الذكاء الاصطناعي والتمازج المعرفي
" نمذجة المحتوى المفاهيمي في المعاجم العربية الرقمية بين آفاق التطوير وتحديات التعريف"

نادية سيد عبد الواحد عبد المجيد*
Nsa02@fayoum.edu.eg

ملخص

تقدم الورقة تحليلًا نظريًا لعملية نمذجة المحتوى المفاهيمي في المعاجم العربية الرقمية، مع التركيز على دور الذكاء الاصطناعي وآليات التمازج المعرفي. وتستكشف كيف يمكن للتقنيات المعاصرة، ولا سيما الذكاء الاصطناعي، أن تسهم في تطوير بنية المداخل المعجمية، وتحسين سبل الاسترجاع، وتعزيز دقة التعريفات وسياقاتها. وعبر فحص نقدي لعينة من الأوراق ذات الشأن، يكشف البحث عن الفجوات القائمة في تمثيل المحتوى المفاهيمي، ويقترح أطرًا منهجية لتطوير أكثر فاعلية. كما يستشرف البحث آفاق بناء معاجم عربية رقمية أكثر ذكاءً ووعيًا دلاليًا، قادرة على تلبية احتياجات المستخدمين في بيئات رقمية متسارعة. وتعتمد الدراسة المنهج الوصفي التحليلي، سعيًا إلى الإسهام في ترقية الصناعة المعجمية الرقمية ضمن سياق التمازج المعرفي والتطور التقني

الكلمات المفتاحية: الذكاء الاصطناعي، المعاجم العربية الرقمية، التحديات اللغوية، الفرص التقنية، نمذجة المحتوى المفاهيمي، التمازج المعرفي.

* مدرس بقسم اللغة العربية وآدابها – كلية الآداب- جامعة الفيوم.

المقدمة

شهدت العقود الأخيرة تحولات جوهرية في مجال تطوير المعاجم، تمثلت في الانتقال من النسخ الورقية التقليدية إلى النسخ الإلكترونية التفاعلية، مدفوعة بالتطور التكنولوجي وانتشار الإنترنت. ولم يقتصر هذا التحول على تغيير وسيط النشر، وإنما شمل أساليب تنظيم المحتوى اللغوي وعرضه، لا سيما في ظل التقدم في تقنيات الذكاء الاصطناعي التي أسهمت في إعادة صياغة أنماط التواصل وإدارة المعرفة. وفي هذا السياق، تبرز نمذجة المحتوى المفاهيمي بوصفها أداة محورية في تطوير المعاجم الرقمية، ولا سيما تلك المخصصة للغة العربية، نظرًا لما تتميز به من بنية نحوية وصرفية ودلالية فريدة.

وقد أسهم التكامل بين الأساليب اللغوية العميقة (مثل الصرف والنحو) والتقنيات الحاسوبية المتقدمة (مثل التعلم العميق والشبكات الدلالية) في تحويل المعاجم من قوائم ثابتة إلى كيانات ديناميكية قادرة على التفاعل مع مختلف السياقات. ومع ذلك، تفرض طبيعة اللغة العربية (كونها لغة اشتقاقية متعددة التصريفات واللهجات) تحديات معقدة أمام التمثيل الرقمي لمفاهيمها، مما يتطلب توازناً دقيقاً بين الإمكانات التي يوفرها الذكاء الاصطناعي (مثل التحديث التلقائي وتوسيع السياقات الدلالية) والتحديات المرتبطة بغياب المعايير الموحدة وندرة البيانات المهيكلة.

وتواجه المعاجم العربية الرقمية تحديات تتجاوز مجرد تخزين المفردات وتعريفاتها، إذ تتطلب بناء وتطوير نماذج مفاهيمية قادرة على تمثيل العلاقات الدلالية العميقة بين المصطلحات. ومن أبرز الإشكالات صعوبة التعامل مع تنوع الجذور والتراكيب الصرفية والتعدد اللهجي، بالإضافة إلى نقص المشاريع التي تعتمد مقاربات معرفية حديثة في بناء المعاجم.

وتماشيا مع الطابع التحليلي المناسب للبحوث المرتبطة بالسياسات اللغوية والتخطيط التقني تنظر هذه الورقة في إشكالية جوهرية تتمثل في كيفية نمذجة المحتوى المفاهيمي للمعاجم الرقمية العربية ضمن بيئة معرفية ديناميكية تتطلب مرونة في التعريفات، مع الحفاظ على الإطار المعياري للغة. عبر التحقيق في الفرص التي تتيحها تقنيات مثل

الأنطولوجيات والتمثيل الدلالي، ورصد التحديات المرتبطة بالهُويّة اللغوية وصعوبة توحيد المصطلحات بين الدول العربية.

وتهدف إلى تحليل واقع المعجم العربي الرقمي المدعوم بتقنيات الذكاء الاصطناعي، مع تقديم تصور منهجي لرسم خريطة طريق تستند إلى مراجعة الدراسات والأبحاث ذات الصلة وتنظيمها ضمن إطار تحليلي متماسك بما قد يعزّز حضور اللغة العربية في تطبيقات الذكاء الاصطناعي، ويُمكّن من تطوير منصات قائمة على الذكاء الاصطناعي أكثر كفاءة، مثل المعاجم التخصصية، وأنظمة الترجمة الآلية الدقيقة.. كما تُبرز أهمية تضافر جهود الباحثين في علوم اللغة والتقنيات الحاسوبية ضمن تعاون مؤسسي منظم، مع الدعوة إلى اعتماد مقاربة تكاملية بين أداء الإنسان وإمكانات الآلة في تطوير النماذج اللغوية، بما يسهم في بناء منظومات رقمية دقيقة وفعالة لمعالجة اللغة العربية.

وتتبع أهمية هذه الورقة من كونها تسعى إلى الإسهام في مقاربة إشكاليات متعددة ذات أبعاد نظرية، مجتمعية، وتطبيقية.

فعلى المستوى النظري، تُقدّم تصورًا أوليًا لإطار مفاهيمي يساعد على دمج تقنيات الذكاء الاصطناعي في علم المعاجم العربية الرقمية، مع إبراز التحديات المرتبطة بالتوازن بين المرونة التقنية والدقة الأكاديمية.

وعلى الصعيد المجتمعي، تُثير الورقة قضيةً جوهريةً تتعلق بصون الهوية اللغوية العربية في ظل هيمنة اللغات الأخرى في الفضاء الرقمي، عبر دعم التوجه نحو بناء معاجم ذكية قادرة على عكس الخصوصية الثقافية. أما عمليًا، فهي تطمح إلى دعم تطوير أنظمة معالجة اللغة الطبيعية (NLP) لفهم السياقات العربية بعمق.

ومن ثم تعتمد الورقة منهجية تحليلية تجمع بين النظرية اللغوية والتطبيقات الحاسوبية، في محاولة للإجابة عن السؤال الآتي: كيف يمكن للذكاء الاصطناعي والعلوم المعرفية أن يسهما في إعادة تشكيل المعاجم العربية بوصفها منصات ذكية مع المحافظة على الخصوصية اللغوية؟

ولتحقيق أهداف الدراسة، قُسمت الورقة إلى مقدمة تقفوها ستة مباحث رئيسة: أولها تأسيس للإطار المفاهيمي وجاء ثانيها وثالثها لمعالجة تحديات التعريف، أما رابعها وخامسها فلمعالجة آفاق التطوير، ثم جاء سادسها استراتيجيات مقترحة لتعزيز نمذجة المحتوى المفاهيمي العربي.

وأخيرا، لا تدعي الباحثة لنفسها تميزًا خاصا، بل تضع بين يدي القارئ الكريم هذا الجهد المتواضع، الذي سعت فيه إلى معالجة موضوع البحث برؤية علمية متأنية، آملة أن تكون قد وُفقت في تسليط الضوء على جوانبه وإثراء النقاش حوله. وإن كان في العمل من صواب، فذلك من فضل الله وتوفيقه، وإن كان فيه خطأ أو تقصير، فمردّه إلى ضعف البشر، والكمال لله وحده

والشكر أجزله إلى قارئ هذا البحث على صبره، وعلى ما يقدمه من إرشادات وإضاءات تكمل نواقصه وتسد هناته وتضيء عتماته. وعلى الله قصد السبيل

المبحث الأول: المفاهيم الأساسية:

١- المعجم الرقمي:

في سياق النقاش الذي أثارته Müller-Spitzer (2014) حول مدى ملاءمة مصطلح 'المعجم الإلكتروني للاستخدام في سياق الحديث عن المنتجات الإلكترونية التي تعطي معلومات حول التهجئة أو المعنى أو استخدام الكلمات...إلخ؛ حيث إن الكائن المحدد منتجًا مبتكرًا للغاية لا يشبه المعجم المطبوع ولا يوجد كتاب بالمعنى المعروف، وفضلت تجنب مصطلح "القاموس"؛ لتوضيح الاختلاف المصطلحي بين القواميس المطبوعة وأنظمة المعلومات المعجمية الرقمية [١]

سيعتمد البحث مصطلح" المعجم الرقمي"؛ لأنه أكثر ملاءمة لما يرنو إليه البحث، فيشير "الرقمي" هنا إلى عملية إنتاج المعجم باستخدام الحاسوب، سواء بتحويل المعاجم الورقية إلى صورة إلكترونية أو بإنشاء معجم جديد يولد آليًا، في حين يرتبط مصطلح "الإلكتروني" بالمنتج النهائي بوصفه أداة تفاعلية لها وظائف ومحتويات محددة...إلخ

تعريف مصطلح " المعجم الرقمي":

هناك عدد لا يحصى من المظاهر والسمات المفاهيمية المتصورة للنص الفائق [٢] (معظم المعجمات الإلكترونية الأصل مؤسسة على فكرة النص الفائق). ومن ثم تعددت المصطلحات والمفاهيم، ولا يسعى البحث إلى تفنيد ما قدم من تعريفات لهذا المصطلح في الأدبيات البحثية المختلفة، وإنما يقترح تعريفا يأتي في سياق التمييز بين المعاجم الرقمية المُولَّدة آليًا وتلك التي حررها البشر (أُسميها: الفرق بين إنتاج آلي/رقمي وعرض آلي/رقمي) فالمعجم الرقمي: عمل مرجعي لغوي يُعرض ويخزن(في قاعدة بيانات كبرى) على الآلة وقد يُنشأ ويعرض ويخزن بواسطتها، ويستخدمه البشر وقد

(1)see:Müller-Spitzer, C. (2014). P3

(٢) النص الفائق أو التشعيبي hypertextنظام رقمي يتيح للمستخدمين التنقل بين أجزاء المعلومات باستخدام الروابط. تربط هذه الارتباطات، المعروفة باسم الارتباطات التشعبية، أقسامًا مختلفة من النص أو المستندات أو عناصر الوسائط المتعددة. عندما ينقر المستخدم على رابط تشعبي، يُوجه إلى المحتوى المرتبط، والذي قد يكون في الصفحة ذاتها أو في صفحة مختلفة أو موقع ويب مختلف.

تستخدمه الآلة(الروبوت) وأتوقع أن هذا التعريف استوفى: ١- الخاصية الأساس والبعد الوظيفي للمعجمات وهي أنها عمل مرجعي يلجأ إليها البشر للحصول على معلومات دقيقة خاصة اللغوية منها.٢- المجال الموضوعي (اللغة في المعجمات العامة أو المتخصصة)٣-الوسيط الناقل وهو الآلة، وغالبًا ستكون الشابكة.٤- المُنتِج: قد يكون البشر وقد تكون برامج معالجة اللغات الطبيعية أو نماذج الذكاء الاصطناعي مستقبلا، إضافة إلى دور البشر من المعجميين في تقييم وتحسين ما أنتجته الآلة ونقلته إلى مرحلة النشر(وإن كان هذا محل نظر حال سيطرة أدوات الذكاء الاصطناعي).٥- المستخدم: بشري أو نظام ذكي يعتمد علىNLP ،وغالبا ما يكون المستخدم للمعجم بشري يبحث عن معلومات لغوية في موقف من المواقف التالية: تواصلي؛ حيث يحتاج إلى حل مشكلة تواصلية مثل إنتاج نص أو استقباله بلغته الأصل أو لغة أجنبية. أو موقف معرفي؛ حيث يبحث عن معرفة لغوية خاصة بموضوع. أو موقف تفسيري؛ حيث الحاجة إلى تفسير وفهم رموز وعلامات لغوية وغير لغوية.

ويمكننا توسيع هذا التعريف بإضافة تفصيلات ذات صلة بنوع المعجم والهدف، أو نوع المستخدم وأشياء أخرى مُحددة مما لا مجال لذكرها.

فالبحث إذن ينظر إلى المعجم الرقمي بوصفه قواعد بيانات آلية للوحدات اللغوية، تشمل النطق والأصول الصرفية، والعلاقات الدلالية، ويرتكز على الأخيرة؛ لأنها محور نمذجة المحتوى المفاهيمي.

٢ -*التمازج المعرفي:*(Cognitive Blending)

يقصد به التكامل النظري والمفاهيمي بين اللغويات والحوسبة والمنطق، وقد حدث هذا التمازج المعرفي على مراحل متدرجة (من التأسيس النظري إلى الأطر الحاسوبية إلى تعددية التخصصات وقابلية التوسيع) وعلى نطاق واسع عبر الأطر الرسمية مثل الإطار النحوي، فها هو آرتي رانتا يستخدم نظرية النوع (المنطق) لنمذجة بناء الجملة والدلالات، رابطا المجالات (اللغة والمنطق والحوسبة) بصورة واضحة وصريحة مع

النمذجة الدلالية الرسمية لينشئ محللا لغويّا، مما يتيح أنظمة ترجمة ومعالجة اللغات الطبيعية متعددة اللغات. [٣]

وكثيرة هي الدراسات الموجودة في هذا الشأن، يتمثل البحث لبعضها: فمنها أبحاث المرحلة التأسيسية وكان مدار اهتمامها على التمثيل المنطقي لمعنى اللغة الطبيعية من طريق حساب التفاضل والتكامل المسند، والدلالات القصدية intensional semantics. [٤]. ومنها ما وضع المنطق مركزا للتحليل الدلالي والتمثيل النحوي، [٥] وكان التمثيل الحاسوبي ناشئا. [٦]. وبعد ذلك سلطت اللغويات الحاسوبية الضوء على الجمع بين المعطى النظري اللغوي وخوارزميات تحليلية قابلة للتطوير باستخدام النهج الاحصائي. [٧]. وفي السنوات الأخيرة (تقريبا من ٢٠٢٠) ظهرت أطر متعددة

(3)RANTA, A. (2004). Grammatical Framework. *Journal of Functional Programming*, *14*(2), 145–189. doi:10.1017/S0956796803004738.

(4)Jün-Tin Wang. 1973. On logical formulation of the computation process in semantical systems. In Proceedings of the 5th conference on Computational linguistics - Volume 1 (COLING '73). Association for Computational Linguistics, USA, 87–100. https://doi.org/10.3115/992532.992540 & . Robert M. Schwarcz. 1969. Towards a computational formalization of natural language semantics. In Proceedings of the 1969 conference on Computational linguistics (COLING '69). Association for Computational Linguistics, USA, 1–53. https://doi.org/10.3115/990403.990432

(5)Polzin, T., & Rieser, H. (1991). Parsing with Situation Semantics. *Nordic Journal of Linguistics*, *14*(2), 141–189. doi:10.1017/S0332586500002432

(6)Raymond Reiter. 1975. Formal reasoning and language understanding systems. In Proceedings of the 1975 workshop on Theoretical issues in natural language processing (TINLAP '75). Association for Computational Linguistics, USA, 175–179. https://doi.org/10.3115/980190.980238

(7)Zettlemoyer, L., & Collins, M. (2007). Online Learning of Relaxed CCG Grammars for Parsing to Logical Form. *Conference on Empirical Methods in Natural Language Processing*.PP:678-687.

التخصصات تدمج التفكير الرمزي لفهم اللغات الطبيعية، مع استكشاف أسس منطقية متقدمة. [٨]

صفوة القول إن عملية التمازج المعرفي أعمق وأعقد من أن تستوفيها هذه المساحة، لكن المحاور الأساسية لها تتمثل في:

–انتقال التركيز الأولي من التجريد النظري إلى الأطر الحاسوبية وقابلية التوسع متعدد التخصصات.

