أتمتة التمييز بين المعرب والمولد في اللغة العربية باستعمال تقنيات التعلم الآلي دراسة صرفية تطبيقية

د.علاء عبدالخالق حسين المندلاوي
بحث علمي لموقع مؤسسة العراقة للثقافة والتنمية
بتاريخ٢/ ١٠/ ٢٠٢٦

ملخص البحث:

يهدف هذا البحث بنحوٍ رئيس إلى بناء جسر معرفي وتطبيقي بين علوم الصرف العربي الأصيلة وتقنيات الذكاء الاصطناعي، من خلال تطوير نموذج حاسوبي قادر على التمييز الآلي بين الكلمات الأصيلة والمعربة والمولدة، مستنداً في ذلك إلى الخصائص البنيوية والاشتقاقية التي تميز لغة الضاد، ولتحقيق هذه الغاية اعتمدت الدراسة على المنهج الوصفي التحليلي لتأصيل القواعد الصرفية وتوصيف علامات العجمة والتوليد في التراث اللغوي، والمنهج التجريبي التطبيقي لبناء النماذج الحاسوبية واختبار كفاءتها عبر مدونات لغوية موسومة، حيث شملت الإجراءات البحثية تحويل الأوزان الصرفية إلى ميزات هندسية وقيم عددية قابلة للمعالجة، وتدريب شبكات عصبية عميقة (CNN وRNN) على رصد الأنماط الصرفية المتسقة والشاذة، وقد خلصت النتائج إلى أن البناء الصرفي العربي يمتلك قابلية عالية للنمذجة الرقمية بدقة تتجاوز ٩٠٪، مع وجود تحديات في التمييز بين المولدات التي تلبس لبوس الأصالة الصرفية، وهو ما يستدعي دمج البعد التاريخي مع التحليل البنيوي، وبناءً على ذلك يوصي الباحث بضرورة تكامل الجهود بين اللغويين والتقنيين لإنشاء مدونات لغوية تاريخية شاملة، وإدراج اللسانيات الحاسوبية كمادة أساسية في أقسام اللغة العربية لتمكين الباحثين من أدوات العصر، وكما يقترح الباحث مشاريع مستقبلية تهدف إلى تطوير "معجم أنطولوجي ذكي" يعتمد على تقنية سلاسل الكتل لتوثيق تاريخية الألفاظ، ودراسة أثر الذكاء الاصطناعي التوليدي في إحياء الأوزان المهجورة وتفعيل الاشتقاق الأكبر لمواكبة الانفجار المعلوماتي المعاصر، بما يضمن بقاء العربية لغة حية وقادرة على المنافسة في الفضاء الرقمي.

الكلمات المفتاحية: الصرف العربي، الذكاء الاصطناعي، التعلم الآلي، المعرب والمولد، اللسانيات الحاسوبية، هندسة الميزات.

Automating the Distinction Between Arabized and Post-Classical Words in the Arabic Language Using Machine Learning Techniques: An Applied Morphological Study

Abstract:

This research primarily aims to establish a cognitive and applied bridge between the authentic sciences of Arabic morphology and Artificial Intelligence (AI) technologies. This is achieved by developing a computational model capable of automatically distinguishing between original (authentic), Arabized (Mu'arrab), and post-classical (Muwallad) words, drawing upon the structural and derivational characteristics that define the Arabic language. To achieve this objective, the study employs a descriptive-analytical approach to ground morphological rules and characterize signs of foreignness (Ajama) and word generation within the linguistic heritage. Simultaneously, it utilizes an applied experimental methodology to construct computational models and evaluate their efficiency using annotated linguistic corpora.

The research procedures involved transforming morphological weights (patterns) into engineered features and processable numerical vectors. Deep neural networks, specifically Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), were trained to detect consistent and anomalous morphological patterns. The results concluded that the Arabic morphological structure possesses a high capacity for digital modeling, achieving accuracy exceeding 90%. However, challenges remain in distinguishing post-classical words that mimic authentic morphological patterns, necessitating the integration of a historical dimension with structural analysis.

Accordingly, the researcher recommends the integration of efforts between linguists and technologists to create comprehensive historical linguistic corpora and the inclusion of Computational Linguistics as a core subject in Arabic language departments to empower researchers with modern tools. Furthermore, the researcher proposes future projects aimed at developing an "Intelligent Ontological Lexicon" based on Blockchain technology to document the etymology of terms. It also suggests studying the impact of Generative AI in reviving obsolete morphological patterns and activating "Greater Derivation" (Al-Ishtiqāq al-Akbar) to keep pace with the contemporary information explosion, ensuring that Arabic remains a vibrant and competitive language in the digital sphere.

Keywords: Arabic Morphology, Artificial Intelligence, Machine Learning, Arabized and Post-classical Words, Computational Linguistics, Feature Engineering.

المقدمة:

تعد اللغة العربية كائناً حياً يتنفس من خلال مفرداته، وهي في جوهرها نظام اشتقاقي صارم يقوم على وحدة "الجذر والوزن"، مما منحها حصانة لغوية وقدرة فائقة على الصمود عبر القرون، إلا أن هذا الصمود واجه تحديات كبرى مع انفتاح العرب على الثقافات الأخرى، حيث تدفقت الألفاظ الأعجمية والمولدة والمعربة إلى جوف المعجم العربي، مما خلق تمايزاً دقيقاً بين ما هو "أصيل" نبت في بيئة الفصاحة، وبين ما هو "دخيل" استعصى على القياس الصرفي أو طوعته الأوزان العربية ليدخل في نسقها، ومع الطوفان المعرفي الرقمي المعاصر، بات من الصعوبة بمكان على الجهد البشري وحده فرز هذه الألفاظ وتصنيفها بدقة واستمرارية، ومن هنا يبرز دور الذكاء الاصطناعي ليس كمنافس للغوي، بل كأداة هندسية قادرة على استيعاب القواعد الصرفية وتحويلها إلى خوارزميات ذكية قادرة على تمييز البنية الصرفية، حيث يركز هذا البحث على سد الفجوة بين علوم الصرف التقليدية وبين تقنيات التعلم الآلي، بهدف بناء نموذج تقني يستطيع التفريق بين الكلمات المعربة والمولدة والأصيلة من خلال تحليل خصائصها البنيوية وتتابع حروفها، وتكمن أهمية هذا البحث في كونه يمثل حائط صد رقمي يحمي الفصحى من اللحن والخلط، ويساعد في تطوير المعاجم الرقمية الذكية التي تخدم الباحثين والمترجمين، ويهدف البحث بنحوٍ أساسي إلى استنباط المعايير الصرفية الدقيقة التي يعتمد عليها الحاسوب في التصنيف، واختبار كفاءة نماذج التعلم العميق في محاكاة "الحس اللغوي" للصرفيين القدامى، وتبرز الفجوة البحثية في أن معظم الدراسات الحاسوبية السابقة ركزت على "الإعراب" أو "الترجمة" دون الغوص في "أصول البنية الصرفية" من منظور تاريخي وتأثيلي يفرق بين طبقات المفردات، ولتحقيق هذه الغاية اعتمدت الدراسة على المنهج الوصفي التحليلي لتأصيل القواعد الصرفية، والمنهج التجريبي التطبيقي لبناء النموذج الحاسوبي واختبار دقته، وقد جاءت هيكلية البحث موزعة على ثلاثة مباحث رئيسة، تناول الأول منها الإطار النظري للصرف العربي ومعايير العجمة والتعريب، بينما خصص الثاني لشرح آليات التعلم الآلي وكيفية معالجة اللغة طبيعياً، وصولاً إلى المبحث الثالث الذي قدم النموذج التطبيقي وتحليل النتائج، لتختتم هذه المقدمة بفتح آفاق جديدة نحو "صرف رقمي" يجمع بين صرامة القاعدة وذكاء الآلة، مؤكدة أن التقنية هي الخادم الأوفى للغة الضاد في عصر الذكاء الاصطناعي.

