مجلة الجامعة القاسمية

للغة العربية وآدابها

مجلة علمية محكمة نصف سنوية

المجلد: 5، العدد: 1

ذو الحجة 1447هـ / يونيو 2026م

الترقيم الدولي المعياري للدوريات: 2788-5542

AQU Journal of Arabic Language and Literature, Vol. 5 No. 1, pp. 173- 216 (June.2026)
DOI:10.52747/aqujall.5.1.529
E-ISSN: 2958-2318/ ISSN: 2958-230X

اللُّغة العربية والتحوّل الرَّقْمي: قراءة في أثر الذّكاء الاصطناعيّ

ARABIC LANGUAGE AND DIGITAL TRANSFORMATION: A READING INTO THE IMPACT OF ARTIFICIAL INTELLIGENCE[1]

مهند بيازيد

دكتوراه النَّحْو والصَّرْف، كلية الآداب والعلوم الإنسانية، جامعة دمشق، سوريا

Mohanad Byazeed

PhD in Arabic Language and Literature - Grammar and Morphology, Faculty of Arts and Humanities at the University of Damascus, Syria

[1] Article received: March 2026; article accepted: April 2026

الملخص:

يهدف هذا البحث إلى تقويم واقع معالجة اللغة العربية حاسوبيًّا في سياق الابتكار الرَّقْمي والتحوّل نحو نماذج الذكاء الاصطناعي في معالجة اللغات الطبيعية من خلال تحليل منهجي للخصائص البنيوية التي تؤثر في كفاءة النمذجة. ويعتمد البحث مقاربة وصفية تحليلية تستند إلى تفكيك مستويات البنية اللغوية العربية (الصرفي، والتركيبي، والدلالي، والكتابي)، وربطها بمتطلبات التمثيل الحاسوبي. وتبيّن الدراسة أن غياب التشكيل وغنى النظام الصرفي ومرونة الترتيب التركيبي وتعدّد الأدوات الوظيفية إلى جانب التنوّع اللهجي تمثّل محددات رئيسة ترفع درجة الغموض في النص العربي وتؤثر سلبًا في دقة التحليل الآلي. كما تكشف عن قصور نسبي في الموارد اللغوية الرقمية من حيث الحجم والتوسيم والجودة، الأمر الذي ينعكس على أداء النماذج، ولا سيما في مهام الفهم الدلالي والتصنيف والترجمة. في المقابل، يرصد البحث التحسّن الذي حققته النماذج القائمة على التعلّم العميق في تمثيل السياق اللغوي، مع التأكيد أن فاعليتها في العربية تظل مرتبطة بمدى مواءمتها للخصائص البنيوية للغة. وبناءً على ذلك، يقترح البحث توجّهًا تكامليًّا يقوم على دمج المعرفة اللسانية العربية في مستوياتها المختلفة ضمن تصميم النماذج الحاسوبية بهدف تحسين الدقة وتقليل الغموض، وتعزيز قابلية التعميم في التطبيقات اللغوية المعاصرة.

Abstract:

This study aims to evaluate the current state of computational Arabic language processing within the context of digital innovation and the shift toward artificial intelligence models in natural language processing, through a systematic analysis of the structural properties that affect modeling efficiency. The study adopts a descriptive-analytical approach grounded in the decomposition of Arabic linguistic structure across its levels (morphological, syntactic, semantic, and written) and maps these levels onto the requirements of computational

representation. The findings demonstrate that the absence of diacritical marking, the richness of the morphological system, the flexibility of syntactic ordering, the multiplicity of functional particles, and dialectal variation collectively constitute principal constraints that elevate the degree of ambiguity in Arabic text and adversely affect the accuracy of automated analysis. The study further reveals a relative deficiency in digital linguistic resources with respect to size, annotation, and quality, a deficiency that is directly reflected in model performance particularly in tasks involving semantic comprehension, classification, and translation. At the same time, the study documents the progress achieved by deep learning models in representing linguistic context, while affirming that their effectiveness in Arabic remains contingent upon the degree to which they accommodate the language's structural properties. On this basis, the study proposes an integrative approach that embeds Arabic linguistic knowledge, across its various levels, into the design of computational models, with the aim of improving accuracy, reducing ambiguity, and enhancing generalizability across contemporary language applications.

الكلمات المفتاحية: معالجة اللغة العربية، الذكاء الاصطناعي، الابتكار الرَّقمي، التعلّم العميق، الموارد اللغوية.

Keywords: Arabic language processing, Artificial intelligence, Digital innovation, Deep learning, Linguistic resources.

المقدمة

تتميّز اللغة العربية بخصائص بنيوية جعلتها محورًا مهمًّا للدراسة في ميادين اللسانيات الحديثة وتقنيات الذكاء الاصطناعي. فقد حافظت العربية الفصحى على قدرٍ ملحوظ من الثبات الوظيفي عبر العصور، الأمر الذي مكّنها من الاستمرار أداةً للتواصل الثقافي والعلمي رغم تنوّع اللهجات الإقليمية. ويكتسب هذا الثبات بعدًا خاصًّا لارتباطها بالقرآن الكريم، وهو ارتباط أسهم في الحفاظ على بنيتها المعيارية وتعزيز حضورها في المجال الحضاري العربي[1].

ومع هذا الثبات؛ تمتاز العربية بمرونة بنيوية تتيح التجديد داخل نظامها، من خلال آليات اشتقاقية وصيغ قياسية تسمح بتوليد مفردات وتراكيب جديدة دون الإخلال بأصولها اللغوية. وقد أشار علماء العربية إلى هذه السمة مبكرًا؛ إذ قرّر ابن جني أن ما جرى على قياس كلام العرب فهو من كلامهم. وبفضل هذه الميزة الاشتقاقية ظلّت العربية قادرة على استيعاب التحوّلات العلميَّة والمعرفيَّة مع الحفاظ على صلة واضحة بين الألفاظ وأصولها الدلالية[2].

أسهمت جهود التعريب العلمي وتوليد المصطلحات في تعزيز حضور العربية في المجالات العلمية والتقنية. وقد ازداد هذا الحضور مع التحول الرَّقْمي وانتشار تطبيقات الذكاء الاصطناعي ومعالجة اللغة الطبيعية، رغم استمرار تحديات تقنية تتطلَّب تطوير أدوات حاسوبية أكثر كفاءة لدعم العربية في البيئة الرَّقْمية العالمية.

---

(1) عبد التواب، رمضان: بحوث ومقالات في اللغة، مكتبة الخانجي بالقاهرة، ودار الرفاعي بالرياض، 1982م، 179.

(2) انظر: ابن جني، عثمان: الخصائص، تحقيق محمد علي النجار، دار الكتب المصرية، القاهرة، 1952م، 115/1، 358-370. والمنصف، دار إحياء التراث القديم، مصر، 1954م، 180/1 - 183. وأنيس، إبراهيم: من أسرار اللغة، ط6، مكتبة الأنجلو المصرية، القاهرة، 62 - 68.

أسباب اختيار البحث:

ينبع الدافع وراء هذا البحث من الأهمية البالغة للحفاظ على دور اللغة العربية وتعزيزه في العصر الرَّقْمي، خصوصًا في ظل التحديات التي تواجهها ضمن الذكاء الاصطناعي.

أهداف البحث:

يهدف هذا البحث إلى استكشاف سُبل تعزيز استمرارية اللغة العربية في ظل التحولات التكنولوجية المعاصرة، وتعزيز قدرتها على الاندماج الفاعل في البيئات الرقمية. ويهدف إلى تطوير نماذج حاسوبية قادرة على تمثيل البنية اللغوية العربية تمثيلًا أعمق، بما تنطوي عليه من تعقيد نحوي وصرفي ودلالي، ولا سيما ما يتصل بظواهر التشكيل والإعراب وغنى الاشتقاق. ويهدف إلى تحسين كفاءة التطبيقات اللغوية الرقمية ـ مثل الترجمة الآلية ومعالجة الأخطاء ـ عبر رفع مستوى الدقة في النماذج المعتمدة، وتوسيع قدرتها على توليد نصوص عربية سليمة ومتسقة، إلى جانب تقويم أدائها بمقارنته بالأنظمة التقليدية في معالجة النصوص العربية.

أهمية البحث:

تتجلّى القيمة العلمية لهذا البحث في سعيه إلى تضييق الفجوة بين اللغة العربية والتقنيات الرقمية المعاصرة عبر تطوير مستوى معالجتها الحاسوبية وتحسين جودة تحليل النصوص العربية وإنتاجها. ويسهم في تعزيز كفاءة التطبيقات اللغوية في مجالات حيوية مثل الترجمة الآلية والتعليم الإلكتروني، ويتيح توسيع إمكانات التفاعل مع المحتوى العربي في البيئة الرقمية. ومن شأن ذلك أن يدعم حضور العربية في العالم الرَّقْمي العالمي، ويعزّز دورها أداةً فاعلة للتواصل المعرفي في العصر الرقمي.

منهج البحث:

يُعتمد في هذا البحث المنهج الوصفي بوصفه الأنسب لتحقيق أهدافه؛ إذ يقوم على دراسة النماذج اللغوية القائمة وتقويم أدائها، وتحليل الإشكالات التي تطرحها اللغة العربية في معالجة النصوص الرقمية تمهيدًا لتطوير هذه النماذج وتحسين كفاءتها.

المبحث الأول: الذكاء الاصطناعي وتحول مقاربات معالجة اللغة الطبيعية

أفضت التحولات المتسارعة في تقنيات الذكاء الاصطناعي إلى إحداث نقلة نوعية في التعامل مع اللغة بوصفها موضوعًا للتحليل والمعالجة الآلية. ولم يعد الاشتغال اللغوي، في صورته المعاصرة، منفصلًا عن النماذج الذكية التي باتت تشكّل الأساس التقني لفهم الخطاب الإنساني. وفي هذا الإطار، برزت معالجة اللغة الطبيعية أحد أكثر الحقول تأثرًا بهذا التطور؛ إذ أُعيد بناء آلياتها اعتمادًا على خوارزميات قادرة على استيعاب البنية اللغوية وتحليلها بدقة متقدمة. وقد انعكس ذلك بوضوح على كفاءة أنظمة الترجمة الآلية وتقنيات تحليل النصوص، وهذا ما أسهم في توسيع آفاق التواصل بين متحدثي لغات مختلفة، وتعزيز أنماط التفاعل الذكي بين الإنسان والآلة في سياقات متعددة[1].

وتقوم معالجة اللغة الطبيعية على منظومة معقدة من القدرات المعرفية المتقدمة؛ إذ تتطلب تمكين الأنظمة الحاسوبية من التعامل مع مستويات متعددة من التمثيل اللغوي. ويشمل ذلك التعرف على السمات اللفظية والدلالية، وربطها بالخبرات والوقائع المرتبطة بالسياق، إضافة إلى تنظيم هذه المعارف وحفظها واستدعائها عند الحاجة. كما تقتضي هذه المعالجة إنشاء علاقات تفاعلية مرنة بين مكونات النص ما يدعم مهام الفهم والمراجعة ومعالجة التكرار والتكامل بين وحدات التحليل المختلفة. ويُضاف إلى ذلك القدرة على

(1) Jurafsky, Daniel, & James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition". 3rd ed., Stanford: Stanford University Press, (2024), pp. 45-47.