– سيطرة أساليب معينة، مثل: التقنيات الرمزية والأساليب الغامضة، ومحاولة التوازن بين الدقة والتوسع. –وسيادة نماذج معينة، نحو: القواعد النحوية النظرية، (الخاصة بالنوع) وحساب التفاضل والتكامل (Lambda Calculus) [٩] والقواعد الفئوية.

٣–مفهوم نمذجة اللغة / نمذجة المحتوى المفاهيمي:

١–٣: نمذجة اللغة:

عملية إنشاء نموذج رياضي أو حاسوبي يمثل اللغة الطبيعية (Natural Language Modeling)، بهدف فهمها، وتحليلها أو توليدها. وتهدف في سياق المعاجم الرقمية إلى تمكين الحاسوب من فهم اللغة الطبيعية ومعالجتها بدقة (بحثا عن المفردات أو تحليلا للنصوص أو توليدا للغة الطبيعية) وتتضمن نمذجة اللغة في سياق المعجم الرقمي تمثيل مجموعة من الخصائص اللغوية، نحو:

- مفردات المعجم :(Lexicon) قائمة الكلمات ومعانيها.
- القواعد النحوية :(Syntax) هيكل الجمل وترتيب الكلمات.
- الدلالة :(Semantics) معاني الكلمات والجمل.
- التداولية (Pragmatics)كيفية استخدام اللغة في السياقات المختلفة. [١٠]

(8)Schaefer, J.F., & Kohlhase, M. (2020). GLIF: A Declarative Framework for Symbolic Natural Language Understanding. *FCR@KI*.pp:4-11.

(٩) نظام رياضي حاسوبي يستخدم لتمثيل الوظائف والحوسبة وله دور محوري في نمذجة الدلالات اللغوية (Semantics)

ولنمذجة اللغة أنواع عدة، منها: (أ) النمذجة العصبية وتستخدم الشبكات العصبية لتمثيل اللغة. [١١](ب) النمذجة الإحصائية أو تحليل البيانات اللغوية باستخدام الأساليب الأساليب الإحصائية. [١٢] (ج)نماذج معتمدة على القواعد، مثل القواعد الدلالية والنحوية. (د) نماذج هجينة تجمع بين الأنواع الثلاثة السابقة.

ولتمثيل المعرفة اللغوية في المعاجم الرقمية تُستخدم تقنيات مختلفة من نمذجة اللغة تختلف تبعا للغرض من المعجم والتقنيات المتاحة. [١٣]وتعد نمذجة الموضوعات: إحدى التقنيات المستخدمة في معالجة اللغات الطبيعية من أجل فهم وتصنيف الموضوعات الرئيسة في مجموعة من النصوص. وتستخدم في سياق المعاجم الرقمية لتنظيم المفردات والمفاهيم في مجالات دلالية معرفية واستخراج العلاقة بينها. وتعتمد

(10)For more about the statistical modeling of the natural language: Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press. وكذلك Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd Edition). Pearson٣ فصول: (N-gram Language Models) و ١٨ (Semantics and Discourse)

(11)An example :Transformers ،RNN (Recurrent Neural Networks) ،LSTM (Long Short-Term Memory. pp1-15 .For detail see Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. الفصل العاشر وعنوانه Sequence Modeling: Recurrent and Recursive Nets.pp:367-415.

(12)Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd Edition). Pearson https://web.stanford.edu/~jurafsky/slp3/.) الفصلين الثالث N-gram Language Models. (3.pdf) والرابع Naive Bayes and Sentiment Classification(4.pdf

(١٣) من هذه التقنيات: الشبكات الدلالية (نحو WordNet الذي يمثل المفردات في مجموعات دلالية) ، والاطارات المعجمية (وتمثل الكلمات في سياقاتها النحوية والدلالية)، والتضمين الكلمي (تضمين الكلمات في فضاء متعدد الابعاد) وتحليل السياقات (مثل توفير معاني ومفاهيم معتمدة على تحليل السياق) وقواعد البيانات المعجمية(تخزين الملومات اللغوية في قواعد بيانات منظمة)

نمذجة الموضوعات على خوارزميات مثل LDA (Latent Dirichlet Allocation) لفهم توزيع الكلمات عبر الموضوعات المختلفة. (١٤)

٢-٣: نمذجة المحتوى المفاهيمي:

هي تمثيل تجريدي عالي المستوى للمفاهيم والعلاقات بينها في مجال معين. وفي السياق المعجمي حدد جرار معيارًا لأنطولوجيا العربية عبر التوصيف الصريح ل "نمذجة المحتوى المفاهيمي" على أنها إبداع نموذج معرفي نظيف أنطولوجيًا للمعجم. وعرّف المفهوم بأنه" فئة نظامية تمثل التجريد – ليس مجرد معنى، وإنما وحدة دلالية مميزة مع التزام وجودي (القسم ٣.٢، الشكل ١). Lemma & Sense: متباينة بوضوح عن المفهوم". يرسم النموذج هذه الطبقات وعلاقاتها. العلاقات الدلالية الرسمية (Subtype Of، PartOf)، وتحديد كل مفهوم ببديهيات منطقية. تم قياسه مقابل الأنطولوجيات العليا (BFO و DOLCE) و WordNet و Wikidata، مما يضمن التوافق بين الموارد والصرامة المفاهيمية. (١٥)

وتتكون النمذجة المفاهيمية بصورة عامة من عناصر ثلاثة أساسية: أولها: الكيانات، وتمثل الأشياء أو المفاهيم الرئيسة للمدروس. ثانيها: السمات، وتحدد خصائص الكيانات. ثالثها: العلاقات، وتمثل الارتباطات بين الكيانات المختلفة. (١٦) ويمكننا

---

(١٤) Steyvers, M., & Griffiths, T. (2007). *Probabilistic Topic Models*. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 427-448). Lawrence Erlbaum Associates. حيث يقدم نظرة على نمذجة الموضوعات وتطبيقاتها.

(15)https://2u.pw/Od4RBxYH**Data modeling in RDBMS: From Concept Design to implementation ٢٠٢٤/١٢/١٣Review date. &** Jarrar, M. (2022). The Arabic Ontology -- An Arabic Wordnet with Ontologically Clean Content. *Applied Ontology Journal, 16:1, 1-26. IOS Press. (2021)*. https://doi-org.sdl.idm.oclc.org/10.3233/AO-200241.p6.

(16)Jürgenson, S. (2023). نموذج لبيانات في RDBMS [A data model in RDBMS]. AppMaster. https://appmaster.io/ar/blog/nmdhj-lbynt-fy-rdbms

استخدام النمذجة المفاهيمية في سياق العمل المعجمي لتنظيم المعلومات اللغوية وتمثيل العلاقات بين الكلمات والمعاني أو بين المصطلحات والمفاهيم بطريقة منهجية ومنظمة، مما يسهل عملية البحث واسترجاع المعلومات في المعاجم الإلكترونية الحديثة. وتوضيح العلاقة بين مختلف المفاهيم داخل المعجم وإنشاء روابط دلالية بين مختلف المصطلحات. (١٧)

وتعتمد نمذجة المحتوى المفاهيمي على تقنيات معالجة اللغة الطبيعية والتعلم الآلي لتحليل النصوص وتحديد العلاقات بين الكلمات والمفاهيم، مساعدة على تطوير معاجم رقمية أكثر ذكاءً وقدرة على فهم السياق اللغوي والدلالي للمصطلحات. وتعد النمذجة المفاهيمية أولى الخطوات في عملية نمذجة البيانات؛ فتشكل أساسا متينا للمراحل التالية من طريق فهم متطلبات العمل بوضوح. وكذلك تعد جوهر نموذج النشر، حيث تصنف سلاسل البيانات في قاعدة المعلومات على أساس محتواها والغرض منها، وتصاغ في هيكل يعبر عن العلاقات بينها. وتعتمد هذه النمذجة على تحليل المواد المتاحة (سواء كانت نصوصًا مطبوعة أو مصادر أخرى) لفصل المضمون عن البنية الشكلية مثل الترتيب والطباعة (١٨)

ويمكننا تلخيص خطوات النمذجة المفاهيمية في السياق المعجمي بعد تحديد الهدف من المعجم (تعليم لغوي، ترجمة آلية، بحث أكاديمي...إلخ) وتحديد الفئة المستهدفة من المستخدمين، فيما يلي:

(17)Murel, J., & Kavlakoglu, E. (n.d.). Topic modeling. IBM. Retrieved [14l9l2024تاريخ الوصول] from https://www.ibm.com/sa-ar/think/topics/topic-modeling **وكذلك** Boguraev, B., & Briscoe, T. (1989). Computational lexicography for natural language processing. *Language, 66*, 626.

(18)Lemberg, I., Schröder, B., & Storrer, A. (2001). *Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Max Niemeyer Verlag.P.47.

(أ) تحديد المجالات الدلالية، حيث تصنيف الوحدات المعجمية (كلمات أو تعابير) ضمن مجالات مترابطة، ويساعد ذلك في إبراز العلاقات بين الألفاظ والمعاني المتقاربة أو المتشابهة. (١٩)

(ب) فجمع البيانات اللغوية من مصادر اللغة، مثل قواعد البيانات اللغوية، والمعاجم الورقية والإلكترونية الموثوقة، والمتون النصية المختلفة، وغيرها. فتنوع المصادر يضمن تمثيل السياقات المختلفة، وتحاشيا للبيانات غير الدقيقة تطبق خوارزميات التنقية من الأخطاء أو التكرار.

(ج) ثم تحديد المفاهيم والعلاقات، فبعد تحديد الكلمات والمفاهيم الأساسية وتحديد العلاقات المختلفة بينها (كالتضاد والترادف والترابط الدلالي... إلخ) نحلل العلاقات والمفاهيم باستخدام نماذج وأدوات مثل الإطارات المفاهيمية (Conceptual Frames).لوصف سياقات الاستخدام والشبكات(Semantic Networks) الدلالية لتمثيل العلاقات بين المفاهيم بشكل مرئي وهكذا.

(ج) وأخيرا يكون تصميم النموذج المفاهيمي رقميّا وتمثيل البيانات ثم تقييمها وتحسينها، حيث تحويل المفاهيم والعلاقات السابقة إلى تنسيق رقمي يعالج بواسطة الحاسوب، هذه البيانات تمثل بواسطة لغات مثل XMLأوRDF وفي هذه المرحلة قد تُستخدم بعض تقنيات الذكاء الاصطناعي (مثل المحولاتTransformers) لتحسين البحث الدلالي عبر استخراج السياقات الدلالية من النصوص الكبيرة، وتوليد النصوص بشكل سياقي متماسك وذلك من طريق تحليل النصوص واستخراج الأنماط الدلالية بين المفردات، وبعد ذلك تُقيّم وتحسن لضمان الشمول والدقة. (٢٠)

(١٩) ينظر في ذلك، صبير، عبد الناصر عثمان عبد الله" صناعة معجم تعليمي للناطقين بغير العربية باستخدام نظرية الحقول الدلالية"، مجلة اللغة العربية بالجامعة الإسلامية بالمدينة المنورة، ٢٠٢٢، ٧٩٠-٧٩١

(20)For detail see: Patrick Hanks. (2013). *Lexical Analysis: Norms and Exploitations*. The MIT Press. . الذي يتحدث عن ترميز وتمثيل المعاني خاصة الفصلين الخامس (١١٥-١٤٣) والثامن (٢١١-٢٥٠) & Pustejovsky, J., & Boguraev, B. (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence, 63*(1–2), 193–223. Lexical knowledge representation and natural language processing - ScienceDirect

الخلاصة: تُعد العلاقة بين **نمذجة اللغة، نمذجة الموضوعات، ونمذجة المحتوى المفاهيمي** علاقة تكاملية تفاعلية؛ حيث تركز الأولى على فهم وتحليل البنية اللغوية(نحوًا وصرفا) و الأنماط الإحصائية للغة، وتهدف الثانية إلى اكتشاف المفاهيم الرئيسة في مجموعة من النصوص، بينما تعمل الثالثة على فهم وتحليل المفاهيم الدقيقة والعلاقات بينها داخل النصوص، ورغم اختلاف أدوات كل نوع إلا أنها تتداخل وتتكامل في التطبيق العملي (حيث تعتمد نمذجة الموضوعات على نتائج نمذجة اللغة مثل التحليل الإحصائي للكلمات لاكتشاف المفاهيم الرئيسة، *بينما* تستخدم نمذجة المحتوى المفاهيمي مخرجات كلا النموذجين لرسم العلاقات الدقيقة بين المفاهيم) لتعمل معا على تمثيل اللغة الطبيعية وفهمها بشكل شامل. إذ يركّز كل نوع على زاوية معينة من اللغة، لكنها جميعًا تسهم في مهام مثل :تحليل النصوص وتنظيم المعرفة وفهم المعاني.

٤-الذكاء الاصطناعي: (Artificial Intelligence)

"أحد مجالات علوم الحاسب يهدف إلى إنشاء أنظمة يمكنها تنفيذ المهام التي تحتاج عادةً إلى الإدراك البشري، مثل التعلّم وصنع القرار والتطوير الذاتي، ويُشار إليه غالباً باسم "ذكاء الآلة". (٢١)

وتوجد مجموعة من التعريفات الفنية التي طرحت للذكاء الاصطناعي، منها: تعريف مكارثي: نظام قادر على التفكير بطريق التلاعب الرمزي". (٢٢) وتتنوع تعريفات الذكاء

(٢١)مكتب وزير الدولة للذكاء الاصطناعي والاقتصاد الرقمي وتطبيقات العمل عن بعد- الإمارات العربية المتحدة(د-ت). [المعجم العربي للذكاء الاصطناعي]. بوابة حكومة الإمارات العربية المتحدة للذكاء الاصطناعي. تم الاسترداد في ٢٣ أكتوبر ٢٠٢٣، من AI Dictionary | Artificial Intelligence Office, UAE

(22)Chaudhary, J., Parmar, N., & Mehta, A. (2024). *Artificial intelligence and expert systems. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 4*(4). ويقصد بالتلاعب الرمزي استخدام رموز مجردة لتمثيل المعارف واتخاذ القرارات مثل القواعد المنطقية ومعادلات الرياضيات، لمحاكاة العقل البشري

الاصطناعي حسب السياقات التطبيقية، بين: معالجة اللغة، [٢٣] أو النماذج (الأساليب الرمزية مقابل الاتصالية)، وغيرها [٢٤]. وتختلف التعريفات تبعا للمنهجية البحثية المستخدمة (منها ما يركز على جانب اللغة والآخر على النماذج الحاسوبية)، وبناء على هذه المنهجيات يمكن تصنيف عدة أنواع من الذكاء الاصطناعي، منها: الذكاء الرمزي والذكاء التواصلي، والأنظمة الهجينة الجامعة بينهما. [٢٥]

(23)IBID: (Chaudhary et al., 2024pp:536-546.

(24)Allen Newell. 1983. Intellectual issues in the history of artificial intelligence. The study of information: interdisciplinary messages. John Wiley & Sons, Inc., USA, 187–294.

(25)F.. Saini, T. Sharma, and S. Madan, "A Comparative Analysis of Expert Opinions on Artificial Intelligence: Evolution, Applications, and Its Future", *Adv. J. Grad. Res.*, vol. 11, no. 1, pp. 10–22, Sep. 2021. وG, K., Mohan, V., & Senthilkumar, S. (2023). A BRIEF REVIEW OF THE DEVELOPMENT PATH OF ARTIFICIAL INTELLIGENCE AND ITS SUBFIELDS. *International Journal of Engineering Technologies and Management Research*, *10*(6), 1–12. https://doi.org/10.29121/ijetmr.v10.i6.2023.1331 وAmendola, G. (2023). Special Issue on Logic-Based Artificial Intelligence. *Algorithms*, *16*(2), 106. https://doi.org/10.3390/a16020106 يعتمد الذكاء الرمزي على التمثيل المنطقي بينما يعتمد التواصلي على التعلم الآلي والبيانات الضخمة.

