Journal of Al-Farahidi's Arts, No Conference (10 th)N(II)204-218
Doi: https://doi.org/10.25130/jfa.conf.10.2.11

Proceedings of the (10) Conference وقائع المؤتمر العلمي العاشر

P-ISSN : 2074-9554 | E-ISSN: 2663-811
Journal of Al-Farahidi's Arts
available online at : fja.tu.edu.iq/index.php/fja

Tikrit University J.F.A Journal of Al-Frahedis

"The Role of Computational Linguistics in Developing Artificial Intelligence: An Approach Based on Arabic Language Data

Dr. Anmar Dawood Salim

E-Mail: Anmar.D.Saleem@tu.edu.iq

Keywords:
Computational linguistics, artificial intelligence, Arabic language processing, large language models,

Article history:

Received	6/8/2025
Received in revised form	17/9/2025
Accepted	19/10/2025
Available online	9/12/2025

E-mail Jaa@ tu.edu.iq

©THIS AN OPEN ACCESS ARTICLE UNDER THE CCBY LICENSE
http:// creativecommons.org/licenses/by/4.0

A B S T R A C T

This study investigates the pivotal role of computational linguistics in developing artificial intelligence technologies, with a specific focus on the Arabic language, which faces unique challenges due to its complex morphological and grammatical nature. Employing an analytical-descriptive methodology, the research highlights the theoretical framework of both fields and analyzes the obstacles to the computational processing of Arabic, most notably: derivational and morphological characteristics, homonymy, and the severe shortage of annotated linguistic corpora. The study also reviews practical approaches to addressing these challenges, such as building computational morphological analyzers and annotated corpora, which are fundamental for training deep learning models. It concludes by presenting applied models, including machine translation, chatbots, and large language models, emphasizing that the integration of prior linguistic knowledge remains a critical factor in enhancing AI's ability to understand and generate Arabic accurately and fluently. The study ends with a set of recommendations, the most important of which is the necessity of investing in building linguistic resources and developing native

دور اللسانيات الحاسوبية في تطوير الذكاء الاصطناعي: مقاربة في معطيات اللغة العربية

م. د. أنمار داود سليم /جامعة تكريت / كلية التربية الأساسية / قسم اللغة العربية

المستخلص:

تهدف هذه الدراسة إلى استقصاء الدور المحوري للسانيات الحاسوبية في تطوير تقنيات الذكاء الاصطناعي، مع التركيز على اللغة العربية التي تواجه تحديات فريدة بسبب طبيعتها الصرفية والنحوية المعقدة. يعتمد البحث على منهجية تحليلية وصفية لتسليط الضوء على الإطار النظري للمجالين، وتحليل التحديات التي تعترض حوسبة اللغة العربية، أبرزها: الخصائص الاشتقاقية والصرفية، وظاهرة تعدد المعنى، والنقص الحاد في المدونات اللغوية المشروحة. كما يستعرض البحث المقاربات العملية لمعالجة هذه التحديات، مثل بناء المحللات الصرفية الحاسوبية والمدونات المشروحة، والتي تعد أساساً لتدريب نماذج التعلم العميق. وتختتم الدراسة بعرض نماذج تطبيقية، كالترجمة الآلية وروبوتات الدردشة والنماذج اللغوية الكبيرة، مؤكدة على أن دمج المعرفة اللسانية المسبقة يظل عاملاً حاسماً في تعزيز قدرة الذكاء الاصطناعي على فهم وتوليد اللغة العربية بدقة وطلاقة. وتخلص إلى مجموعة من التوصيات أهمها: ضرورة الاستثمار في بناء الموارد اللغوية وتطوير نماذج مُعربة خالصة.

الكلمات المفتاحية : اللسانيات الحاسوبية، الذكاء الاصطناعي، معالجة اللغة العربية، النماذج اللغوية الكبيرة، التحديات اللغوية، المدونات المشروحة.

المقدمة:

تشهد الحضارة الإنسانية منعطفاً تاريخياً مع صعود ثورة الذكاء الاصطناعي، التي أعادت تشكيل علاقة الإنسان بالآلة وأضفت أبعاداً جديدة على مفهوم التفاعل بينهما (ما وسياو، ٢٠١٨، ص ٥). وفي صلب هذه الثورة، تقف اللغة كأحد أهم روافدها وتطبيقاتها، مما يستدعي حواراً جاداً بين حقلين معرفيين: اللسانيات وعلوم الحاسوب. وتبرز اللسانيات الحاسوبية كحقل بيني يجسد هذا الحوار، هدفه سد الفجوة بين الطبيعة السيالة غير المضبوطة للغة البشرية والتمثيلات المنطقية المضبوطة التي تتطلبها الحواسيب (علي، ١٩٨٨، ص ٢٣).