استخلاص المفاهيم الجوهرية، مثل المصطلحات الإجرائية، وبناء تمثيلات دقيقة للمفاهيم المجردة التي لا تُدرك مباشرة على مستوى السطح اللغوي[1].

وتُعد هذه القدرات شرطًا أساسًا للانتقال من المستوى الإجرائي لمعالجة اللغة الطبيعية إلى المستوى الأعمق المتمثّل في فهم اللغة الطبيعية. فرغَم أن التقنيات المعاصرة في هذا المجال تعتمد بدرجة كبيرة على تحليل البنى النحوية والعلاقات الشكلية داخل النص، فإنها لا تزال محدودة الأفق؛ إذ تشتغل في الغالب ضمن الإطار النصي الظاهر، دون امتلاك القدرة على استحضار المعارف والسياقات الخارجية التي يعتمد عليها الإنسان في تفسير الخطاب. ويترتب على ذلك قصور في استيعاب المقاصد والمعاني الضمنية التي تتجاوز حدود النص المباشر.

ويزداد هذا التحدي تعقيدًا بسبب الطبيعة المفتوحة للغات الطبيعية، التي لا تخضع لحصر عددي في تراكيبها أو جملها، إضافة إلى اتّسامها بدرجة عالية من التعدد الدلالي. فكثير من المفردات تحمل أكثر من معنى، ولا يُحسم المقصود منها إلا بالرجوع إلى السياق، كما في الكلمات التي قد تُستخدم للدلالة على أفعال أو أسماء أو مفاهيم مختلفة بحسب المقام. ويؤدي هذا الغموض الدلالي إلى احتمالات متعددة في تفسير الجملة الواحدة؛ وذلك مثل كلمة (Book) التي تعني (كتاب/اسم) أو (يحجز/فعل)، وكلمة (Watch) التي تعني (ساعة/اسم) أو (يشاهد/فعل)، وكلمة (Train) التي تعني (قطار/اسم) أو (يدرّب/فعل)، ما يجعل عملية الفهم الآلي عرضة للالتباس والانحراف عن المعنى المقصود. ولهذا السبب، يظل تطوير أنظمة قادرة على فهم اللغة الطبيعية فهمًا دقيقًا وشبيهًا بالفهم البشري من أكثر المهام تعقيدًا في مجال الذكاء الاصطناعي[2].

(1) Chowdhary, K. R. “Fundamentals of Artificial Intelligence”. New Delhi: Springer, (2020), p. 604.
(2) Chowdhary, “Fundamentals of Artificial Intelligence”, p. 605.

وينصرف حقل معالجة اللغة الطبيعية إلى تطوير منظومة من التطبيقات الوظيفية التي تهدف إلى تمكين الحاسوب من التعامل مع اللغة البشرية في صور متعددة، من أبرزها الترجمة، واسترجاع البيانات، واستخلاص المعلومات، وتكثيف المحتوى عبر التلخيص، والإجابة عن الاستفسارات، إضافة إلى تحليل الموضوعات وبناء نماذجها. غير أن هذا المجال، رغم اتساع تطبيقاته وتقدم تقنياته، لا يخلو من إشكالات جوهرية تعيق الوصول إلى مستويات عالية من الدقة والفهم. ويمكن الوقوف على أبرز هذه التحديات في النقاط الآتية[1]:

- **تحدّيات التمثيل اللغوي:**

1- الاعتماد على البنية السطحية وإغفال المستوى العميق:

اتجهت كثير من مقاربات معالجة اللغة الطبيعية إلى تمثيل البنية السطحية للنص مع إغفال البعد الدلالي العميق القائم على الاستدلال وفهم العلاقات الكامنة بين المعاني. وقد اقتصر هذا الاتجاه على نماذج شكلية محدودة، لا تستوعب السياق التداولي ولا تعالج ظاهرة التعدد الدلالي معالجة كافية. ويُعد هذا القصور سببًا رئيسًا في استمرار الغموض الدلالي وصعوبة رفع اللَّبس عن الألفاظ متعددة المعاني[2].

2- التغليب النحوي وإهمال البعد الدلالي:

اتسمت البدايات الأولى لمعالجة اللغة الطبيعية بالتركيز على البنية النحوية للجملة، مع إهمال نسبي لتمثيل المعنى. وقد شكّل هذا الخلل أحد أبرز العوائق أمام تطور المجال؛

(1) Alagar, R., "Fundamentals of Artificial Intelligence", pp. 606-608. Alagar, R. "The Role of Natural Language Processing (NLP) in AI Applications". Skill floor, (2023, October 2) Retrieved from: https://skillfloor.com/blog/the-role-of-natural-languageprocessing-nlp-in-ai-applications.

(2) انظر: أبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، تحرير يوسف العريان، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019، 178.

إذ تبيّن أن التحليل التركيبي وحده لا يحقق فهمًا لغويًّا حقيقيًّا. ومن ثمّ أكدت دراسات لاحقة ضرورة إدماج الدلالة في النماذج الحاسوبية، وهو ما أشار إليه (Minsky) من خلال دعوته إلى اعتماد مقاربات تقوم على تمثيل المعنى والحقول الدلالية[1].

- **تحدّيات موارد البيانات:**

1- فجوة الموارد اللغوية:

تركّزت معظم التطورات في تقنيات معالجة اللغة الطبيعية على اللغات ذات الحضور الكثيف رَقْميًّا، في حين ما تزال اللغات واللهجات محدودة الموارد تعاني ضعف الدعم التقني وقلّة البيانات. ويترتب على ذلك صعوبة بناء نماذج عالية الكفاءة لهذه اللغات، الأمر الذي يقلّص شمولية تقنيات الذكاء الاصطناعي ويحدّ من استفادة المجتمعات اللغوية الأقل تمثيلًا في العالم الرَّقْمي.

2- التحيز الإحصائي:

قد تنطوي نماذج معالجة اللغة الطبيعية على أنماط من التحيز، وهذا ما يفضي إلى مخرجات غير منصفة أو ذات طابع تمييزي. وتبرز إشكالات مرتبطة بآليات جمع البيانات واستخدامها، ولا سيما عند غياب الضبط والدقة الكافيين، وهو ما يستدعي عناية منهجية للحد من الآثار السلبية المحتملة وضمان الاستخدام المسؤول لهذه التقنيات[2].

- **تحدّيات الأداء الحاسوبي:**

(1) Minsky, Marvin, ed. "Semantic Information Processing". (Cambridge, MA: MIT Press, 1968), pp. 17 – 20.

وانظر: عطية، محمد، وراغب، أحمد وآخرون: العربيَّة والذكاء الاصطناعي، تحرير المعتز بالله السعيد، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019، 120، 200.

(2) Blodgett, S., Barocas, S., Daumé III, H., & Wallach, H. "Language (Technology) is Power: A Critical Survey of 'Bias' in NLP". arXiv, (2020): pp. 1-23. Retrieved from: https://arxiv.org/abs/2005.14050.

- **قيود التوسع وكفاءة المعالجة:**

تحتاج النماذج الكبيرة لمعالجة اللغة الطبيعية إلى موارد حوسبية هائلة، وهذا ما يقلل من إمكانية نشرها واستخدامها في البيئات ذات الإمكانات المحدودة. ويمثّل تحقيق التوازن بين قابلية التوسع وسرعة المعالجة عند تشغيل هذه النماذج تحديًّا مستمرًا، ويظل أحد أبرز المعوقات أمام تعميم تطبيقات الذكاء الاصطناعي على نطاق واسع.

❖ إسهامات الذكاء الاصطناعي في تطوير معالجة اللغات الطبيعية:

أحدث الذكاء الاصطناعي تحوّلًا نوعيًّا في مقاربات معالجة اللغة الطبيعية؛ إذ انتقلت النماذج اللغوية من التحليل الشكلي إلى أنظمة تعتمد التعلم الآلي والشبكات العصبية العميقة لفهم المعنى والسياق بصورة أعمق. وقد أفضى هذا التطور إلى اتساع تطبيقات المعالجة اللغوية في مجالات معرفية وخدمية متعددة، وأسهم في تطوير أنماط التفاعل بين الإنسان والآلة داخل البيئات الرقمية الذكية. ويمكن إجمال أبرز هذه التحولات في المسارات الآتية:

- **معالجة قيود التمثيل اللغوي:**

1- النماذج العميقة في معالجة البنية اللغوية:

تعدُّ النماذج العميقة من الركائز الأساسية في تطوير أنظمة معالجة اللغة الطبيعية، فهي تقوم على توظيف المعطيات النصية لاستخلاص البنى والأنماط اللغوية بصورة تُمكّن النماذج من تحسين أدائها تدريجيًّا. وأسهم هذا المنهج في تعميق قدرة الأنظمة على تمثّل السياق اللغوي، وهذا ما انعكس مباشرة على جودة الترجمة، ولا سيما في النصوص ذات التركيب الدلالي المركّب[1].

(1) Jurafsky & Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", (2024).

ولم يقتصر أثر التطور في تقنيات المعالجة اللغوية على اللغات الأوروبية، بل امتدّ إلى لغات كانت محدودة الحضور في النماذج التقليدية، ومنها العربية وعدد من اللغات الآسيوية[1]. وتجسّد أنظمة الترجمة العصبية مثل (**GNMT**) هذا التحول؛ إذ تعالج الجملة بوصفها وحدة دلالية متكاملة بدل الاعتماد على الكلمات المفردة، وهو ما أسهم في رفع دقة الترجمة، ولا سيما في اللغات منخفضة الموارد. كما مكّنت تقنيات تمثيل الكلمات النماذج من إدراك العلاقات السياقية بين الألفاظ والتراكيب، الأمر الذي انعكس في تحسن أداء مهام لغوية معقدة، مثل التصنيف النصي والترجمة الآلية[2].

2- الترجمة العصبية متعددة السياقات:

أحدثتِ الترجمة العصبية (GNMT) نقلة نوعية في مسار الترجمة الآلية؛ فقد تجاوزت منطق المعالجة التجزيئية إلى المعالجة الكلّية للنص من خلال فهم الجملة في سياقها الدلالي والتركيبي المتكامل. وقد أتاح ذلك قدرة أعلى على تمثيل المعنى، وإنتاج تراكيب لغوية أكثر انسجامًا مع أنظمة اللغة الهدف، إضافة إلى قابلية النظام للتطوّر المستمر عبر التعلم التراكمي من البيانات، وهذا ما يعزّز دقته وأداءه بمرور الزمن[3].

(1) Nassif, A. B., Elnagar, A., Shahin, I., & Henno, S. "Deep Learning for Arabic Subjective Sentiment Analysis: Challenges and Research Opportunities". ScienceDirect, (January 2021), p. 5. Retrieved from: https://www.sciencedirect.com/science/article/abs/pii/S1568494620307742.

(2) Mikolov, T., Chen, K., Corrado, G., & Dean, J. "Efficient Estimation of Word Representations in Vector Space". arXiv preprint, (2013). Retrieved from: https://arxiv.org/abs/1301.3781.