المبحث الثاني: التحديات اللغوية في بناء التعريفات المعجمية العربية

يستعرض المبحث التحديات اللغوية للتعريف أي الصعوبات والعقبات التي تواجه نمذجة المحتوى المفاهيمي في المعاجم العربية الرقمية عبر تحليل لصعوبة التعامل مع الصرف والنحو والدلالة واللهجات في صياغة التعريفات الدقيقة، موضحا محاولات تذليل بعض العقبات والتحديات التي تواجه نمذجة النظام اللغوي العربي، مثل تحليل النصوص العربية، وفهم السياقات، وتوليد تعاريف ،وأمثلة دقيقة وغيرها.

١-تحسين البحث الدلالي وتحليل النص العربي (التشكيل والاشتقاق)

التشكيل: يمثل تشكيل النصوص العربية تحديًا كبيرًا مع كثير من التطبيقات؛ لأسباب أهمها: الأثر الكبير لعلامات التشكيل على معنى المفردة والجملة، وبدورها تتأثر بالسياق إضافة إلى أن إمكانية تشكيل النص العربي جزئيا أو كليًا وإلى عوامل أخرى، مثل ازدواجية العربية وتنوعها بين فصحى (تراثية ومعاصرة) وعامية، وقلة الموارد اللغوية المتاحة للباحثين والمطورين. وعلى الرغم من أهميته إلا أن الاهتمام بتطبيقه آليًا محدود، وهناك منهجان شائعان في الأدبيات المتعلقة بمعالجة مشكلة التشكيل الآلي في اللغة العربية: النهج القائم على القواعد (يعتمد على القواعد اللغوية والمعالجات الإحصائية) ونهج التعلم الآلي [٢٦]. وقد تفوق الثاني على الأول تبعا لإحدى الدراسات بمعدل خطأ تشكيلي٢.٨٨% مقارنة ب ١٣.٧٨% [٢٧] وقد اعتمدت الدراسة على استخدام آليات التعلم العميق (الخوارزميات العصبية العميقة) التي تعد فرعا متخصصا من الذكاء الاصطناعي، وقد أسفرت تجارب المؤلفين عن أن نظام Shakkalaالعصبي تفوق على الأساليب التقليدية مع قيم منخفضة من معدل الخطأ

(٢٦)أحد فروع الذكاء الاصطناعي، ويعني دراسة خوارزميات الحاسوب التي تتحسن تلقائيا بواسطة التجربة ويقوم بتطوير خوارزميات تسمح للحواسيب بالتعلم التلقائي من البيانات دون سابق برمجة لأداء المهمة، وله أنواع عدة: موجه، وغير موجه، وتعزيزي. للمزيد ينظر Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill. Machine Learning textbook

(٢٧)للمزيد عن الدراسات التي اعتمدت التعلم العميق لمعالجة مشكلة التشكيل في النص العربي ينظر Zeroual, I., & Lakhouaja, A. (2018). *Arabic Diacritization Using Deep Learning*. Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC).

فـي التشـكيل (DER) ومعـدل خطـأ الكلمـات (WER)فـي محاولـة التعامـل مـع تحـد التشكيل للغة العربية

أما عن الاشتقاق: فيمثل تحديّا في طريق المعالجة الآلية للغة العربية لا يقل تعقيدا عن التشكيل، وبالنظر إلى الأبحاث والدراسات العربية والإنجليزية التي أجريت لمعالجة الجانب الصرفي عامة والاشتقاق على وجه الخصوص، يمكن تقسيمها إلى:

–أبحاث: حول التحديات الأساسية، منها ما دار حول التشكيل الآلي ودمجه مع التحليل الصرفي ضمانا للدقة، خاصة في النصوص غير المشكولة. [٢٨] ومنها ما كان حول الخصوصية اللغوية مشيرا إلى تعقيدات البنية الصرفية العربية خاصة الاشتقاق والتجريد. [٢٩]

–أبحاث عن تطوير المحللات الصرفية سواء تطوير النماذج القائمة على القاعدة[٣٠] أو النماذج الاحصائية [٣١]

–أبحاث تهدف إلى دراسة التكامل مع تقنيات الذكاء الاصطناعي دراسة نظرية مثل تقنية الشبكات العصبية لتحسين دقة التمثيل الصرفي (خاصة في التعامل مع التلافيف اللغوية) مطبقة نماذج المحولات Transformer[٣٢]

---

(٢٨)رباحي، محمد، وحاكم، عمارية. (٢٠٢١). تحديات المعالجة الآلية للغة العربية *اللغة العربية*، مج٢٣، ع٢ ، ٢٤٩ .268 - مسترجع من

http://search.mandumah.com.sdl.idm.oclc.org/Record/1154667).

(٢٩) شعلان، خالد (٢٠١١) المعالجة الآلية للغة العربية، محاضرة مستضافة، المؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية، جامعة الامام محمد بن سعود، المملكة العربية السعودية، الرياض، https://goo.su/nHArmFG

(٣٠) ينظر على سبيل المثال، سوسن، معافة(٢٠٢٢): تقنيات معالجة اللغة العربية آليًّا دراسة مقارنة لنماذج من المحللات الصرفية العربية، رسالة دكتوراة، كلية الآداب واللغات، جامعة ٨ مايو، الجزائر.١٢٨-١٦١

(٣١) شعلان، خالد (٢٠١١) المعالجة الآلية للغة العربية، مرجع سابق فعالية النماذج الإحصائية في معالجة الكلمات المبتورة أو المشوشة

(٣٢) قبي أحمد. "الذكاء الاصطناعي والمعالجة الآلية للغة العربية: بحث في أنظمة التعرف الآلي على الكلام ."*المجلة العلمية البحثية*، م ١، عدد ٢، أكتوبر، ٢٠٢٤، ص ٨. https://ojs.ahumi.edu.vn/index.php/SRJ/article/view/27 ,

-أبحاث تطبيقية عملية: مثل أنظمة التدقيق اللغوي [٣٣]وإنشاء محللات صرفية محسنة [٣٤]

٢-**<u>توليد التعريفات والأمثلة</u>**: يُعد توفير تعريفات دقيقة، وتوليد أمثلة توضيحية تعكس استخداماتها في سياقات متنوّعة، من أهم التحديات التي تواجه المعالجة الآلية للغة العربية لا سيّما في عمليات النمذجة اللغوية. وقد أجريت في هذا الإطار بعض الأبحاث لمحاولة الاستفادة من تقنيات الذكاء الاصطناعي، مثل استخدام نماذج توليد النصوص مثل GPT-3 لتوليد تعاريف بناءً على السياق. [٣٥] بينما استخدامBERT لفهم السياقات وتوليد أمثلة. [٣٦]

(٣٣) مثل صحصلي من شركة صخر وقلم وغيرها من أنظمة تدقيق صرفي تعتمد على تحليل الوزن والجذر الصرفي.

(34)Pasha, A., Al-Badrashiny, M., Diab, M.T., Kholy, A.E., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. (2014). MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. *International Conference on Language Resources and Evaluation*. مزج بين) محتفظا بميزاتهما مضيفا إليهما.Amira) و(Madaتطبيقي (

(35)Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language Models are Few-Shot Learners. *ArXiv, abs/2005.14165*. GPT-3 هو نموذج لغة انحداري ذاتي يحتوي على ١٧٥ مليار معلمة حاول المؤلفون عبره توسيع نطاق نماذج اللغة في أداء المهام دون ضبط تقليدي دقيق، وتوضيح قدرتها على التعلم من أمثلة قليلة بكفاءة تفوق الأساليب السابقة

(36)Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

٣- **تحسين القدرة على فهم السياق**، سواء أكان فهما لمتعدد المعاني أم تحليلا للعلاقات الدلالية، فالتحدي الذي يواجه الأولى هو تعدد المعنى تبعا للسياق، وقد أجريت محاولات لاستخدام نماذج لفهم السياقات المختلفة، نحو BERT [٣٧] والذي يواجه الثانية تحديد العلاقات الدلالية بين المفردات نحو التفرع والترادف وغيرها، وقد استخدمت عدة تقنيات مثل Word2Vec أو GloVe لتمثيل الكلمات كمتجهات. [٣٨]

وبالنظر العام إلى أحدث ما أجري من أبحاث في معالجة الجانب الدلالي للغة العربية آليّا، نجد أبحاثا متنوعة منها ما عالج اللبس الدلالي الناتج عن تعدد معنى الكلمة [٣٩] ومنها ما استخدم تقنيات المعالجة الحاسوبية في محاولة توظيف الذكاء الاصطناعي لمحاولة معالجة الانزياحات الدلالية في النصوص الأدبية. [٤٠]

تعقيب: على الرغم من الجهود المبذولة والتقدم الملحوظ، ما تزال المعالجة الألية للغة العربية وتكاملها مع تقنيات الذكاء الاصطناعي تواجه كثير من التحديات في كل من المنهجيات التقليدية والنماذج اللغوية الكبرى LLMs، من أبرز هذه التحديات: التشكيل الغني، ومرونة الترتيب في المفردات (SVO / VSO)، والتنوع اللهجي. [٤١] إضافة

(37)Antoun, W., Baly, F., & Hajj, H. (2020). *AraBERT: Transformer-based Model for Arabic Language Understanding*. arXiv preprint arXiv:2003.00104.

(38)Al-Rfou, R., et al. (2013). *Polyglot: Distributed Word Representations for Multilingual NLP*. Proceedings of the 17th Conference on Computational Natural Language Learning (CoNLL).

(٣٩) حمادة، سلوى السيد وهمر مهديوي(٢٠٢٤) : المعالجة الدلالية الآلية للغة العربية، السان العربي،المعالجة الدلالية الآلية للغة العربية - # مارس ٢، ٢٠٢٤ مسترجع بتاريخ ٢٠٢٤/٩/١١.

(٤٠) كامل، وفاء، محسن رسوان، وعبد العاطي هواري (٢٠١٤): معالجة المحتوى المعجمي الدلالي في المعجم العربي الحاسوبي " مقاربة لغوية حاسوبية"، مجلة الجمعية المصرية لهندسة اللغة، ع٦٣، ١-٧٨.

(٤١) ينظر تفصيل ذلك في المرجع السابق Aljanabi, Mohammad. (2024). Assessing the Arabic Parsing Capabilities of ChatGPT and Cloude: An Expert-Based Comparative Study . وكذلك Abdelali, A., Durrani, N., Dalvi, F., & Sajjad, H. (2022). مرجع سابق

إلى التحدي الخفي المتمثل في الفروق الدقيقة داخل اللهجة الواحدة حتى مع استخدام النماذج المتقدمة، مما يفرض الحاجة إلى موارد لهجية مخصصة أو ضبطا دقيقا، خاصة عند التعامل مع الأصناف ذات الموارد المحدودة. (٤٢)

المبحث الثالث: التحديات التقنية في تمثيل المحتوى المفاهيمي في المعاجم الرقمية العربية:

تعد قلة البيانات المُعلّمة وقلة وجود معايير دلالية موحدة إضافة إلى قيود أدوات التحليل اللغوي من التحديات التقنية الرئيسية في تمثيل المحتوى المفاهيمي في المعجم الرقمي العربي. وقد عولجت جزئيًّا عبر الإثراء شبه التلقائي، لا سيما في WordNet العربية والنمذجة المعتمدة علىLMF وكذلك تصميم الأنطولوجيا، وقد انقسم النهج الرئيس في معالجة هذه التحديات إلى:

١- قلة البيانات المُعلّمة، و تُعد من أبرز التحديات في معالجة اللغة العربية آليًا، إذ تندر مجموعات البيانات المصنفة يدويًا بشكل كافٍ (تحديد معنى الكلمة وصيغها الصرفية والعلاقات الدلالية وسياقات التداول وغيرها) وقد ظهرت بعض المبادرات البحثية لمعالجة هذه الفجوة، أبرزها على سبيل المثال محاولة لإضافة مجموعة من التطويرات إلى مشروع الشبكة العربية (AWN)(مشروع عربي لإنشاء شبكة مفاهيمية على غرار Word net) والقضايا الجوهرية المتعلقة به(مثل ضعف تغطية مجموعات المترادفات(synsets) ، وغياب بعض العلاقات الدلالية)؛ إذ يُعد –وفقا ل Yasser Regragui – أول " مورد متوافق مع WordNet. ولقد أتاح إمكانيات جديدة من حيث بناء تطبيقات معالجة اللغات الطبيعية المتطورة المتعلقة بهذه اللغة الساميّة" (٤٣) وبعد أن رصد المؤلفون الأعمال السابقة حول إثراء الشبكة العربية

(٤٢)المرجع السابق Habash. 2022. Morphosyntactic Tagging with Pre-trained Language Models for Arabic and its Dialects

(43)Yasser Regragui, et al., 2016. Arabic WordNet: New Content and New Applications. In *Proceedings of the 8th Global WordNet Conference*

واستخدامها في تطبيقات معالجة اللغات الطبيعية قدموا تلخيصا للتقنيات الإثرائية التي استخدموها في الشبكة العربية. وفي الإطار البحثي نفسه كانت محاولة Mohamed Ali Batita وآخرون حول تحديثات أضيفت إلى النسخة الموسعة من WordNetالعربية، وقد أظهرت نتائجهم أن العلاقات الدلالية المضافة حديثا يمكن أن تحسن أداء نظام إزالة الغموض بمعاني الكلمات. (٤٤)

وقد أسفرت هذه المحاولات عن تحسين تغطية AWN المعجمية والدلالية لدعم المهام التطبيقية، لكن هذه الآليات تظل بحاجة إلى تدقيق يدوي من قِبَل خبراء أو ناطقين بالعربية نتيجة وجود بيانات غير دقيقة، مما يفرض تكاليفا ووقتًا كبيرين، ومن ثم يُبطئ عملية تطوير المحتوى المفاهيمي الدقيق.

٢- أما عن **تحدي قلة وجود معايير دلالية موحدة** فقد رصدت محاولات بارزة في مجال التوحيد القياسي والنمذجة المفاهيمية. وفي هذا السياق، أظهرت الدراسات ذات الصلة اتجاهين رئيسيين:الأول: توحيد الهيكلة باعتماد معيار LMF وحدوده وهو إطار معياري دولي يهدف إلى توحيد هيكلة المعاجم لتسهيل تبادلها وتكاملها بين الأنظمة، حيث نُمذجت بعض المعاجم العربية الكبرى(كليًّا أو جزئيّا) باستخدام هذا الإطار(إطار عمل الترميز المعجمي24613 LMF–ISO) محققة درجة أساسية من interoperabilityالتشغيل البيني (أي قدرة المعاجم على العمل معا من دون تعارض)ومن الهيكلة الموحدة للمدخلات والمعاني والعلاقات، وكان مدار التطبيق

(GWC), pages 333–341, Bucharest, Romania. Global Wordnet Association.P.2

(44)Mohamed Ali Batita and Mounir Zrigui. 2019. The Extended Arabic WordNet: a Case Study and an Evaluation using a Word Sense Disambiguation System. In *Proceedings of the 10th Global Wordnet Conference*, pages 46–53, Wroclaw, Poland. Global Wordnet Association. Abstract

على الشبكة العربية ومعجم المدار[٤٥] والثاني: مشروع الأنطولوجيا العربية الذي حاول التصدي لعدم الاتساق في الشبكة العربية عبر منهج أنطولوجي يعتمد على المبادئ ويُعرّف المفاهيم الأصلية(native) وعلاقاتها الشكلية الدقيقة، مقدما تمثيلا رسميا للمفاهيم التي تنقلها المصطلحات العربية [٤٦]

ونستنتج مما سبق اعتماد محاولات التوحيد القياسي في الطريق الأول على الإطار الدولي LMF ولكنه غير كاف لطبيعة الموارد اللغوية العربية؛ حيث غياب التمثيل الواضح للعلاقات الدلالية بين المفاهيم (السببية وجزء من كل وغيرها) وضعف التعامل مع ظواهر لغوية كالمشترك اللفظي إضافة إلى ضعف تحديد المجال التخصصي للمفهوم (طب أو قانون أو لغويات...إلخ) أو التكيف مع التطبيقات الحديثة [٤٧]مما يؤدي إلى–صعوبة تكامل المعاجم مع أنظمة الذكاء الاصطناعي المتقدمة ومحدودية الاستفادة من هذه المعاجم في تطبيقات مختلفة. في

(45)KHEMAKHEM, A., GARGOURI, B., BEN HAMADOU, A., & FRANCOPOULO, G. (2016). ISO standard modeling of a large Arabic dictionary. *Natural Language Engineering*, *22*(6), 849–879. doi:10.1017/S1351324915000224. Abstract وعن الإطار فقد صدرت منه عدة إصدارات منها مسودة Frontini, F., Khan, A. F., & Romary, L. (n.d.). *ISO LMF 24613-6: A revised syntax semantics module for the Lexical Markup Framework*. CNR-Istituto di Linguistica Computazionale "Antonio Zampolli" & Inria. Retrieved from https://www.iso.org/standard/68516.html

(46)Jarrar M. The Arabic ontology – an Arabic wordnet with ontologically clean content. Applied Ontology. 2020;16(1):1-26. doi:10.3233/AO-200241المقدمة

(47)Shaalan, Khaled & Siddiqui, Sanjeera & Alkhatib, Manar & Monem, Azza. (2018). Challenges in Arabic Natural Language Processing. 10.1142/9789813229396_0003. للمزيد عن التحديات التي تواجه معالجة اللغة الطبيعية العربية ينظر صفحات٦٢-٧٨

حين انطوت الطريقة الثانية على مزايا متعلقة بالأنطولوجيا نحو قدرتها على التكامل مع تقنيات الذكاء الاصطناعي. وعليه، فإن الدمج بين إطار LMF والمقاربات الأنطولوجية قد يشكل مسارًا واعدًا نحو بناء معاجم رقمية عربية دقيقة وشاملة.