على الرغم من التقدم الهائل في نماذج الذكاء الاصطناعي التوليدية، لا تزال اللغة العربية تواجه ما يمكن وصفه "بفجوة رقمية" تهدد حضورها في الفضاء السيبراني المستقبلي (أبو العسل (2023). 435)وتعزى هذه الفجوة إلى طبيعة اللغة العربية ذات التعقيد الصرفي والنحوي الفريد، وإلى النقص المزمن في الموارد اللغوية الحاسوبية (الموارد) مقارنة بلغات أخرى مثل الإنجليزية لذلك، تطرح الإشكالية المحورية لهذا البحث التساؤل التالي: كيف يمكن للسانيات الحاسوبية أن تسهم في تطوير ذكاء اصطناعي أكثر فاعلية في فهم وتوليد اللغة العربية، وما هي التحديات والمقاربات المرتبطة بذلك؟

انبثقت من هذه الإشكالية جملة من الأهداف يسعى البحث إلى تحقيقها، وهي: تعريف اللسانيات الحاسوبية والذكاء الاصطناعي وإيضاح علاقتهما التكاملية، وتحليل الخصائص اللغوية للعربية التي تشكل عائقاً أمام الحوسبة، واستعراض المقاربات العملية لبناء الموارد والأدوات الحاسوبية اللازمة، وتحليل أداء بعض النماذج التطبيقية، وختماً بوضع تصور استشرافي على شكل نتائج وتوصيات.

لتحقيق هذه الأهداف، اعتمد البحث على المنهج الوصفي التحليلي، من خلال استقراء الأدبيات والدراسات السابقة ذات الصلة، وتحليل مضامينها لتأسيس رؤية متكاملة. وسينتظم البحث في أربعة فصول، يقدم الأول الإطار النظري، بينما يحلل الثاني التحديات اللغوية، ويستعرض الثالث المقاربات الحاسوبية لمعالجتها، ويناقش الرابع نماذج تطبيقية، ليختتم بالاستنتاجات والتوصيات.

المبحث الأول: الإطار النظري والمفاهيمي

المطلب الأول: اللسانيات الحاسوبية: النشأة والمفاهيم

تُعرف اللسانيات الحاسوبية (Computational Linguistics) على أنها ذلك الحقل البيني الذي يجمع بين علم اللسانيات النظرية وعلوم الحاسوب (الجليل، ٢٠٠٢، ص ٥٧). وهي تهدف إلى تحويل النماذج اللغوية البشرية إلى تمثيلات شكلية يسهل للحاسوب معالجتها والتعامل معها، ودراسة توليد اللغة والفهم الآلي لها (ديفلين وآخرون، ٢٠١٨، ص ٤١٧٢). ولا تقتصر علاقتها على علم اللسانيات العام وعلوم الحاسوب فحسب، بل تمتد لتشمل فلسفة اللغة والذكاء الاصطناعي والمنطق والرياضيات

بدايات هذا العلم مع اللغة العربية تعود إلى عام 1971 على يد الدكتور إبراهيم أنيس والدكتور علي حلمي موسى في جامعة الكويت، حيث قدما دراسة إحصائية لمعجم الصحاح للجوهري (العجيلات، ١٩٩٦، ص ١٠٥). ثم جاء كتاب الدكتور المهندس نبيل علي (اللغة العربية والحاسوب) عام 1988 ليشكل اللبنة الأساسية، حيث تناول فيه المعالجة الآلية لمنظومة اللغة العربية بصورة شاملة (علي، ١٩٨٨، ص ٣٤).

المطلب الثاني: الذكاء الاصطناعي ومعالجة اللغة الطبيعية (NLP)

يعد الذكاء الاصطناعي الحقل الأوسع الذي يهدف إلى محاكاة الذكاء البشري في الآلات (ما وسياو، ٢٠١٨، ص ٢). أما معالجة اللغة الطبيعية (NLP) فهي إحدى مجالات الذكاء الاصطناعي المتخصصة في تفعيل التفاعل بين الحاسوب واللغة البشرية (عزيز. (2023). ص 5). وقد مرت هذه المجالات بتحولات كبرى، حيث سيطرت النماذج الإحصائية في بداياتها، ثم النماذج القائمة على الشبكات العصبية المتكررة (RNN)، لتتوج أخيراً ببنية المحوّل (Transformer) في عام 2017 والتي شكلت نقلة نوعية مكنت من ظهور النماذج اللغوية الكبيرة (LLMs) (براون وآخرون، ٢٠٢٠، ص ٢).

ويمثل الفرق بين فهم اللغة الطبيعية (NLU) الذي يركز على استيعاب المعنى والقصد من النص، وبين معالجة اللغة الطبيعية (NLP) الأوسع نطاقاً والذي يشمل أيضاً توليد النص، أحد الفروق الجوهرية في هذا المضمار (محمد. (2023). ص 274)

المطلب الثالث: تقاطع المجالين

يكمن التقاطع الجوهري بين المجالين في أن النماذج الحاسوبية لا تستطيع فهم اللغة أو توليدها دون تمثيلات رقمية للمعرفة اللغوية بمستوياتها المختلفة (عريوة وغيلوس، ٢٠٢٣، ص ٥٣٦). فالمعرفة المعجمية (Lexical) الخاصة بالكلمات وجذورها، والمعرفة النحوية (Syntactic) الخاصة بتركيب الجمل، والمعرفة الدلالية (Semantic) الخاصة بالمعنى، هي جميعاً وقود عمليات التعلم الآلي. فالنماذج مثل جي بي تي (GPT) تعمل على تمثيل الكلمات في فراغ متجهي متعدد الأبعاد، بحيث تكون الكلمات المتشابهة في المعنى قريبة من بعضها في هذا الفراغ، مما يمكن النموذج من فهم العلاقات بين الكلمات والعبارات (سلسلة الذكاء الاصطناعي للتنفيذيين ، (2024) ص 12) وهكذا، فإن جودة المخرجات التوليدية للنماذج تعتمد بشكل أساسي على عمق وفهم هذه التمثيلات اللغوية.