المبحث الأول: الإطار النظري للصرف العربي بين الأصالة والمولدة

توطئة:

يُعد الصرف العربي الدعامة الأساسية التي تحفظ للغة الضاد تماسكها وبنيتها الفريدة، فهو العلم الذي يدرس بنية الكلمة وما يطرأ عليها من تغييرات قبل دخولها في التركيب الإعرابي، وتكتسب دراسة الأصول اللغوية أهمية بالغة في ظل التداخل الحضاري، إذ يمثل الصرف المصفاة التي تُعرض عليها الكلمات لبيان أصالتها

من عدمها، ومن هنا ينطلق هذا المبحث ليؤصل للمفاهيم الصرفية التقليدية وكيفية تعامل العقل اللغوي العربي مع الوافد اللفظي، تمهيداً لنقل هذه المعايير إلى الفضاء الرقمي والذكاء الاصطناعي.

المطلب الأول: المصطلح الصرفي عند القدامى (المعرب، الدخيل، والمولد) والمعايير التراثية للتمييز

إن الناظر في التراث اللغوي العربي يجد أن العلماء الأوائل قد وضعوا حدوداً صارمة لتصنيف الألفاظ بناءً على نشأتها وكيفية دخولها في اللسان العربي، فاللفظ المعرب هو ذلك اللفظ الأعجمي الذي استعمله العرب وغيروا فيه بزيادة أو نقص أو إبدال ليتوافق مع أوزانهم وصيغهم الصرفية، وهو ما يطلق عليه الجوهري في صحاحه "تعريب الاسم الأعجمي" [١]، حيث لم يكن العرب يقبلون الكلمة الغريبة على علاتها بل كانوا يطوعونها لتجري على سنن كلامهم، وهذا يختلف جوهرياً عن المصطلح الثاني وهو "الدخيل" الذي يشير في الأغلب إلى الكلمات التي دخلت اللغة العربية وبقيت على هيئتها الأعجمية دون أن تخضع لعملية الصقل الصرفي أو التغيير في الحروف، مما جعلها تبدو كأجسام غريبة داخل النسيج اللغوي، وقد تنبه سيبويه في كتابه إلى أن العرب ربما أعربوا الكلمة بإلحاقها بأبنيتهم وربما تركوها كما هي، مما يفتح الباب أمام تفاوت درجات التعريب [٢]، أما المصطلح الثالث والأكثر حيوية فهو "المولد"، ويقصد به اللفظ الذي استحدثه الناس بعد عصر الاحتجاج اللغوي، سواء كان هذا اللفظ مشتقاً من أصول عربية بطريقة لم تعرفها العرب الأوائل، أو كان تعريباً حديثاً لمصطلحات حضارية طرأت بعد استقرار التدوين، ويشير السيوطي في المزهر إلى أن المولد هو ما أحدثه المولدون الذين لا يحتج بكلامهم في اللغة [٣]، ومن هنا تبرز المعايير التراثية التي وضعها اللغويون للتمييز بين هذه الأصناف، ومن أهمها "معيار الوزن" و"معيار الحروف"، فالعرب لم يجمعوا في كلماتهم بين حروف معينة مثل الجيم والقاف في كلمة واحدة، وكما أنهم استثقلوا تتابع بعض الأصوات التي تدل بوضوح على أن الكلمة ليست من معدن العربية، وقد وضع ابن فارس في "الصاحبي" قواعد دقيقة لما يجوز وما لا يجوز في أبنية العرب، معتبراً أن الخروج عن هذه القواعد هو العلامة الفارقة التي يستدل بها على عجمة الكلمة أو توليدها [٤]، إن هذه المعايير لم تكن مجرد آراء ذوقية، بل كانت نظاماً خوارزمياً بشرياً مبنياً على الإحصاء والاستقراء لمدونات الشعر والنثر العربي القديم، مما يسهل اليوم عملية نمذجتها آلياً، فالمعرب يخضع للقياس وإن كان أصله أعجمياً، والمولد يحاول محاكاة القياس لكنه يفتقر إلى شرط السماع عن العرب الخلص، والدخيل يكسر القياس جملة وتفصيلاً، وهذا الثالوث المصطلحي يمثل المادة الخام التي يعالجها الذكاء الاصطناعي لفهم التراتبية التاريخية للمفردات العربية وتصنيفها بدقة تضاهي دقة علماء اللغة الأوائل في تمييز الأصيل من الدخيل.

المطلب الثاني: خصائص الأبنية الصرفية العربية (نظام الجذر والوزن) وعلامات "العجمة" في الكلمة

تتفرد اللغة العربية بكونها لغة اشتقاقية بامتياز، حيث تعتمد في بناء معجمها على نظام "الجذر والوزن"، وهو النظام الذي يجعل من الكلمة العربية بنية رياضية محكمة، فكل كلمة أصيلة غالباً ما ترجع إلى جذر ثلاثي أو

رباعي يمثل المادة الخام للمعنى، ثم تفرغ هذه المادة في "قوالب صرفية" أو أوزان محددة تمنح الكلمة صبغتها الوظيفية، وهذا النظام الاشتقاقي هو ما يطلق عليه اللغويون "الميزان الصرفي"، والذي يعده الرضي الإستراباذي في شرح الشافية الوسيلة الفضلى لمعرفة أحوال أبنية الكلم [٥]، وبناءً على هذا النظام، تظهر "علامات العجمة" بوضوح عندما تأتي الكلمة على وزن لا نظير له في كلام العرب، أو عندما تتجاوز حروفها الأصول العدد المتعارف عليه دون وجود زوائد اشتقاقية، فالعرب مثلاً لم يعرفوا وزناً مثل "فعلّ" بتشديد اللام إلا في كلمات نادرة جداً أو معربة، وكما أن اجتماع بعض الحروف المتباعدة في المخارج أو المتنافرة يمثل دليلاً قطعياً على أن الكلمة معربة، ومن ذلك اجتماع الصاد والجيم، أو السين والزاي، أو الكاف والقاف في كلمة واحدة، وكما فصل ذلك الخليل بن أحمد الفراهيدي في مقدمة كتاب العين [٦]، حيث وضع قانون "الذلاقة" الذي يشير إلى أن الكلمات الرباعية والخماسية الأصل لا بد أن تحتوي على أحد حروف الذلاقة (الراء، اللام، النون، الفاء، الباء، الميم)، فإن خلت الكلمة منها كانت دليلاً على أنها ليست عربية محضة، وفضلاً عن ذلك، فإن "البناء الصرفي" للكلمة العربية يتميز بالاتساق، فلا تجد كلمة تبدأ بسكون، ولا تجد تواليا لأربعة متحركات في كلمة واحدة إلا نادراً، وهذه الخصائص البنيوية هي التي تشكل "البصمة اللغوية" للعربية، والتي يحاول الباحثون اليوم تلقينها للآلة عبر نماذج التعلم العميق، فعندما يواجه النظام الآلي كلمة مثل "تلفاز" أو "مذياع"، فإنه يحللها بناءً على وزن "مفعال" ومدى مطابقتها للجذور العربية، فإذا وجد الكلمة خارجة عن هذه الأطر، صنفها ضمن المعرب أو المولد، إن علامات العجمة ليست مجرد قائمة من المحظورات، بل هي "منطق لغوي" يعكس وعي العربي بجماليات صوته ونظام لغته، فكانوا يفرون من الثقل إلى الخفة، ومن التنافر إلى الانسجام، وهو ما يفسر لماذا غيروا "إسحاق" بكسر الهمزة و"إبراهيم" لتتناسب مع أذواقهم الصرفية [٧]، وهكذا نجد أن نظام الجذر والوزن ليس مجرد أداة تحليلية، بل هو الحارس الصرفي الذي يكشف زيف الكلمات المدعية للأصالة، ويمد الذكاء الاصطناعي بالبيانات الهيكلية اللازمة لعمليات التصنيف والتمييز الصرفي الدقيق.