٣- **قيود أدوات التحليل اللغوي:** وتمثل أحد أبرز التحديات التقنية في بناء الموارد المعجمية العربية الرقمية، ويكمن الخلل والضعف- في حدود الإطار المعرفي الذي أمكن رصده- في محدودية دمج أدوات التحليل الصرفي/ النحوي المستخدمة في أثناء بناء المعجم الرقمي لتحليل المفردات العربية إلى جذورها الصرفية وفئاتها النحوية، فيكون تحديد الجذر (مثل الجذر كتب العربي للمفردات مكتوب وكاتب وكتابة ...إلخ) وتصنيف المفردة إلى فئة نحوية وغيرها. غير أن حدوث أخطاء في هذا التحليل (كأن تحدد جذرا خاطئا أو تصيفا نحويا غير دقيق) قد يؤدي إلى تشويش في الشبكة المفاهيمية للمعجم الرقمي، مما يؤثر سلبًا على موثوقية البيانات ويقلل من جودة التطبيقات المعتمدة عليه، مثل أنظمة الترجمة الآلية أو البحث الدلالي على سبيل المثال. ويكمن التحدي (فيما وقفت عليه الباحثة) في نقص الدراسات الموضحة لأثر حدوث الخطأ في تلك الأدوات.

وهنا تجدر الإشارة إلى أن التحديات غير مقتصرة على الجانب التقني، وإنما تمتد إلى الفجوة بين البحث والتطبيق؛ فمعظم التحديثات على AWN تبقى حبيسة الأوراق البحثية دون توظيفها في تطبيقات صناعية، ربما بسبب عدم توافقها مع حاجات السوق (مثل دعم اللهجات أو التعامل مع البيانات غير الرسمية).

المبحث الرابع: إمكانات الذكاء الاصطناعي في تطوير نمذجة المحتوى المفاهيمي للمعاجم الرقمية" مقاربة تقنية لغوية"

يهدف المبحث إلى استعراض دور تقنيات الذكاء الاصطناعي، لا سيما التعلم العميق، والشبكات الدلالية، والأنطولوجيات، في تحسين نمذجة المحتوى المفاهيمي للمعاجم الرقمية، مع التركيز على الآليات التقنية الكفيلة بتطوير التعاريف واسترجاعها. ويتناول المبحث هذا الدور من طريق محورين رئيسين

المحور الأول: بعض أطر معالجة اللغة الطبيعية (NLP) الخاصة باللغة العربية، مع التركيز على كيفية توظيف الذكاء الاصطناعي للفهم الدلالي وتحليل السياقات، عبر خطوات تشمل: تفكيك التشكيل والتحليل الصرفي والتحليل النحوي، تمهيدًا لإدخال النصوص إلى نماذج الذكاء الاصطناعي.

ونظرًا لاتساع نطاق الأبحاث في هذا المجال، سأنظم عرضها إلى بندين: الأول: تسلسل تاريخي لهذه الأفكار البحثية، والثاني: تصنيف تمثيلي لبعض الأبحاث التي تتماس مع موضوع المحور.

١: **<u>البند الأول</u>:** التسلسل الزمني لتطور الأطر البحثية في مجال المعالجة الطبيعية للغة العربية المدعومة بالذكاء الاصطناعي، مع التركيز على تطور تقنيات الفهم الدلالي وتمثيل العلاقات بين الكلمات في سياقات مختلفة.

(أ)المرحلة المبكرة (في ضوء الإطار المعرفي الذي توفر للباحثة) (٢٠١٤-٢٠١٦)

نلحظ سيطرة الأساليب المعتمدة على القواعد، [٤٨] وكذلك الأنظمة المعتمدة على الأنطولوجيا، [٤٩] وقد ركزت على وضع العلامات اليدوية والرسوم المفاهيمية

(٤٨) W. Bakari, O. Trigui and M. Neji, "Logic-based approach for improving Arabic question answering," *2014 IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, India, 2014, pp. 1-6, doi: 10.1109/ICCIC.2014. 7238319.الذي وفر نهجا يعتمد على المنطق للإجابة على الأسئلة العربية باستخدام الاستلزام النصي

(٤٩) Abouenour L, Nasri M, Bouzoubaa K, Kabbaj A, Rosso P. Construction of an ontology for intelligent Arabic QA systems leveraging the Conceptual Graphs

والمسندات المنطقية,logical predicates [٥٠] في محاولة معالجة التحديات التي تواجه نمذجة اللغة العربية وعلى الرغم من تبني هذه النماذج لمقاربة دلالية صريحة إلا أن قابليتها للتنوع والقدرة على التكيف كانت محدودة.

(ب)المرحلة الثانية: الانتقال إلى النماذج العصبية (٢٠١٧-٢٠١٩) وكانت محاولات مبكرة تمثلت في استخدام تضمينات *(Embeddings from Language Models)* ELMo [٥١] للتشابه الدلالي والعمل على مجموعات البيانات مثل ARCD [٥٢] ويمكننا أن نعد – في حدود الإطار المعرفي الذي أمكن رصده– نموذج سوكال SOQAL [٥٣]تحولا إلى البنى القائمة على المحولات(Transformers ؛ لأنه استخدم

---

representation. *Journal of Intelligent & Fuzzy Systems*. 2014;27(6):2869-2881. doi:10.3233/IFS-141248 الذي وفر نهجا معتمدا على الأنطولوجيا لتحسين أنظمة الإجابة على الأسئلة باللغة العربية، جامعا بينWordN و VerbNet العربية للتمثيل الدلالي باستخدام الرسوم البيانية المفاهيمية ، مما يتيح الفهم الدلالي للأسئلة والمستندات.

(٥٠) تعبير يستخدم لتمثيل المعنى الدلالي للجمل في صيغ قابلة للحوسبة

(٥١) H. Al-Bataineh, W. Farhan, A. Mustafa, H. Seelawi and H. T. Al-Natsheh, "Deep Contextualized Pairwise Semantic Similarity for Arabic Language Questions," *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, Portland, OR, USA, 2019, pp. 1586-1591, doi: 10.1109/ICTAI.2019.00229. وقد قد المرجع نموذجا للتشابه الدلالي من سؤال إلى آخر في العربية مستخدما تضمينات (إلمو) ومعالجة تحديات التشكيل واللهجة محققا درجات F1 عالية على المعايير. وهو نموذج يولد تمثيلات تلقائية للكلمات استنادا إلى السياق.

(٥٢) Hussein Mozannar, Elie Maamary, Karl El Hajal, and Hazem Hajj. 2019. Neural Arabic Question Answering. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 108–118, Florence, Italy. Association for Computational Linguistics. ركز على الفهم الدلالي عبر فهم القراءة العصبي، والاستفادة من مجموعات بيانات ARCD و Arabic-SQuADلمهام الإجابة عن الأسئلة الواقعية ذات المجال المفتوح.

(53)Ibid: Hussein Mozannar, Elie Maamary, Karl El Hajal, and Hazem Hajj. 2019.P:1-11.

عمليات تضمين قائمة علىBERT مقترنة بعمليات استرجاع المستنداتTF–IDF

(ج) المرحلة الثالثة: بداية صعود نماذج المحولات المدربة مسبقا (٢٠٢٠ وما بعدها). ومثّلت هذه المرحلة انتقالًا نوعيًا، تمثل في تبني نماذج مثالها:AraBERT أربرات(٢٠٢٠) [٥٤] الذي عُد أول نموذج مبني علىBERT ، وقد كُيّف مع الميزات اللغوية الخاصة باللغة العربية. وكذلك ظهرت في هذه المرحلة مشاريع ضخمة مثلArTrivia أرتريفيا [٥٥] و AQAD [٥٦]اللذان أسهما مساهمة واسعة النطاق في مجموعات البيانات بتوفيرهما موارد قوية للتدريب والتقييم ومعالجة ندرة البيانات السابقة.

(د) المرحلة المعاصرة (٢٠٢٢–٢٠٢٥) ويمكن أن أطلق عليها بداية الضبط الدقيق للمجال؛ حيث نجد أنواعا من الضبط: أولها: دقة وضبط النماذج مسبقة التدريب في مجالات محددة مثل: القرآن الكريم [٥٧]، والجوانب التاريخية [٥٨]، والتعليم، [٥٩] وغيرها.

---

(٥٤) Antoun, W., Baly, F., & Hajj, H.M. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. *ArXiv, abs/2003.00104*. ,r] ulg ugn على تحسين الفهم الدلالي باستخدام التضمينات السياقية ، ومعالجة التشكيل الخاص باللغة العربية.و ضمان الدقة هو مجال واحد تم اختباره.

(٥٥) Sultan Alrowili and K Vijay-Shanker. 2023. ArTrivia: Harvesting Arabic Wikipedia to Build A New Arabic Question Answering Dataset. In *Proceedings of ArabicNLP 2023*, pages 191–207, Singapore (Hybrid). Association for Computational Linguistics. نموذج مقدم من قبل ويكيبديا العربية ويقيم نماذج الذكاء الاصطناعي الخاصة باللغة العربية، ويستكشف تحديات مثل تطبيع الإجابة وتنوع مجموعات البيانات.

(٥٦) A. Atef, B. Mattar, S. Sherif, E. Elrefai and M. Torki, "AQAD: 17,000+ Arabic Questions for Machine Comprehension of Text," *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, Antalya, Turkey, 2020, pp. 1-6, doi: 10.1109/AICCSA50499.2020.9316526. مجموعة بيانات عربية كبيرة (AQAD) للقراءة الآلية والإجابة على الأسئلة.

(٥٧) Alsaleh, A., Althabiti, S., Alshammari, I.K., Alnefaie, S., Alowaidi, S., Alsaqer, A., Atwell, E., Altahhan, A., & Alsalka, M.A. (2022). LK2022 at Qur'an QA 2022: Simple Transformers Model for Finding Answers to Questions from

وثانيها: التركيز على التمثيل الدلالي للنصوص وتحسين التعامل مع السياقات وتضميناتها embeddings للتعامل مع التشكيل، والتعدد اللهجي ،والغموض الدلالي وغيرها. (٦٠)

٢: **البند الثاني:** تصنيف تمثيلي لبعض الأبحاث التي تتماس مع موضوع المحور(المعالجة الآلية للغة العربية).

(أ) أبحاث حول النماذج الدقيقة الخاصة باللغة العربية: ومنها ما يدور حول دقة الأداء مثل AraBERTv2 و CamelParser2.0الذي حقق وفقا لمؤلفيه أحدث

Qur'an. *OSACT*. تدور الورقة حول الإجابة عن الأسئلة العربية باستخدام نماذج الذكاء الاصطناعي المدربة مسبقا على النصوص القرآنية العربية الفصحى. وكذلك ElKomy, M., & Sarhan, A. M. (2022). TCE at Qur'an QA 2022: Arabic Language Question Answering Over Holy Qur'an Using a Post-Processed Ensemble of BERT-based Models. *arXiv preprint arXiv:2206.01550* الذي ركز على استخدام الفهم الدلالي من النماذج القائمة على *BERT*.

(٥٨) S. Maged *et al.*, "HistoryQuest: Arabic Question Answering in Egyptian History with LLM Fine-Tuning and Transformer Models," *2024 Intelligent Methods, Systems, and Applications (IMSA)*, Giza, Egypt, 2024, pp. 135-140, doi: 10.1109/IMSA61967.2024.10652824 ركزعلى الفهم الدلالي والتطبيق العملي للإجابة عن الأسئلة في التاريخ المصري، باستخدام مجموعات البيانات والنماذج مثل AraBERT و.LLaMa

(٥٩) Sammoudi, M., Habaybeh, A., Ashqar, H. I., & Elhenawy, M. (2024, July). Question-answering (qa) model for a personalized learning assistant for arabic language. In *International Conference on Intelligent Systems, Blockchain, and Communication Technologies* (pp. 356-367). Cham: Springer Nature Switzerland. ركز على الفهم الدلالي من طريق الإجابة على أسئلة من كتب اللغة العربية، وكما ذكر فقد حقق أداء معتدلا (20 EM٪، 51 F1٪).

(٦٠) S. M. AlAwawdeh and G. A. Abandah, "Improving the Accuracy of Semantic Similarity Prediction of Arabic Questions Using Data Augmentation and Ensemble," *2021 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, Amman, Jordan, 2021, pp. 272-277, doi: 10.1109/JEEIT53412.2021.9634095.

النتائج لتحليل التبعية ووضع العلامات النحوية [٦١] من طريق الاستفادة من بنى محولات مضبوطة أحادية اللغة. وقد صَممت ArabianGPT الرمز المميز (AraNizer) وفقا للتشكيل العربي، مما يسهم في تحقيق نتائج قوية في المهام الخاصة بتحليل المشاعر والتلخيص وغيرها. [٦٢]

(ب) – أبحاث حول تطوير بنوك الأشجار Treebank وأطر التعليقات التوضيحية اليدوية، ويمكن توضيح محورها في محاولة إنتاج محلل التبعية العربي CamelParser2.0 وتعليقات توضيحية نحوية في أطر عمل CATiB و UD التي تستفيد من الترميز المتوقع، بتغطية ١٨٨ ألف كلمة عربية قياسية حديثة وعربية فصحى، مصححة يدويا وفقا لإرشادات CATiB ، لتحليل نحوي متنوع. وقد حاول Camel Treebank أن يوفر بنكا شجريا غنيا متعدد الأنواع للتبعية، مما يدعم تحليل التباينات المعجمية والنحوية. [٦٣] وكذلك حاول Palmyra 3.0 توفير منصة قائمة

(٦١)على سبيل المثال:- Al-Ghamdi, S., Al-Khalifa, H., & Al-Salman, A. (2023). Fine-Tuning BERT-Based Pre-Trained Models for Arabic Dependency Parsing. *Applied Sciences*, *13*(7), 4225. https://doi.org/10.3390/app13074225 وقد وفر البحث تقييما مقارنا لنماذج BERT العربية المضبوطة بدقة لتحليل التبعية باستخدام تسعة نماذج مدربة مسبقا، وثلاثة بنوك شجرة ، واستراتيجيات ضبط دقيقة ، وتحديد التحديات الحرجة مثل أخطاء الترميز و Go Habash. 2022. Morphosyntactic ينظر كذلك Inoue, Salam Khalifa, and Nizar Tagging with Pre-trained Language Models for Arabic and its Dialects. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1708–1719, Dublin, Ireland. Association for Computational Linguistics. وقد حاول استكشاف إعدادات الضبط الدقيق والموارد الخارجية محققا أحدث النتائج على العربية الفصحى الحديثة واللهجات، والبحث متصل بالمعالجة اللغوية العربية لكنه افتقر إلى التركيز الصريح على التحليل النحوي أو التحليلات النحوية المشروحة أو المقارنات بين اللغات.