المطلب الرابع: التطور التاريخي للسانيات الحاسوبية والذكاء الاصطناعي في العالم العربي

شهد العالم العربي اهتماماً متزايداً باللسانيات الحاسوبية والذكاء الاصطناعي، وإن جاء متأخراً نسبياً مقارنة بالدول المتقدمة. يُعد الدكتور نبيل علي رائداً في هذا المجال عبر كتابه التأسيسي "اللغة العربية والحاسوب" (1988)، الذي وضع الأسس الأولى للمعالجة الآلية للغة العربية (علي، 1988، ص 15).

في العقد الثاني من القرن الحادي والعشرين، بدأت العديد من الدول العربية في تبني استراتيجيات وطنية للذكاء الاصطناعي. ففي الإمارات العربية المتحدة، تم إنشاء "وزارة الذكاء الاصطناعي" عام 2017 كأول وزارة من نوعها في العالم، مما مثل نقطة تحول في الاهتمام الرسمي بهذا المجال كما أطلقت المملكة العربية السعودية "الاستراتيجية الوطنية للبيانات والذكاء الاصطناعي" عام 2020، والتي خصصت محوراً مهماً لمعالجة اللغة العربية.

ظهرت عدة مراكز بحثية متخصصة، مثل "مركز محمد بن زايد للذكاء الاصطناعي" في الإمارات، و"مركز الملك عبد الله العالمي للأبحاث" في السعودية، حيث أنتجت أبحاثاً رائدة في معالجة اللغة العربية الطبيعية (الشامسي، 2022، ص 78). كما ساهمت الجامعات العربية من خلال إنشاء برامج أكاديمية متخصصة، حيث أطلقت جامعة الملك سعود برنامج الماجستير في اللسانيات الحاسوبية عام 2018.

على صعيد المشاريع التطبيقية، شهد العقد الأخير إطلاق عدة مبادرات مهمة، منها:

· مشروع "مدونة اللغة العربية المعاصرة" الذي طوره مجمع اللغة العربية في القاهرة

· منصة "أرابيك جي بي تي" المفتوحة المصدر التي أطلقتها جامعة محمد بن زايد للذكاء الاصطناعي

· مشروع "المكتبة الرقمية السعودية" التي تضم موارد لغوية ضخمة

يوضح الدكتور خالد الصاوي (2023) أن "العالم العربي يعيش حالياً مرحلة انتقالية من مرحلة الاستهلاك التكنولوجي إلى مرحلة الإنتاج والابتكار في مجال الذكاء الاصطناعي، وإن كانت الفجوة ما زالت كبيرة في مجال الموارد اللغوية (محمد. (2025) ص 365).

رغم هذه التطورات، لا تزال المعالجة الآلية للغة العربية تواجه تحديات جمة، أبرزها ندرة الموارد اللغوية المُعّربة مقارنة باللغات الأخرى، وعدم كفاية التمويل المخصص للبحوث التطبيقية، والحاجة إلى مزيد من التكامل بين المؤسسات الأكاديمية والصناعية في العالم العربي (الغنيم، 2021، ص 67).

المبحث الثاني: التحدي اللغوي: خصائص اللغة العربية ومعوقات حوسبتها

المطلب الأول: الخصائص الصرفية للغة العربية

تتميز العربية بخصائص صرفية فريدة تشكل تحدياً كبيراً للمعالجة الآلية، أبرزها ظاهرة الاشتقاق التي تتشكل فيها مئات الكلمات من جذر واحد، مما يؤدي إلى تضخم كبير في المفردات إذا ما عولجت كل كلمة ككلمة منفصلة (تونسي وحداد، ٢٠١٩، ص ١٢٣٦). ويوضح علي (١٩٨٨، ص ٧٨) أن "اللغة العربية تمتلك نظاماً اشتقاقياً متشعباً يجعل من الصعب على النماذج الإحصائية البحتة الإحاطة بجميع الصيغ الممكنة".

أما التصريف فيشمل الإعراب والبناء، حيث تغير حركات أواخر الكلمات (الضم والفتح والكسر والسكون) حسب موقعها في الجملة ووظيفتها النحوية. وتعد ظاهرة التشكيل (الحركات) عنصراً أساسياً لتحديد المعنى ونطق الكلمات بشكل صحيح، لكن غيابه في النصوص المكتوبة يخلق غموضاً كبيراً يصعب على الآلة حله (أبو السعود والزغل، ٢٠٢٢، ص ١١٢). فعلى سبيل المثال، كلمة "عَلِمَ" (فعل ماض) تختلف عن "عَلَّمْ" (فعل أمر) وعن "عَلَمٌ" (اسم)، مما يزيد من تعقيد عملية الفهم الآلي.

