الذكاء الاصطناعي في بناء الموارد اللغوية العربية ودعم البحث اللساني

Artificial intelligence in building Arabic language resources and supporting linguistic research

أ.قراوي دلال*

جامعة تيبازة (الجزائر)

dallalgraoui91@gmail.com

تاريخ الإرسال: 2025/12/26	تاريخ: القبول: 2025/12/28	تاريخ النشر: 2026/06/30

الملخص: تهدف هذه الدراسة إلى إبراز دور الذكاء الاصطناعي في بناء الموارد اللغوية العربية، وبيان إسهام هذه الموارد في دعم البحث اللساني المعاصر بمستوياته النظرية والتطبيقية، من خلال تحليل الظواهر الصرفية والنحوية والدلالية، وتطوير اللسانيات التطبيقية، خاصة في مجالات تعليم اللغة العربية والترجمة الآلية وتحليل الخطاب. وفي المقابل، ناقشت الدراسة جملة من التحديات التي تواجه بناء الموارد اللغوية العربية، مثل ندرة البيانات المشروحة، وغياب التوحيد المعياري، وتعقيد البنية اللغوية للعربية.

وخلصت الدراسة إلى أن الذكاء الاصطناعي يشكّل رافعة أساسية لتطوير الموارد اللغوية العربية ودعم البحث اللساني، شريطة اعتماد مقاربة تكاملية تجمع بين التحليل الآلي والرؤية اللسانية، وتعزيز التعاون بين التخصصات والمشاريع المؤسسية في هذا المجال.

كلمات مفتاحية: الذكاء الاصطناعي، الموارد اللغوية العربية، البحث اللساني، المدونات اللغوية، الشبكات الدلالية، التنمية اللغوية الرقمية.

Abstract: This study aims to highlight the role of artificial intelligence in building Arabic linguistic resources and to demonstrate the contribution of these resources to supporting contemporary linguistic research at its theoretical and applied levels.

This is achieved through the analysis of morphological, syntactic, and semantic phenomena, and the development of applied linguistics, particularly in the fields of Arabic language teaching, machine translation, and discourse analysis. Conversely, the study discussed a number of challenges facing the development of Arabic language resources, such as the scarcity of annotated data, the lack of standardization, and the complexity of the Arabic language structure.

Keywords :Artificial intelligence, Arabic language resources, linguistic research, linguistic corpora, semantic networks, digital language development.

1.مقدمة :

تشهد الدراسات اللغوية في الربع الأول من القرن الحادي والعشرين تحولاً جذرياً بفعل التطورات المتسارعة في مجال الذكاء الاصطناعي **(AI)** وتطبيقاته في معالجة اللغات ليشمل العلوم الإنسانية واللغوية، وبخاصة مجال اللسانيات. فقد أتاح الذكاء الاصطناعي إمكانات غير مسبوقة في بناء الموارد اللغوية الرقمية، وتحليل البيانات اللغوية الضخمة، واستخلاص الأنماط اللسانية بدقة وسرعة تفوق العمل البشري التقليدي.

وتُعد الموارد اللغوية – مثل المدونات النصية، والمعاجم الحاسوبية، وقواعد البيانات الصرفية والنحوية، والمستودعات الدلالية – العمود الفقري لأي بحث لساني معاصر، إذ تمثل المادة الخام التي تُبنى عليها التحليلات اللسانية النظرية والتطبيقية.

في هذا السياق، تواجه اللغة العربية تحديات خاصة ناتجة عن خصائصها البنيوية، من اشتقاق صرفي غني، وتعدد في المستويات الأسلوبية، وازدواجية لغوية، وتنوع لهجي واسع، مما يجعل بناء موارد لغوية عربية دقيقة وشاملة أمرًا معقدًا، لكنه في الوقت ذاته مجالًا خصبًا لتوظيف تقنيات الذكاء الاصطناعي.

الإشكالية الرئيسية:

تتمحور الإشكالية حول التفاوت الواضح بين التطور الهائل في الموارد اللغوية للغات العالمية المدعومة بتقنيات الذكاء الاصطناعي المتقدمة، وبين الوضع المتأخر نسبياً للموارد اللغوية العربية. هذا التفاوت يخلق تحديات مزدوجة :بنيوية (طبيعة اللغة العربية المعقدة) وتقنية (نقص البيانات اللغوية المعلمة).