وانظر: أبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 150-151، 157. والقنيعير، فارس، وآخرون: خوارزميات الذكاء الاصطناعي في تحليل النص العربي، تحرير عبد الله الفيفي، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019، 63.

(3) Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … Dean, J. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation". Computer Science, (2016): pp. 1-23. Fix Ai Editor, F. A. "Revolutionizing Language

3- إعادة الصياغة المكثفة للنصوص:

تتمثّل هذه الآلية في إعادة بناء النص بصورة أكثر تكثيفًا، مع الحفاظ على بنيته الدلالية ومحوره الموضوعي. وتبرز أهميته في البيئات الإعلامية والتعليمية التي تتطلّب معالجة سريعة للمحتوى المطوّل دون استيعابه كاملًا. وتعتمد هذه الأنظمة على نماذج عميقة تستخلص الجمل المحورية أو تعيد صياغة المضمون استنادًا إلى تمثيل سياقي للنص، وهذا ما يفضي إلى ملخص دقيق ومتماسك[1].

- **معالجة قيود البيانات اللغوية:**

1- استكشاف البيانات النصية:

يشير استكشاف البيانات النصية إلى عملية تحليل النصوص الكبيرة آليًّا لاستنباط البيانات ذات الفائدة. وتكتسب هذه التقنية أهمية واسعة في مجالات متعددة، من أبرزها البحث العلمي حيث تُسرّع جمع المعلومات ذات الصلة، والصحافة حيثُ تتيح تحليل التقارير الإخبارية بفعالية. وتعتمد النظم في ذلك على أدوات معالجة اللغة الطبيعية لاستخراج عناصر محددة مثل الأسماء أو الأماكن أو التواريخ، بما يقلل الحاجة إلى التدخل اليدوي ويعزز سرعة التحليل. ويُعدّ (Factiva) مثالًا عمليًّا على تطبيقات هذه التقنية

Translation: Google's Neural Machine Translation System". FXIS.ai, (6 September 2024): pp. 1-2. Retrieved from: https://fxis.ai/edu/revolutionizinglanguage-translation-googles-neural-machine-translation-system/.

وانظر: أبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 168.

(1) Nallapati, R., Zhou, B., dos Santos, C., Gülçehre, Ç., & Xiang, B. "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond". Conference on Computational Natural Language Learning, (Berlin: Association for Computational Linguistics ,2016): pp. 280-282.

وانظر: عطية، محمد، وراغب، أحمد وآخرون: العربيَّة والذكاء الاصطناعي، 198 - 199. وأبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 169.

في تحليل الأخبار والبيانات الإعلامية، و(PubMed) في استخراج المعلومات من الأدبيات الطبية[1].

2- استكشاف التوجّه العاطفي للنصوص:

يرتكز استكشاف التوجّه العاطفي للنصوص على توظيف تقنيات الذكاء الاصطناعي في فحص المحتوى النصي بغية استنباط الموقف العاطفي الكامن فيه. وقد وجد هذا التوجّه تطبيقات واسعة في مجالات متعدّدة، من أبرزها التسويق وتحليل البيانات، وهذا يتيح للمؤسسات قراءة اتجاهات الجمهور واستجلاء تصوّراته حيال المنتجات والخدمات عبر دراسة التعليقات والنصوص المتداولة، إضافة إلى حضوره الكثيف في منصّات التواصل الاجتماعي[2].

ويعتمد هذا النمط من التحليل اعتمادًا أساسًا على نماذج التعلّم العميق، ولا سيما الشبكات العصبية المتكرّرة، القادرة على معالجة السلاسل النصية الطويلة والكشف عن البنى العاطفية المركّبة. وتمتاز هذه النماذج بقدرتها على التعلّم من مدوّنات واسعة النطاق، الأمر الذي أكسبها درجة أعلى من الدقة والموثوقية مقارنة بالمقاربات التقليدية في تصنيف المشاعر[3].

---

(1) Sarawagi, S. "Information Extraction". Foundations and Trends® in Databases, (2008): pp.12-15.

(2) وانظر: عطية، محمد، وراغب أحمد وآخرون: العربيَّة والذكاء الاصطناعي، 122. وأبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 105، 108-109، 170-171.

(3) Zhang, L., Wang, S., & Liu, B. "Deep Learning for Sentiment Analysis: A Survey". Wiley Interdisciplinary Reviews, (2018): pp. 2-4. Medhat, W., Hassan, A., & Korashy, H. "Sentiment Analysis Algorithms and Applications: A Survey". Ain Shams Engineering Journal, (2014): pp. 13-14. Kalluri, S. "Deep Learning Based Sentiment Analysis". Master Thesis, (Sweden: Blekinge Institute of Technology, 2023), p. 13.

وانظر: القنيعير، فارس، وآخرون: خوارزميات الذكاء الاصطناعي في تحليل النص العربي، 62.

3- التبويب الآلي للمحتوى النصي:

يشير التبويب الآلي إلى إسناد النصوص آليًّا إلى فئات محددة استنادًا إلى خصائصها اللغوية والدلالية. ويعتمد هذا الإجراء على نماذج تعلّم آلي تستخلص الأنماط المتكررة في البنية المعجمية والتركيبية، ثمَّ توظّفها في تقرير الانتماء التبويبي للنص. وتتجلّى تطبيقاته بوضوح في محركات البحث الأكاديمية التي تصنّف المقالات حسب الموضوع أو المجال، كما هو الحال في قاعدة بيانات (Scopus)؛ إذ تستخدم خوارزميات التعلم الآلي لتصنيف الأوراق البحثية ضمن مجالاتها العلمية تلقائيًا[1].

- **معالجة قيود التعدد اللغوي:**

أسهم التقدم المتسارع في تقنيات الذكاء الاصطناعي في إحداث تحوّلات نوعية في معالجة اللغات الطبيعية، ولا سيما فيما يتعلَّق بالتعامل مع التعدد البنيوي والاختلافات اللسانية بين اللغات. فقد باتت النماذج الحديثة تعتمد على استثمار المدونات اللغوية الضخمة وتحليلها آليًّا، وهذا ما يتيح للأنظمة اكتشاف الأنماط اللغوية المركبة والتكيف مع خصائص لغوية متعددة عبر لغات مختلفة. وقد أسفرت هذه الإمكانات المتقدمة عن نقل تطبيقات الذكاء الاصطناعي إلى مستويات أعلى من الكفاءة في فهم النصوص ومعالجتها وترجمتها؛ إذ بات الأداء يتحسن بصورة تراكمية منتظمة، مع انخفاض واضح في نسب الأخطاء التي كانت سمة ملازمة للأنظمة التقليدية في الترجمة الآلية في مراحلها السابقة. وبهذا، أصبح التفاعل الآلي مع التعدد اللغوي قائمًا على نماذج تعلم تكيفية قادرة على استيعاب التعقيد اللساني أكثر عمقًا وكفاءةً.

- **معالجة قيود الإدخال اللغوي:**

(1) Palanivinayagam, A., El-Bayeh, C. Z., & Damaševičius, R. "Twenty Years of Machine Learning-Based Text Classification: A Systematic Review". Algorithms, (2023): pp. 1-2.

1- التقنيات الذكية لفهم الكلام المنطوق:

تمثّل تقنية فهم الكلام المنطوق أحد التحولات المتقدّمة للذكاء الاصطناعي في معالجة اللغات؛ إذ تقوم على التحويل الآلي للكلام المنطوق إلى صيغة نصية مكتوبة. وقد أفضى هذا التطور إلى توسيع نطاق التطبيقات الذكية، ولا سيما في المساعدات الصوتية وبرمجيات الترجمة الفورية. وترتكز هذه النظم على نماذج متقدمة في معالجة اللغة الطبيعية تمكّنها من استيعاب السياق الدلالي للخطاب والتمييز الدقيق بين الوحدات الصوتية المتقاربة، وهذا تجاوز واضح للأنظمة التقليدية التي كانت تكتفي بمعالجة الإشارة الصوتية بمعزل عن السياق اللغوي[1].

2- الواجهات الذكية للتواصل البشري – الآلي:

يمثّل التواصل بين الإنسان والآلة أحد المسارات المتقدمة في تطوّر تقنيات الذكاء الاصطناعي، فقد أتاح إمكان التواصل مع الأنظمة الرَّقْمية عبر مدخلات لغوية صوتية أو كتابية. وتتجلّى هذه المقاربة بوضوح في المساعدات الافتراضية التي تعتمد على تقنيات تفسير الأوامر اللغوية بما يسمح بإجراء تفاعل أقرب إلى الطابع الطبيعي. وترتكز هذه الأنظمة على تكامل معالجة اللغة الطبيعية مع التعرّف الآلي على الكلام، إلى جانب نماذج التعلّم العميق التي تمكّنها من تحليل السياق وتوليد استجابات أكثر دقة وملاءمة، إضافة إلى قدرتها على تنفيذ مهام عملية متنوعة مثل تشغيل الوسائط أو تقديم الإرشاد المكاني والملاحي[2].

---

(1) Manning, C. D., & Schütze, H. “Foundations of Statistical Natural Language Processing”. (Cambridge, MA: MIT Press, 1999), pp. 101–105. وانظر: عطية، محمد، وراغب، أحمد وآخرون، العربيَّة والذكاء الاصطناعي، 169 – 170. وأبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 157. والقنيعير، فارس، وآخرون: خوارزميات الذكاء الاصطناعي في تحليل النص العربي، 63.

(2) Luger, E., & Sellen, A. “Like Having a Really Bad PA: The Gulf between User Expectation and Experience of Conversational Agents”. Proceedings

وبناء على ما سلف؛ يتضح أنَّ الذكاء الاصطناعي قد أعاد تشكيل حقل معالجة اللغة الطبيعية، وقد أفضى هذا التحول إلى تقليص الحواجز اللغوية وتوسيع إمكانات التفاعل بين البيئات الثقافية المختلفة. وعلى الرغم من استمرار بعض الإشكالات التقنية والمنهجية، فإنَّ مسار التطور المتسارع ينبئ بترسيخ دور هذه التقنيات في تعميق الفهم المتبادل وتحسين قنوات التواصل الإنساني.

المبحث الثاني: إشكالات المعالجة الحاسوبية للغة العربية

تنبُع صعوبة المعالجة الحاسوبية للغة العربية من خصائصها البنيوية المميّزة، وفي مقدمة ذلك الصرف القائم على الجذر والوزن، وتعدّد المستويات اللهجية، وتشعّب البنى التركيبية. وقد حدّت هذه السمات من فاعلية النماذج المصمّمة للغات أخرى عند إسقاطها على العربية دون تكييف منهجي، وهذا ما استدعى تطوير مقاربات متخصصة تراعي طبيعتها اللسانية. وأفضى ذلك إلى نماذج أكثر قدرة على تمثّل السياق، وتحليل البنية الصرفية، واستيعاب العلاقات النحوية، إضافة إلى معالجة الدلالات التقويمية مثل المشاعر والاتجاهات.