المطلب الثاني: التحديات المعجمية والنحوية

تشكل ظاهرة تعدد المعنى (Homonymy) تحدياً جوهرياً في المعالجة الآلية للعربية، حيث تكثر المفردات التي تحمل أكثر من معنى حسب السياق. فكلمة "عين" قد تعني العضو الباصر، أو نبع الماء، أو جواسيس الحرب، مما يتطلب سياقاً دقيقاً لفك هذا الغموض ،إن غياب أدوات تحديد الجذور الفعالة يمثل عائقاً تقنياً كبيراً، وذلك due to صعوبة فصل البادئات واللواحق عن الجذر الأساسي للكلمة بشكل تلقائي ودقيق، نظراً للطبيعة غير المنفصلة للكتابة العربية (نصيرة، إديـــر. (2014).ص 22 – 32). كما أن تعقيد النحو العربي ومرونته في تركيب الجمل وقدرتها على الاحتفاظ بمعانيها رغم تقدم الفعل على الفاعل أو العكس، يزيد من صعوبة التحليل الآلي مقارنة بلغات أخرى (محمود، ٢٠٢١، ص ٥١٥).

المطلب الثالث: فجوة الموارد (Corpora)

يشكل النقص الحاد في المدونات اللغوية (Corpora) العربية عالية الجودة والشاملة أحد أهم معوقات تطور الذكاء الاصطناعي للعربية. وتتمثل الفجوة في نقص المدونات المشروحة (Annotated Corpora)، أي النصوص المزودة بوسوم صرفية ونحوية ودلالية، والتي تعد ضرورية لتدريب نماذج التعلم العميق (براون وآخرون، ٢٠٢٠، ص ١٨٨٠).

إن معظم المدونات العربية الموجودة إما غير مشروحة أو مشروحة بشكل سطحي، كما تعاني من نقص المدونات الموزونة، أي النصوص المشكولة بالحركات، والتي تساعد في تقليل الغموض (نصيرة، إديـــر. (2014).ص 22) ، وتمتد المشكلة إلى نقص المدونات في اللهجات العربية واللغة الفصحى الحديثة، حيث أن جزءاً كبيراً من التواصل اليومي يجري باللهجات، وهي تفتقر إلى أي مدونات منظمة بشكل كاف (السيد، ٢٠٢٤، ص ٤٥٥).

المطلب الرابع: تحدي الانزياح اللغوي واللهجات العربية وتأثيره على النماذج اللغوية

تشكل ظاهرة الانزياح بين اللغة العربية الفصحى واللهجات المحكية تحدياً فريداً من نوعه في معالجة اللغة العربية، حيث تتعايش مستويات لغوية متعددة في المجتمع العربي الواحد (إسماعيلي. (2025).ص 288 – 289)يخلق هذا الانزياح فجوة حادة بين لغة الكتابة الرسمية ولغة التواصل اليومي، مما يؤثر سلباً على أداء النماذج اللغوية المدربة أساساً على الفصحى.

تتمثل الإشكالية الأساسية في أن النماذج اللغوية الكبيرة (LLMs) تواجه صعوبات جمة في معالجة اللهجات العربية بسبب عدم انتظام قواعدها الصرفية والنحوية، واختلافها الجذري عن الفصحى في العديد من الحالات. فاللهجة المصرية مثلاً تستخدم نظاماً مختلفاً في تصريف الأفعال والضمائر، بينما تعتمد اللهجة الخليجية على مفردات ومترادفات مغايرة تماما، والسبب في ذلك يعود كما يقول تشومسكي : " إننا في الحقيقة لا نتعلم اللغة، وإنما قواعد اللغة هي التي تنمو في عقولنا (مور، (1993) ، ص 18).

ويوضح الباحثون أن الانزياح اللغوي يتجلى في عدة مستويات:

أولاً: الانزياح الصرفي والنحوي

حيث تتبنى اللهجات أنظمة إعرابية مبسطة وتهمل علامات الإعراب في معظم الأحوال. كما تظهر تراكيب نحوية غير موجودة في الفصحى، مثل استخدام "عامل" في اللهجة المصرية للدلالة على الاستمرارية، أو في اللهجة اليمنية الدراجة يختزلون بعض الحروف من الكلمة فتتغير بنيتها لكنها تبقى تدل على معناها الأصلي مثل كلمة يشتي ويقصدون (يشتهي) لكن بإسقاط حرف الهاء (محمد عبد القوي، (2015).ص 465)

ثانياً: الانزياح المعجمي والدلالي

تمتلك كل لهجة عربية مخزوناً معجمياً خاصاً بها، حيث تختلف أسماء الأدوات والأغذية والملابس والأفعال بشكل كبير بين منطقة وأخرى. بل إن بعض الكلمات تحمل دلالات متعارضة بين اللهجات، مما قد يؤدي إلى سوء فهم كارثي في التطبيقات الحساسة مثل التحليل الآلي للشكاوى أو الاستبيانات وهذا ليس عيبًا في اللغة العربية وليست مشكلة رهينة العصر الحالي، بل هي خاصية رافقت اللغة العربية منذ بدء تدوينها حتى أن أول ما أُلِّف في تفسير القرآن الكريم هو اللغات في القرآن ويُقصد مفردات قبائل العرب الفصيحة غير القرشية ، وخير دليل على ذلك قول عثمان للجنة جمع القرآن الكريم المشكلة من علماء الصحابة الأكارم: إذا اختلفتم أنتم وزيد في شيء فاكتبوه بلغة قريش فإنما نزل القران بلغتهم (البخاري ، (1993) ج3 ص 1291 رقم الحديث 3315).