وهنا يبرز **السؤال الجوهري** الذي تطرحه هذه الدراسة هو:

كيف يمكن تسخير تقنيات الذكاء الاصطناعي المتقدمة لبناء موارد لغوية عربية شاملة، وكيف يمكن لهذه الموارد أن تدعم وتطور البحث اللساني النظري والتطبيقي؟

تتفرع عن الإشكالية الرئيسة **الأسئلة الفرعية** الآتية:

1. ما المقصود بالموارد اللغوية في الدرس اللساني المعاصر؟
2. ما هي أنواع الموارد اللغوية العربية التي يمكن بناؤها أو تطويرها باستخدام الذكاء الاصطناعي؟
3. كيف تؤثر الموارد اللغوية المُعزَّزة بالذكاء الاصطناعي على منهجيات البحث اللساني العربي؟
4. ما أبرز التحديات العلمية والمنهجية التي تعيق هذا التوظيف؟
5. ما هي آفاق المستقبل وآثار هذه التطورات على حفظ اللغة العربية وتطويرها؟

أهداف الدراسة:

تهدف هذه الدراسة إلى:

- ✓ إبراز الدور المحوري للذكاء الاصطناعي في بناء الموارد اللغوية العربية.
- ✓ تحليل أثر هذه الموارد في دعم البحث اللساني بمستوياته المختلفة.
- ✓ رصد أهم الإشكالات التي تواجه تطوير الموارد اللغوية العربية آليًا.
- ✓ اقتراح آفاق بحثية مستقبلية لتطوير موارد لغوية عربية ذكية.

أهمية الدراسة:

تكمن أهمية الدراسة في:

- ✓ سد فجوة بحثية في الدراسات العربية التي تكمن حول تقاطع الذكاء الاصطناعي واللسانيات.
- ✓ تقديم رؤية عملية لتطوير أدوات لغوية عربية متقدمة.
- ✓ المساهمة في الحفاظ على اللغة العربية والعمل على تعزيز حضورها الفعال في المجال الرقمي.
- ✓ إفادتها للباحثين والاختصاصيين في اللسانيات، واللغة العربية، والمعالجة الآلية للغة.

منهج الدراسة:

اعتمدت الدراسة المنهج الوصفي التحليلي، مستندة إلى تحليل الأدبيات اللسانية والحاسوبية الحديثة، وإلى استقراء نماذج ومشاريع معاصرة في مجال المعالجة الآلية للغة العربية. وقد تناولت الدراسة الإطار المفاهيمي للذكاء الاصطناعي واللسانيات الحاسوبية، وأنواع الموارد اللغوية العربية، ثم ركزت على آليات بناء هذه الموارد باستخدام تقنيات التعلم الآلي والتعلم العميق، مثل بناء المدونات اللغوية، والمعاجم الحاسوبية، والموارد المشروحة.

1. الإطار المفاهيمي والنظري للدراسة:

1.1. الذكاء الاصطناعي: المفهوم والنشأة والتطور:

يعد الذكاء الاصطناعي من أبرز مجالات البحث العلمي المعاصر، وقد ارتبط ظهوره بمحاولة محاكاة القدرات العقلية البشرية عبر أنظمة حاسوبية قادرة على التعلم والاستنتاج واتخاذ القرار. ويعرّفه راسل ونورفيغ بأنه "علم وهندسة صناعة آلات ذكية، وبخاصة البرامج الحاسوبية القادرة على أداء مهام تتطلب عادة ذكاءً بشريا" [1]

وقد مرّ الذكاء الاصطناعي بمراحل تطور متعددة، بدءًا من النماذج الرمزية القائمة على القواعد، وصولًا إلى تقنيات **التعلم الآلي (Machine Learning)** و**التعلم العميق**

(Deep Learning) التي أحدثت نقلة نوعية في معالجة البيانات الضخمة، ومن بينها البيانات اللغوية. [2]

ومع تطور القدرة الحاسوبية وتوفر البيانات الرقمية، أصبح الذكاء الاصطناعي أكثر قدرة على معالجة اللغات الطبيعية، الأمر الذي انعكس مباشرة على تطور اللسانيات الحاسوبية وبناء الموارد اللغوية.

2.1. الذكاء الاصطناعي ومعالجة اللغات الطبيعية:

معالجة اللغات الطبيعية **(Natural Language Processing)** أحد الفروع التطبيقية الأساسية للذكاء الاصطناعي، وتهدف إلى تمكين الحاسوب من فهم اللغة البشرية وتحليلها وإنتاجها. وتشمل هذه المعالجة عدة مستويات لغوية: الصرف، النحو، الدلالة، والتداولية. [3]

وترتبط معالجة اللغات الطبيعية ارتباطًا وثيقًا بعدة مجالات علمية، في مقدمتها اللسانيات التي تمدّها بالنظريات اللغوية.