(62)see: Koubaa, A., Ammar, A., Ghouti, L., Najar, O., & Sibaee, S. (2024). Arabiangpt: Native arabic gpt-based large language model. arXiv preprint arXiv:2402. 15313. وقد دمج البحث نماذج ArabianGPT مع AraNizer tokenizer لتمثيل النص بدقة، مركزا على التشكيل وبناء الجملة الفريدين للغة العربية.

(63)Habash, N., AbuOdeh, M., Taji, D., Faraj, R., Gizuli, J.E., & Kallas, O. (2022). Camel Treebank: An Open Multi-genre Arabic Dependency Treebank. *International Conference on Language Resources and Evaluation*.

على السحابة للصرف، والتعليقات التوضيحية بناء الجملة، قابلة للتكيف مع مهام بنوك الأشجار العربية مركزا على أدوات التعليقات التوضيحية بدلا من LLMs أو التحديات النحوية / الدلالية المقارنة. (٦٤)

(ج) – أبحاث دارت حول البيانات التركيبية للتعميم وتصحيح الأخطاء، حيث ناقشت توليد واستخدام البيانات التركيبية لتعزز تصحيح الأخطاء التوضيحية في اللغة العربية، منها على سبيل المثال: ما دار حول تطوير نماذج علامات الأخطاء (٢٦ نوع من الأخطاء) والترجمة العكسية لمجموعة متنوعة منGEC ، مما يحقق أحدث النتائج، ولكنه يفتقر إلى التحليل النحوي الصريح المستند إلى LLM أو التعليق التوضيحي. (٦٥)ومنها ما استخدم البيانات التركيبية للزيادة في أداء GPT–4 على GEC العربي، مركزا على الضبط الدقيق للتعليم ومداره مورفولوجيا العربية مما يدل على تحسينات مقارنة بالمعايير القياسية. (٦٦)

(د) أبحاث تتعلق باستخدام المحولات للتحليل اللهجي والمورفولوجي للغة العربية، فانطلقت إلى استكشاف نماذج المحولات للصرف العربي وتمييز اللهجات والتعميم النحوي، منها على سبيل المثال ما يوضح فائدة الموارد الخارجية لوضع العلامات

(64)AbuOdeh, M., Phan, L., Elshabrawy, A.F., & Habash, N. (2024). Palmyra 3.0: A User-Friendly Cloud-Based Platform for Morphology and Dependency Syntax Annotation. *International Conference on Language Resources and Evaluation*.

(65)Alrehili, A., & Alhothali, A. (2025). Towards the Development of Balanced Synthetic Data for Correcting Grammatical Errors in Arabic: An Approach Based on Error Tagging Model and Synthetic Data Generating Model. *arXiv preprint arXiv:2502.05312*.

(66)Sang Kwon, Gagan Bhatia, El Moatez Billah Nagoudi, and Muhammad Abdul-Mageed. 2023. Beyond English: Evaluating LLMs for Arabic Grammatical Error Correction. In *Proceedings of ArabicNLP 2023*, pages 101–119, Singapore (Hybrid). Association for Computational Linguistics.وكذلكKwon, S. Y., Bhatia, G., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). Beyond english: Evaluating llms for arabic grammatical error correction. *arXiv preprint arXiv:2312.08400*.

النحوية اللهجية،[٦٧] ومنها ما عمل على وسم الكلام المدرب على مجموعة بيانات مشروحة من التغريدات العربية في أربع مجموعات لهجات رئيسية: المصرية والشامية والخليجية والمغاربية، بتنفيذ نهجين مختلفين لوضع علامات التسلسل. يستخدم الأول الحقول العشوائية الشرطية(CRFs) ، بينما يجمع الثاني بين التمثيلات المستندة إلى الكلمات والأحرف في شبكة عصبية عميقة مع طبقات مكدسة من الشبكات التلافيفية والمتكررة مع طبقة إخراج .CRF ووفقا للمؤلفين فقد نجحوا في استغلال مجموعة متنوعة من الميزات التي تساعد في تعميم النماذج المنتقاة، مثل مجموعات البنية وقوالب الجذع. أيضا، طوروا نماذج مشتركة قوية تضع علامة على التغريدات متعددة اللهجات وتتفوق على علامات اللهجة الواحدة. محققة دقة مجمعة تبلغ ٩٢.٤٪ عبر جميع اللهجات، مع نتائج لكل لهجة تتراوح بين ٩٠.٢٪ و ٩٥.٤٪. حصلنا على النتائج باستخدام تقسيم القطار / التطوير / الاختبار ٢٠/١٠/٧٠ لمجموعة بيانات من ٣٥٠ تغريدة لكل لهجة. [٦٨]ومنها ما وفر رؤى حول ترميز المحولات العربية للمعلومات اللغوية مركزا على استكشاف النماذج العربية سابقة التدريب باستخدام المهام الجوهرية مثل وضع العلامات المورفولوجية وتحديد اللهجة لتحليل التمثيلات الداخلية.[٦٩]

(هـ)أبحاث عن قيود الذكاء الاصطناعي: تطرقت هذه الدراسات إلى حدود قدرة النماذج الكبرى على فهم البنية العميقة للغة العربية والتعامل مع بناء الجملة والصرف والدلالات الخاصة بها ومن هذه الأبحاث: ما سلط الضوء على صعوبة تعامل

(67) *bid*. Habash. 2022. Morphosyntactic Tagging with Pre-trained Language Models for Arabic and its Dialects.

(68)Darwish, K., Attia, M., Mubarak, H., Samih, Y., Abdelali, A., Màrquez, L., … Kallmeyer, L. (2020). Effective multi-dialectal arabic POS tagging. *Natural Language Engineering*, *26*(6), 677–690. doi:10.1017/S1351324920000078.

(69)Abdelali, A., Durrani, N., Dalvi, F., & Sajjad, H. (2022). Interpreting Arabic Transformer Models. *ArXiv, abs/2201.07434*.

ChatGPTمع التعابير الاصطلاحية مقدما تحليلا مقارنا لأدوات الذكاء الاصطناعي ودوها في التحليل العربي، كأن يقارن ويفحص Cloude و ChatGPT حال تعاملهما مع بناء الجملة العربية والتراكيب الغامضة والصرف والتعابير الاصطلاحية مسلطا الضوء على اختلاف الأداء. (٧٠) ومنها ما ناقش القيود في رسم خرائط rab'iوالتعامل مع القواعد النحوية الدقيقة، محللا أداء ChatGPT و Geminiو Perplexityعلى rab'iوعلامات الترقيم (الحركات) والترجمة، مع التركيز على تحديات رسم خرائط القواعد النحوية، مفتقرا إلى التعليقات التوضيحية النحوية التفصيلية القائمة على LLM (٧١)

<u>تعقيب:</u> في حدود الدراسات المستشهد بها في هذه الورقة سجل نموذج AraBERT تفوقا على كثير من الأساليب العصبية القائمة على القواعد، لا سيما في مهام الإجابة عن الأسئلة(Question Answering) ، بفضل تضميناته السياقية الدقيقة. (٧٢)

---

(70)Aljanabi, Mohammad. (2024). Assessing the Arabic Parsing Capabilities of ChatGPT and Cloude: An Expert-Based Comparative Study. Mesopotamian Journal of Arabic Language Studies. 16-23. 10.58496/MJALS/2024/002.. https://doi.org/10.58496/MJALS/2024/002

(71)Siyam, Fera & Hidayat, Rahmat & Rochmat, Cecep & Maulaya, Rosendah & Avilya, Annisa & Maulidi, Muhammad. (2024). Accuracy Analysis of Artificial Intelligence in Arabic Language Translation and Grammatical Rule Mapping. Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab. 16. 558. 10.24042/albayan.v16i2.24588.

(72)**see:** Afnan H. Alshehri. AraQA-BERT: Towards an Arabic Question Answering System using Pre-trained BERT Models. WSEAS Transactions on Information Science and Applications. 2024; 21:361-373. الذي وفر نظاما للإجابة عن الأسئلة باللغة العربية باستخدام النماذج 10.37394/23209.2024.21.34 ويمتاز MSA) للغة العربية الفصحى الحديثة AraBERT مركزا على الفهم الدلالي عبر BERT سابقة التدريب بتناولـه التطبيـق العملـي والتحليـل الـدلالي باللغـة العربيـة بشـكل مباشـر. وكـذلك **And see**: Hussein Mozannar, Elie Maamary, Karl El Hajal, and Hazem Hajj. 2019. Neural Arabic Question Answering. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 108–118, Florence, Italy. Association for Computational Linguistics. والتضمينات السياقية، مركزا على الفهم الدلالي عبر BERT الذي استخدم القراءة العصبي مسلطا الضوء على التحديات اللغوية المتعلقة بالعربية مثل التشكيل ومحققا تقدما وجودة ملحوظه

وكذلك حقق تفوقا ملحوظا على نماذج المحولات الأخرى مثلAraGPT[٧٣] و) T5[٧٤] وتعدLLaMA [٧٥] فعالة أيضا، لا سيما عند إخضاعها لعمليات ضبط دقيقة على مجموعات بيانات متخصصة.. في المقابل، ورغم دقة الأساليب القائمة على المنطق اليدوي، فإن فعاليتها تبقى محدودة في السياقات المرنة، مما يعزز الحاجة إلى دمجها بتقنيات مثل التعلم الآلي لتحقيق أداء أكثر توافقًا مع خصائص اللغة العربية..

Alsubhi, K., Jamal, A.T., & Alhothali, A.M. (2021). Pre-trained Transformer-Based Approach for Arabic Question Answering: A Comparative Study. *ArXiv, abs/2111.05671*. و**أيضا** Arabic-SQuADو ARCD بالاستفادة من مجموعات بيانات ElKomy, M., & Sarhan, A. M. (2022). TCE at Qur'an QA 2022: Arabic Language Question Answering Over Holy Qur'an Using a Post-Processed Ensemble of BERT-based Models. *arXiv preprint arXiv:2206.01550*. الذي ركز على الإجابة عن الأسئلة في معالجة بعض pRR: 56.6٪) درجة) الألفاظ القرآنية، مع المعالجة اللاحقة لتحسين التنبؤات الدلالية.

(٧٣) J. Aljebreen, A. Alhargan and S. Nagro, "Enhancing Zero-Shot Question Answering in Arabic Language with Fine-Tuned AraGPT Model," *2025 8th International Conference on Data Science and Machine Learning Applications (CDMA)*, Riyadh, Saudi Arabia, 2025, pp. 48-53, doi: 10.1109/CDMA61895.2025.00014.وقد وفر تحسين التنبؤ بالتشابه الدلالي للأسئلة العربية باستخدام طرق AraBERT والمجموعة.

(٧٤) Mellah, Y., Touahri, I., Kaddari, Z., Haja, Z., Berrich, J., & Bouchentouf, T. (2022). LARSA22 at Qur'an QA 2022: Text-to-Text Transformer for Finding Answers to Questions from Qur'an. *OSACT*. الذي استفاد من mT5 مع تحليل النص الدلالي، مركزا على النص القرآني الكريم محاولا نمذجة العلاقات الدلالية باللغة العربية للحصول على إجابات دقيقة لأزواج من الأسئلة والأجوبة المشروحة من الخبراء.

(٧٥) A. Tamer, A. -A. Hassan, A. Ali, N. Salah and W. Medhat, "Fine Tuning of large language Models for Arabic Language," *2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, Giza, Egypt, 2023, pp. 1-4, doi: 10.1109/AICCSA59173.2023.10479346.

المحور الثاني: الأطر الفنية لتطوير أدوات المعجم العربي المعززة الذكاء الاصطناعي

يقصد هذا المحور إلى توضيح بعض الآليات التقنية لبناء المعاجم الذكية؛ إذ لا يزال دمج الأدوات المعتمدة على الذكاء الاصطناعي في المعجم العربي الرقمي أمراً غير مكتشف في الوقت الحالي فمعظم الأبحاث تركّز على توحيد إطار الترميز المعجمي (LMF) والأساليب المبنية على القواعد للإغناء الدلالي والنحوي. بينما تظل منهجيات الذكاء الاصطناعي المتطورة، مثل الشبكات العصبونية والتعلم العميق، غائبة إلى حد كبير. ويمكنني تقسيم الأعمال في حدود الإطار المعرفي الذي أمكن رصده، إلى أقسام متنوعة مترابطة على النحو التالي:

القسم الأول: أوراق بحثية حاولت إثراء المعاجم العربية الرقمية من عدة منهجيات، منها ما اقترح الإثراء بواسطة إنشاء روابط نحوية دلالية، مسلطة الضوء على الإثراء النحوي والدلالي عبر استخدام المحتوى النصي في قواميس LMF وتصنيف Gaston Grossللفئات الدلالية، مطبقة على معجم المدار. [٧٦]وهناك من حاول من طريق منهجية ثلاثية الخطوات تشمل: تحديد الحجج النحوية، والربط الدلالي ثم التكامل النحوي الدلالي. وهذا البحث وإن كان لا يتناول بشكل مباشر تقنيات أو أطر الذكاء الاصطناعي إلا ان أهميته تكمن في إثراء المعاجم العربية الرقمية لمهام معالجة

(76)Elleuch, I., Gargouri, B., & Ben Hamadou, A. (2021). Lexical data mining-based approach for the self-enrichment of LMF standardized dictionaries: Case of the syntactico-semantic knowledge. *CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE*, *33*(17). https://doi-org.sdl.idm.oclc.org/10.1002/cpe.6312تصنيف جروس هو إطارٌ نظري يُصنِّف الأفعال والكلمات بناءً على الدور الدلالي للعناصر المرتبطة بها في الجملة، مع التركيز على العلاقة بين التركيب النحوي والدلالة. يعتمد التصنيف على تحليل "الحجج النحوية" (مثل الفاعل، المفعول به) وكيفية تفاعلها مع المعنى العام للجملة.

اللغات الطبيعية. (٧٧)ومنها ما استخدم تصنيف جاستون جروس وحقول الموضوعات في المعاجم العربية الموحدة. (٧٨) ومنها ما ركز على استخراج السلوك النحوي من المحتوى النصي العربي في قواميس LMF وربطها بالمعاني المعجمية. ينطبق على ٩,٨٠٠ فعل عربي في المدار. (٧٩)

القسم الثاني: أعمال ركزت على استخدام معايير ISO لإطار الترميز المعجمي (LMF)لتطبيع المعجم وقابلية التوسيع: منها ما دعا إلى الاستخدام التحريري للمعاجم العربية مع نمذجة LMF الموحدة، مما يتيح وظائف الاستجواب العامة لاحتياجات المستخدم عبر نظام.ADIQTQ (٨٠)، ومنها ما طبق معايير LMF–ISO 24613 على معجم المدار لهيكلة المعاجم العربية متضمنا أنماطا صرفية غنية وعلاقات نحوية

(77)Elleuch, I., Gargouri, B., & Hamadou, A.B. (2015). Self Syntactico-Semantic Enrichment of LMF Normalized Dictionaries. *Pacific Asia Conference on Language, Information and Computation*.

(٧٨) Gargouri, B., Elleuch, I., & Hamadou, A.B. (2014). Towards automatic enrichment of standardized electronic dictionaries by semantic classes. *ROCLING/IJCLCLP*. هذه الورقة تعزز العمق الدلالي وإن كانت مفتقرة إلى تعريفات صريحة للذكاء الاصطناعي أو التركيز الأوسع على تطوير المعجم بما يتجاوز الإثراء الدلالي.

(٧٩) Elleuch, I., Gargouri, B., & Hamadou, A.B. (2015). Using the Textual Content of the LMF-Normalized Dictionaries for Identifying and Linking the Syntactic Behaviors to the Meanings. *Recent Advances in Natural Language Processing*. هذه الورقة ركزت على التحليل اللغوي بدلا من التحسين القائم على الذكاء الاصطناعي أو تطوير القواميس العربية الرقمية..