وقد انعكس هذا التطور على أداء تطبيقات عملية مثل التّرجمة الآلية، والتعرّف على الكلام، وتحليل محتوى المنصّات الرقمية؛ إذ أسهمت الأدوات المستحدثة في تجاوز عوائق كانت تعيق النمذجة الدقيقة للعربية. ويستعرض هذا المبحث أبرز هذه المقاربات المتخصصة وأثرها في تعزيز كفاءة الذكاء الاصطناعي في التعامل مع الخصوصية اللسانية للعربية.

- **إشكالات الموارد والبيانات اللغوية:**

1- قلة الموارد اللغوية:

of the 2016 CHI Conference on Human Factors in Computing Systems, (2016): pp. 5289-5292.

تشكّل قلة الموارد اللغوية المتاحة أبرز العقبات أمام تطوير أنظمة معالجة اللغة العربية. فالنماذج الحديثة تحتاج إلى كميات كبيرة من النصوص لتعلّم البنى الصرفية والنحوية والدلالية، غير أن هذه النصوص غير متوفرة بشكل متوازن بين الفصحى واللهجات المحلية المختلفة. كما يزداد التحدي بسبب قلة التزام النصوص بالتشكيل والحركات وعلامات الترقيم، ما يعرقل قدرة النماذج على التعلّم الدقيق والتنبؤ الصحيح. نتيجة لذلك، يواجه المطوّرون صعوبة في إنشاء تطبيقات لغوية موثوقة وفعّالة[1].

2- التنوع اللهجي:

يُعدّ التنوع اللهجي من أبرز السمات الاجتماعية – اللسانية للعربية؛ إذ تتعايش الفصحى بوصفها معيارًا كتابيًّا وتعليميًّا مع لهجات محلية مثل السورية والمصرية والخليجية والمغربية متباينة في المعجم والصوت والبنية التركيبية. ويؤدي هذا التباين إلى تفاوت ملحوظ في قابلية الفهم المتبادل، إضافة إلى تعقيد مهمة النمذجة الحاسوبية[2].

وتزداد الإشكالية حدةً لأنَّ اللهجات لا تخضع لضبط إملائي أو نحوي مستقر مثل الفصحى، كما أنَّ حضورها الكتابي محدود قياسًا إلى تداولها الشفهي، وهذا ما يفضي إلى ندرة المدونات المعيارية اللازمة لتدريب النماذج الإحصائية. لذلك اتجهت الجهود البحثية إلى بناء موارد خاصَّة وتمثيلات عميقة قادرة على استيعاب الخصوصيات اللهجية[3].

---

(1) النابلسي، محمد: تحديات معالجة اللغة العربية في تطبيقات الذكاء الاصطناعي، مجلة الدراسات اللغوية الحديثة، المجلد 12، العدد 4، (2024م)، 215-245.

(2) Habash, N. "Introduction to Arabic Natural Language Processing". (San Rafael, CA: Morgan & Claypool Publishers, 2010). pp. 20–21.

(3) Previous reference, p. 21.

وانظر: عطية، محمد، وراغب أحمد وآخرون: العربيَّة والذكاء الاصطناعي، 200.

وعليه، فإنَّ الاقتصار على أدوات مصممة للفصحى لا يلبّي متطلبات المعالجة الدقيقة للنصوص اللهجية؛ بل يستلزم تطوير نماذج تكيفية تراعي التغاير الداخلي للعربية وتتعامل مع محدودية البيانات المكتوبة لهذه اللهجات[1].

- **قيود النظام الكتابي:**

1- النظام الكتابي للعربية:

يمثّل النظام الكتابي للعربية أحد أبرز مواطن التعقيد في المعالجة الآلية من حيث خصائص التمثيل الخطّي. فأوَّل ما يواجه النمذجة الحاسوبية هو غياب التشكيل؛ إذ تُكتب الصوامت مجردة في الغالب، فكلمة (درس) يُحتمل أن تكون (دَرَسَ أو دُرِسَ أو دَرَّسَ)، وكذلك كلمة (علم) يحتمل أن تكون (عَلِمَ أو عَلَّمَ أو عِلْم أو عَلَم)، وهذا ما يفتح المجال لتعدّد القراءات الدلالية للصيغة الواحدة، ويجعلها معتمدة على التحليل السياقي[2].

ويُضاف إلى ذلك رسم الحرف؛ إذ يتبدّل شكله تبعًا لموقعه في أول الكلمة أو وسطها أو آخرها، وهو ما يفرض على الأنظمة التعرّف على الأشكال المتنوعة بوصفها تمثيلات لوحدة حرفية واحدة. أما على مستوى التنظيم النصي، فإن التفاوت في استعمال علامات

(1) Farghaly, A., & Shaalan, K. "Arabic Natural Language Processing: Challenges and Solutions". ACM Transactions on Asian Language Information Processing, 8(4), (2009): Article 14, pp. 5-6. Khalati, M., & Al-Romany, S. "Dialectal Arabic Processing: Current State and Future Directions". International Journal of Advanced Computer Science and Applications, 11(8), (2020): pp. 921. Al Khatib, M., & Shaalan, K. "The Key Challenges for Arabic Machine Translation". In: Intelligent Natural Language Processing: Trends and Applications. Studies in Computational Intelligence, (2018): pp. 143.

(2) Habash, N. "Introduction to Arabic Natural Language Processing", p. 7.

الترقيم، يعقّد مهام تقسيم الجمل وبناء العلاقات التركيبية مقارنة باللغات الأخرى التي تلتزم به[1].

وعليه، فإن التحدي في النص العربي المكتوب يكمن في التمثيل الكتابي وكثافته الدلالية، وما يقتضيه ذلك من نماذج قادرة على فهم السياق لاستعادة المعلومات المحذوفة أو غير المصرّح بها خطيًّا.

2- خصوصية الاتجاه الكتابي في اللغة العربية:

يتميّز النظام الكتابي في العربية باتجاهه من اليمين إلى اليسار، وهو ما يطرح تحديات ملحوظة في مجال المعالجة اللغوية الطبيعية وفي تصميم البرمجيات اللغوية. ويبرز أثر هذا الاختلاف خصوصًا عند تطوير التطبيقات متعددة اللغات؛ إذ ينبغي تهيئة الواجهات وخيارات التنسيق بما يتوافق مع اتجاه النص العربي لضمان عرضه بصورة صحيحة. كما يفرض هذا الاتجاه اعتبارات إضافية في عمليات التحليل الصرفي والنحوي، وفي تصميم خوارزميات معالجة النصوص. وتزداد هذه الصعوبات عند التعامل مع النصوص المختلطة التي تجمع بين العربية ولغات أخرى أو الأرقام، نحو قولنا: «التقيتُ بالباحث David في المؤتمر رَقْم 3» و«ناقشتُ الأستاذ Robert في موضوع البحث رَقْم 2»، الأمر الذي يتطلَّب أنظمة قادرة على إدارة اتجاهات الكتابة المختلفة بدقة داخل النص الواحد[2].

---

(1) Farghaly, A., & Shaalan, K. "Arabic Natural Language Processing: Challenges and Solutions", p. 7.

وانظر: عطية، محمد، وراغب أحمد وآخرون: العربيَّة والذكاء الاصطناعي، 96-99. وأبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 158.

(2) Habash, N. "Introduction to Arabic Natural Language Processing", pp. 16-17. Farghaly, A., & Shaalan, K. "Arabic Natural Language Processing: Challenges and Solutions", p. 18. Attia, M. "Arabic Tokenization System". Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages, Prague: Association for Computational Linguistics ,(2007): p67.

وانظر: أبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، 158.

- **القيود الدلالية والتقابلية:**

تُعَدّ مسألة التكافؤ اللغوي من أبرز التحديات في معالجة اللغة العربية؛ إذ لا تقابِل بعض المفردات العربية مكافئات دقيقة في لغات أخرى، أو تختلف دلالاتها باختلاف اللهجات والسياقات الثقافية. ويظهر ذلك بوضوح في المصطلحات المرتبطة بالثقافة الإسلامية أو الاجتماعية مثل: حجّ وفتوى، وشَرَفَ وقبيلة، التي تحمل أبعادًا دينية وثقافية يصعب نقلها حرفيًّا. كما تبرز المشكلة في الألفاظ التي لا تمتلك مقابلًا مباشرًا في اللغات الأخرى، مثل كلمة: (عقيقة) المرتبطة بممارسات دينية إسلامية. إضافة إلى ذلك، قد تتعدد ترجمة بعض المفردات مثل كلمة (عين) تبعًا للسياق أو المجال التخصصي، وهذا ما يخلق صعوبة في توحيدها ومعالجتها حاسوبيًّا بدقة[1].

- **القيود الصرفية في العربية:**

تتميّز اللغة العربية ببنى صرفية متعددة قائمة على الجذر والوزن. فالجذر، غالبًا ثلاثي أو رباعي، يحمل المعنى الأساسي، بينما تولّد الأوزان المختلفة منه مفردات متعددة، فمثلًا: الجذر (د–ر–س) يأتي منه: دارس، مدروس، دراسة، مدرسة، وهذا ما يزيد ثراء المفردات ويصعّب التمييز بين الصيغ بحسب السياق[2].

ويتعزز هذا التعقيد بظواهر صرفية تحويلية، أبرزها ظاهرة الإعلال والإبدال؛ إذ تتغير حروف الكلمة عند التصريف، مثل: باع يبيع؛ إذ تتحول الألف إلى ياء، وهذا ما يزيد

A., & Fellbaum, C. "Building a WordNet for Arabic". Proceedings of the Fifth International Conference on Language Resources and Evaluation, (Genoa, Italy: European Language Resources Association ,2006): pp. 29-34.

(2) حسان، تمّام: اللغة العربية معناها ومبناها، عالم الكتب، القاهرة، 2006 م، 136-137.

صعوبة معالجة الصرف تلقائيًّا ويستدعي نماذج قادرة على التعامل مع تغيّر أشكال الكلمات وفهم سياقاتها[1].

ومن جهة أخرى، تتّسم العربية بتنوّع الضمائر بين متصلة ومنفصلة، مع قابلية الضمير للاتصال بالفعل أو الاسم أو الحرف نحو: ضَرَبَهُ، وضَارِبُهُ، وله. ويترتب على هذا الاتصال تحوّل في بنية الكلمة، نحو: شربْتُ، وشربُوا، ويشربونه، وشرابُكَ، وشرابُها. ويسهم هذا في توسيع إمكانات التعبير، لكنه يفرض في الوقت نفسه تحديات على المعالجة الحاسوبية؛ إذ يتعيّن على النماذج اللغوية تحليل أثر الضمير في بنية الكلمة ووظيفتها داخل الجملة. ويضاف إلى ذلك، إلى أنَّ ضمائر المتكلم أو المخاطب مفتقرة إلى القرائن باعتبارها شرطًا أساسًّا لدلالتها على معيّن، وهو ما يعقّد عمليات الاسترجاع الآلي للمعلومات، ويستلزم نماذج قادرة على استنتاج العناصر المحذوفة من البنية اللغوية نفسها[2].

ويُفضي هذا التنوع الصرفي إلى تحديات ملحوظة في المعالجة الحاسوبية للعربية، ويجعل تمثيل البنية الصرفية للعربية مهمة معقدة نسبيًّا في النماذج اللغوية.