وقد ساهم الذكاء الاصطناعي في تجاوز المقاربات التقليدية لمعالجة اللغة، التي كانت تعتمد على القواعد اليدوية، إلى مقاربات تعتمد على التعلم من البيانات اللغوية الضخمة، مما أتاح بناء موارد لغوية أكثر دقة ومرونة. [4]

وفي هذا الإطار، برزت النماذج اللغوية الضخمة **(Large Language Models)** التي تعتمد على الشبكات العصبية العميقة، وأسهمت في تحسين جودة التحليل اللغوي، خاصة في اللغات ذات البنية المعقدة مثل العربية.

3.1. اللسانيات الحاسوبية: المفهوم والمجال:

تعد اللسانيات الحاسوبية مجالًا بينيًا يجمع بين اللسانيات النظرية وعلوم الحاسوب، ويهدف إلى دراسة اللغة الطبيعية من خلال نماذج حسابية قادرة على محاكاة الظواهر اللغوية. [5]

ولا تقتصر اللسانيات الحاسوبية على الجانب التطبيقي فقط، بل تسهم أيضًا في تطوير البحث اللساني النظري، من خلال اختبار الفرضيات اللغوية على مدونات واسعة، وتحليل الترددات والأنماط اللغوية بصورة كمية دقيقة. (6)

وقد أدت هذه المقاربة إلى تحول في منهجية البحث اللساني، حيث لم يعد الباحث يعتمد فقط على الأمثلة المصطنعة أو الحدسية، بل أصبح يستند إلى بيانات لغوية واقعية مستمدة من موارد رقمية ضخمة.

4.1. الموارد اللغوية: المفهوم والأهمية في البحث اللساني:

تشير الموارد اللغوية إلى مجموع الأدوات والبيانات الرقمية التي تستخدم في تحليل اللغة ومعالجتها، وتشكل الأساس المادي للبحث اللساني الحديث. ويؤكد هاباش أن الموارد اللغوية تمثل "العنصر الحاسم في نجاح أي مشروع لمعالجة اللغة العربية آليا".

وتكمن أهمية الموارد اللغوية في: (7)

- دعم البحث اللساني النظري عبر تحليل الظواهر اللغوية واقعيا
- تطوير التطبيقات اللغوية (الترجمة الآلية، التعليم، التحليل النصي)
- توحيد المعايير اللغوية وتحسين جودة الدراسات المقارنة

5.1. أنواع الموارد اللغوية:

1.5.1. المدونات اللغوية (Corpora)

المدونات اللغوية هي مجموعات نصية رقمية كبيرة، تُجمع وفق معايير علمية محددة، وتُستخدم لتحليل اللغة في استعمالها الفعلي. وقد أصبحت المدونات أداة مركزية في البحث اللساني المعاصر. (8)

وتنقسم المدونات إلى:

- مدونات عامة
- مدونات متخصصة
- مدونات مشروحة لغويا (Morphological, Syntactic Annotation)

وتعد المدونات العربية أقل عددًا وأصغر حجمًا مقارنة بنظيراتها في اللغات الأخرى، مما يشكل تحديًا أمام البحث اللساني العربي.

2.5.1. المعاجم الحاسوبية:

تمثل المعاجم الحاسوبية موارد لغوية رقمية تحتوي على معلومات صرفية ودلالية وتركيبية، وتُستخدم في التحليل اللغوي الآلي وفي البحث اللساني. وقد أسهم الذكاء الاصطناعي في تطوير هذه المعاجم من خلال تقنيات الاستخراج الآلي للمفردات والعلاقات الدلالية. [9]

3.5.1. الموارد المشروحة (Annotated Resources):

تشمل الموارد المشروحة تلك النصوص التي أُضيفت إليها طبقات من التحليل اللغوي، مثل الوسم الصرفي والنحوي والدلالي. وتُعد هذه الموارد ذات قيمة عالية في الدراسات اللسانية التطبيقية والحاسوبية. [10]

6.1. مكانة الموارد اللغوية في البحث اللساني المعاصر:

أصبح الاعتماد على الموارد اللغوية الرقمية سمة أساسية في البحث اللساني الحديث، حيث أتاحت هذه الموارد إمكانات تحليلية جديدة، وأسهمت في تطوير مناهج البحث الكمي والنوعي على حد سواء. ويؤكد بيبر وآخرون أن "اللسانيات المعاصرة أصبحت علمًا قائمًا على البيانات بامتياز". [11]

2. تقنيات الذكاء الاصطناعي في خدمة اللغة العربية:

1.2. نماذج اللغة الضخمة وتطبيقاتها العربية:

أحدثت نماذج اللغة الكبيرة **(LLMs)** ثورة في مجال معالجة اللغات الطبيعية. في السياق العربي، ظهرت عدة نماذج مهمة:

-AraBERT : [12] أحد أول النماذج المسبقة التدريب المخصصة للعربية، مبني على هندسة**BERT**، وحقق نتائج ممتازة في مهام تصنيف النصوص واستخراج المعلومات.