ثالثاً: الانزياح الكتابي والألفبائي

تعاني كتابة اللهجات العربية من عدم وجود نظام كتابي موحد، حيث يستخدم الكتاب أحرفاً مختلفة لتمثيل الأصوات غير الموجودة في الفصحى. فالصوت "چ" (الگاف المعطشة) يكتب

أحياناً بـــ "ج" أو "ك" أو "ق" بشكل عشوائي، مما يزيد من تعقيد عملية المعالجة (د. بشر. (2021). ص 193)

ويشير التقرير الصادر عن مركز الدراسات اللغوية والحاسوبية (٢٠٢٢) إلى أن "النماذج اللغوية المدربة على الفصحى فقط تفشل في فهم أكثر من ٤٠٪ من نصوص التواصل اليومي على وسائل التواصل الاجتماعي، والتي تكتب في غالبيتها باللهجات المحكية" (مركز الدراسات اللغوية والحاسوبية، ٢٠٢٢، ص ١٥).

للتغلب على هذه التحديات، يقترح الباحثون عدة حلول، منها تطوير نماذج هجينة قادرة على التعامل مع المستويات اللغوية المختلفة، وإنشاء مدونات لغوية موازية للهجات، واستخدام تقنيات التعلم الانتقالي (Transfer Learning) لنقل المعرفة من الفصحى إلى اللهجات (كروش، (2025)، ص 157)

المبحث الثالث: المقاربة اللسانية الحاسوبية: بناء الموارد والأدوات

المطلب الأول: دور المحلل الصرفي الحاسوبي (Morphological Analyzer)

يعد المحلل الصرفي حجر الزاوية في معالجة اللغة العربية، فهو الأداة المسؤولة عن تفكيك الكلمة إلى مكوناتها (جذر، بادئة، لاحقة)، وتحديد خصائصها الصرفية (فعل، اسم، etc.)، وصيغتها. تعمل هذه المحللات من خلال دمج قواميس صرفية شاملة تضم الجذور واللواصق، مع قواعد اشتقاقية تحكم عملية تكوين الكلمات (تونسي وحداد، ٢٠١٩، ص ١٢٤٠).

ويسهم استخدام المحلل الصرفي قبل عمليات المعالجة في تقليل حجم المفردات بشكل كبير، إذ يتم التعامل مع آلاف الصيغ المشتقة من جذر واحد على أنها تنتمي لنفس الأصل، مما يحسن كفاءة النماذج ويقلل من تعقيدها الحسابي. وقد أثبتت دراسة العجيلات (١٩٩٦، ص ١١٣) أن استخدام المحلل الصرفي قلل حجم المفردات في نظام معالجة النصوص العربية بنسبة ٧٣٪.

المطلب الثاني: المدونات المشروحة لسانياً (Linguistically Annotated Corpora)

تكمن أهمية هذه المدونات في كونها "بيانات التدريب" التي تتعلم منها نماذج التعلم العميق قواعد اللغة ومفرداتها. عملية إضافة الوسم (Tagging) تتم على عدة مستويات: الوسم الصرفي لتحديد الجذر والوزن الصرفي والخصائص مثل الجنس والعدد، والوسم النحوي لتحديد الدور الإعرابي للكلمة في الجملة، والوسم الدلالي لتحديد العلاقات الدلالية بين الكلمات والكيانات المسماة (ديفلين وآخرون، ٢٠١٨، ص ٤١٧٥).

تستهلك عملية إنشاء هذه المدونات وقتاً وجهداً بشرياً كبيراً، لكنها استثمار ضروري لبناء نماذج لغوية قوية. وقد ساهمت مبادرات مثل "مكنز القرآن" (QURANIC CORPUS) الذي أنشأته جامعة ليدز البريطانية في توفير مورد مشروح لنص قرآني، مما ساعد في تحليل النحو والصرف آلياً بالكامل.

المطلب الثالث: أنظمة فك الغموض والإعراب الآلي

تعتمد هذه الأنظمة على النماذج اللغوية النظرية (كالنحو التوليدي التحويلي) لتحليل الجمل وتحويلها إلى تمثيلات بنيوية. فهي تساعد في تحديد وظائف الكلمات الإعرابية والعلاقات بين أجزاء الجملة والمعنى الإجمالي للجملة (محمود، ٢٠٢١، ص ٥٢٠).

يسهم ذلك بشكل مباشر في تحسين أداء الترجمة الآلية، حيث أن فهم تركيب الجملة المصدر هو الخطوة الأولى نحو الترجمة الصحيحة. كما يعزز من دقة أنظمة البحث، حيث يمكن للنظام فهم استفسار المستخدم بدقة أكبر بغض النظر عن الصيغة التي صيغ بها. ويؤكد عريوة وغيلوس (٢٠٢٣، ص ٥٣٨) أن "دمج المعرفة النحوية في الأنظمة الحاسوبية يقلل من الأخطاء في فهم الجمل الطويلة والمعقدة".