- **القيود التركيبية في العربية:**

تنتظم التراكيب النحوية في العربية ضمن نسق وظيفي يضبط ترابط مكوّنات الجملة. وفيما يأتي بيان أبرز سماتها المميزة[3]:

1- النسق التركيبي:

(1) السامرائي، إبراهيم: الفعل زمانه وأبنيته، مؤسسة الرسالة، بيروت، 2013م،217.

(2) انظر: عطية، محمد، وراغب أحمد وآخرون: العربيَّة والذكاء الاصطناعي، 102 - 109. والقنيعير، فارس، وآخرون: خوارزميات الذكاء الاصطناعي في تحليل النص العربي، 102. وحسان، تمّام. اللغة العربية معناها ومبناها، 110.

(3) انظر: عطية، محمد، وراغب أحمد وآخرون: العربيَّة والذكاء الاصطناعي، 109-116.

تتسم العربية بتنوع أنماط الإسناد والترتيب في عناصر الجملة؛ فالنسق التركيبي يظلّ قابلًا للتقديم والتأخير تبعًا لمقتضيات السياق وأغراض المتكلم التداولية. فالعلاقات النحوية تُستمدّ من العلامات الإعرابية التي تضبط وظائفها داخل التركيب، وهو ما يسمح بتنوّع الصيغ مع بقاء البنية الإسنادية محفوظة. ويتضح ذلك في أمثلة نحو: ضَرَبَ زيدٌ عمروًا، وضَرَبَ عمروًا زيدٌ، وزيدٌ ضَرَبَ عمروًا، وعمروًا ضَرَبَ زيدٌ؛ فجميع هذه التراكيب صحيحة نحويًّا، غير أنَّ اختلاف ترتيب عناصرها يفضي إلى تباين في موضع العناية أو التركيز الدلالي داخل الجملة[1].

وتنبني الجملة العربية على نمطين إسناديين رئيسين: الاسمي القائم على علاقة المبتدأ والخبر، والفعلي المؤسس على الفعل والفاعل وما يتصل به من متممات. غير أنّ هذا البناء لا يظلّ دائمًا على صورته الأصلية؛ إذ قد يطرأ عليه حذف أو تقديم أو فصل بين عناصره تبعًا لمقتضيات البلاغة أو السياق، وهو ما تعضده العلامات الإعرابية التي تحفظ العلاقات النحوية رغم تغيّر الترتيب[2].

ويُفضي هذا النسق التركيبي إلى تحديات ملحوظة في المعالجة الحاسوبية؛ إذ يتعيّن على النماذج اللغوية استيعاب احتمالات الترتيب المتعددة واستنباط العلاقات النحوية من القرائن الصرفية والسياقية.

2- توافق التركيب النحوي:

يقوم التركيب النحوي في العربية على مبدأ التوافق بين الفعل وفاعله؛ إذْ ينسجم الفعل مع الفاعل في العدد والجنس. ويعزز هذا التوافق دقة التعبير العربي في ضبط العلاقات

(1) الجرجاني، عبد القاهر: دلائل الإعجاز في علم المعاني، تحقيق محمود محمد شاكر، مكتبة الخانجي، القاهرة، 1992م، 106.

(2) الغضية، اعتدال بنت محمد: الجملة العربية أنماطها وتحولاتها، مجلة العلوم الإنسانية والطبيعة، المجلد 3، العدد 10، (2022)، 511، 514.

الإسنادية بين عناصر الجملة، نحو قولنا: صمَّمَ المهندسُ المشروعَ، وصمَّمتِ المهندسة المشروعَ، والمهندسون صمّموا المشروع[1]. غير أنّ هذا التقييد التركيبي يفرض تحديات على النماذج الحاسوبية؛ إذ يتطلَّب استيعاب البنية السياقية كاملة لضبط علاقات التوافق على نحو صحيح.

3- تنوّع الصيغ الزمنية:

تستوعب اللغة العربية ترتيبًا زمنيًّا يضم الأزمنة الأساسية: الماضي، والمضارع، والمستقبل، إلى جانب صيغ مركبة مثل: الفعل بعد (كان) للدلالة على استمرار الماضي، و(لام الأمر) و(لن) لتحويل المضارع إلى المستقبل، و(لم) لتحويله إلى الماضي. كما يمكن للسياق أن يوجه زمن الفعل، مثل استخدام المضارع لوصف أحداث وقعت فعليًّا لإبراز الحالة أو سردها. هذا التنوع الزمني والدلالي يطرح تحديات كبيرة أمام معالجة اللغة العربية آليًّا[2].

4- علامات الإعراب الأصلية والفرعية:

تتسم العربية ببنية تركيبية منظمة، قوامها إسناد العلاقات النحوية إلى نسق إعرابي صريح يحمّل أواخر الكلمات وظائفها داخل الجملة. فالدلالة النحوية تُستمدّ من العلامة الإعرابية التي تُحدِّد موقع العنصر ووظيفته. وتشمل هذه العلامات الأصلية (الفتحة والضمة والكسرة)، إلى جانب علامات فرعية (الألف والواو والياء) تؤدي الوظيفة نفسها في صيغ مخصوصة ما يجعل البنية العربية أكثر مرونة وأغنى في إمكانات التصرّف التركيبي[3].

(1) السامرائي، فاضل: معاني النحو، دار الفكر، الأردن، 2000م، 2/59-69.
(2) السامرائي، فاضل: المرجع السابق، 328/3.
(3) حسن، عباس: النحو الوافي، ط15، دار المعارف، مصر، 2010، 104/1.

غير أنّ هذا النسق، على دقته، يطرح إشكالات تمثيلية؛ إذ تتداخل بعض العلامات في صيغ معينة مثل المثنى وجمع المذكر السالم، وتظهر أنماط خارجة عن القياس العام كالأسماء الخمسة، والأفعال الخمسة، والممنوع من الصرف. ومن ثمّ فإن خصوصية التركيب العربي تفرض على المعالجة الحاسوبية نماذج قادرة على استيعاب هذا التداخل بين الشكل والإعراب والوظيفة.

5- تعدّد وظائف الأدوات النحوية:

تحتوي اللغة العربية على أدوات النحوية تؤدي وظائف متعددة، منها: أدوات العطف، والتوكيد، والجر، والقسم، والشرط، وغيرها. وقد تحمل الأداة الواحدة أكثر من وظيفة بحسب السياق؛ فمثلًا: (لو) قد تُستخدم للتمني أو الشرط، و(اللام) للجر أو الأمر. كما أن اختلاف الأدوات المتماثلة وظيفيًّا مثل (من) و(في) و(الباء) و(إلى) و(حتى) في الجر، يغيّر المعنى المقصود. وتؤدي هذه الكثرة وتشابُه الوظائف إلى توليد احتمالات دلالية متعددة، وهذا ما يزيد صعوبة تفسير النصوص العربية تلقائيًّا[1].

المبحث الثالث: نماذج الذكاء الاصطناعي في معالجة اللغة العربية وسُبل تطويرها

واجهت نماذج المعالجة التقليدية صعوبة في التعامل مع النصوص العربية بسبب خصائصها اللغوية المميّزة وغناها الصرفي وتعدّد لهجاتها وتراكيبها النحوية المركّبة، وقد أدى ذلك إلى تطوير نماذج وتطبيقات للذكاء الاصطناعي صُمِّمت خصيصًا لمعالجة هذه الخصائص اللغوية.

وأسهمت هذه النماذج في تحسين فهم السياق وتحليل البنية الصرفية والنحوية للنصوص العربية، كما مكّنت من تطوير تطبيقات متقدمة مثل الترجمة الآلية، والتعرّف

(1) خرباط، علي طرخان: تعدد وظيفة الأداة النحوية في القرآن الكريم وأثره في المعنى في القرآن الكريم، مجلة القادسية في الآداب والعلوم التربوية، العراق، العدد 3، الجزء 1، (2023)، 1-30.

على الكلام، وتحليل محتوى وسائل التواصل الاجتماعي. ويتناول هذا المبحث أبرز نماذج وتطبيقات الذكاء الاصطناعي، وكيف تسهم في تطوير إمكانيات الذكاء الاصطناعي لمعالجة النصوص العربية وفهمها، وهي على النَّحْو الآتي:

- **نماذج التمثيل اللغوي العميق:**

1- نموذج (AraBERT)[1]:

يُعَدّ AraBERT أحد أبرز النماذج المتخصّصة في معالجة اللغة العربية، ويستند إلى BERT (Bidirectional Encoder Representations from Transformers) الذي طوّرته شركة Google في مجال معالجة اللغة الطبيعية. وقد جرى تكييف هذا النموذج ليتلاءم مع خصائص اللغة العربية من خلال تدريبه على النصوص العربية المأخوذة من مصادر متنوّعة مثل ويكيبيديا والمقالات الإخبارية ومنصّات التواصل الاجتماعي[2].

ويتميّز AraBERT بقدرته على فهم السياق اللغوي، وهذا ما يجعله فعّالًا في عدد من المهام، مثل تصنيف النصوص، وتحليل المشاعر، والترجمة الآلية. كما يسهم في التعامل مع خصوصية اللغة العربية المركبة وتنوّع اللهجات في العربية، الأمر الذي جعله أداة شائعة الاستخدام في الأبحاث والتطبيقات المرتبطة بتحليل البيانات والنصوص العربية.

ومن أبرز استخداماته دعم معالجة النصوص العربية الطويلة والمركّبة، وتحسين أداء النماذج اللغوية في تطبيقات المعالجة الآلية للنصوص.

(1) https://huggingface.co/aubmindlab/bert-base-arabert.

(2) Antoun, W., Baly, F., & Hajj, H. "AraBERT: Transformer-Based Model for Arabic Language Understanding". Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille: European Language Resource Association, (2020): pp. 10 - 12.

2- نموذج (Jais)[1]:

تُعَدّ Jais نموذجًا لغويًّا عربيًّا متقدمًا يندرج ضمن فئة النماذج التوليدية واسعة النطاق (Large Language Models)، وقد طُوِّر في سياق السعي إلى بناء بنية تحتية لغوية رقمية تدعم العربية. ويقوم هذا النموذج على بنى Transformer، ويُدرَّب على مدونات ضخمة تشمل العربية الفصحى وأنماطًا من العاميات، وهذا ما يتيح له إنتاج تمثيلات سياقية عميقة ومعالجة طيف واسع من المهام، مثل توليد النصوص، وتحليل المعنى، والإجابة عن الأسئلة. وتكمن أهميته في كونه محاولة لتجاوز التبعية للنماذج غير العربية عبر تقديم نموذج مكيَّف بنيويًا وثقافيًا مع خصائص العربية. غير أنَّه لا يزال يواجه التحدي نفسه الذي تعانيه النماذج العميقة عمومًا، وهو غياب التمثيل الصريح للبنية النحوية التفسيرية؛ إذ يشتغل في إطار التعلّم الاحتمالي القائم على الأنماط. ومن ثمّ تبرز الحاجة إلى إدماجه ضمن مقاربات هجينة تعيد وصل التمثيل العميق بالبنية النحوية العربية، ليحقق توازنًا بين الكفاءة التوليدية والدقة التفسيرية[2].