-CAMeL Tools : [13] مجموعة أدوات مفتوحة المصدر طورها فريق جامعة نيويورك، تضم أدوات للتنميط الصرفي والتشكيل التلقائي.

JASMINE : [14] نموذج عربي متعدد اللغات طورته شركة stc، أحد النماذج العربية في حجم البيانات المستخدمة في التدريب.

2.2. التعلم العميق في المعالجة الصرفية والنحوية للعربية:

تستخدم تقنيات التعلم العميق بنجاح في معالجة التحديات الصرفية والنحوية للعربية. تشمل التطبيقات:

- **التشكيل التلقائي**: استخدام الشبكات العصبية المتكررة (RNNs) والشبكات الانتباهية للتشكيل التلقائي للنصوص العربية، مع دقة تصل إلى 97% في النصوص الفصحى. [15]
- **التنميط الصرفي** :تحليل الكلمات العربية إلى مكوناتها الصرفية باستخدام نماذج Sequence-to-Sequence
- **التحليل النحوي التلقائي** :بناء أشجار نحوية تلقائية للنصوص العربية، مع تحديات خاصة ناتجة عن حرية الترتيب النسبي في العربية.

3.2. تقنيات توليد اللغة وترجمتها الآلية للعربية:

حققت تقنيات الترجمة الآلية للعربية تقدماً ملحوظاً بفضل النماذج التحويلية (Transformers) تشير دراسة Rogers وآخرون (2023) إلى أن جودة الترجمة من وإلى العربية تحسنت بنسبة 40% في السنوات الخمس الماضية. [16]

في مجال توليد النصوص، بدأت النماذج العربية تظهر قدرات ملحوظة في كتابة المقالات والتقارير وتلخيص النصوص الطويلة والإجابة على الأسئلة باللغة العربية.

3. بناء الموارد اللغوية العربية بالذكاء الاصطناعي:

1.3. قواعد البيانات والمدونات اللغوية المُعلمة آليا:

شهدت السنوات الأخيرة ظهور عدة مدونات لغوية عربية مهمة:

-**Arabic Treebank** مشروع مستمر لتوسيع قاعدة البيانات النحوية العربية المُعلَّمة، يحتوي حالياً على أكثر من 100,000 جملة مُحلَّلة نحويا. [17]

-**ArCOV-19** مدونة ضخمة جمعت تلقائياً من وسائل التواصل الاجتماعي خلال جائحة كوفيد-19، تحتوي على أكثر من 90 مليون تغريدة عربية، وتُستخدم في دراسة اللغة الطبيعية في السياقات الطارئة. [18]

-**SANAD** مدونة عربية كبيرة جمعت من مصادر إخبارية عربية، تحتوي على أكثر من 200 مليون كلمة، وتستخدم في تدريب نماذج اللغة العربية. [19]

2.3. المعاجم والشبكات الدلالية الذكية:

تستخدم تقنيات الذكاء الاصطناعي في بناء وتطوير المعاجم العربية:

-**Arabic WordNet** شبكة دلالية للغة العربية، تربط بين المفاهيم والكلمات، وقد تم توسيعها مؤخراً باستخدام تقنيات التعلم الآلي لاكتشاف العلاقات الدلالية الجديدة.

-**ArSenL** معجم للمشاعر (**Sentiment Lexicon**) للغة العربية، بُني باستخدام تقنيات التعلم العميق لتصنيف الكلمات العربية حسب مشاعرها الإيجابية أو السلبية أو المحايدة.

3.3. أنظمة التخاطب الآلي وتحليل المشاعر للغة العربية:

تشهد تطبيقات التخاطب الآلي وتحليل المشاعر تطوراً سريعاً في السياق العربي:

- ظهور مساعدين افتراضيين باللغة العربية في قطاعات الخدمات المصرفية والرعاية الصحية.
- تطور أدوات تحليل المشاعر التي تأخذ في الاعتبار الخصوصيات الثقافية واللغوية للعالم العربي.
- تطبيقات في مراقبة الرأي العام وتحليل اتجاهات المستهلكين.