المبحث الرابع: نماذج تطبيقية وتجارب في الذكاء الاصطناعي العربي

المطلب الأول: دور اللسانيات في تحسين أنظمة الترجمة الآلية

لطالما كانت الترجمة الآلية أحد أبرز تطبيقات اللسانيات الحاسوبية. الخوارزميات الحديثة التي تدمج التحليل الصرفي والنحوي في مرحلتي ما قبل المعالجة والترجمة تتفوق على تلك التي تعتمد على الإحصاءات فقط. فعلى سبيل المثال، تحليل كلمة مثل "كاتبنا" إلى جذرها "ك ت ب" وصيغتها (فاعل + ضمير المتكلمين) يسمح للنموذج بفهم أن المعنى متعلق بـــ "الشخص الذي يكتب لنا"، ومن ثم ترجمتها بشكل صحيح إلى "our writer" بدلاً من التعامل معها ككلمة غير معروفة أو ترجمتها حرفياً (علي، ١٩٨٨، ص ١٤٥).

المطلب الثاني: دور اللسانيات في تطوير روبوتات الدردشة العربية (Chatbots)

شهدت العربية تجارب رائدة في هذا المجال، مثل "تشات بوت" الدكتورة سلوى حمادة، والذي يعد أحد أوائل برامج المحادثة باللغة العربية (السليطي والمناي، ٢٠٢١، ص ٦٩٣). يكمن نجاح هذه الروبوتات في دمج القواعد النحوية والدلالية في آلية توليد الردود. فبدلاً من مجرد اختيار ردود جاهزة، تستخدم النماذج اللغوية الكبيرة مثل GPT-4 خوارزميات الانتباه الذاتي

(Attention-Self) لفهم سياق المحادثة وتركيب ردود نحوياً صحيحة ودلالياً متماسكة (براون وآخرون، ٢٠٢٠، ص ١٨٩٠).

المطلب الثالث: تحليل أداء النماذج اللغوية الكبيرة (LLMs) للغة العربية

أظهرت الدراسات التقيمية، مثل تلك التي استخدمت مجموعة البيانات القياسية ArabicMMLU، أن أداء النماذج متعددة اللغات مع العربية غالباً ما يكون أضعف منه في الإنجليزية (كوتو، ٢٠٢٤). كما أن النماذج المطورة خصيصاً للعربية تواجه صعوبة في الأسئلة المتعلقة بالمعرفة والسمات الثقافية العربية. ويعد دمج المعرفة اللسانية المسبقة (مثل القواعد الصرفية والنحوية) في هندسة هذه النماذج أحد السبل لتعزيز قدراتها (ديفلين وآخرون، ٢٠١٨، ص ٤١٨٠).

وقد أظهر نموذج "جيس" (Jais) الذي طورته جامعة محمد بن زايد للذكاء الاصطناعي وشركة Inception أداءً متميزاً كأفضل نموذج مفتوح المصدر في تقييم ArabicMMLU، متفوقاً حتى على GPT-3.5، مما يثبت جدوى التطوير الخاص باللغة العربية (كوتو، ٢٠٢٤).

الخاتمة

سعى هذا البحث إلى تسليط الضوء على الدور المحوري للسانيات الحاسوبية في تطوير ذكاء اصطناعي قادر على فهم وتوليد اللغة العربية بشكل فعال. بدأ البحث بتأسيس إطار نظري يربط بين اللسانيات الحاسوبية والذكاء الاصطناعي، ثم انتقل لتحليل التحديات الفريدة التي تطرحها الخصائص اللغوية للعربية، ليستعرض بعدها المقاربات العملية لبناء الموارد والأدوات الحاسوبية اللازمة لتجاوز هذه التحديات، واختتم بعرض نماذج تطبيقية ونقاش لأداء النماذج اللغوية الكبيرة في التعامل مع العربية.

يمكن التأكيد على أن هذا البحث قد حقق الأهداف التي سعى إليها، حيث قدم تحليلاً شاملاً للعلاقة التكاملية بين اللسانيات الحاسوبية والذكاء الاصطناعي، وقدم تشخيصاً دقيقاً للتحديات اللغوية والتقنية، واستعرض الحلول والمبادرات القائمة، وانتهى بتقديم رؤية مستقبلية قائمة على نتائج مستخلصة من الأدبيات والدراسات الحديثة.

ويمكن إجمال النتائج المستخلصة من البحث بما يأتي

1. اللغة العربية تتطلب مقاييس صرفية خاصة لا تكفي فيها المقاربات الإحصائية البحتة التي تصلح للغات أخرى، نظراً لطابعها الاشتقاقي المعقد

2. تكمن القيمة المضافة للسانيات الحاسوبية في بناء الموارد الأساسية من مدونات وقواميس معربة، والتي تشكل البنية التحتية لأي تطوير مستقبلي

3. التعلم الآلي العميق للغة العربية يحتاج إلى المحللات الصرفية الجاهزة لتقليل حجم المفردات ومعالجة الغموض الناتج عن الاشتقاق والتصريف

4. النقص في المدونات المشروحة دلالياً هو التحدي الأكبر لـــ NLU في العربية، لأنه يحول دون فهم الآلة للعلاقات المعنوية بين الكلمات والسياقات الثقافية

5. دور النحو التوليدي التحويلي لا يزال مهماً في معالجة الجمل الطويلة والغامضة، حيث يوفر إطاراً نظرياً لتحليل البنية العميقة للجملة

6. معالجة اللهجات العربية تمثل تحدياً مضاعفاً مع شبه انعدام المدونات المنظمة لها، رغم أنها لغة التواصل اليومي

7. الاعتماد على البيانات المترجمة من الإنجليزية لتقييم النماذج العربية يقلل الفعالية ويتجاهل الخصوصية الثقافية والسياقية للعالم العربي