(٨٠) Baccar Ben Amar, F., Khemakhem, A., Gargouri, B., Haddar, K. and Ben Hamadou, A. (2008). LMF Standardized Model for the Editorial Electronic Dictionaries of Arabic. In *Proceedings of the 5th Internationa l Workshop on Natural Language Processing and Cognitive Science (ICEIS 2008) - NLPCS*; ISBN 978-989-8111-45-6, SciTePress, pages 64-73. DOI: 10.5220/0001737800640073 ن الشريط العلوي، اضغط على **"تحديد " (Select)**ورغم أن هذه الورقة لم تذكر الذكاء الاصطناعي صراحة ، فإن الإطار الموحد قد يدعم التكامل التكنولوجي المستقبلي ، بما في ذلك التحسينات المعجمية القائمة على الذكاء الاصطناعي.

دلاليـة، مـوفرا واجهـة مسـتخدم علـى شـبكة الإنترنـت [٨١]ومنهـا مـا اقتـرح تطبيـق استراتيجيات تطبيع LMF لربط الهياكل المعجمية بالمعرفة النحوية والدلالية. [٨٢]

القسم الثالث: أبحاث عملية للتوليد التلقائي للمعجم العربي الرقمي سلطت الضوء علـى أنظمـة إنشـاء المعجـم الرقمـي محاولـة الاسـتفادة مـن الجانـب النظـري اللغـوي والعمليات الخوارزمية، فمنها ما وصف تطوير معجم رقمي عربي متخذا من المعجم الوسيط وبياناته المحوسبة مجالا للتطبيق، محاولا دمج التحليل المورفولوجي والدلالي مع دعم واجهة المستخدم الرسومية للبحث من دون تحديد جذر مسبق [٨٣]، ومنها ما استخدم نظرية" شروط التركيب المورفيمية" ليشرح عملية التوليد التلقائي للمعجم، وكان هذا جزء من مشروع مختبر RIADI [٨٤]

أما القسم الرابع، فكانت أوراق ركزت على تعزيز تعلم اللغة وتفاعل المستخدم عبر مناقشـة واقتـراح اسـتراتيجيات وأدوات معجميـة وميـزات متعـددة اللغـات؛ فنجـد مقترحـا

---

(٨١) KHEMAKHEM, A., GARGOURI, B., BEN HAMADOU, A., & FRANCOPOULO, G. (2016). ISO standard modeling of a large Arabic dictionary. *Natural Language Engineering*, *22*(6), 849–879. doi:10.1017/S1351324915000224 ركزت هذه الورقة على تحسين المعجم الرقمي العربي ولكنها تفتقر إلى ذكر محدد للتقنيات القائمة على الذكاء الاصطناعي أو النماذج المتقدمة.

(٨٢)المرجع السابق Elleuch, I., Gargouri, B., & Ben Hamadou, A. (2021). وقد طبق يطبق أساليب الذكاء الاصطناعي لتحسين محتوى المعجم العربي، بما يتماشى مع أهداف تحسين المعجم.

(٨٣) O. Al Dakkak and A. Zein, "Towards Arabic Electronic Dictionary," *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, Damascus, Syria, 2008, pp. 1-6, doi: 10.1109/ICTTA.2008.4530327. وقد عالج البحث التحديات المعجمية باللغة العربية ويتوافق جزئيا مع تحسين المعجم، لكنه يفتقر إلى التركيز المباشر على أساليب الذكاء الاصطناعي.

(٨٤) Haddad, A., Zrigui, M., & Ahmed, M.B. (2005). Un système de génération automatique de dictionnaires linguistiques de l'arabe. *JEPTALNRECITAL*. الورقة كذلك لم تركز على الأساليب القائمة على الذكاء الاصطناعي أو التطبيقات العربية الفصحى الحديثة.

لمعجم عربي متعدد اللغات معتمد على الجذر مع أدوات التقييم الذاتي لتعزيز مساهمة المستخدم. كما نجد من قدم الشفافية المورفولوجية لتسهيل البحث وتكامل البيانات. (٨٥)

وعن **القسم الخامس** فسلط الضوء على النمذجة الدلالية الصرفية عبر تعزيز التصنيف الدلالي والتمثيل المورفولوجي في التركيب المعجمي بوجه خاص، ومن أمثلتها – في ضوء ما توفر من الدراسات والبحوث ذات الصلة– يعالج أشكال الكلمات المنحرفة والنمذجة المورفولوجية المصممة خصيصا لميزات اللغة العربية في القواميس الموحدة ل.LMF (٨٦)، ومنها استخدم طرق التصنيف الدلالي Gaston Gross لإضافة عمق دلالي إلى إدخالات المعجم. (٨٧)

تعقيب: عند النظر إلى الأقسام والمراحل السابق ذكرها نلحظ أنه تقريبا كل الجهود انصبت على النمذجة التأسيسية، ثم الانتقال إلى التوحيد القياسي والاثراء المعجمي، ثم منذ عام (٢٠٢٢) تقريبا بدأت محاولات محدودة لدمج الذكاء الاصطناعي من أجل تحسين العمل المعجمي تناولت بشكل غير مباشر منهجيات الذكاء الاصطناعي معتمدة على النظريات اللغوية والإثراء اليدوي، ولكنها تفتقر إلى أدوات الذكاء الاصطناعي المتقدمة الصريحة مثل الشبكات العصبية أو نماذج المحولات.

(٨٥)المرجع السابق O. Al Dakkak and A. Zein, "Towards Arabic Electronic Dictionary," *2008*

(٨٦)المرجع السابق KHEMAKHEM, A., GARGOURI, B., BEN HAMADOU, A., & FRANCOPOULO, G. (2016).

(٨٧)المرجع السابق Gargouri, B., Elleuch, I., & Hamadou, A.B. (2014).

المبحث الخامس: التطبيقات والمبادرات العملية، نحو نماذج ديناميكية للمفاهيم.

يستعرض المبحث بعض المبادرات والمشروعات العربية لتحسين وتطوير المعجم العربي الرقمي باستخدام الذكاء الاصطناعي. مع التركيز على تكنولوجيا الذكاء الاصطناعي والجوانب اللغوية المرتبطة به، وبالمقارنة بين الجهود العربية والجهود اللغوية الأخرى نجد ما يلي:

١- دراسات حالة ومشروعات عملية تطبيقية استخدمت الذكاء الاصطناعي في بناء المعاجم أو تحليل المحتوى المفاهيمي، وتتضمن العربية.

١-١: مشروع "BabelNet" (جامعة روما) (٨٨)

١-٢: مشروع "Lexicala" (شركة K Dictionaries) (٨٩)

٢- دراسات حالة ومشروعات عربية عملية تطبيقية:

بالاطلاع على المبادرات والمشروعات المؤسسية العربية لتطوير وتحسين المعجم الرقمي العربي المدعوم بالذكاء الاصطناعي نتبين تقدما ملحوظا في إثراء المعجم الرقمي العربي، وخاصة WordNet العربية(AWN) باستخدام الأساليب شبه الآلية وتقنيات الذكاء الاصطناعي مثل تضمين الكلمات والشبكات العصبية، ومع ذلك توجد أدلة محدودة على وجود مبادرات شاملة يقودها العرب تدمج الذكاء الاصطناعي المتقدم عبر مجالات متعددة، وقد رصدت الباحثة بعضا منها متمثلة للمشاريع الرئيسة فيما يلي:

١-٢: ثروة: معجم لهجة ثلاثي اللغات يدمج العربية المصرية وMSAوالإنجليزية(٩٠)

---

(٨٨)وفقا للموقع الرسمي فهي قاعدة بيانات ضخمة متعددة اللغات تربط بين المفردات والمفاهيم. وتتكون من ٢٣مليون إدخال على وجه التقريب https://babelnet.org/

(٨٩)معجم متعدد اللغات باستخدام الذكاء الاصطناعي يستخدم التعلم العميق لاستخراج التعريفات والتصريفات https://lexicala.com/

(90)Diab, M.T., Al-Badrashiny, M., Aminian, M., Attia, M.A., Elfardy, H., Habash, N., Hawwari, A., Salloum, W., Dasigi, P., & Eskander, R. (2014). Tharwa: A Large

٢-٢: مكنون: معجم للغة العربية الفلسطينية، مع شرح يدوي للملامح اللغوية.[٩١]

٢-٣: المعجم التونسي: التوليد التلقائي للأسماء اللفظية باستخدام أدوات MSA [٩٢]

٢-٤: مشروع الأنطولوجيا العربية: هو محرك بحث معجمي من إنتاج جامعة بيرزيت. يتيح للباحث استرجاع ترجمة مترادفات ومعاني كلمة معينة من ١٥٠ (مئة وخمسين) معجما عربيا ومتعدد اللغات. وقد حصل الموقع على جائزة محمد بن راشد للغة العربية لسنة ٢٠١٨ لأفضل مشروع في فئة أفضل مبادرة في استعمال شبكات التواصل الاجتماعي أو تطبيق تقني ذكي لتعلم اللغة العربية ونشرها [٩٣]

٢-٥: مشروع "Arabic WordNet" (جامعة نيويورك أبو ظبي) [٩٤]

تعقيب: بالنظر إلى المبادرات السابقة نلحظ ما يلي:

١- يغلب على المشاريع العربية معاجم للهجات تعطي الأولوية للحفظ اللغوي، ولكنها لا تستفيد استفادة كافية من منهجيات الذكاء الاصطناعي، وتعتمد بشكل ملحوظ على الجهد اليدوي وتقريب المجموعة.

Scale Dialectal Arabic - Standard Arabic - English Lexicon. *International Conference on Language Resources and Evaluation*.

(91)Dibas, S., Khairallah, C., Habash, N., Sadi, O. F., Sairafy, T., Sarabta, K., & Ardah, A. (2022). Maknuune: A large open palestinian arabic lexicon. *arXiv preprint arXiv:2210.12985*.

(92)Ahmed Hamdi, Núria Gala, and Alexis Nasr. 2014. Automatically building a Tunisian Lexicon for Deverbal Nouns. In *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*, pages 95–102, Dublin, Ireland. Association for Computational Linguistics and Dublin City University

(93)https://2u.pw/riUTf

(٩٤) قاعدة بيانات عربية تربط بين المفردات ومعانيها وعلاقاتها مع نظيرتها في WordNet الإنجليزي باستخدام التعلم الآلي، وتستخدم في الترجمة الآلية وتحليل النصوص الأدبية والدينية العربية. ويستخدم المشروع الأنطولوجيا العليا المدمجة المقترحة بوصفها لغة وسيطة لربط شبكة وورد نت العربية بشبكات وورد نت المطورة سابقًا https://2u.pw/Tgynp

٢-تتسم معظم المبادرات والمشروعات بطابع جغرافي أو لهجي معزول، مما يعطي انطباعا بوجود فجوة أو نقص المبادرات العربية المنسقة والشاملة.

٣-قليلة هي المبادرات المستفيدة من تكامل تقنيات الذكاء الاصطناعي مع نمذجة اللغة.

٤-هيمنة التنوع اللهجي والإثراء الدلالي على المجال البحثي.

٥-البيانات هي الوقود، فكل مشروع يعتمد على بيانات ضخمة ودقيقة.

٦-يؤدي التكامل بين المنهجيات مثل الجمع بين NLPوالتعلم العميق إلى نتائج ملحوظة.

وبالنظر إلى الأبحاث التي تناولت علاقة الذكاء الاصطناعي ومعالجة اللغة العربية الطبيعية نظريا وعمليا يمكننا استنتاج ما يلي:

- حققت النماذج العربية الأصلية المضبوطة بدقة مثل AraBERTv2 و ArabianGPT تفوقا مقارنة بالنجاح المحدود الذي حققته النماذج اللغوية الكبرى (LLMs)حتى وقت هذه الورقة (على سبيل المثال، ChatGPT و4-GPT). نظرا لقدرتها على معالجة التحديات المورفولوجية والنحوية الفريدة للغة العربية بشكل أكثر فعالية.
- تؤدي الموارد المستندة إلى Corpus مثل Camel Treebank ومنصات مثل Palmyra 3.0 دورا مهما في تطوير التعليقات التوضيحية النحوية.
- ما تزال معالجة اللهجات والمقارنات عبر اللغات معقدة، وتُعالج عبر التدريب المسبق المستهدف ووضع علامات على الأخطاء وأطر التبعيات العالمية. ومع ذلك، ما تزال الفجوة في دمج هذه التقنيات في نموذج مفاهيمي شامل للمعاجم الرقمية قائمة.
- تشير المراجع إلى مساعي التوحيد القياسي، وإنشاء معجم مع التركيز على اللهجات، والإثراء القائم على الذكاء الاصطناعي لتحليل المشاعر والمعالجة الصرفية. ومع ذلكما تزال هناك فجوات بارزة من أبرزها، الاستخدام المحدود للذكاء الاصطناعي في أطر المعجم الأوسع وأولوية استخدام موارد العربية الفصحى الحديثة(MSA) على حساب التوازن بين اللهجات والفصحى.

المبحث السادس: استراتيجيات مقترحة لتعزيز نمذجة المحتوى المفاهيمي العربي

يتضح بناءً على التحليل السابق، وفي حدود الإطار المعرفي الذي أمكن رصده وجود فجوة تقنية في نمذجة المحتوى المفاهيمي العربي؛ حيث غياب الأعمال الجامعة بين التقنيات المتقدمة (شبكات عصبية وتقنيات تعلم عميق وأنظمة استخراج العلاقات الثلاثية من النصوص...إلخ) والتي نجح استخدامها في لغات أخرى. وغياب المعايير الموحدة؛ فلا توجد مجموعة بيانات لقياس أو تقييم الأنطولوجيات العربية... إلخ. ومن الاستراتيجيات المقترحة لتعزيز نمذجة المحتوى المفاهيمي العربي

١-دمج التقنيات الحديثة مثل استخدام نماذج المحولات (Transformers) مثل ChatGPT المُدرَّبة على العربية لاستخراج العلاقات الدلالية تلقائيًا.

٢- معالجة اللهجات وتحسين التكامل مع الويب الدلالي.

هيكل نظري مقترح لنمذجة المحتوى المفاهيمي لمصطلحات اللسانيات الإدراكية باستخدام الذكاء الاصطناعي.

فنظرًا للطبيعة المركّبة لإشكالية نمذجة المحتوى المفاهيمي في المعاجم العربية الرقمية، وما يتطلبه ذلك من موارد تطبيقية وتجريبية واسعة النطاق، فقد ارتأت الباحثة تقديم مقترح نظري أولي، يهدف إلى وضع إطار مفاهيمي منظم. بدافع الحاجة إلى بناء تصور متماسك للظاهرة قيد الدراسة قبل الانتقال إلى الاختبار العملي، فضلًا عن أهمية سد فجوة معرفية قائمة في هذا المجال.

وهذا الإطار النظري المبدئي لم يتحقق من فاعليته بالتطبيق العملي، إلا أنه يستند إلى أساس نظري راسخ عبر الاطلاع على عدد كبير من الدراسات والأوراق البحثية المعتمدة في هذا المجال [٩٥]. وتأمل الباحثة أن يشكل النموذج فرضية جديدة قابلة

(95)Including, for example, Kungurtsev, O. B., Mileiko, I. I., & Novikova, N. O. (2024). TECHNOLOGY FOR AUTOMATED CONSTRUCTION OF DOMAIN DICTIONARIES WITH SPECIAL PROCESSING OF SHORT DOCUMENTS. *Radio Electronics, Computer Science, Control*, (4), 148. https://doi.org/10.15588/1607-

للتجريب، أو أن يسهم في فتح آفاق جديدة للبحث والتطوير، ولا سيما في مجال حوسبة المصطلحات، مع تركيز خاص على مصطلحات اللسانيات الإدراكية بما تتسم به من تشعب مفاهيمي وتجاذب بين عدة تخصصات.