4.3. منصات الذكاء الاصطناعي مفتوحة المصدر للباحثين العرب:

ساهمت منصات مفتوحة المصدر مثل **Hugging Face** في ديمقراطية الوصول إلى نماذج الذكاء الاصطناعي للغة العربية. تحتوي المنصة حالياً على أكثر من 500 نموذج ومورد عربي، متاحة للباحثين والمطورين العرب مجانا. [20]

4. **دعم البحث اللساني بالذكاء الاصطناعي**

1.4. تحولات في المنهج اللساني

أدت تقنيات الذكاء الاصطناعي إلى تحول جذري في منهجية البحث اللساني:

1.1.4. التحليل الكمي الكبير :(Big Data Analysis) إمكانية تحليل ملايين النصوص في وقت قصير، مما يسمح باكتشاف أنماط لغوية كانت خفية سابقا.

2.1.4. اللسانيات التكوينية :(Corpus Linguistics) تطورت بشكل كبير مع توفر مدونات لغوية ضخمة وأدوات تحليل متقدمة.

3.1.4. اللسانيات الحاسوبية النظرية :استخدام النمذجة الحاسوبية لاختبار النظريات اللغوية وتطويرها.

2.4. تطبيقات في فروع اللسانيات المختلفة:

1.2.4. في الصوتيات والمختبرات الافتراضية: تسمح تقنيات التعلم العميق بمحاكاة وتحليل الخصائص الصوتية للعربية بدقة عالية.

2.2.4. في التحليل النحوي التلقائي للنصوص التاريخية: استخدام نماذج اللغة المدربة على العربية الحديثة لفهم وتحليل النصوص التراثية.

3.2.4. في دراسة التغير الدلالي: استخدام تقنيات تضمين الكلمات لدراسة تطور معاني الكلمات العربية عبر الزمن.

5. التحديات والآفاق المستقبلية لبناء الموارد اللغوية العربية بالذكاء الاصطناعي:

1.5. ندرة الموارد العربية المشروحة عالية الجودة:

تعد ندرة الموارد اللغوية العربية المشروحة **(Annotated Resources)** من أبرز العوائق أمام توظيف الذكاء الاصطناعي في البحث اللساني. فالنماذج الذكية، ولا سيما

القائمة على التعلم العميق، تتطلب كميات كبيرة من البيانات الدقيقة والمتوازنة، وهو ما لا يتوفر بالقدر الكافي في السياق العربي. [21]

وتشير دراسات حديثة إلى أن كثيرًا من الموارد المتاحة تعاني من:

- محدودية الحجم
- ضعف التغطية الأسلوبية والمجالية
- غياب التحديث الدوري

وهو ما يؤثر مباشرة في دقة النماذج اللغوية ونتائج التحليل اللساني. [22]

2.5. إشكالية التوحيد المعياري:

تعاني مشاريع بناء الموارد اللغوية العربية من غياب معايير موحّدة في:

- وسم الظواهر الصرفية والنحوية
- توصيف البنى الدلالية
- تمثيل اللهجات العربية

ويؤدي هذا التباين إلى صعوبة دمج الموارد المختلفة أو مقارنتها، مما يحدّ من إمكانات البحث اللساني المقارن ويؤثر في قابلية إعادة استخدام البيانات.

3.5. التعقيد البنيوي للعربية:

يمثل التعقيد الصرفي والنحوي للغة العربية تحديًا مستمرًا أمام النماذج الذكية. فظواهر مثل الاشتقاق، والإعراب، والتعدد الدلالي، والازدواجية اللغوية تتطلب نماذج قادرة على التعامل مع مستويات لغوية متداخلة. [23]

وعلى الرغم من التطور الملحوظ في النماذج العصبية، لا تزال بعض الظواهر اللغوية العربية عصيّة على المعالجة الآلية الدقيقة، ما يستدعي تدخلًا لسانيًا بشريًا في مراحل التصميم والتقييم.

4.5. محدودية البيانات المتو افرة :

لا تزال البيانات اللغوية العربية المُعلَّمة محدودة مقارنة بالإنجليزية، خاصة في المجالات التخصصية. [24]

5.5. قضايا الخصوصية والأخلاقيات

:تحديات جمع البيانات من وسائل التواصل الاجتماعي مع الحفاظ على خصوصية المستخدمين، ومشاكل التحيز في النماذج اللغوية.