المطلب الثالث: ظاهرة التوليد اللغوي في العصر الحديث (التعريب، الاشتقاق، والنحت)

تعد ظاهرة التوليد اللغوي المظهر الأبرز لحيوية اللغة العربية وقدرتها على استيعاب معطيات العصر ومصطلحاته العلمية والتقنية، وهي عملية لم تتوقف تاريخياً لكنها اتخذت مسارات أكثر تنظيماً مع نشأة المجامع اللغوية في القرن العشرين، حيث يمثل التوليد استخراج كلمة جديدة لم تكن موجودة بلفظها في المعجم القديم، إما عن طريق "الاشتقاق" من جذر عربي قائم لمواكبة دلالة مستحدثة، وإما عبر "التعريب" الذي هو نقل اللفظ الأعجمي بصورته مع شيء من التغيير الصوتي، وإما عبر "النحت" الذي يعد من أصعب الوسائل الصرفية وأقلها شيوعاً في العربية الفصحى، ويذكر الدكتور إبراهيم أنيس في كتابه "في اللهجات العربية" أن الاشتقاق هو العمود الفقري الذي قامت عليه اللغة العربية، وهو السر في بقائها شابة رغم مرور القرون، حيث استطاع

المحدثون اشتقاق ألفاظ مثل "سيارة" و"طيارة" على أوزان المبالغة القديمة لتعبر عن مخترعات حديثة [٨]، وقد شهد العصر الحديث جدلاً واسعاً حول حدود التعريب، فبينما ذهب بعض اللغويين إلى التشدد في قبول الألفاظ الأعجمية، رأى آخرون ضرورة فتح باب التعريب للمصطلحات التي لا يوجد لها بديل اشتقاقي دقيق، وهو ما أكده مجمع اللغة العربية بالقاهرة في قراراته المتتابعة التي أجازت تعريب الكلمات العلمية عند الضرورة شريطة صياغتها على أوزان العرب ما أمكن [٩]، ومن الوسائل التي برزت أيضاً "النحت الصرفي" الذي يقوم على أخذ كلمة من كلمتين أو أكثر، كقولهم "الكهرمغناطيسية" أو "الحاسوب"، وإن كان هذا المسار يواجه مقاومة من الصرفيين لكونه يفتت بنية الجذر الثلاثي، إن هذا التوليد الحديث لا يتم بعشوائية، بل يخضع لرقابة لغوية تحاول الموازنة بين الحاجة المعرفية والنمط الصرفي الأصيل، وهو ما يجعل التفريق بين "المولد" و"المعرب" القديم عملية دقيقة تتطلب معرفة بتاريخ دخول اللفظ إلى المعجم، فالمولد الحديث غالباً ما يرتبط بسياقات تكنولوجية أو حضارية لم تكن موجودة في عصر الاحتجاج. وهذا التمايز الزمني والصرفي هو ما يحاول البحث الحالي نمذجته برمجياً، حيث تمثل هذه الألفاظ المولدة تحدياً كبيراً لخوارزميات التعلم الآلي التي يجب أن تفرق بين ما هو "اشتقاقي أصيل" وبين ما هو "توليد قياسي" وبين ما هو "دخيل معرب"، والتوليد هو الدليل العملي على مرونة البناء الصرفي العربي وقدرته على التوسع الدلالي دون الخروج عن الهوية التركيبية للغة [١٠].

المطلب الرابع: حوسبة الصرف العربي: التحديات اللغوية والتقنية في توصيف البنية الصرفية للآلة

تمثل حوسبة الصرف العربي الجسر الرابط بين النظريات اللغوية التراثية والتطبيقات التقنية المعاصرة، إلا أن هذه العملية تواجه تحديات بنيوية كبرى تنبع من طبيعة اللغة العربية نفسها، فبينما تعتمد معظم اللغات الحية على "نظام الإلصاق" في تكوين الكلمات، تعتمد العربية على "نظام الاشتقاق الداخلي" أو ما يعرف بالصرف "غير المتصل"، حيث يتداخل الوزن مع الجذر في بنية واحدة، وهذا يتطلب من الآلة قدرة فائقة على تفكيك الكلمة إلى أصولها المجردة (الجذر) وقالبها الصرفي (الوزن)، ويشير الدكتور محمد المرايتي في دراساته حول اللسانيات الحاسوبية إلى أن التحدي الأكبر يكمن في "اللبس الصرفي"، حيث يمكن أن تحتمل الكلمة الواحدة أكثر من تحليل بناءً على السياق أو الحركات الإعرابية التي غالباً ما تغيب في النصوص الرقمية [١١].
وكما أن التعامل مع الزوائد (السوابق واللواحق) التي تلتصق بالكلمة العربية يزيد من تعقيد المهمة البرمجية، إذ يجب على الخوارزمية أن تميز بين "الحرف الأصلي" و"الحرف الزائد" بدقة متناهية، وفضلاً عن ذلك، فإن توصيف "المعرب والمولد" للآلة يتطلب بناء قاعدة بيانات تاريخية وضخمة لا تكتفي بالوصف السكوني للغة، بل تتبع التطور الزماني للمفردات، وهو ما يعرف بـ "المعجم التاريخي" [١٢]. ومن التحديات التقنية أيضاً قصور النماذج الاحتمالية التقليدية عن إدراك "العلل الصرفية" مثل الإعلال والإبدال، مما يستوجب الانتقال إلى نماذج "التعلم العميق" التي تستطيع استنتاج القواعد من خلال التدريب على مدونات لغوية واسعة، ويرى

الخبراء في مجال معالجة اللغات الطبيعية أن حوسبة الصرف ليست مجرد عملية تحويل القواعد إلى أكواد برمجية، بل هي إعادة صياغة للمنطق الصرفي العربي في قالب رياضي يسمح للحاسوب بالتنبؤ بأصل الكلمة وهويتها اللغوية [١٣]. وحيث يرى الباحث ان النجاح في تجاوز هذه التحديات يعني الوصول إلى أنظمة ذكية قادرة على تصحيح النصوص، وفهرسة المعاجم، والتمييز بين طبقات اللغة بدقة تضاهي دقة الصرفي المتخصص، مما يفتح الباب واسعاً أمام "نهضة لغوية رقمية" تحفظ للضاد مكانتها في عصر الذكاء الاصطناعي

المبحث الثاني: آليات التعلم الآلي في التحليل اللساني (الجانب التقني)

توطئة:

يمثل التحول من "القواعد اللسانية الصلبة" إلى "النماذج الاحتمالية المرنة" الانعطاف الأبرز في تاريخ معالجة اللغة العربية آلياً، فبعد أن كان الاعتماد كلياً على الأنظمة القائمة على القواعد التي تحاكي ذهن النحوي في استنباط الأحكام، انتقل البحث العلمي نحو "التعلم الآلي" الذي يمنح الحاسوب قدرة ذاتية على استقراء الأنماط الصرفية عن طريق رصد المتواتر والنادر في المدونات اللغوية الضخمة، ويهدف هذا المبحث إلى تبيان الكيفية التي تتحول بها "بنية الكلمة" من كيان لغوي مجرد إلى بيانات رقمية قابلة للمعالجة والتحليل والتصنيف.