- **أدوات التحليل اللغوي:**

1- أداة (Farasa)[3]:

تُعَدّ Farasa من الأدوات المتخصصة في معالجة النصوص العربية، وقد طوّرها معهد قطر لبحوث الحوسبة (QCRI). وتُستخدم أساسًا في مهام التحليل الصرفي وتجزئة الكلمات والتشكيل الآلي للنصوص، كما تتميّز بسرعة عالية في معالجة البيانات اللغوية،

(1) JAIS – Inception.
(2) Sengupta, Neha, Sunil Kumar Sahu, Bokang Jia, et al. "Jais and Jais-chat: Arabic-Centric Foundation and Instruction-Tuned Open Generative Large Language Models." arXiv, September 29 (2023). https://doi.org/10.48550/arXiv.2308.16149.
(3) https://farasa.qcri.org.

وهذا ما يجعلها مناسبةً للتطبيقات التي تتعامل مع كميات كبيرة من النصوص أو تتطلب معالجة فورية[1].

وتبرز أهمية Farasa في قدرتها على تقديم تحليل صرفي، وإجراء تجزئة للكلمات العربية، إضافة إلى توفير خاصية التشكيل التلقائي، وهو ما يجعلها أداة مفيدة في التطبيقات التعليمية، وأنظمة القراءة الآلية، ومختلف تطبيقات معالجة اللغة العربية.

2- محلّل (MADAMIRA)[2]:

يُعَدّ MADAMIRA من الأدوات المتقدمة في تحليل النصوص العربية حاسوبيًّا، فهو يوفّر منظومة متكاملة للتحليل الصرفي والنحوي. ويعمل على تقسيم النص إلى وحدات لغوية، وتحديد السمات الصرفية للكلمات، إضافة إلى استخراج الجذور وتوليد التشكيل الآلي للنصوص. وقد حظي هذا النموذج باهتمام واسع في الدراسات اللغوية وتطبيقات معالجة اللغة العربية لما يتمتع به من دقة وفعالية في تحليل النصوص الفصيحة وبعض أنماط اللهجات العربية[3].

(1) Darwish, K., & Mubarak, H. "Farasa: A Fast and Furious Segmenter for Arabic". Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, San Diego, California: Association for Computational Linguistics (2016): pp. 11-16.

(2) https://github.com/CAMeLLab/WIDH_2020_Arabic_Text_Analysis/tree/master.

(3) Pasha, A., Al-Badrashiny, M., Diab, M., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic". Proceedings of the Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland: European Language Resources Association (2014): pp.1095-1097.

وتبرز أهميته في قدرته على تقديم تحليل صرفي للكلمات العربية، والتعامل مع أنماط لغوية متعددة، إضافة إلى توفير خاصية التشكيل التلقائي للنصوص، وهو ما يجعله أداة مفيدة في التطبيقات البحثية والبرمجية المرتبطة بمعالجة اللغة العربية.

3- إطار (CAMeL Tools)[1]:

تُعَدّ CAMeL Tools مجموعة متكاملة من أدوات معالجة اللغة العربية الطبيعية طوّرها مختبر CAMeL في جامعة نيويورك بأبوظبي، وتهدف إلى توفير حلول حاسوبية متقدمة لتحليل النصوص العربية بمختلف أنماطها. وتدعم الأداة معالجة اللغة العربية الفصحى وعددًا من اللهجات، كما تتيح تنفيذ مهام لغوية متعددة من أبرزها تحليل البنية الصرفية والتركيبية للنصوص، وتصنيف المحتوى النصي، والتعرّف على العناصر المسماة مثل الأسماء والأماكن والمؤسسات، كما أنها أداة مفتوحة المصدر، وهذا ما يسهّل على الباحثين والمطورين استخدامها ودمجها في التطبيقات المختلفة. وتُستخدم هذه الأداة في تحليل أنواع متعددة من النصوص، بما في ذلك النصوص الإعلامية والأكاديمية ومحتوى وسائل التواصل الاجتماعي[2].

- **أدوات التشكيل والصوتيات:**

1- أداة (Tashkeel)[3]:

(1) https://github.com/CAMeL-Lab/camel_tools.

(2) Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Go, F., Inoue, N., Eryani, F., Erdmann, A., & Habash, N. "CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing". Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France: European Language Resources Association (2020): pp. 7022-7026.

(3) https://rdi-tashkeel.com/ar/home.

يُعَدّ Tashkeel أداة تعتمد على تقنيات الذكاء الاصطناعي لتشكيل النصوص العربية تلقائيًا. ويسهم ذلك في تحسين وضوح النصوص ودقة قراءتها، ولا سيما في التطبيقات التي تتطلب فهمًا صحيحًا للبنية الصوتية للكلمات العربية[1].

وتبرز أهمية هذه الأداة في دعم التطبيقات التعليمية وأنظمة القراءة الآلية؛ إذ تساعد على تقديم نصوص عربية مشكولة تسهم في تسهيل التعلم وتحسين الفهم، كما تُستخدم في الدراسات اللغوية التي تحتاج إلى نصوص مضبوطة بالشكل لضمان دقة التحليل.

2- نظام (Kaldi)[2]:

يُعَدّ Kaldi نظامًا مفتوح المصدر متخصِّصًا في التعرف الآلي على الكلام (ASR)، وقد طُوِّر ليكون مرنًا وقابلًا للتخصيص في تطبيقات تحويل الصوت إلى نص. ويعتمد على تقنيات التعلم العميق في تحليل الإشارات الصوتية وفهمها، كما استُخدم في عدد من المشاريع التي تدعم اللغة العربية بهدف تحسين دقة التعرف على الكلام المنطوق وتحويله إلى نصوص مكتوبة[3].

وتبرز أهمية Kaldi في قدرته على توفير دقة جيدة في تحويل الكلام إلى نص، مع إمكانية تكييفه وَفق متطلبات التطبيقات المختلفة. لذلك يُستخدم في مجالات متعددة، مثل الأنظمة الصوتية الذكية، وتحويل المحاضرات والتسجيلات الصوتية إلى نصوص، إضافة إلى تطبيقات خدمة العملاء المعتمدة على التفاعل الصوتي.

---

(1) Alasmary, Faris, Zaafarani, Orjuwan, Ghannam, Ahmad. "CATT: Character-based Arabic Tashkeel Transformer". Association for Computational Linguistics. Bangkok, Thailand. (2024). arXiv:2407.03236v1

(2) https://kaldi-asr.org.

(3) Manohar, V., Povey, D., & Khudanpur, S. "JHU Kaldi System for Arabic MGB-3 ASR Challenge Using Diarization, Audio-Transcript Alignment and Transfer Learning". IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, (2017): pp. 347-350.

3– أداة (Diacritization Tool)[1]:

تُعَدّ Diacritization Tool أداة متخصصة في التشكيل الآلي للنصوص العربية، طوّرها معهد قطر لبحوث الحوسبة (QCRI)، وتهدف إلى إضافة الحركات إلى النصوص غير المشكولة اعتمادًا على تقنيات المعالجة الحاسوبية للغة (Language Technology Group, 2025). وتسهم هذه الأداة في تحسين دقة قراءة النصوص العربية وفهمها، خصوصًا في البيئات التعليمية والبحثية.

وتبرز أهميتها في توفير نصوص مضبوطة بالشكل، الأمر الذي يدعم الدراسات اللغوية وتطبيقات التعليم الإلكتروني والأنظمة التي تعتمد على القراءة الآلية للنصوص العربية.

- **منصات للمعالجة:**

1– خدمة (Google Cloud Natural Language API) لدعم العربية[2]:

يُعَدّ Google Cloud Natural Language API خدمة سحابية متقدمة لمعالجة اللغة الطبيعية، وتوفّر دعمًا لمعالجة النصوص العربية باستخدام تقنيات الذكاء الاصطناعي. وتتيح هذه الخدمة تنفيذ عدد من المهام مثل تحليل المشاعر، واستخراج الكيانات المسماة كالأشخاص والأماكن، وفهم محتوى النصوص بصورة آلية[3].

وتبرز أهميته في قدرته على تحليل كميات كبيرة من البيانات النصية بسرعة ودقة، وهذا ما يجعله أداة مفيدة في التطبيقات التجارية والبحثية، ولا سيما في تحليل بيانات وسائل التواصل الاجتماعي ودراسة آراء المستخدمين.

---

(1) https://github.com/qcri/dialectal_arabic_tools.
(2) https://cloud.google.com/natural-language.
(3) Restak, M. "Cloud Natural Language API Documentation". (Mountain View, California: Google Cloud, 2025). Retrieved from: https://cloud.google.com/natural-language/docs.

2- محرك (Elasticsearch) وأداة (Kibana) في دعم معالجة النصوص العربية[1]:

يُستخدَم Elasticsearch محرِّكًا متقدّمًا للبحث وتحليل البيانات، فهو يتيح معالجة كميات كبيرة من النصوص بسرعة وكفاءة، مع توفير أدوات لغوية تساعد على تحسين البحث في النصوص العربية. وتتكامل معه أداة Kibana التي تمكّن من عرض نتائج التحليل في صور مرئية مثل الرسوم البيانية ولوحات البيانات، وهذا ما يسهم في تسهيل فهم المعلومات واستخلاص الأنماط منها[2].

وتُستعمل هاتان الأداتان على نطاق واسع في تحليل البيانات النصية، ولا سيما في مجالات دراسة محتوى وسائل التواصل الاجتماعي، والتقارير المؤسسية، والأبحاث.

3- مكتبة (spaCy) لمعالجة اللغة العربية[3]:

تُعَدّ spaCy مكتبة مفتوحة المصدر في معالجة اللغة الطبيعية، تُستخدم على نطاق واسع في تحليل النصوص وتصنيفها واستخراج المعلومات منها. وقد طُوِّرت فيها أدوات ونماذج تدعم اللغة العربية، وهذا ما يتيح معالجة النصوص بكفاءة وسرعة في تطبيقات متعددة مثل تحليل البيانات النصية والتصنيف الآلي واستخراج الكيانات[4].

---

(1) https://www.elastic.co.

(2) Kathare, A., Reddy, P. V., & Prabhu, S. "Analysis of Elasticsearch in Comparison with MongoDB and Hadoop". International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, India,(2017): pp. 718-719.

(3) https://spacy.io.

(4) Otten, M. "Advanced Natural Language Processing with spaCy: A Guide to Understanding and Building NLP Applications". (Birmingham, UK: Packt Publishing, 2023). spaCy: Industrial-Strength NLP. "spaCy Usage Documentation". (Berlin: Explosion AI, 2025). Retrieved from: https://spacy.io/usage.

وتبرز أهمية هذه المكتبة في أدائها العالي وقدرتها على معالجة كميات كبيرة من النصوص، الأمر الذي يجعلها مناسبة للاستخدام في الأبحاث الأكاديمية والتطبيقات العملية، مثل تحليل محتوى وسائل التواصل الاجتماعي وتصنيف المقالات والبيانات النصية.