6.5. الفجوة بين الفصحى والعاميات :

يقصد بالفجوة بين الفصحى والعاميات هو الحاجة إلى تطوير نماذج قادرة على التعامل مع التنوع اللغوي في العالم العربي

6. الآفاق المستقبلية لتوظيف الذكاء الاصطناعي في البحث اللساني العربي:

1.6. تطوير موارد لغوية عربية مفتوحة:

يمثل التوجه نحو **الموارد اللغوية المفتوحة (Open Linguistic Resources)** أحد أهم الآفاق المستقبلية، إذ يتيح للباحثين الوصول إلى بيانات لغوية موثوقة وقابلة لإعادة الاستخدام، ويسهم في تسريع وتيرة البحث اللساني. [25]

2.6. تعزيز التكامل بين التحليل الآلي والرؤية اللسانية:

تؤكد الدراسات الحديثة ضرورة عدم الفصل بين الذكاء الاصطناعي والتحليل اللساني، بل اعتماد مقاربة تكاملية تجمع بين: [26]

- دقة النماذج الحاسوبية
- عمق التحليل اللساني النظري

وهو ما يضمن نتائج علمية أكثر موثوقية وتفسيرا.

3.6. إدماج اللهجات العربية في الموارد اللغوية:

يمثل إدماج اللهجات العربية في الموارد اللغوية المبنية بالذكاء الاصطناعي أفقا بحثيا واعدا، لما له من أثر في:

- توسيع قاعدة البيانات اللغوية

- دعم الدراسات السوسيولسانية
- توفير أدوات لغوية ذكية
- تطوير تطبيقات لغوية أكثر واقعية

وتشير الأبحاث الحديثة إلى إمكانات كبيرة لتوظيف التعلم العميق في هذا المجال.

4.6. توظيف الموارد في البحث اللساني الأكاديمي:

من الآفاق المستقبلية المهمة تعميم استخدام الموارد اللغوية الرقمية في البحوث اللسانية الجامعية، سواء في المذكرات أو الأطروحات أو المقالات العلمية، بما يسهم في تحديث مناهج البحث اللساني العربي.

7. الخاتمة العامة:

خلصت هذه الدراسة إلى أن الذكاء الاصطناعي لم يعد مجرد أداة تقنية مساعدة، بل أصبح عنصرا بنيويا في بناء الموارد اللغوية العربية ودعم البحث اللساني المعاصر. وقد أظهرت الفصول الأربعة أن الموارد اللغوية المبنية بالذكاء الاصطناعي تشكل قاعدة معرفية أساسية لتطوير اللسانيات النظرية والتطبيقية والدلالية والتداولية.

كما أبرزت الدراسة أن اللغة العربية، رغم ما تتسم به من تعقيد بنيوي، تمتلك إمكانات كبيرة للاستفادة من تقنيات الذكاء الاصطناعي، شريطة توفر موارد لغوية دقيقة، وتعاون علمي متعدد التخصصات، ودعم مؤسسي مستدام.

وتؤكد الخاتمة أن مستقبل البحث اللساني العربي مرهون بمدى قدرته على استثمار الموارد اللغوية الرقمية وتوظيف الذكاء الاصطناعي توظيفًا علميًا واعيًا، يراعي الخصوصية اللغوية ويخدم التطور المعرفي.

النتائج:

أسفرت هذه الدراسة عن مجموعة من **النتائج العلمية** يمكن إجمالها فيما يأتي:

-تأكيد الدور المحوري للذكاء الاصطناعي في بناء الموارد اللغوية العربية:

أظهرت الدراسة أن تقنيات الذكاء الاصطناعي، ولا سيما التعلم الآلي والتعلم العميق، أصبحت عنصرًا أساسيًا في بناء الموارد اللغوية العربية الحديثة، مثل المدونات النصية، والمعاجم الحاسوبية، والموارد المشروحة صرفيًا ونحويًا ودلاليًا. وقد مكّن هذا التطور من تجاوز حدود العمل اليدوي التقليدي، وتسريع عمليات البناء والتحليل.

-تعزيز البحث اللساني القائم على البيانات :

بيّنت النتائج أن الموارد اللغوية المبنية بالذكاء الاصطناعي أسهمت في إحداث تحول منهجي في البحث اللساني العربي، من المقاربات الحدسية والوصفية إلى مقاربات كمية وتجريبية قائمة على تحليل بيانات لغوية واسعة النطاق، مما عزز دقة النتائج وموضوعيتها.

-دعم اللسانيات النظرية والتطبيقية معا :

أثبتت الدراسة أن هذه الموارد لا تقتصر على خدمة التطبيقات اللغوية، بل تؤدي دورًا مهمًا في دعم اللسانيات النظرية، من خلال اختبار الفرضيات الصرفية والنحوية والدلالية، فضلًا عن دورها في اللسانيات التطبيقية، مثل تعليم اللغة العربية، والترجمة الآلية، وتحليل الأخطاء اللغوية.