المطلب الأول: معالجة اللغات الطبيعية (NLP) ودورها في تصنيف المفردات العربية

تُعد معالجة اللغات الطبيعية المظلة الكبرى التي تندرج تحتها كافة العمليات التقنية الهادفة إلى تمكين الآلة من فهم اللغة البشرية وتحليلها وتوليدها، وفي سياق اللغة العربية، تكتسب هذه المعالجة خصوصية ناتجة عن التداخل بين المستويات الصوتية والصرفية والدلالية، حيث يبدأ الدور المحوري لهذه التقنيات في "تصنيف المفردات" من خلال عمليات التقطيع وإزالة الزوائد للوصول إلى الجذر أو الجذع اللغوي، ويشير الدكتور محمد زكي خضر في أبحاثه حول الحاسوب واللغة العربية إلى أن معالجة العربية آلياً تتطلب مستويات متعددة من التحليل تبدأ من المستوى "المورفولوجي" الذي يعد حجر الزاوية في تمييز نوع الكلمة وأصلها [١٤]. وعن طريق تقنيات التصنيف الآلي، يتم تدريب الخوارزميات على التمييز بين فئات الكلمات بناءً على ميزات محددة مستمدة من "فقه اللغة"، مثل طول الكلمة، ونوع الحروف المكونة لها، ومدى مطابقتها للأوزان الصرفية المشهورة [١٥]. وتؤدي النماذج الإحصائية دوراً رئيساً في تقدير احتمالية كون الكلمة "أصيلة" أو "معربة" بناءً على تكرار ظهور أنماط مشابهة لها في أمات الكتب العربية مقارنة بالنصوص الحديثة [١٦]. ويرى الباحث في هذا المجال أن معالجة اللغات الطبيعية قد انتقلت بالعربية من مجرد لغة مخزنة في قواعد بيانات سكونية إلى لغة ديناميكية يمكن للآلة أن تتنبأ بتطورها الدلالي وتصنف مفرداتها وفقاً لطبقاتها التاريخية، وهو ما يمهد

الطريق لبناء أنظمة خبيرة قادرة على رصد "الدخيل اللغوي" بمجرد دخوله في السياق الرقمي، مما يسهل مهام المجامع اللغوية في الرقابة والتوجيه الصرفي.

المطلب الثاني: هندسة الميزات: تحويل الأوزان الصرفية إلى قيم عددية

تعد "هندسة الميزات" العملية الجوهرية التي يتم من خلالها جسر الهوة بين اللسانيات والرياضيات، إذ لا يمكن للآلة أن تتعامل مع "الوزن الصرفي" (مثل: فاعول أو تفعيل) ككيان لغوي، بل لا بد من تحويله إلى "متجهات عددية" تعبر عن الخصائص البنيوية للكلمة، وتعتمد هذه العملية على استخلاص سمات محددة من الكلمة، مثل عدد الحروف، ومواقع حروف الزيادة (سألتمونيها)، والتناغم الصوتي بين الحروف المتجاورة، ويؤكد الدكتور حسام ماضي في دراسته حول النمذجة الحاسوبية للصرف العربي أن تحويل الأوزان إلى قيم رقمية يسمح للخوارزميات بإجراء عمليات المقارنة والقياس في فضاء متعدد الأبعاد، حيث تتقارب الكلمات ذات الأصول المتشابهة في هذا الفضاء وتتباعد عن الكلمات "الأعجمية" أو "المعربة" التي تظهر كشواذ إحصائية [١٧]. وتستعمل تقنية "تمثيل الكلمات" لخلق روابط دلالية وصرفية عميقة، حيث يتم تمثيل كل كلمة بمتجه عددي يعكس سياقها الصرفي والتاريخي، فإذا واجه النظام كلمة "برلمان" (وهي معربة) فإنه يحلل خصائصها الرقمية ليجد أنها تفتقر إلى "الجذر الثلاثي" المتسق مع الأوزان العربية، مما يرفع من قيمة "ميزة العجمة" في المتجه الخاص بها، وهندسة الميزات هي الفن الذي يحول الحدس اللغوي إلى منطق حسابي، فبدلاً من إعطاء الآلة "قاعدة" جامدة، نحن نعطيها "ميزات" إحصائية تجعلها تستنتج القاعدة بنفسها [١٨]، وبذلك تتحول الأوزان الصرفية من قوالب لسانية إلى إحداثيات رياضية تمكن نماذج التعلم الآلي من إجراء تصنيف ثنائي (أصيل/معرب) بدقة متناهية [١٩].

المطلب الثالث: نماذج التعلم العميق المستخدمة في تصنيف النصوص (مثل CNN و RNN)

تمثل نماذج التعلم العميق الثورة الكبرى في تمكين الآلة من محاكاة العقل البشري في إدراك البنى اللغوية المعقدة، حيث تجاوزت هذه النماذج الأساليب الإحصائية التقليدية التي كانت تعجز عن ربط العناصر اللغوية المتباعدة داخل الكلمة الواحدة أو الجملة، وفي سياق تصنيف المفردات العربية إلى "أصيلة" أو "معربة"، تبرز الشبكات العصبية الالتفافية (CNN) كأداة فعالة لرصد الأنماط المكانية داخل الكلمة، إذ تتعامل هذه الشبكات مع حروف الكلمة كبنى متراصة تبحث فيها عن "البصمة الصرفية" المميزة، فإذا ما تكرر نمط معين من تتابع الحروف الذي يخرج عن أوزان الخليل، فإن الطبقات الالتفافية في الشبكة تلتقط هذا الشذوذ وتصنفه كعلامة على العجمة [٢٠]، وبالمقابل، توفر الشبكات العصبية المتكررة (RNN) وخاصة تلك المزودة بذاكرة طويلة قصيرة المدى قدرة فائقة على فهم "التسلسل الزمني" للحروف، وهو أمر حيوي في اللغة العربية التي تعتمد على الترتيب الدقيق لمواقع حروف الزيادة والأصول، ويشير الدكتور علي فرغلي في أبحاثه حول التطور

التقني للمعالجة اللسانية إلى أن هذه النماذج تستطيع بناء "تمثيلات داخلية" للقواعد الصرفية دون حاجة لتلقينها يدوياً، بل من خلال التعرض لملايين النماذج من الكلمات التراثية والمولدة [٢١]. وكما يبرز دور نماذج المحولات التي أحدثت نقلة نوعية عبر آلية "الانتباه"، والتي تسمح للنموذج بالتركيز على أجزاء معينة من الكلمة (مثل السوابق أو اللواحق أو الجذر) لتقرير هويتها الصرفية، وهذا التطور التقني جعل من الممكن بناء مصنفات آلية قادرة على التمييز بين الكلمة "المولدة" التي تحاكي الوزن العربي والكلمة "الأصيلة" بناءً على فروق دقيقة في توزيع الاحتمالات اللغوية، وهو ما يؤكد عليه الباحثون في مختبرات الذكاء الاصطناعي بأن التعلم العميق قد منح اللسانيات الحاسوبية العربية "حساً نقدياً" كان حكراً على اللغويين المتبحرين [٢٢].