يتجه هذا المبحث إلى تقديم مقترح لتصميم نموذج حاسوبي متقدم يعتمد على تقنيات الذكاء الاصطناعي(AI) ، بغرض معالجة المحتوى المفاهيمي لمصطلحات اللسانيات الإدراكية .(Cognitive Linguistics) ويرتكز النموذج المقترح على تحليل المدونات النصية باستخدام خوارزميات التعلم العميق (Deep Learning) وتقنيات المعالجة الآلية للغة الطبيعية(NLP) ، بهدف إنشاء معجم ديناميكي قادر على استيعاب العلاقات الدلالية والتركيبية بين المصطلحات وتمثيلها في صيغ قابلة للتوظيف في تطبيقات الذكاء الاصطناعي، مثل الفهم الآلي، والترجمة الآلية، وأنظمة الحوار الذكية.

وتود الباحثة، قبل المضي في عرض تفاصيل هذا المقترح، التنبيه إلى عدد من الاحترازات المنهجية، على النحو الآتي:

١. **الاحتراز الأول** :إن تقديم هذا المقترح لا يعني التزام الباحثة بتنفيذه أو تطبيقه العملي الكامل؛ وإنما هو تصور نظري مبني على قناعة بأهمية استثمار تقنيات الذكاء الاصطناعي في جمع المصطلحات وتوليدها وتنظيمها ضمن مجالات

وكذلك Sampritha H. Manjunath and John P. McCrae. Encoder-Attention-Based Automatic Term Recognition (EA-ATR). In 3rd Conference on Language, Data and Knowledge (LDK 2021). Open Access Series in Informatics (OASIcs), Volume 93, pp. 23:1-23:13, Schloss Dagstuhl – Leibniz-Zentrum für Informatik (2021) https://doi.org/10.4230/OASIcs.LDK.2021.23 وأيضا Nazar, R., & Acosta, N. (2023). Termout: a tool for the semi-automatic creation of term databases. *CONTENTS*. Kungurtsev, O., Zinovatnaya, S., Potochniak, I., وكذلك Kutasevych, M. (2018). Development of information technology of term extraction from documents in natural language. *Eastern-European Journal of Enterprise Technologies*, *6*(2 (96), 44–51. https://doi.org/10.15587/1729-4061.2018.147978 والمرجع السابق كامل، وفاء، محسن رشوان، وعبد العاطي هواري (٢٠١٤

معرفية متعددة، واستنتاج الآليات الكفيلة بتحقيق معايير جودة عالية في تصميم معاجم مصطلحية مبنية بصورة تلقائية.

٢. **الاحتراز الثاني** :ينتمي النموذج المقترح إلى المقاربات المعرفية في اللغويات الحاسوبية، لا إلى المقاربات الإحصائية أو النحوية. ففي الوقت الذي اعتمدت فيه نماذج ATR التقليدية على أساليب خالية من السياق وقائمة على القواعد، مما أدى إلى تجاهل البعد المفاهيمي للمصطلحات، فإن هذا النموذج يسعى إلى مراعاة الخاصية السياقية، مما يفضي إلى دقة أعلى في تحديد المصطلحات. وقد شهدت الآونة الأخيرة تطورًا ملحوظًا في هذا السياق، لاسيما مع ظهور تقنيات مثل نموذج BERT، الذي يلتقط سياقات الكلمات بصورة ديناميةـ، ويُحسن تصنيف المصطلحات المرشحة، مما يرفع من دقة نتائج استخراج المصطلحات مقارنة بالطرائق التقليدية.

٣. **الاحتراز الثالث** :يُعد تقليل أبعاد المدخلات أحد العوامل الحاسمة في تحسين أداء تقنيات التعلم العميق، مثل الشبكات العصبية التكرارية (RNN) والشبكات العصبية الالتفافية(CNN) ، إلى جانب الكفاءة والدقة. وتتصور الباحثة أنه، حال تجريب هذا النموذج، يمكن تعميم منهجيته ليُوظف على مجموعات بيانات مماثلة في مجالات علمية أخرى.

١- عنوان المقترح: نموذج ذكي لتمثيل مصطلحات اللسانيات الإدراكية.

يركز المقترح على جميع مصطلحات اللسانيات الإدراكية ضمن المناهج الهجينة للنمذجة الذكية (نحو الرسوم البيانية المعرفية والتضمينات العصبية وغيرها). مع إعطاء الأولوية للتشابه الدلالي والعناصر الهيكلية للمعنى، وقد نضع له مقياس تقييم متخيل مثل استخدامه للمحاكاة المعرفية أو مهام معالجة اللغات الطبيعية أو التحليل اللغوي.

٢- مبررات المقترح:

٢-١:اختيار مصطلحات اللسانيات الإدراكية:

ينبع اختيار مصطلحات اللسانيات الإدراكية – إضافة إلى اهتمام الباحثة بمجالها البحثي . من أن الصعوبة في ترجمة Cognitive Linguisticsإلى العربية ليست فقط ترجمة كلمة وإنما نقل منظومة فكرية كاملة تتقاطع مع علم النفس، والذكاء، والإدراك، والثقافة، والتكنولوجيا. لهذا السبب يُعد من أصعب المصطلحات نقلاً إلى العربية بدقة ووضوح دون شرح مصاحب أو حاشية تفسيرية.. وتعد المصطلحات بوجه عام محورًا أساسياً في عدد من التطبيقات المرتبطة بمعالجة اللغات الطبيعية، مثل محركات البحث، والتوليد التلقائي للمترادفات، واستخراج المعلومات، والتجريد الآلي، والترجمة الآلية، وبناء الأنطولوجيات (علم الوجود).أما في السياق العربي، فيواجه الدرس المصطلحي تحديًا خاصًا يتمثل في تعدد المصطلحات المستخدمة للدلالة على المفهوم الواحد، بالإضافة إلى تباين الترجمات بين الدول العربية، مما يجعل اللجوء إلى خوارزميات المعالجة الآلية للغة الطبيعية (NLP) لتحليل السياقات الشائعة وتوحيد الترجمات خيارًا واعدًا وفعالًا. ومع تطور تقنيات الذكاء الاصطناعي، تزايدت الحاجة إلى تمثيل هذه المصطلحات بطريقة آلية تتيح للحواسيب فهمها ومعالجتها. ويقوم هذا المقترح على محاولة استثمار أحدث تقنيات التعلم الآلي (ML) والشبكات العصبية (Neural Networks) لتحليل مصطلحات اللسانيات الإدراكية وربطها بسياقاتها المعرفية بصورة ديناميكية.

٢-٢:تحديات النمذجة المفاهيمية في ظل التطورات الحديثة:

تواجه النمذجة المفاهيمية كثيرًا من التحديات، لا سيما في سياق التطبيقات الحديثة مثل البيانات الضخمة، وأنظمة دعم القرار، والتطبيقات الصحية الرقمية، وعلم الوجود. إضافة إلى استمرار الصعوبات التقليدية المرتبطة بالمنهجيات، والأدوات، وتطوير النظريات.

وقد أسهمت الأبحاث الحديثة في دمج تخصص النمذجة المفاهيمية مع الذكاء الاصطناعي عبر مسارين رئيسين: أولهما: كيفية توظيف النمذجة المفاهيمية في

تصميم خوارزميات الذكاء الاصطناعي والتعلم الآلي. وثانيهما: كيفية تطبيق تقنيات الذكاء الاصطناعي والتعلم الآلي في تطوير الحلول القائمة على النماذج، مثل الهندسة المعتمدة على النماذج، لتطوير النماذج وتحسينها.

٢-٣: مبررات توظيف الذكاء الاصطناعي في المشروع المقترح:

تتعدد مبررات الاعتماد على الذكاء الاصطناعي في بناء هذا النموذج المقترح، وأبرزها:

(أ)التعقيد الدلالي لمصطلحات اللسانيات الإدراكية: تتسم مصطلحات هذا الحقل، مثل الفضاءات الذهنية، والاستعارة المفاهيمية، والأطر المعرفية، بدرجة عالية من التجريد والتشابك الدلالي، مما يتطلب نماذج لغوية ذكية قادرة على فهم العلاقات غير المباشرة بين المفاهيم. ومن هنا، تبرز أهمية تدريب هذه النماذج على مدونات ضخمة من النصوص الأكاديمية والدراسات الحديثة.

(ب)القدرة على التكيف مع التطور اللغوي: يمثل التكيف الذاتي مع ظهور مصطلحات جديدة إحدى أهم ميزات النماذج القائمة على الذكاء الاصطناعي، إذ يمكنها تحديث نفسها باستمرار لمواكبة المستجدات اللغوية والمعرفية.

(ج) إمكانية التكامل مع تطبيقات الذكاء الاصطناعي: يُتيح هذا التوجه دمج النموذج مع مجموعة واسعة من التطبيقات، مثل الترجمة الآلية العصبية(NMT) ، والإجابة عن الأسئلة(QA) ، والويب الدلالي(Semantic Web) ، مما يفتح آفاقًا رحبة لتوظيفه في مختلف ميادين المعرفة الحديثة.

٣-أهداف المقترح: يهدف المقترح-نظريًّا- إلى ما يلي:

يتمثل الهدف الرئيس في التعرف على مصطلحات اللسانيات الإدراكية بأكبر قدر ممكن من الدقة، لذا من المهم فهم سياق تسلسل الكلمات. وعلى الرغم من أن برامج التضمين مثل Glove وword2vec تتجاهل هذه المعلومات. إلا أنه يستعان بها إضافة إلى استخدم BERT (تمثيلات المشفر ثنائي الاتجاه من المحولات) لالتقاط المعلومات السياقية.

٣-١: تطوير نموذج تعلم آلي Machine learning Model لتصنيف مصطلحات اللسانيات الإدراكية بناءً على سياقاتها.

٣-٢: تصميم شبكة عصبونية (Neural Network) لفهم العلاقات بين المفاهيم (مثل: التماثل، التباين، التضمين).

3-3 بناء قاعدة معرفية (Knowledge Graph) لربط المصطلحات بمفاهيمها في إطار اللسانيات الإدراكية وما يتصل بها من علوم لغوية وغير لغوية.

٣-٤: تطوير واجهة برمجية (API) تتيح للمطورين الوصول إلى النموذج في تطبيقات NLP.

٤-المنهجية المقرحة:

٤-١: مصادر جمع البيانات: في هذا الصدد يمكن اللجوء إلى عدة مصادر، مثل: الأبحاث الأكاديمية في اللسانيات الإدراكية (عربي/إنجليزي). وكتب ومقالات متخصصة في العلوم المعرفية. وقواعد بيانات المصطلحات اللغوية (مثل: WordNet, BabelNet). ونتصور أن حجم البيانات: ٥٠,٠٠٠ وثيقة (نصوص، تعريفات، سياقات).

٤-٢: تصور لبعض التقنيات المستخدمة:

التقنية	وظيفتها
Word2Vec / GloVe	تمثيل الكلمات كمتجهات (Word Embeddings) للكشف عن العلاقات الدلالية.
BERT / AraBERT	فهم السياق الدلالي للمصطلحات باستخدام النماذج اللغوية العميقة.
GNN (Graph Neural Networks	تحليل العلاقات بين المفاهيم في الشبكات المعقدة (مثل: الاستعارة المفاهيمية والمعجم الذهني وغيرها)
Transformer Models	توليد تعريفات آلية للمصطلحات بناءً على السياق (مثل: GPT-3.5)
ChatGPT	لتوليد تعريفات بلغة طبيعية.
Ontology Learning	لبناء هياكل مفاهيمية ديناميكي
Explainable AI (XAI)	توضيح أسباب قرارات النموذج لضمان الشفافية (أسباب توافق او عدم توافق النصوص)

٤-٣: خطوات المعالجة

٤-٣-١: جمع البيانات وتنقيتها:

- استخدم Web Scraping الأكاديمي لجمع النصوص الأكاديمية؛ لتميزه عن غيره (مثل Google Scholar و Semantic Scholar: Microsoft Academicوغيرها من محركات البحث الأكاديمية أو قواعد البيانات المؤسسية) بميزات نسبية أهمها: المرونة العالية في الوصول إلى البيانات وإمكانية التحديث الفوري لها دون الاعتماد على المزود.

- استخدم المعالجة المسبقة للغة الطبيعية NLP Preprocessing لتنقية البيانات، نحو إزالة العناصر المشوّشة أو الزائدة (مثل الروابطURLs والأرقام وعلامات الترقيم الزائدة ونحوها) والترميز (تقسيم النص إلى وحدات أصغر tokens) وتوحيد شكل المصطلحات(التطبيعNormalization) وغيرها.

٤-٣-٢: تحليل المصطلحات:

- استخدام خوارزميات Random Forest [٩٦] وSVM(Support Vector Machines) للتصنيف الدلالي ومن ميزاتها (رغم بطئها مع البيانات الكبرى جدا) فعالية عالية في تصنيف النصوص ومقاومة للحمل الزائدOverfitting حال ضبط معلماتها

- استخراج الكيانات المعرفية (NER) Named Entity Recognition[٩٧] (المصطلحات في حالتنا) لتحديد المفاهيم الأساسية.

(٩٦) يرتكز أساسها العلمي على بناء مجموعة من أشجار القرار (Ensemble of Decision Trees) وتدرب كل شجرة على عينة عشوائية من البيانات ومجموعة فرعية من الخصائص. ونتيجة التصويت بين الأشجار هي المحدد لفئة النهائية. وتتكامل هذه الخوارزمية مع خوارزمية SVM حيث تتميز الأولى بقدرتها على التعامل مع البيانات الخطية بكفاءة وسرعة التعامل مع البيانات الكبيرة.

(٩٧) نموذج خوارزمي يكتشف المصطلحات ويصنفها آليا داخل النصوص غير المهيكلة معتمدا على الأدوات المناسبة للتعلم الآلي والتعلم العميق

٤-٣-٣: تمثيل المصطلحات:

- استخدام Word Embeddings لرسم خريطة دلالية للمفاهيم. (٩٨)
- استخدام Neo4j لبناء شبكة معرفية (Knowledge Graph) (٩٩)

٤-٣-٤: التدريب والتحسين:

- تدريب نموذج Deep Learning على تصنيف المصطلحات.
- تحسين النتائج باستخدام Transfer Learning.

٥- البنية النظرية العامة للنموذج المقترح (تصور هيكلي):

يتصور أن يتكون من طبقات هي: طبقة الإدخال (Input Layer) وتتمثل في مدونة نصية مدخلة (تعريفات وسياقات). وطبقة المعالجة (Processing Layer) وفيها يستخدم نموذج BERT لاستخلاص السمات الدلالية. وGNN لتحليل العلاقات بين المفاهيم. أما طبقة التمثيل (Representation Layer) ففيها توليد Embeddings للمصطلحات وبناء Knowledge Graph. وأخيرًا طبقة المخرجات (Output Layer) التي تظهر فيها تعريفات آلية للمصطلحات وإجابات عن استفسارات لغوية. وفي كل طبقة تفصيلات عميقة ودقيقة عند التطبيق.

(٩٨) إحدى تقنيات معالجة اللغة الطبيعية تحول الكلمات إلى متجهات رياضية في فضاء ذي أبعاد عالية وتعكس هذه المتجهات العلاقات الدلالية والتركيبية بين الكلمات.

(٩٩) Neo4jهو نظام إدارة قواعد بيانات والشبكات المعرفية رسوم بيانية تمثل المصطلحات (العُقد Nodes) والعلاقة بينها (الروابط)

الخاتمة والنتائج:

خلصت الورقة إلى أن نمذجة المحتوى المفاهيمي في المعاجم العربية الرقمية تمثل مسارًا حيويًّا لإعادة بناء العلاقة بين اللغة العربية والفضاء الرقمي المعاصر. وقد أظهرت التحليلات أن التكامل بين الذكاء الاصطناعي والمنهجيات اللغوية التقليدية يفتح آفاقًا واسعة لتطوير المعاجم، سواء عبر تحسين بنية المداخل، أو عبر تقديم تعريفات أكثر مرونة ودقة استجابة لتعدد السياقات. في المقابل، أبرز البحث تحديات جوهرية تتعلق بصعوبة تمثيل البنى الدلالية والصرفية للغة العربية بصورة تضمن الحفاظ على أصالتها من جهة، وتمكّن من توظيفها بكفاءة في بيئات المعالجة الآلية من جهة أخرى.