❖ نحو نموذج بنيوي متقدم لمعالجة النصوص العربية

ما تزال النماذج المخصّصة لمعالجة اللغة العربية تواجه تحديات جوهرية في فهم النصوص والتعامل مع تنوّعها، وسنعرض أبرز هذه الإشكالات مع اقتراح مسارات تطوير تعزّز كفاءة النماذج العربية ودقتها.

- **تشخيص أوجه القصور في أدوات المعالجة وسُبل تطويرها:**

تبدأ عملية تطوير نماذج معالجة اللغة العربية بتحديد أوجه القصور في الأدوات المتاحة؛ إذ تواجه كثير من التطبيقات صعوبات تتعلق بالتعامل مع اللهجات العامية، وتحليل النصوص غير المشكولة، وفهم السياقات اللغوية المركّبة[1].

فعلى سبيل المثال، يواجه (AraBERT) تحديات في معالجة النصوص التي تجمع بين الفصحى واللهجات، نظرًا لاعتماده أساسًا على بيانات فصيحة، مما يحدّ من قدرته على تحليل التنوع اللغوي. ويمكن تحسين أدائه من خلال تدريبه على (corpora) أوسع تضم نصوصًا عامية مأخوذة من وسائل التواصل الاجتماعي والمصادر الإعلامية، مع توظيف تقنيات التعلم المعزز لتعزيز دقة التحليل.

وعلى الرغم من سرعة (Farasa) العالية، فإنَّ دقته تتأثر عند معالجة النصوص غير المشكولة، وهو ما يشكّل أحد أبرز التحديات في العربية. ويمكن معالجة ذلك بتطوير

(1) النابلسي، محمد: تحديات معالجة اللغة العربية في تطبيقات الذكاء الاصطناعي.

نماذج قائمة على تقنيات المحوّلات (Transformers) وتدريبها على بيانات تجمع بين النصوص المشكولة وغير المشكولة لتحسين التمييز الدلالي والصرفي.

وأمّا (MADAMIRA)، فإنه يتميَّز بقدرات تحليل صرفي ونحوي متقدمة، غير أنَّ أداءه قد يتراجع عند التعامل مع النصوص المطوّلة أو التي تمزج بين الفصحى والعامية. ويمكن تعزيز كفاءته عبر تطبيق أساليب التعلم متعدد المهام وتدريبه على بيانات لغوية متنوعة تمثل سياقات مختلفة.

أما (Kaldi) الخاص بالتعرف الآلي على الكلام، فيواجه صعوبات في تمييز اللهجات العربية والتشابهات الصوتية بينها، مما يؤثر في دقة تحويل الكلام إلى نص. ويمكن تحسين أدائه من خلال توسيع مجموعات البيانات الصوتية لتشمل لهجات متعددة، إضافة إلى توظيف تقنيات التعلم المعزز في تصحيح الأخطاء الصوتية المتكررة.

ويمكن القول: إن الاتجاه الإستراتيجي في معالجة العربية ينبني على تجاوز الثنائية التقليدية بين النماذج القاعدية والنماذج التعلّمية، عبر تطوير نماذج هجينة (Hybrid Models) تُدمج فيها المعرفة النحوية داخل البنية الخوارزمية ذاتها. ويقتضي ذلك إعادة صياغة المفاهيم التراثية والعلاقات التركيبية في صورة تمثيلات قابلة للترميز والتعلّم، بحيث تتفاعل مع القدرات التعميمية للنماذج العميقة دون أن تُختزل فيها. وبهذا يتحول مسار البحث من مجرد تحسين الأداء إلى بناء نماذج تفسيرية قادرة على الجمع بين الصرامة القاعدية والمرونة الإحصائية، ليؤسس لنقلة نوعية في حوسبة العربية، قوامها استعادة البنية اللغوية بوصفها نظامًا معرفيًّا منتجًا داخل الفضاء الرَّقمي.

- **توظيف النماذج التوليدية المتطورة:**

يمكن توظيف النماذج التوليدية المتطورة مثل (GPT) و(T5) في تطوير معالجة اللغة العربية، لما تمتلكه من قدرة على فهم السياق اللغوي وتوليد النصوص. ويعتمد ذلك

أساسًا على تدريب هذه النماذج على بيانات لغوية متنوعة تجمع بين العربية الفصحى واللهجات العامية، بما يسهم في تحسين قدرتها على تحليل النصوص العربية غير القياسية.

وتتم عملية التطوير عبر إعادة ضبط النماذج المدربة مسبقًا باستخدام مجموعات بيانات متخصصة، وهذا ما يمكّنها من التعامل مع مهام متعددة مثل تحليل المشاعر، واستخراج الكيانات، وتوليد النصوص المختلطة بين الفصحى والعامية. كما يمكن الاستفادة من التعلم متعدد المهام لتدريب النموذج على أكثر من مهمة لغوية في الوقت نفسه، كتحليل النصوص وتوليدها.

ومن التطبيقات المهمة لهذه النماذج إمكان الترجمة بين اللهجات العربية والفصحى، وهو ما يوفّر أداة فعّالة لفهم التنوع اللغوي في المحتوى العربي وتحسين أداء أنظمة معالجة اللغة الطبيعية[1].

- **التقييم الدائم:**

وهي خطوة أساسية لقياس مدى فاعلية التحسينات المدخلة على أنظمة معالجة اللغة العربية. ويعتمد هذا التقييم على مجموعة من المعايير التي تقيس أداء النموذج في فهم النصوص العربية وإنتاجها بدقة.

ومن أهم هذه المعايير: التماسك اللغوي لقياس قدرة النموذج على إنتاج نصوص مترابطة وواضحة، والتنوع الأسلوبي لقياس محافظته على خصائص الأسلوب العربي، والاستمرارية الأسلوبية في الحفاظ على النمط اللغوي المطلوب. كما يشمل التقييم الدقة النحوية والصرفية في تحليل الكلمات غير المشكولة، ودقة التصنيف في تصنيف النصوص،

(1) Murhaf, Fares and Touileb Samia. "BabelBot at AraFinNLP2024: Fine-Tuning T5 for Multi-Dialect Intent Detection with Synthetic Data and Model Ensembling." In Proceedings of ArabicNLP,. Bangkok, Thailand: Association for Computational Linguistics, (2024): pp. 433–440. https://aclanthology.org/2024.arabicnlp-1.40.

إضافة إلى الاستجابة السياقية التي تقيس قدرة النموذج على فهم المعنى ضمن سياق الجمل والفقرات.

كذلك يتضمن التقييم التحليل الدلالي للعلاقات المعنوية داخل النص، وتحليل المشاعر لفهم الدلالات العاطفية، إضافة إلى القدرة على التعامل مع اللهجات العربية المختلفة. وتوفّر هذه المعايير مجتمعة صورة شاملة عن مدى كفاءة النماذج في تحليل النصوص العربية وفهمها وإنتاجها.

ومن أجل تحقيق مزيد من التطوير في هذا المجال، تبرز أهمية تعزيز التعاون بين الجامعات والشركات التقنية، وهذا ما يسهم في تطوير نماذج أكثر قدرة على التعامل مع التنوع اللغوي الغني في اللغة العربية.

الخاتمة

سعى هذا البحث إلى تحليل التحديات التي تواجه معالجة اللغة العربية في تطبيقات الذكاء الاصطناعي، من خلال تقويم عدد من النماذج المعاصرة. وقد أظهرت الدراسة أن هذه النماذج حققت تقدّمًا ملحوظًا في معالجة العربية الفصحى، غيرَ أنَّها ما تزال تواجه قيودًا تتصل بالتعامل مع التعدد اللهجي، والنصوص المختلطة بين الفصحى والعامية، إضافة إلى محدودية الفهم الدلالي العميق للسياق.

وقد خلص البحث إلى جملة نتائج رئيسة، من أهمها: الحاجة إلى تطوير النماذج الحالية عبر توظيف تقنيات التعلم العميق والتعلم متعدد المهام لزيادة قدرتها على تحليل البنى اللغوية المعقدة؛ وتوسيع المدونات اللغوية المستخدمة في التدريب لتشمل أنماطًا متنوعة من النصوص الفصيحة والعامية؛ ومعالجة إشكالات الضبط والتنوع الصرفي والنحوي التي تؤثّر في دقة التحليل، إضافة إلى تطوير آليات أكثر كفاءة لفهم السياق الدلالي. كما يتطلَّب تقدم هذا المجال تعزيز التعاون بين المؤسسات الأكاديمية والجهات التقنية لتطوير أدوات أكثر قدرة على تمثيل التنوع اللغوي العربي.

ويوصي البحث بتوجيه الدراسات المستقبلية نحو قضايا تطبيقية، من أبرزها: تحليل المحتوى العربي في البيئات الرقمية، ومعالجة التعدد اللهجي باستخدام تقنيات الذكاء الاصطناعي، وتطوير نماذج سياقية متقدمة لتحليل النصوص، وبناء أنظمة تعلم لغوي مستمر تعتمد تحديث المدونات اللغوية، إضافة إلى دراسة أثر الظواهر الدلالية كالاستعارة وتعدد الألفاظ في كفاءة النماذج الحاسوبية للغة العربية.

المصادر والمراجع

أنيس، إبراهيم: من أسرار اللغة، ط6، مكتبة الأنجلو المصرية، القاهرة.

أبو جبارة، أمجد، والحايك، أحمد، وآخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، تحرير يوسف العريان، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019.

الجرجاني، عبد القاهر: دلائل الإعجاز في علم المعاني، تحقيق محمود محمد شاكر، مكتبة الخانجي، القاهرة، 1992م.

ابن جني، عثمان: الخصائص، تحقيق محمد علي النجار، دار الكتب المصرية، القاهرة، 1952م.

ابن جني، عثمان، المنصف، دار إحياء التراث القديم، مصر، 1954م.

حسان، تمّام: اللغة العربية معناها ومبناها، عالم الكتب، القاهرة، 2006 م.

حسن، عباس: النحو الوافي، ط15، دار المعارف، مصر، 2010م.

خرباط، علي طرخان: تعدد وظيفة الأداة النحوية في القرآن الكريم وأثره في المعنى في القرآن الكريم، مجلة القادسية في الآداب والعلوم التربوية، العراق، العدد 3، الجزء 1، (2023)، 1-30.

السامرائي، إبراهيم: الفعل زمانه وأبنيته، مؤسسة الرسالة، بيروت، 2013م.

السامرائي، فاضل: معاني النحو، دار الفكر، الأردن، 2000م.

عبد التواب، رمضان: بحوث ومقالات في اللغة، مكتبة الخانجي بالقاهرة، ودار الرفاعي بالرياض، 1982م.

عطية، محمد، وراغب، أحمد وآخرون: العربيَّة والذكاء الاصطناعي، تحرير المعتز بالله السعيد، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019.

الغضية، اعتدال بنت محمد: الجملة العربية أنماطها وتحولاتها، مجلة العلوم الإنسانية والطبيعة، المجلد 3، العدد 10، (2022)، 494-522.

القنيعير، فارس، وآخرون: خوارزميات الذكاء الاصطناعي في تحليل النص العربي، تحرير عبد الله الفيفي، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019.