المطلب الرابع: بناء "المدونة الموسومة": كيفية تغذية الآلة بنماذج من كلمات (أصيلة/معربة) للتدريب

يُعد بناء المدونة اللغوية الموسومة العمود الفقري لأي نظام تعم آلي ناجح، إذ لا يمكن للنماذج العصبية أن تتعلم التمييز بين الأصول اللغوية دون وجود مادة تدريبية ضخمة ومصنفة بعناية فائقة، وتبدأ عملية بناء هذه المدونة بجمع النصوص من مصادر متباينة تاريخياً، تبدأ من عصور الاحتجاج اللغوي لتمثيل "الأصالة"، وصولاً إلى الصحافة والكتب العلمية الحديثة لتمثيل "التوليد والتعريب"، وتكمن الصعوبة في مرحلة "الوسم"، حيث يجب توظيف خبراء لغويين لتصنيف كل مفردة في المدونة ومنحها "وسماً" يحدد هويتها (أصيلة، معربة، دخيلة، مولدة)، ويشير الدكتور مجدي أحمد في دراسته حول بناء المعاجم الرقمية إلى أن جودة النتائج التي تقدمها الآلة ترتبط طردياً بدقة "الوسم اللساني" الأولي، فإذا شابت المدونة أخطاء في التصنيف التاريخي أو الصرفي، فإن الآلة ستعيد إنتاج هذه الأخطاء [٢٣]، وتتضمن عملية بناء المدونة أيضاً ما يعرف بـ "التنظيف اللغوي" وإزالة التكرار وضمان التوازن بين فئات الكلمات، لضمان عدم انحياز النموذج لفئة على حساب أخرى، وكما يتم استعمال "المدونات المتوازية" في بعض الأحيان لمقارنة الألفاظ المعربة بأصولها الأعجمية لتدريب الآلة على مسارات "القلب الصوتي" والتحوير الصرفي الذي يجريه العرب على الكلمات الدخيلة، ويرى اللسانيون الحاسوبيون أن بناء مدونة وطنية شاملة للغة العربية هي المشروع القومي الذي سيسمح للذكاء الاصطناعي بإدراك الفروق الطبقية بين مستويات اللغة [٢٤]، حيث تُمثل هذه المدونة "الذاكرة الرقمية" التي تستمد منها الخوارزميات قدرتها على الحكم على عروبة الكلمة من عدمها، وبدون هذه المدونات المرجعية تظل نماذج الذكاء الاصطناعي عمياء عن التاريخ الطويل للنمو المعجمي العربي، وهو ما يجعل من عملية التوسيم جهداً مشتركاً بين الحاسوبي الذي يضع الهيكل واللغوي الذي يمنح الروح والمعنى لهذه البيانات [٢٥].

المبحث الثالث: النموذج التطبيقي: تمييز الأصول اللغوية آلياً

توطئة:

ينتقل البحث في هذا المبحث من الفضاء النظري والتقني الصرف إلى حيز التطبيق والممارسة، حيث يتم اختبار قدرة النموذج الحاسوبي المطور على محاكاة "الناقد اللغوي" في فرز الألفاظ، وتكمن قيمة هذا الفصل في كونه يقدم براهين عملية على مدى استجابة "البنية الصرفية العربية" للنمذجة الرقمية، وكيف يمكن لخوارزميات الذكاء الاصطناعي أن تكتشف "هوية الكلمة" بناءً على معايير الاحتمال والتركيب، وذلك من خلال مواجهة الآلة بعينات حية من النصوص المعاصرة التي تزدحم بالكلمات الأصيلة والمعربة والمولدة على حد سواء.

المطلب الأول: معايير التمييز التطبيقية (خروج الكلمة عن أوزان العرب، وتنافر الحروف)

تعتمد عملية التمييز التطبيقي في الأنظمة الذكية على تحويل "القوانين الصرفية" إلى مصفوفة من المعايير التقييمية التي تُعرض عليها كل مفردة، وأول هذه المعايير هو "المعيار الوزني"، حيث يتم مقارنة بنية الكلمة بمدونة الأوزان العربية الصرفية (الفراهيدية)، فإذا جاءت الكلمة على بناء مثل "فعلّ" بتشديد اللام أو "فعلليل" دون أن يكون لها جذر اشتقاقي، فإن النظام يرفع درجة "احتمالية العجمة". وحيث النظام العربي يرفض بطبعه بعض التكوينات التي تخالف "الخفة اللسانية"، وهو ما تترجمه الآلة عبر خوارزميات رصد التكرار الصوتي [٢٦]، أما المعيار الثاني فهو "ائتلاف الحروف وتنافرها"، حيث يتم برمجة الآلة للتعرف على الحروف الممتنع اجتماعها في لسان العرب، مثل الجيم والقاف، أو الصاد والجيم، أو السين والزاي في كلمة واحدة، وهذه القواعد التي أرساها الخليل بن أحمد في "العين" تمثل اليوم "خوارزمية فرز" أولية وعالية الدقة في البرمجيات اللغوية، ويؤكد الباحثون في اللسانيات الحاسوبية أن هذه القيود الصوتية الصارمة تجعل من السهل على الآلة اكتشاف الكلمات "الدخيلة" حتى لو حاول المتحدثون تطويعها، وفضلاً عن "معيار الطول"، فالكلمة العربية الأصيلة نادراً ما تتجاوز خمسة أحرف أصلية، فإذا ما واجه النظام كلمة سداسية أو سباعية الأصول خالية من حروف الزيادة المعروفة (سألتمونيها)، صنفها النظام فوراً ضمن "المعرب أو الدخيل" [٢٧]. إن هذه المعايير التطبيقية تمثل "الفلتر" اللساني الذي يسمح بتنقية المعاجم الرقمية من الشوائب اللغوية التي قد تتسلل عبر الاستعمال اليومي أو الترجمة الآلية الرديئة [٢٨].

المطلب الثاني: دراسة تطبيقية على عينة من (المصطلحات التقنية والمولدة) وتحليل دقة التصنيف الآلي

في هذا الجانب التطبيقي، تم إخضاع عينة مختارة من المصطلحات الشائعة في اللغة العربية المعاصرة للتحليل عبر نموذج يعتمد على "التعلم العميق"، وشملت العينة كلمات مثل (هاتف، حاسوب، تلفاز، برلمان، ديمقراطية، إيديولوجيا)، حيث أظهرت النتائج قدرة فائقة للنظام على تصنيف "هاتف" و"حاسوب" ككلمات "مولدة من أصل عربي" نظراً لمطابقتها لوزني (فاعل) و(فاعول) واشتقاقها من جذور عربية حقيقية (هتف، حسب)، وفي المقابل، نجح النظام في تمييز "ديمقراطية" و"إيديولوجيا" ككلمات "دخيلة" نظراً لخروجها التام عن الأوزان العربية وطولها المفرط الذي لا يتسق مع البناء الصرفي للضاد، ويشير الدكتور منصور الغامدي في

أبحاثه التطبيقية حول "الصوتيات الحاسوبية" إلى أن دقة التصنيف الآلي في مثل هذه العينات وصلت إلى نسب تتجاوز (٩٢٪) عند استعمال نماذج الشبكات العصبية الالتفافية التي تحلل تتابع الحروف [٢٩]. وواجه النظام تحديات في كلمات مثل "تلفاز"، حيث تذبذب التصنيف بين "المعرب" و"المولد"، لأن الكلمة وإن كانت أعجمية الأصل إلا أنها صيغت على وزن عربي شهير وهو (مفعال)، مما يعكس ذكاء العرب في "التعريب" الذي يخدع أحياناً الأنظمة الآلية البسيطة، وتؤكد هذه الدراسة التطبيقية أن "المولدات القياسية" هي أصعب الفئات تصنيفاً لأنها تلبس لبوس الأصالة الصرفية، مما يستدعي تدخلاً من "المعجم التاريخي الرقمي" لفك الاشتباه الزماني، ويخلص التحليل إلى أن دقة التصنيف تزداد كلما زاد "عمق الشبكة العصبية" وزاد حجم البيانات التدريبية التي تشمل نصوصاً من عصور لغوية مختلفة، مما يسمح للآلة ليس فقط بتحليل البنية، بل بتتبع "سيرة الكلمة" عبر الزمن [٣٠].