-كشف الفجوة بين الإمكانات التقنية والو اقع العربي :

أظهرت النتائج وجود فجوة واضحة بين الإمكانات التي تتيحها تقنيات الذكاء الاصطناعي الحديثة وبين واقع الموارد اللغوية العربية المتاحة، سواء من حيث الحجم أو الجودة أو التوحيد المعياري، وهو ما يحدّ من الاستفادة الكاملة من هذه التقنيات في البحث اللساني العربي.

-استمرار التحديات البنيوية والمؤسسية:

خلصت الدراسة إلى أن تعقيد البنية اللغوية للعربية، إلى جانب ضعف التنسيق بين المشاريع البحثية، ومحدودية الدعم المؤسسي، ما زال يشكل عائقًا أمام بناء موارد لغوية عربية شاملة ومستدامة، رغم التقدم التقني الملحوظ.

-أهمية التكامل بين الذكاء الاصطناعي والتحليل اللساني البشري :

أثبتت النتائج أن الذكاء الاصطناعي، رغم قدرته العالية على المعالجة والتحليل، لا يمكن أن يحل محل اللساني، بل يظل أداة داعمة تتطلب توجيهًا نظريًا وتأويليًا من الباحث اللساني لضمان سلامة النتائج ودقتها.

التوصيات:

في ضوء النتائج المتوصل إليها، توصي الدراسة بما يأتي:

-تعزيز بناء الموارد اللغوية العربية المفتوحة:

توصي الدراسة بضرورة دعم مشاريع وطنية وإقليمية تهدف إلى بناء موارد لغوية عربية مفتوحة، عالية الجودة، وقابلة لإعادة الاستخدام، بما يسهم في تطوير البحث اللساني والتطبيقات اللغوية على حد سواء.

-توحيد المعايير اللغوية والمنهجية:

تؤكد الدراسة أهمية العمل على توحيد معايير بناء الموارد اللغوية العربية، خاصة في مجالات الوسم الصرفي والنحوي والدلالي، لضمان قابلية التكامل بين الموارد المختلفة، وتعزيز البحث المقارن.

-تعزيز التكامل بين التخصصات:

توصي الدراسة بضرورة تعزيز التعاون بين اللسانيين، وخبراء الذكاء الاصطناعي، والمعلوماتيين، من خلال فرق بحثية متعددة التخصصات، لضمان بناء موارد لغوية دقيقة ومراعية للخصوصية اللغوية العربية.

-إدماج اللهجات العربية في الموارد اللغوية:

تشدد الدراسة على أهمية إدراج اللهجات العربية ضمن الموارد اللغوية المبنية بالذكاء الاصطناعي، لما لذلك من أثر في توسيع قاعدة البيانات اللغوية، ودعم الدراسات السوسيولسانية، وتطوير تطبيقات لغوية أكثر واقعية.

-إدماج الموارد اللغوية الرقمية في البحث الأكاديمي:

توصي الدراسة بتشجيع الباحثين وطلبة الدراسات العليا على اعتماد الموارد اللغوية الرقمية في بحوثهم اللسانية، وتدريبهم على استثمار أدوات الذكاء الاصطناعي في التحليل اللغوي.

-تطوير برامج تكوين متخصصة:

تقترح الدراسة إدراج مساقات وبرامج تكوين جامعية تجمع بين اللسانيات والذكاء الاصطناعي، بما يسهم في إعداد جيل من الباحثين القادرين على توظيف التقنيات الحديثة في خدمة اللغة العربية.

الاحالات والهوامش:

1. Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2018). *Deep Learning*.
3. Jurafsky, D., & Martin, J. (2023). *Speech and Language Processing*.
4. Manning, C., et al. (2019). *Introduction to Information Retrieval*.
5. AlKhalil, M., & El-Haj, M. (2022). Arabic Natural Language Processing: Challenges and Solutions. ACM Computing Surveys.
6. Biber, D., Reppen, R., & Conrad, S. (2021). *Corpus Linguistics*.
7. Habash, N. (2022). Introduction to Arabic Natural Language Processing.
8. Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. Proceedings of the 12th Language Resources and Evaluation Conference.
9. Atkins, B. T., & Rundell, M. (2018). The Oxford Guide to Practical Lexicography.
10. Pustejovsky, J., & Stubbs, A. (2019). Natural Language Annotation for Machine Learning.
11. Biber, D., Reppen, R., & Conrad, S. (2021). Corpus Linguistics.
12. Antoun, W., et al. (2020). AraBERT: Transformer-based Model for Arabic.
13. Obeid, O., et al. (2020). CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. Proceedings of the 12th Language Resources and Evaluation Conference.
14. Zhang, S., et al. (2022). OPT: Open Pre-trained Transformer Language Models. arXiv preprint.