النابلسي، محمد: تحديات معالجة اللغة العربية في تطبيقات الذكاء الاصطناعي، مجلة الدراسات اللغوية الحديثة، المجلد 12، العدد 4، (2024م)، 215-245.

References:

ʿAbd al-Tawwāb, Ramaḍān. *Buḥūth wa Maqālāt fī al-Lugha (Studies and Articles in Language)*. Cairo/Riyadh: Maktabat al-Khānjī and Dār al-Rifāʿī, 1982.

Abū Jabbāra, Amjad, Aḥmad al-Ḥāyik, et al. *Taṭbīqāt al-Dhakāʾ al-Iṣṭināʿī fī Khidmat al-Lugha al-ʿArabiyya (Artificial Intelligence Applications in the Service of the Arabic Language)*. Edited by Yūsuf al-ʿAryān. Riyadh: King Abdullah bin Abdulaziz International Center for the Arabic Language, 2019.

Al Khatib, M., & Shaalan, K. "The Key Challenges for Arabic Machine Translation". In: Intelligent Natural Language Processing: Trends and Applications. Studies in Computational Intelligence, (2018): pp. 139-156.

Alagar, R. "The Role of Natural Language Processing (NLP) in AI Applications". Skill floor, (2023, october 2) Retrieved from: **https://skillfloor.com/blog/the-role-ofnaturallanguageprocessing-nlp-in-ai-applications.**

Alasmary, Faris, Zaafarani, Orjuwan, Ghannam, Ahmad. "CATT: Character-based Arabic Tashkeel Transformer". Association for Computational Linguistics. Bangkok, Thailand. (2024). arXiv:2407.03236v1

al-Ghuḍayya, Iʿtidāl bint Muḥammad. "al-Jumla al-ʿArabiyya: Anmāṭuhā wa Taḥawwulātuhā (The Arabic Sentence: Its Patterns and Transformations)." *Journal of Humanities and Nature*, vol. 3, no. 10 (2022): 494–522.

al-Jurjānī, ʿAbd al-Qāhir. *Dalāʾil al-Iʿjāz fī ʿIlm al-Maʿānī (Proofs of Inimitability in the Science of Meanings)*. Edited by Maḥmūd Muḥammad Shākir. Cairo: Maktabat al-Khānjī, 1992.

al-Nābulusī, Muḥammad. "Taḥaddiyāt Muʿālajat al-Lugha al-ʿArabiyya fī Taṭbīqāt al-Dhakāʾ al-Iṣṭināʿī (Challenges of Arabic Language Processing in AI Applications)." *Journal of Modern Linguistic Studies*, vol. 12, no. 4 (2024): 215–245.

al-Qunayʿīr, Fāris, et al. *Khawārizmiyyāt al-Dhakāʾ al-Iṣṭināʿī fī Taḥlīl al-Naṣṣ al-ʿArabī (AI Algorithms in Arabic Text Analysis)*. Edited by ʿAbd Allāh al-Fayfī. Riyadh: King Abdullah bin Abdulaziz International Center for the Arabic Language, 2019.

al-Sāmarrāʾī, Fāḍil. *Maʿānī al-Naḥw (Meanings of Grammar)*. Jordan: Dār al-Fikr, 2000.

al-Sāmarrāʾī, Ibrāhīm. *al-Fiʿl: Zamānuhu wa Abniyatuhu (The Verb: Its Tense and Forms)*. Beirut: Muʾassasat al-Risāla, 2013.

Anīs, Ibrāhīm. *Min Asrār al-Lugha (Secrets of the Language)*. 6th ed. Cairo: Maktabat al-Anjلو al-Miṣriyya.

Antoun, W., Baly, F., & Hajj, H. "AraBERT: Transformer-Based Model for Arabic Language Understanding". Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille: European Language Resource Association, (2020): pp. 9-15.

ʿAṭiyya, Muḥammad, Aḥmad Rāghib, et al. *al-ʿArabiyya wa al-Dhakāʾ al-Iṣṭināʿī (Arabic and Artificial Intelligence)*. Edited by al-Muʿtazz bi-llāh al-Saʿīd. Riyadh: King

Abdullah bin Abdulaziz International Center for the Arabic Language, 2019.

Attia, M. “Arabic Tokenization System”. Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages, Prague: Association for Computational Linguistics ,(2007): pp. 65-72.

Blodgett, S., Barocas, S., Daumé III, H., & Wallach, H. “Language (Technology) is Power: A Critical Survey of 'Bias' in NLP”. *arXiv*, (2020): pp. 1-23. Retrieved from: https://arxiv.org/abs/2005.14050.

Chowdhary, K. R. “Fundamentals of Artificial Intelligence”. New Delhi: Springer, (2020).

Darwish, K., & Mubarak, H. “Farasa: A Fast and Furious Segmenter for Arabic”. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, San Diego, California: Association for Computational Linguistics (2016): pp. 11-16.

Elkateb, S., Black, W., Rodríguez, H., Alkhalifa, M., Vossen, P., Pease, A., & Fellbaum, C. “Building a WordNet for Arabic”. Proceedings of the Fifth International Conference on Language Resources and Evaluation, Genoa, Italy: European Language Resources Association ,(2006): pp. 29-34.

Farghaly, A., & Shaalan, K. “Arabic Natural Language Processing: Challenges and Solutions”. ACM Transactions on Asian Language Information Processing, 8(4), (2009): Article 14, pp. 1-22.

Fix Ai Editor, F. A. “Revolutionizing Language Translation: Google’s Neural Machine Translation System”. FXIS.ai, (6 September 2024). Retrieved from: https://fxis.ai/edu/revolutionizinglanguage-translation-googles-neural-machine-translation-system/.

Habash, N. “Introduction to Arabic Natural Language Processing”. (San Rafael, CA: Morgan & Claypool Publishers, 2010).

Ḥasan, ʿAbbās. *al-Naḥw al-Wāfī (Comprehensive Grammar)*. 15th ed. Egypt: Dār al-Maʿārif, 2010.

Ḥassān, Tammām. *al-Lugha al-ʿArabiyya: Maʿnāhā wa Mabnāhā (The Arabic Language: Its Meaning and Structure)*. Cairo: ʿĀlam al-Kutub, 2006.

Ibn Jinnī, ʿUthmān. *al-Khaṣāʾiṣ (The Characteristics)*. Edited by Muḥammad ʿAlī al-Najjār. Cairo: Dār al-Kutub al-Miṣriyya, 1952.

Ibn Jinnī, ʿUthmān. *al-Munṣif (The Fair One)*. Egypt: Dār Iḥyāʾ al-Turāth al-Qadīm, 1954.

Jurafsky, Daniel, & James H. Martin. “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition”. 3rd ed., Stanford: Stanford University Press, (2024).

Kalluri, S. “Deep Learning Based Sentiment Analysis”. Master Thesis, (Sweden: Blekinge Institute of Technology, 2023).

Kathare, A., Reddy, P. V., & Prabhu, S. “Analysis of Elasticsearch in Comparison with MongoDB and Hadoop”. International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, India ,(2017): pp. 717-722.

Khalati, M., & Al-Romany, S. “Dialectal Arabic Processing: Current State and Future Directions”. International Journal of Advanced Computer Science and Applications, 11(8), (2020): pp. 920-928.

Kharbāṭ, ʿAlī Ṭarkhān. “Taʿaddud Waẓīfat al-Adāt al-Naḥwiyya fī al-Qurʾān al-Karīm wa Atharuh fī al-Maʿnā (Multiplicity of Grammatical Particle Functions in the Qurʾān).” *Majallat al-Qādisiyya fī al-Ādāb wa al-ʿUlūm al-Tarbawiyya*, no. 3, pt. 1 (2023): 1–30.

Luger, E., & Sellen, A. “Like Having a Really Bad PA: The Gulf between User Expectation and Experience of

Conversational Agents". Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, (2016): pp. 5286-5297.

Manning, C. D., & Schütze, H. "Foundations of Statistical Natural Language Processing". (Cambridge, MA: MIT Press, 1999).

Manohar, V., Povey, D., & Khudanpur, S. "JHU Kaldi System for Arabic MGB-3 ASR Challenge Using Diarization, Audio-Transcript Alignment and Transfer Learning". IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan,(2017): pp. 346-352.

Medhat, W., Hassan, A., & Korashy, H. "Sentiment Analysis Algorithms and Applications: A Survey". *Ain Shams Engineering Journal*, (2014): pp. 1-22.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. "Efficient Estimation of Word Representations in Vector Space". arXiv preprint, (2013). Retrieved from: https://arxiv.org/abs/1301.3781.

Minsky, Marvin, ed. "Semantic Information Processing". (Cambridge, MA: MIT Press, 1968).

Murhaf, Fares and Touileb Samia. "BabelBot at AraFinNLP2024: Fine-Tuning T5 for Multi-Dialect Intent Detection with Synthetic Data and Model Ensembling." In Proceedings of ArabicNLP,. Bangkok, Thailand: Association for Computational Linguistics, (2024): pp. 433–440. https://aclanthology.org/2024.arabicnlp-1.40.

Nallapati, R., Zhou, B., dos Santos, C., Gülçehre, Ç., & Xiang, B. "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond". Conference on Computational Natural Language Learning, (Berlin: Association for Computational Linguistics ,2016): pp. 280-290.

Nassif, A. B., Elnagar, A., Shahin, I., & Henno, S. "Deep Learning for Arabic Subjective Sentiment Analysis: Challenges and Research Opportunities". ScienceDirect,

(January 2021). Retrieved from: https://www.sciencedirect.com/science/article/abs/pii/S1568494620307742.

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Go, F., Inoue, N., Eryani, F., Erdmann, A., & Habash, N. "CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing". Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France: European Language Resources Association (2020): pp. 7022-7032.

Otten, M. "Advanced Natural Language Processing with spaCy: A Guide to Understanding and Building NLP Applications". (Birmingham, UK: Packt Publishing, 2023).

Palanivinayagam, A., El-Bayeh, C. Z., & Damaševičius, R. "Twenty Years of Machine Learning-Based Text Classification: A Systematic Review". Algorithms, (2023): pp. 1-28.

Pasha, A., Al-Badrashiny, M., Diab, M., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic". Proceedings of the Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland: European Language Resources Association (2014): pp.1094-1101.

Restak, M. "Cloud Natural Language API Documentation". (Mountain View, California: Google Cloud, 2025). Retrieved from: https://cloud.google.com/natural-language/docs.

Sarawagi, S. "Information Extraction". Foundations and Trends® in Databases, (2008).

Sengupta, Neha, Sunil Kumar Sahu, Bokang Jia, et al. "Jais and Jais-chat: Arabic-Centric Foundation and Instruction-Tuned Open Generative Large Language Models." arXiv,

September 29 (2023). https://doi.org/10.48550/arXiv.2308.16149.

spaCy: Industrial-Strength NLP. “spaCy Usage Documentation”. (Berlin: Explosion AI, 2025). Retrieved from: https://spacy.io/usage.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … Dean, J. “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation”. Computer Science, (2016): pp. 1-23.

Zhang, L., Wang, S., & Liu, B. “Deep Learning for Sentiment Analysis: A Survey”. Wiley Interdisciplinary Reviews, (2018): pp. 1-25.