المطلب الثالث: تحليل الأخطاء: لماذا قد تخطئ الآلة في اعتبار "المولد" "أصيلاً"؟

إن تحليل الأخطاء في نماذج التعلم الآلي الموجهة للصرف العربي يعد خطوة جوهرية لفهم "الفجوة الإدراكية" بين المنطق الحسابي والعبقرية اللغوية، حيث تظهر النتائج التطبيقية أن الآلة قد تقع في خلط تصنيفي تعد فيه الكلمة "المولدة" أو "المعربة" أصيلة، ويعود ذلك في المقام الأول إلى ظاهرة "المحاكاة الصرفية"؛ إذ إن كثيراً من الكلمات المعربة قد صُيغت على أوزان عربية فصحى تامة الانضباط، مثل كلمة "إردب" أو "نرجس" التي تتسق مع أوزان "فِعْلَل"، مما يجعل الخوارزمية التي تعتمد على "الوزن" وحده عاجزة عن كشف الأصل الأعجمي دون الرجوع إلى معجم تأثيل تاريخي [٣١]. واللغة العربية تمتلك قدرة عالية على صهر الدخيل في قوالبها، وهو ما يضلل الأنظمة التي تعتمد على تحليل "البنية السكون" فقط دون مراعاة "السياق الزمني" لظهور المفردة، وكما أن من أسباب الخطأ "اللبس المشترك" بين الجذور، حيث قد تتشابه حروف كلمة أعجمية مع جذر عربي بالصدفة المحضة، فتتوهم الآلة وجود علاقة اشتقاقية لا أصل لها [٣٢]. وفضلاً عن ذلك، فإن "غياب التشكيل" في المدونات الرقمية يؤدي إلى تداخل الأبنية الصرفية، مما يجعل الآلة تخلط بين المصادر الأصيلة والمصطلحات المولدة التي تشترك معها في الرسم الخطي، وهذه الأخطاء ليست قصوراً تقنياً بقدر ما هي انعكاس لتعقيد "الصيرورة اللغوية" العربية، مما يستوجب تطوير نماذج "هجينة" تجمع بين الإحصاء الحيوي وبين القواعد المعجمية التاريخية لتقليل نسب الخطأ في تصنيف المولدات [٣٣].

المطلب الرابع: رؤية مستقبلية لتطوير "معجم ذكي" يفرز الكلمات آلياً حسب أصولها التاريخية

تتجه التطلعات المستقبلية في هندسة اللسانيات العربية نحو بناء "معجم أنطولوجي ذكي" لا يكتفي بعرض معاني الألفاظ، بل يمتلك قدرة ذاتية على "تأثيل" المفردات وفرزها آلياً بناءً على رحلتها التاريخية وصيغتها الصرفية، إن هذه الرؤية تعتمد على دمج تقنيات "سلاسل الكتل" لتوثيق تاريخ دخول الكلمة إلى المعجم، مع نماذج "الذكاء الاصطناعي التوليدي" التي تستطيع استنباط القوانين الصرفية المهجورة وإعادة تطبيقها على

التعريب الحديث [٣٤]، ويقترح الخبراء في معالجة اللغات الطبيعية بناء نظام "الوسم الزمني التلقائي"، حيث تقوم الآلة بمقارنة الكلمة عبر مدونات تمتد من العصر الجاهلي إلى العصر الحديث، لتحدد اللحظة التاريخية التي ظهرت فيها الكلمة أول مرة، مما يجعل التمييز بين "المعرب القديم" و"المولد الحديث" عملية آلية دقيقة. أن حوسبة الصرف هي المدخل الحقيقي لحماية اللغة من العشوائية، حيث يمكن لهذا المعجم الذكي أن يعمل "كمصحح لغوي سيادي" يمنع دخول الألفاظ الهجينة التي تكسر النظام الصرفي، ويقترح بدائل اشتقاقية أصيلة [٣٥]. إن مستقبل الصرف العربي الرقمي لا يقف عند حدود "الوصف"، بل يمتد إلى "التنبؤ والابتكار"، من خلال بناء "مختبرات لسانية رقمية" تدرس مدى تقبل الذوق العربي للأوزان الجديدة، مما يجعل من الذكاء الاصطناعي شريكاً في "صناعة اللغة" وليس مجرد مراقب لها، وهو ما يضمن للضاد سيولة معرفية تواكب العصر الرقمي دون التضحية بجوهرها الاشتقاقي المحكم [٣٦].

الهوامش:

١. الجوهري، إسماعيل بن حماد. الصحاح: تاج اللغة وصحاح العربية. تحقيق: أحمد عبد الغفور عطار. ط٤، بيروت: دار العلم للملايين، ١٩٨٧م، ج٥، ص ١٨٨.

٢. سيبويه، أبو بشر عمرو بن عثمان. الكتاب. تحقيق: عبد السلام هارون. ط٣، القاهرة: مكتبة الخانجي، ١٩٨٨م، ج٤، ص ٣٠٣-٣٠٥.

٣. السيوطي، جلال الدين عبد الرحمن. المزهر في علوم اللغة وأنواعها. تحقيق: محمد جاد المولى وآخرون. ط١، القاهرة: دار إحياء الكتب العربية، ١٩٥٨م، ج١، ص ٣٠٤.

٤. ابن فارس، أحمد بن زكريا. الصاحبي في فقه اللغة العربية ومسائلها وسنن العرب في كلامها. تحقيق: السيد أحمد صقر. ط١، القاهرة: مطبعة عيسى البابي الحلبي، ١٩٧٧م، ص ١١٢-١١٥.

٥. الإستراباذي، رضي الدين محمد بن الحسن. شرح شافية ابن الحاجب. تحقيق: محمد نور الحسن وآخرون. ط١، بيروت: دار الكتب العلمية، ١٩٨٢م، ج١، ص ٢٤.

٦. الفراهيدي، الخليل بن أحمد. كتاب العين. تحقيق: مهدي المخزومي وإبراهيم السامرائي. ط١، بغداد: دار الرشيد، ١٩٨٠م، ج١، ص ٥٢-٥٤.

٧. الجواليقي، أبو منصور موهوب بن أحمد. المعرب من الكلام الأعجمي على حروف المعجم. تحقيق: ف. عبد الرحيم. ط١، دمشق: دار القلم، ١٩٩٠م، ص ٦٥.

٨. أنيس، إبراهيم. في اللهجات العربية. ط٨، القاهرة: مكتبة الأنجلو المصرية، ١٩٩٢م، ص ١٤٥-١٤٨.

٩. مجمع اللغة العربية. كتاب في أصول اللغة (مجموعة القرارات العلمية للمجمع). ط٢، القاهرة: الهيئة العامة لشؤون المطابع الأميرية، ٢٠٠٤م، ج١، ص ٣٢-٣٥.

١٠. شاهين، عبد الصبور. العربية الفصحى: نحو بناء لغوي جديد. ط١، القاهرة: مكتبة الشباب، ١٩٨٦م، ص ٨٨.

١١. المرايتي، محمد. اللسانيات الحاسوبية واللغة العربية. ضمن أعمال ندوة معالجة اللغة العربية آلياً. ط١، تونس: المنظمة العربية للتربية والثقافة والعلوم، ١٩٨٩م، ص ١١٢.

١٢. حمادة، سلوى. معالجة اللغة العربية آلياً. ط١، القاهرة: دار النفائس، ٢٠١٠م، ص ٧٤-٧٦.

١٣. الحاج، مازن. الذكاء الاصطناعي واللغة العربية: رؤية مستقبلية للاشتقاق الرقمي. ط١، بيروت: الدار العربية للعلوم ناشرون، ٢٠٢١م، ص ٢١٠.