15. القحطاني، عبدالله. (2021). المدونات اللغوية الحاسوبية العربية. مجلة الحاسب الآلي واللغة، 8(3)، 112-130.

16. Rogers, A., et al. (2023). The People's Language: A Large-Scale Study of Arabic Dialects. Computational Linguistics.
17. Zaghouani, W. (2020). Critical Survey of the Freely Available Arabic Corpora. Proceedings of the 2nd Workshop on Open-Source Arabic Corpora and Processing Tools.
18. Darwish, K., et al. (2021). Arabic Natural Language Processing in the Wild: A Case Study on Twitter. Transactions on Asian and Low-Resource Language Information Processing.

19. الجهني، محمد. (2023). الذكاء الاصطناعي واللغة العربية: التحديات والفرص. مجلة علوم اللغة، 15(2)، 45-67.

20. Hugging Face للنماذج العربية. (2024). تم الاسترداد من https://huggingface.co/models?filter=ar
21. Habash, N. (2020). Arabic Natural Language Processing: History and Future. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.
22. Pustejovsky, J., & Stubbs, A. (2019). Natural Language Annotation for Machine Learning.
23. Goldberg, Y. (2023). Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies, 16(2), 1-309.
24. Zaghouani, W. (2020). Critical Survey of the Freely Available Arabic Corpora. Proceedings of the 2nd Workshop on Open-Source Arabic Corpora and Processing Tools.
25. Rogers, A., et al. (2023). The People's Language: A Large-Scale Study of Arabic Dialects. Computational Linguistics.

26. السبيعي، خالد. (2022). معالجة اللغة الطبيعية العربية: الواقع والطموح. الرياض: دار جامعة الملك سعود.

قائمة المراجع:

1-المراجع باللغة العربية:

- الجهني، محمد. (2023). الذكاء الاصطناعي واللغة العربية: التحديات والفرص. مجلة علوم اللغة، 15(2)، 45-67.
- القحطاني، عبدالله. (2021). المدونات اللغوية الحاسوبية العربية. مجلة الحاسب الآلي واللغة، 8(3)، 112-130.

- السبيعي، خالد. (2022). معالجة اللغة الطبيعية العربية: الواقع والطموح. الرياض: دار جامعة الملك سعود.
- النماذج العربية (2024) من الموقع: https://huggingface.co/models?filter=ar

2-المراجع باللغة الأجنبية:

- **AlKhalil, M & ,.El-Haj, M** .(2022) .*Arabic Natural Language Processing: Challenges and Solutions* .ACM Computing Surveys.
- **Antoun, W., et al** .(2020) .*AraBERT: Transformer-based Model for Arabic*.
- **Antoun, W., Baly, F & ,.Hajj, H** .(2020) .*AraBERT: Transformer-based Model for Arabic Language Understanding* .Proceedings of the 12th Language Resources and Evaluation Conference.
- **Atkins, B. T & ,.Rundell, M** .(2018) .*The Oxford Guide to Practical Lexicography*.
- **Darwish, K., et al** .(2021) .*Arabic Natural Language Processing in the Wild: A Case Study on Twitter* .Transactions on Asian and Low-Resource Language Information Processing.
- **Goldberg, Y** .(2023) .*Neural Network Methods for Natural Language Processing* .Synthesis Lectures on Human Language Technologies, 16(2), 1–309.
- **Goodfellow, I., Bengio, Y & ,.Courville, A** .(2018) .*Deep Learning*.
- **Habash, N** .(2020) .*Arabic Natural Language Processing: History and Future* .Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.
- **Jurafsky, D & ,.Martin, J** .(2023) .*Speech and Language Processing*.
- **Obeid, O., et al** .(2020) .*CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing* .Proceedings of the 12th Language Resources and Evaluation Conference.
- **Pustejovsky, J & ,.Stubbs, A** .(2019) .*Natural Language Annotation for Machine Learning*.
- **Rogers, A., et al** .(2023) .*The People's Language: A Large-Scale Study of Arabic Dialects* . Computational Linguistics.
- **Russell, S & ,.Norvig, P** .(2021) .*Artificial Intelligence: A Modern Approach*.
- **Zaghouani, W** .(2020) .*Critical Survey of the Freely Available Arabic Corpora* .Proceedings of the 2nd Workshop on Open-Source Arabic Corpora and Processing Tools.

- **Zhang, S., et al** .(2022) .*OPT: Open Pre-trained Transformer Language Models* .arXiv preprint.