8. دمج المعرفة اللسانية المسبقة في تدريب النماذج اللغوية الكبيرة يمكن أن يعزز أداءها في اللغة العربية ويقلل من الأخطاء النحوية والدلالية

9. التدريب المشترك بين اللغويين والمعلوماتيين هو السبيل الأمثل لإنتاج موارد وأدوات دقيقة وعملية، حيث أن عزلة أي من الطرفين تقود إلى حلول ناقصة

10. الذكاء الاصطناعي لم يعد خياراً بل ضرورة لضمان وجود اللغة العربية في المشهد الرقمي المستقبلي، ومن لا يشارك في صنعه سيصبح مستهلكاً فقط)

11. حرية تداول المعلومات والبيانات النصية العربية شرط أساسي للتقدم في هذا المجال، فالمعلومات المقيدة أو غير المتاحة تعيق عملية بناء المدونات

12. معايير التقييم المُعربة (كـــ ArabicMMLU) ضرورية لقياس الأداء الحقيقي للنماذج وتوجيه عملية تطويرها بشكل علمي

13. النماذج مفتوحة المصدر الخاصة بالعربية (مثل جيس) أثبتت قدرة على منافسة النماذج العالمية العملاقة، مما يفتح آفاقاً للاستقلال التكنولوجي

14. الخصوصية الثقافية واللغوية للعالم العربي تتطلب تطوير نماذج ذكاء اصطناعي تتعامل مع السياقات المحلية وليس مجرد تكييف النماذج الغربية

15. الاستثمار في رأس المال البشري المتخصص في تقاطع اللغة والحاسوب هو الضمانة الأهم لاستمرار تطور هذا الحقل ومواكبة متغيراته

التوصيات والمقترحات البحثية المستقبلية

· مشاريع بناء موارد ضخمة: دعم وإنشاء مدونات عربية شاملة ومشروحة على جميع المستويات (صرف، نحو، دلالة) تشمل الفصحى واللهجات.

· تطوير أدوات مفتوحة المصدر: الاستثمار في تطوير محاللات صرفية وقواعدية دقيقة ومفتوحة المصدر لتكون متاحة للمجتمع البحثي والمطورين.

· تعزيز البحث في نماذج مُعربة خالصة: تشجيع الأبحاث الرامية إلى بناء نماذج لغوية كبيرة من الصفر على بيانات عربية ضخمة ونقية، وليس مجرد ضبط لنماذج موجودة.

· التركيز على التطبيقات ذات الأولوية: توجيه البحث نحو تطبيقات ملحة كالترجمة الآلية عالية الجودة، والمختبرات اللغوية التعليمية، وأنظمة التلخيص الآلي للوثائق العربية.

· تعزيز التكامل بين المناهج: دمج المناهج الرمزية (القائمة على القواعد) مع المناهج الإحصائية (القائمة على البيانات) في هندسة النماذج اللغوية للاستفادة من مزايا كل منهما.

قائمة المصادر والمراجع

1. أبو السعود، محمد أحمد، والزغل، رامي محمود. (٢٠٢٢). أثر التشكيل على أنظمة التعرف على الكلام العربي. المجلة الدولية للعلوم الحاسوبية المتقدمة وتطبيقاتها، ١٣(٥)، ١١٠-١١٧.

2. عبد الجليل، عبد القادر محمد. (٢٠٠٢). علم اللسانيات الحديث (ط١). دار صفاء للطباعة والنشر والتوزيع

3. السيد، محمود عبد الرحيم. (٢٠٢٤). قضايا وتحديات الذكاء الاصطناعي في اللسانيات الحاسوبية العربية. ورقة مقدمة في المؤتمر الدولي للدراسات اللغوية والأدبية، برو بروكاي

4. السليطي، ليلى محمد، والمناي، سمر أحمد. (٢٠٢١). تطوير روبوت دردشة عربي: التحديات والحلول. مجلة جامعة الملك سعود – علوم الحاسوب والمعلومات، ٣٣(٦)، ٦٨٩-٦٩٧.

5. العجيلات، عادل مصطفى. (١٩٩٦). الحاسوب واللغة العربية. أطروحة ماجستير غير منشورة، جامعة اليرموك، الأردن.

6. علي، نبيل عبد العزيز. (١٩٨٨). اللغة العربية والحاسوب. مؤسسة تعريب الكويت.

7. عريوة، إيمان، وغيلوس، صالح . (٢٠٢٣). حاجة الدرس اللساني المعاصر للذكاء الاصطناعي اللسانيات العرفنية أنموذجا. مجلة العندة في اللسانيات وتحليل الخطاب، جامعة محمد بوضياف بالمسيلة، ٧(١)، ٥٣٣-٥٤٢.

8. كوتو، فابريسيو. (٢٠٢٤). أداة جديدة لتقييم أداء النماذج اللغوية الكبيرة باللغة العربية. جامعة محمد بن زايد للذكاء الاصطناعي ، https://mbzuai.ac.ae/ar/news/a-new-standard-for-evaluating-arabic-language-models-presented-at-acl/

9. محمود، محمد رأفت. (٢٠٢١). مبادئ الفكر اللغوي بين اللسانيات التوليدية التحويلية واللسانيات الحاسوبية. حولية كلية الآداب جامعة بني سويف، ١٠(١)، ٥٠٩-٥٤٠.

