doi.org/10.5339/qproc.2015.coe.40>
- 8- Andi- Pallawa, B., & Alam, A. F. A. (2013). *A comparative analysis between English and Indonesian phonological systems*. *International Journal of English Language Education*, 1(3), 103- 129.
- 9- Arifin, Z., & Latif, M. R., & Halim, Z. (2014)
- 10- Catford, J., Palmer, J. D., Dew, J. E., Barry, R. G., Cheng, H. L., Hsu, V. L., & Li, Y. (1974). *A contrastive study of English and Mandarin Chinese*.
- 11- Clements, G. N., & Sezer, E. (1982). *Vowel and consonant disharmony in Turkish*. *The structure of phonological representations*, 2, 213- 255.

- 12- (DOĞAN) Yusuf ،(AYDIN) Tahirhan) التي تواجه المشكلات (2013)
Cumhuriyet Üniversitesi İlahiyat
Fakültesi Dergisi, 2013, Cilt: XVII, Sayı: 1.
- 13- George Quinn, *The Learner's Dictionary of Today's Indonesian*.
Sydney: Allen & Unwin 2001 ISBN 1864485434

ثالثا- المواقع الإلكترونية والبرمجيات:

- 1- <http://www.internationalphoneticalphabet.org>
- 2- <https://www.britishcouncil.org/sites/default/files/languages-for-the-future-report-v3.pdf>
- 3- <https://www.theguardian.com/education/gallery/2015/jan/23/a-language-family-tree-in-pictures#img-1>
- 4- <https://www.theguardian.com/education/gallery/2015/jan/23/a-language-family-tree-in-pictures#img-1>