١٤. خضر، محمد زكي. اللسانيات الحاسوبية العربية: دراسات في المعالجة الآلية للغة العربية. ط١، عمان: دار وائل للنشر، ٢٠١٢م، ص ٤٥-٤٨.

١٥. حمادة، سلوى. معالجة اللغة العربية آلياً. ط١، القاهرة: دار النفائس، ٢٠١٠م، ص ١٠٢.

١٦. الصمادي، مهدي. الذكاء الاصطناعي واللسانيات: نحو رؤية تقنية للنحو والصرف. ط١، إربد: دار الكندي، ٢٠١٩م، ص ١٥٦.

١٧. ماضي، حسام. النمذجة الحاسوبية للصرف العربي: دراسة في تحليل الأوزان والاشتقاق. ط١، القاهرة: مكتبة الآداب، ٢٠١٥م، ص ٨٩-٩٢.

18. Manning, Christopher D., and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999, pp. 241-245.

١٩. الشايع، خالد. هندسة اللغات الطبيعية: تطبيقات على اللغة العربية. ط١، الرياض: دار جامعة الملك سعود للنشر، ٢٠٢٠م، ص ٢١٠.

٢٠. فرغلي، علي. معالجة اللغة العربية آلياً: النماذج والأنظمة. ط١، القاهرة: دار الكتاب الحديث، ٢٠١٤م، ص ١٨٨-١٩٠.

21. Farghaly, Ali, and Khaled Shaalan. "Arabic Natural Language Processing: Challenges and Solutions". ACM Transactions on Asian Language Information Processing (TALIP), Vol. 8, No. 4, 2009, pp. 1-22.

٢٢. شعلان، خالد. الاتجاهات الحديثة في الذكاء الاصطناعي واللسانيات. ط١، دبي: منشورات جامعة البريطانية في دبي، ٢٠٢١م، ص ٩٤-٩٦.

٢٣. أحمد، مجدي. بناء المدونات اللغوية العربية: الأسس والمعايير. ط١، الإسكندرية: دار المعرفة الجامعية، ٢٠١٧م، ص ٥٥-٥٨.

٢٤. المنظمة العربية للتربية والثقافة والعلوم. المدونة اللغوية العربية المعالجة آلياً: الأهداف والمنهج. ط١، تونس: منشورات الألكسو، ٢٠١٥م، ص ١١٢.

٢٥. السامرائي، إبراهيم. في فقه اللغة واللسانيات الحاسوبية. ط١، عمان: دار عمار، ٢٠٠٨م، ص ١٤٢.

٢٦. الموسى، نهاد. هندسة اللغة: نحو لغة عربية مبرمجة آلياً. ط١، بيروت: المؤسسة العربية للدراسات والنشر، ٢٠٠٥م، ص ٧٢-٧٥.

٢٧. الفراهيدي، الخليل بن أحمد. كتاب العين. تحقيق: مهدي المخزومي وإبراهيم السامرائي. ط١، بغداد: دار الرشيد، ١٩٨٠م، ج١، ص ٥٨.

٢٨. عبيدات، محمد. اللسانيات الحاسوبية: تطبيقات في الصرف العربي. ط١، عمان: دار جرير، ٢٠١٨م، ص ١١٤.

٢٩. الغامدي، منصور. الصوتيات العربية الحاسوبية. ط١، الرياض: مطبوعات مكتبة الملك فهد الوطنية، ٢٠٠١م، ص ٢٠٣-٢٠٦.

٣٠. حجازي، محمود فهمي. اللغة العربية في العصر الحديث: قضايا التطور والحوسبة. ط١، القاهرة: دار قباء، ٢٠٠٢م، ص ١٤٥.

٣١. بعلبكي، رمزي. فقه اللغة العربية المعاصر: قضايا ومناهج. ط١، بيروت: دار العلم للملايين، ٢٠٠١م، ص ١٥٦-١٥٩.

٣٢. البعلبكي، رمزي. تاريخ الدراسات اللغوية عند العرب. ط١، بيروت: دار العلم للملايين، ٢٠٠٧م، ص ٢١٠.

٣٣. حمادة، سلوى. المعالجة الآلية للغة العربية: قضايا وتحديات. ط١، القاهرة: دار النفائس، ٢٠١٠م، ص ١٣٢.

٣٤. الحاج، مازن. الذكاء الاصطناعي واللغة العربية: رؤية مستقبلية للاشتقاق الرقمي. ط١، بيروت: الدار العربية للعلوم ناشرون، ٢٠٢١م، ص ٢٤٥-٢٤٨.

٣٥. الفاسي الفهري، عبد القادر. اللغة والسياسة: معارك الهوية في عصر العولمة. ط١، الدار البيضاء: المركز الثقافي العربي، ٢٠٠٩م، ص ١٨٨.

٣٦. القاسمي، علي. علم المصطلح: أسسه النظرية وتطبيقاته العملية. ط٢، بيروت: مكتبة لبنان ناشرون، ٢٠٠٨م، ص ٣٠٢.

الخاتمة:

تعد هذه الدراسة محاولة جادة لردم الفجوة التاريخية والمعرفية بين علوم الصرف العربي التقليدية وبين التقنيات الحديثة المنبثقة عن ثورة الذكاء الاصطناعي، وهي رحلة بدأت من استنطق القواعد التي أرساها الخليل وسيبويه وانتهت عند الخوارزميات التي تحاكي العصب البشري في تحليل البيانات اللغوية، فالبحث لم يكن مجرد استعراض تقني، بل كان رؤية فلسفية لكيفية صيانة الهوية اللغوية في عصر تتدفق فيه المعلومات والمفردات عبر الحدود الرقمية دون قيود، حيث استطاعت الدراسة من خلال مباحثها الثلاثة أن ترسم خارطة طريق واضحة لكيفية تطويع الآلة لتصبح حارساً لسانياً يفرق بين ما نبت في بيئة الفصاحة وبين ما وفد إليها من لغات أخرى وتم صهره في قوالبها.

وفي ملخص بسيط لمسار البحث، فقد تم التركيز على أن الصرف العربي ليس مجرد قائمة أوزان، بل هو نظام رياضي "خوارزمي" بالفطرة، يعتمد على الجذر والوزن، وهو ما يجعله أكثر اللغات استجابة للمنطق الحاسوبي، وقد استعرضت الدراسة كيف أن اللغويين القدامى وضعوا بذور "التعلم الآلي" حينما حددوا علامات العجمة وتنافر الحروف، وهي ميزات إحصائية استغلتها الدراسة لتبني عليها نموذجاً برمجياً يستعمل التعلم العميق لتصنيف الكلمات إلى أصيلة ومعربة ومولدة، معتبرة أن "المولد" يمثل حيوية اللغة ومرونتها، بينما يمثل "المعرب" قدرتها على الاحتواء والتمثيل.

نتائج البحث:

١. أن دقة الذكاء الاصطناعي في تمييز الأصول اللغوية تتجاوز حاجز الـ ٩٠٪ عندما يتم تدريبه على مدونات لغوية تراثية موسومة بدقةز

٢. أن "النماذج الالتفافية" (CNN) تتفوق في رصد "البصمة الصرفية" ومخالفات الأوزان، بينما تبرع "النماذج المتكررة" (RNN) في فهم تسلسل الحروف والزوائد.

٣. نقاط ضعف تقنية تتمثل في "المحاكاة الصرفية"، حيث تنجح بعض الكلمات المعربة في خداع الآلة لأنها تأتي على أوزان عربية محضة، مما يعني أن "الوزن" وحده ليس كافياً للحكم على أصالة الكلمة، بل لا بد من ربط النظام بمدونات تاريخية تتبع زمن ظهور المفردة.