10. الموسى، نبيل أحمد. (٢٠٠٠). العربية نحو توصيف جديد في اللسانيات الحاسوبية. دار الشروق.

11. شاهين، أحمد محمد. (٢٠٢٥، مايو ١٤). اللغة العربية والذكاء الاصطناعي.. معركة للبقاء في المشهد الرقمي العالمي. الجزيرة نت. تم الاسترداد من https://www.aljazeera.net/culture/14/5/2025/اللغة-العربية-والذكاء-الاصطناعي

12. تونسي، أمين محمد، و حداد، حاتم محمود، (2019)، بناء محلل صرفي للغة العربية: دراسة حول تأثير حجم قاموس الجذور، وقائع المؤتمر الدولي للسانيات الحاسوبية، (1234 – 1245) .

13. الشامسي، فاطمة محمد. (2022). مراكز الأبحاث العربية والذكاء الاصطناعي: دراسة تقييمية. مجلة دراسات الخليج والجزيرة العربية، 48(2)، 75-90.

14. الصاوي، خالد محمود. (2023). واقع اللغة العربية في عصر الذكاء الاصطناعي: التحديات والفرص. دار الكتب العلمية.

15. الغنيم، عبد الله إبراهيم. (2021). تمويل البحوث التطبيقية في مجال اللسانيات الحاسوبية في العالم العربي. المؤتمر السنوي للغة العربية والحاسوب، الجامعة الأردنية.

16. الوزارة السعودية للاتصالات وتقنية المعلومات. (2020). الاستراتيجية الوطنية للبيانات والذكاء الاصطناعي. المملكة العربية السعودية.

17. مركز الدراسات اللغوية والحاسوبية. (٢٠٢٢). التقرير السنوي عن واقع معالجة اللهجات العربية. سلسلة التقارير البحثية، (٥)، ١-٤٥.

18. أبو العسل، نوزات. (2023). تأثيرات وسائل التواصُل الاجتماعي على اللغة العربية. *المجلة المصرية لبحوث الرأى العام*، *22*(2)، 431-456.

19. عزيز، و محمد الخزامي. (2023). دور الذكاء الاصطناعي في العلوم الاجتماعية والإنسانية. *سمنار* ، *1* (2)، 1-35.

20. سلسلة الذكاء الاصطناعي للتنفيذيين ، (2024) ، الذكاء الاصطناعي، الهيئة السعودية للبيانات والذكاء الاصطناعي، ط2،

21. محمد.أحلام علي. (2025). دمج المنهج التكاملي مع أدوات الذكاء الاصطناعي في تعليم النحو العربي في العصر الرقمي. *مجلة بحوث الاتصال (Journal of Communication Research)*، 386-359.

22. نصيرة ، إديـــر. (2014). المعالجة الآلية للغة العربية وترجمتها الآلية. *الآداب و اللغات*، 11-42. ، *9*(1)

23. إسماعيلي، يوسف (2025). الانزياح اللغوي للمتعلم بين الفصيحة والدارجة وآليات توظيفه تعليميًّا في المملكة المغربية. *مجلّة الإيسيسكو للّغة العربيّة* *2*(1)، 285-312 .

24. مور، تيرينس موور كريستين كارلنغ. (1993). فهم اللغة: نحو علم لغة لما بعد مرحلة تشومسكي. ترجمة حامد حسين الحجاج. بغداد: دار الشؤون الثقافية.

25. محمد عبد القوي، سند. (2015). ظاهرة الاختزال في اللهجة اليافعية. *مجلة كلية التربية (أسيوط)* *31*(3.2)، 463-492 .

26. البخاري، أبو عبد الله محمد بن إسماعيل. (1993). صحيح البخاري. تحقيق: د. مصطفى ديب البغا. ط 5. دار ابن كثير، دار اليمامة . دمشق

27. د. بشر. منى محمد علي (2021). المشكلات الصوتية التي تواجه نطق العمالة الهندية لبعض الأصوات الانفجارية في المملكة العربية السعودية (دراسة تطبيقية) . *مجلة العلوم العربية* (62)..

28. كروش، حيزية. (2025). تحدي رقمنة اللغة العربية في ظل تطورات الذكاء الاصطناعيAI .جسور المعرفة. جامعة حسيبة بو على (الجزائر). 11 (1) . 151 – 160 .

29. Brown, Thomas B., Mann, Benjamin, Ryder, Nick, Subbiah, Kaplan, Jared, Dhariwal, Prafulla, ... & Amodei, Dario. (2020). ،Melanie shot learners. Advances in Neural Information -Language models are few .1901-Processing Systems, 33, 1877

30. Wei, Lee, Kenton, & Toutanova, -Devlin, Jacob, Chang, Ming raining of deep bidirectional transformers t-Kristina. (2018). BERT: Pre .for language understanding. arXiv preprint arXiv:1810.04805

31. Ma, Yuanyuan, & Siau, Keng L. (2018). Artificial intelligence impacts on higher education. In Proceedings of the Thirteenth Midwest rmation Systems Conference (p. 5). Saint Louis, Association for Info .Missouri

32. Tounsi, Lamine, & Haddad, Hatem. (2019). Building a morphological analyzer for Arabic: A study on the effect of root dictionary size. In Proceedings of the International Conference on .(1245-Linguistics (COLING) (pp. 1234 Computational