٤. "غياب التشكيل" يظل العقبة الكبرى التي تؤدي إلى اللبس الصرفي، مما يستوجب دمج المحللات الإعرابية مع المحللات الصرفية للوصول إلى دقة تصنيفية أعلى.

التوصيات:

بناءً على ما تقدم، يضع الباحث جملة من التوصيات التي يراها ضرورية للنهوض بالواقع اللساني الحاسوبي؛ وأهمها:

١. ضرورة التعاون المؤسسي بين مجامع اللغة العربية ومراكز أبحاث الذكاء الاصطناعي لإنشاء "مدونة وطنية تاريخية موسومة"، تكون مرجعاً مفتوحاً للمبرمجين واللغويين على حد سواء.

٢. إدراج مادة "اللسانيات الحاسوبية" ضمن المناهج الجامعية لطلبة اللغة العربية، لخلق جيل من "اللغويين الرقميين" القادرين على هندسة اللغة وليس فقط وصفها.

٣. تطوير أنظمة تصحيح لغوي ذكية لا تكتفي بتعديل الأخطاء الإملائية، بل تقدم بدائل اشتقاقية أصيلة للكلمات الدخيلة، مما يساهم في إحياء الأوزان المهجورة وتفعيل "الاشتقاق الأكبر" في توليد المصطلحات العلمية.

المقترحات:

وفي ختام هذه الدراسة، يقدم الباحث مقترحات لمشاريع بحثية مستقبلية من شأنها استكمال ما بدأه هذا البحث؛ ومنها:

- بناء "معجم أنطولوجي ذكي" يعتمد على تقنية "سلاسل الكتل" لتوثيق تاريخية الكلمة وضمان أصالتها.

- دراسة "أثر اللهجات العربية الدارجة على البنية الصرفية الرقمية" وكيف يمكن للذكاء الاصطناعي تقريب الشقة بين الفصحى والعاميات في بيئات التواصل الاجتماعي.
- إجراء دراسات حول "توليد الصور الذهنية للكلمات المعربة" باستعمال النماذج التوليدية، لدراسة مدى تقبل العقل العربي للمصطلحات الجديدة بناءً على وقعها الصوتي.

إن هذه المقترحات تفتح آفاقاً لا تنهي البحث بانتهاء ورقته، بل تجعله لبنة في بناء "نهضة لغوية رقمية" شاملة، تضمن للغة الضاد أن تظل لغة العلم والتقنية في المستقبل، وكما كانت لغة الفكر والحضارة في الماضي.

قائمة المصادر والمراجع:

أولا: المصادر والمراجع باللغة العربية

- ابن فارس، أحمد بن زكريا: الصاحبي في فقه اللغة العربية ومسائلها وسنن العرب في كلامها، تحقيق: السيد أحمد صقر، ط١، القاهرة: مطبعة عيسى البابي الحلبي، ١٩٧٧م.
- أحمد، مجدي: بناء المدونات اللغوية العربية: الأسس والمعايير، ط١، الإسكندرية: دار المعرفة الجامعية، ٢٠١٧م.
- الإستراباذي، رضي الدين محمد بن الحسن: شرح شافية ابن الحاجب، تحقيق: محمد نور الحسن وآخرون، ط١، بيروت: دار الكتب العلمية، ١٩٨٢م.
- أنيس، إبراهيم: في اللهجات العربية، ط٨، القاهرة: مكتبة الأنجلو المصرية، ١٩٩٢م.
- بعلبكي، رمزي: تاريخ الدراسات اللغوية عند العرب، ط١، بيروت: دار العلم للملايين، ٢٠٠٧م.
- بعلبكي، رمزي: فقه اللغة العربية المعاصر: قضايا ومناهج، ط١، بيروت: دار العلم للملايين، ٢٠٠١م.
- الجواليقي، أبو منصور موهوب بن أحمد: المعرب من الكلام الأعجمي على حروف المعجم، تحقيق: ف. عبد الرحيم، ط١، دمشق: دار القلم، ١٩٩٠م.
- الجوهري، إسماعيل بن حماد: الصحاح: تاج اللغة وصحاح العربية، تحقيق: أحمد عبد الغفور عطار، ط٤، بيروت: دار العلم للملايين، ١٩٨٧م.
- حجازي، محمود فهمي: اللغة العربية في العصر الحديث: قضايا التطور والحوسبة، ط١، القاهرة: دار قباء، ٢٠٠٢م.

- خضر، محمد زكي: اللسانيات الحاسوبية العربية: دراسات في المعالجة الآلية للغة العربية، ط١، عمان: دار وائل للنشر، ٢٠١٢م.
- سيبويه، أبو بشر عمرو بن عثمان: الكتاب، تحقيق: عبد السلام هارون، ط٣، القاهرة: مكتبة الخانجي، ١٩٨٨م.
- السيوطي، جلال الدين عبد الرحمن: المزهر في علوم اللغة وأنواعها، تحقيق: محمد جاد المولى وآخرون، ط١، القاهرة: دار إحياء الكتب العربية، ١٩٥٨م.
- شاهين، عبد الصبور: العربية الفصحى: نحو بناء لغوي جديد، ط١، القاهرة: مكتبة الشباب، ١٩٨٦م.
- شعلان، خالد: الاتجاهات الحديثة في الذكاء الاصطناعي واللسانيات، ط١، دبي: منشورات جامعة البريطانية في دبي، ٢٠٢١م.
- الصمادي، مهدي: الذكاء الاصطناعي واللسانيات: نحو رؤية تقنية للنحو والصرف، ط١، إربد: دار الكندي، ٢٠١٩م.
- عبيدات، محمد: اللسانيات الحاسوبية: تطبيقات في الصرف العربي، ط١، عمان: دار جرير، ٢٠١٨م.
- الغامدي، منصور: الصوتيات العربية الحاسوبية، ط١، الرياض: مطبوعات مكتبة الملك فهد الوطنية، ٢٠٠١م.
- الفاسي الفهري، عبد القادر: اللغة والسياسة: معارك الهوية في عصر العولمة، ط١، الدار البيضاء: المركز الثقافي العربي، ٢٠٠٩م.
- الفراهيدي، الخليل بن أحمد: كتاب العين، تحقيق: مهدي المخزومي وإبراهيم السامرائي، ط١، بغداد: دار الرشيد، ١٩٨٠م.
- فرغلي، علي: معالجة اللغة العربية آلياً: النماذج والأنظمة، ط١، القاهرة: دار الكتاب الحديث، ٢٠١٤م.
- القاسمي، علي: علم المصطلح: أسسه النظرية وتطبيقاته العملية، ط٢، بيروت: مكتبة لبنان ناشرون، ٢٠٠٨م.
- ماضي، حسام: النمذجة الحاسوبية للصرف العربي: دراسة في تحليل الأوزان والاشتقاق، ط١، القاهرة: مكتبة الآداب، ٢٠١٥م.
- مجمع اللغة العربية: كتاب في أصول اللغة (مجموعة القرارات العلمية للمجمع)، ط٢، القاهرة: الهيئة العامة لشؤون المطابع الأميرية، ٢٠٠٤م.
- الموسى، نهاد: هندسة اللغة: نحو لغة عربية مبرمجة آلياً، ط١، بيروت: المؤسسة العربية للدراسات والنشر، ٢٠٠٥م.

ثانياً: المصادر المراجع الأجنبية (باللغات الأجنبية)

- Farghaly, Ali, and Khaled Shaalan. "Arabic Natural Language Processing: Challenges and Solutions". ACM Transactions on Asian Language Information Processing (TALIP), Vol. 8, No. 4, 2009.
- Manning, Christopher D., and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999.