وقد أكدت الورقة على أهمية تبني أطر منهجية متكاملة تجمع بين المعايير اللغوية الصارمة والمتطلبات التقنية الحديثة، بما يسهم في تجاوز الفجوات المفاهيمية الحالية، ويساعد على بناء معاجم ذكية قادرة على التكيف مع التحولات الرقمية. وتغدو الحاجة ملحّة إلى تضافر جهود اللغويين وعلماء الحوسبة المعرفية تحت مظلات مؤسسية قادرة على وضع معايير موحدة لتطوير المحتوى المفاهيمي العربي.

وكان مما توصل إليه البحث من نتائج، ما يلي:

- لوحظ في الأبحاث التي أجريت في المدة من ٢٠٢٠ إلى ٢٠٢٥ أو قبلها بقليل ما يلي:

- بدأ المجال بالنماذج التقليدية للوسم النحوي وتحول تدريجيا إلى الاستفادة من معماريات المحولات، مع التركيز بشكل كبير على ضبط النماذج الخاصة باللغة العربية.

- اكتسب تطوير النماذج اللغوية الكبرى باللغة العربية زخما في الآونة الأخيرة، وعالج بعض الفجوات التي خلفتها نماذج الأغراض العامة مثل ChatGPT.

- الافتقار إلى المبادرات العربية، وربما تُناقش الأطر التعاونية أو جهود التقييس الإقليمية (على سبيل المثال بواسطة الألكسو) ولكن بصورة غير كافية. وكذلك العزلة الجغرافية لمعظم المشاريع والمبادرات.

- تستفيد دراسات قليلة من مناهج الذكاء الاصطناعي المتطورة، مثل نماذج اللغة مثل AraBERT، مما يقلل من القدرات عبر المجالات للمعاجم الرقمية.
- يهيمن الإثراء الدلالي وتنوع اللهجات على الأوراق البحثية العربية، وإن تناولت أوراق قليلة مجالات خاصة مثل القانون والرعاية الصحية والتاريخ بشكل ضئيل.
- يشكل الذكاء الاصطناعي تهديدًا للغات ذات الموارد التقنية المحدودة (مثل العربية) بسبب هيمنة اللغات الأخرى (خاصة الإنجليزية) على التقنيات الحديثة

وعلى الرغم من كثرة التحديات إلا أن آفاق التحسين والتطوير تظل مفتوحة وقائمة من طريق:

- رأب الصدع وسد الفجوة عبر المشروعات ذات الجهود المؤسسية الدولية.
- تعزيز التعاون بين مهندسي الذكاء الاصطناعي ومتخصصيه واللغويين خاصة من يعملون بالصناعة المعجمية.

وختامًا فأنّ التحديات المطروحة والآفاق التي تناولتها الورقة تُعد انعكاسًا للواقع البحثي والمعرفي في الفترة التي أُجري فيها هذا العمل، وقد تتغير هذه المعطيات مستقبلًا؛ فتطور الأدوات والمنهجيات وزيادة الاهتمام باللغة العربية في مجال الذكاء الاصطناعي إضافة إلى تكامل الجهود بين المختصين اللغويين ومهندسي الذكاء الاصطناعي، يُنذران بتحوُّلات جذرية في المستقبل القريب، قد تعيد تشكيل مشهد المعالجة الآلية للعربية، حيث تؤدي التقنية دور الوسيط بين الأصالة والتحديث، بدلًا من أن تُعد تهديدًا للهوية. بهذه الرؤية، يمكن فتح آفاق جديدة لفهم أعمق للغة العربية بوصفها كيانا حيّا نابضًا بالحياة في فضاء العالم الرقمي. وتجعل من تحديات اليوم منصاتٍ لابتكارات الغد، مما يفتح المجال لمراجعة النتائج ومواكبة المستجدات بشكل مستمر.

المراجع العربية:

١. **الحاكم، عمارية، ورباحي، محمد** .(2021) .تحديات المعالجة الآلية للغة العربية .*اللغة العربية*، ٢٣ (2)، ٢٤٩ .268–مسترجع من http://search.mandumah.com.sdl.idm.oclc.org/Record/1154667

٢. **الشمري، خالد** .(2011) .المعالجة الآلية للغة العربية. محاضرة مُستضافة في *المؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية* .جامعة الإمام محمد بن سعود، الرياض، المملكة العربية السعوديةhttps://goo.su/nHArmFG .

٣. **القاسمي، علي** .(2006) .*المعجم والقاموس: دراسة تطبيقية في علم المصطلح* (ط. ١). الجزائر: المجلس الأعلى للغة العربية.

٤. **كامل، وفاء، رشوان، محسن، والهواري، عبد العاطي** .(2014) . معالجة المحتوى المعجمي الدلالي في المعجم العربي الحاسوبي: مقاربة لغوية حاسوبية .*مجلة الجمعية المصرية لهندسة اللغة*،ع ١ ، ٦٣–٧٨.

٥. **المعافة، سوسن** .(2022) .*تقنيات معالجة اللغة العربية آليًّا: دراسة مقارنة لنماذج من المحللات الصرفية العربية* [رسالة دكتوراه غير منشورة]. كلية الآداب واللغات، جامعة ٨ مايو ١٩٤٥، الجزائر.

٦. **حمادة، سلوى السيد، ومهديوي، عمر** .(2024) .المعالجة الدلالية الآلية للغة العربية. مدونة اللسان *العربي*. تاريخ الاطلاع٢٠٢٤/١٢ https://2u.pw/swJsf

٧. **صبير، عبد الناصر عثمان عبد الله** .(2022) .صناعة معجم تعليمي للناطقين بغير العربية باستخدام نظرية الحقول الدلالية .*مجلة اللغة العربية بالجامعة الإسلامية بالمدينة المنورة*، ٧٩٠.791–

المراجع الأجنبية

1. Abdelali, A., Durrani, N., Dalvi, F., & Sajjad, H. (2022). Interpreting Arabic transformer models. *arXiv preprint*. https://doi.org/10.48550/arXiv.2201.07434
2. Abouenour, L., Nasri, M., Bouzoubaa, K., Kabbaj, A., & Rosso, P. (2014). Construction of an ontology for intelligent Arabic QA systems leveraging the conceptual graphs representation. *Journal of Intelligent & Fuzzy Systems, 27*(6), 2869–2881. https://doi.org/10.3233/IFS-141248
3. AbuOdeh, M., Phan, L., Elshabrawy, A. F., & Habash, N. (2024). Palmyra 3.0: A user-friendly cloud-based platform for morphology and dependency syntax annotation. In *Proceedings of the 14th International Conference on Language Resources and Evaluation (LREC)*. Marseille, France: ELRA.
4. Al Dakkak, O., & Zein, A. (2008). Towards Arabic electronic dictionary. In *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications* (pp. 1–6). Damascus, Syria: IEEE. https://doi.org/10.1109/ICTTA.2008.4530327
5. AlAwawdeh, S. M., & Abandah, G. A. (2021). Improving the accuracy of semantic similarity prediction of Arabic questions using data augmentation and ensemble. In *2021 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 272–277). Amman, Jordan: IEEE. https://doi.org/10.1109/JEEIT53412.2021.9634095
6. Al-Bataineh, H., Farhan, W., Mustafa, A., Seelawi, H., & Al-Natsheh, H. T. (2019). Deep contextualized pairwise semantic similarity for Arabic language questions. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 1586–1591). Portland, USA: IEEE. https://doi.org/10.1109/ICTAI.2019.00229
7. Al-Ghamdi, S., Al-Khalifa, H., & Al-Salman, A. (2023). Fine-tuning BERT-based pre-trained models for Arabic dependency parsing. *Applied Sciences, 13*(7), 4225. https://doi.org/10.3390/app13074225
8. Aljanabi, M. (2024). Assessing the Arabic parsing capabilities of ChatGPT and Cloude: An expert-based

comparative study. *Mesopotamian Journal of Arabic Language Studies*, *2024*(1), 16–23. https://doi.org/10.58496/MJALS/2024/002

9. Aljebreen, J., Alhargan, A., & Nagro, S. (2025). Enhancing zero-shot question answering in Arabic language with fine-tuned AraGPT model. In *2025 8th International Conference on Data Science and Machine Learning Applications (CDMA)* (pp. 48–53). Riyadh, Saudi Arabia: IEEE. https://doi.org/10.1109/CDMA61895.2025.00014

10. Alrehili, A., & Alhothali, A. (2025). Towards the development of balanced synthetic data for correcting grammatical errors in Arabic: An approach based on error tagging model and synthetic data generating model. *arXiv preprint*. https://doi.org/10.48550/arXiv.2502.05312

11. Al-Rfou, R., Perozzi, B., & Skiena, S. (2013). Polyglot: Distributed word representations for multilingual NLP. In *Proceedings of the 17th Conference on Computational Natural Language Learning (CoNLL)* (pp. 183–192). Sofia, Bulgaria: ACL.

12. Alrowili, S., & Vijay-Shanker, K. (2023). ArTrivia: Harvesting Arabic Wikipedia to build a new Arabic question answering dataset. In *Proceedings of ArabicNLP 2023* (pp. 191–207). Singapore: Association for Computational Linguistics.

13. Alsaleh, A., Althabiti, S., Alshammari, I. K., et al. (2022). LK2022 at Qur'an QA 2022: Simple transformers model for finding answers to questions from Qur'an. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT)*. Marseille, France: ELRA.

14. Alshehri, A. H. (2024). AraQA-BERT: Towards an Arabic question answering system using pre-trained BERT models. *WSEAS Transactions on Information Science and Applications*, *21*, 361–373. https://doi.org/10.37394/23209.2024.21.34

15. Alsubhi, K., Jamal, A. T., & Alhothali, A. M. (2021). Pre-trained transformer-based approach for Arabic question answering: A comparative study. *arXiv preprint*. https://doi.org/10.48550/arXiv.2111.05671

16. Amendola, G. (2023). Special issue on logic-based artificial intelligence. *Algorithms, 16*(2), 106. https://doi.org/10.3390/a16020106

17. Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based model for Arabic language understanding. *arXiv preprint*. https://doi.org/10.48550/arXiv.2003.00104

18. Atef, A., Mattar, B., Sherif, S., Elrefai, E., & Torki, M. (2020). AQAD: 17,000+ Arabic questions for machine comprehension of text. In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–6). Antalya, Türkiye: IEEE. https://doi.org/10.1109/AICCSA50499.2020.9316526

19. Baccar Ben Amar, F., Khemakhem, A., Gargouri, B., Haddar, K., & Ben Hamadou, A. (2008). LMF standardized model for the editorial electronic dictionaries of Arabic. In *Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS)* (pp. 64–73). Setúbal, Portugal: SciTePress. https://doi.org/10.5220/0001737800640073

20. Bakari, W., Trigui, O., & Neji, M. (2014). Logic-based approach for improving Arabic question answering. In *2014 IEEE International Conference on Computational Intelligence and Computing Research* (pp. 1–6). Coimbatore, India: IEEE. https://doi.org/10.1109/ICCIC.2014.7238319

21. Boguraev, B., & Briscoe, T. (1989). Computational lexicography for natural language processing. Language, 66(4), 626–652.

22. Chaudhary, J., Parmar, N., & Mehta, A. (2024). Artificial intelligence and expert systems. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 4*(4), 1–9.

23. Darwish, K., Attia, M., Mubarak, H., et al. (2020). Effective multi-dialectal Arabic POS tagging. *Natural Language Engineering, 26*(6), 677–690. https://doi.org/10.1017/S1351324920000078

24. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*. https://doi.org/10.48550/arXiv.1810.04805

25. Diab, M. T., Al-Badrashiny, M., Aminian, M., et al. (2014). Tharwa: A large-scale dialectal Arabic – Standard Arabic – English lexicon. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)* (pp. 3785–3792). Reykjavik,

26. Dibas, S., Khairallah, C., Habash, N., et al. (2022). Maknuune: A large open Palestinian Arabic lexicon. *arXiv preprint*. https://doi.org/10.48550/arXiv.2210.12985

27. ElKomy, M., & Sarhan, A. M. (2022). TCE at Qur'an QA 2022: Arabic language question answering over Holy Qur'an using a post-processed ensemble of BERT-based models. *arXiv preprint*. https://doi.org/10.48550/arXiv.2206.01550

28. Elleuch, I., Gargouri, B., & Ben Hamadou, A. (2015). Self syntactico-semantic enrichment of LMF normalized dictionaries. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation (PACLIC)* (pp. 345–353). Shanghai, China.

29. Elleuch, I., Gargouri, B., & Ben Hamadou, A. (2015). Using the textual content of the LMF-normalized dictionaries for identifying and linking the syntactic behaviors to the meanings. In *Proceedings of the 11th International Conference on Recent Advances in*

30. Elleuch, I., Gargouri, B., & Ben Hamadou, A. (2021). Lexical data mining-based approach for the self-enrichment of LMF standardized dictionaries: Case of the syntactico-semantic knowledge. *Concurrency and Computation: Practice and Experience, 33*(17), e6312. https://doi.org/10.1002/cpe.6312

31. Gargouri, B., Elleuch, I., & Ben Hamadou, A. (2014). Towards automatic enrichment of standardized electronic dictionaries by semantic classes. In *Proceedings of the 26th Conference on Computational Linguistics and Speech Processing (ROCLING)* (pp. 1–10). Taipei, Taiwan.

32. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (1st ed.). Cambridge, USA: MIT Press.

33. Habash, N., AbuOdeh, M., Taji, D., et al. (2022). Camel Treebank: An open multi-genre Arabic dependency treebank. In *Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC)* (pp. 6113–6122). Marseille, France: ELRA.

34. Haddad, A., Zrigui, M., & Ahmed, M. B. (2005). Un système de génération automatique de dictionnaires linguistiques de l'arabe. In *Actes de la 12ème conférence sur le Traitement Automatique des Langues Naturelles (TALN)* (pp. 123–132). Dourdan, France

35. Hamdi, A., Gala, N., & Nasr, A. (2014). Automatically building a Tunisian lexicon for deverbal nouns. In Proceedings of the First Workshop on Applying NLP Tools to Similar Languages (pp. 12–20). Dublin, Ireland: ACL.

36. Hanks, P. (2013). *Lexical analysis: Norms and exploitations* (1st ed.). Cambridge, USA: MIT Press.

37. Jarrar, M. (2022). Arabic ontology: An Arabic WordNet with ontologically clean content. *Applied Ontology Journal, 16*(1), 1–26. https://doi.org/10.3233/AO-200241

38. Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing* (3rd ed.). Hoboken, USA: Pearson.

39. Jürgenson, S. (2023). نموذج لبيانات في RDBMS [A data model in RDBMS]. *AppMaster*. Retrieved April 14, 2024, from https://appmaster.io/ar/blog/nmdhj-lbynt-fy-rdbms

40. Khalifa, S., & Habash, N. (2022). Morphosyntactic tagging with pre-trained language models for Arabic and its dialects. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 1708–1719). Dublin, Ireland: Association for Computational Linguistics.

41. Khemakhem, A., Gargouri, B., Ben Hamadou, A., & Francopoulo, G. (2016). ISO standard modeling of a large Arabic dictionary. *Natural Language Engineering, 22*(6), 849–879. https://doi.org/10.1017/S1351324915000224

42. Koubaa, A., Ammar, A., Ghouti, L., Najar, O., & Sibaee, S. (2024). Arabiangpt: Native Arabic GPT-based large language model. *arXiv preprint*. https://arxiv.org/abs/2402.15313

43. Kungurtsev, A. B., Potocnyak, Y. V., & Silyaev, D. A. (2015). Method of automated construction of explanatory dictionary of subject area. *Technology Audit and Production Reserves, 2*(22), 58–63. https://doi.org/10.15587/2312-8372.2015.40895

44. Kungurtsev, O. B., Mileiko, I. I., & Novikova, N. O. (2024). Technology for automated construction of domain dictionaries with special processing of short documents. *Radio Electronics, Computer Science, Control, 4*, 148–158. https://doi.org/10.15588/1607-3274-2023-4-14

45. Kutasevych, M. (2018). Development of information technology of term extraction from documents in natural language. *Eastern-European Journal of Enterprise Technologies, 6*(2), 44–51. https://doi.org/10.15587/1729-4061.2018.146624

46. Kwon, S. Y., Bhatia, G., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). Beyond English: Evaluating LLMs for Arabic grammatical error correction. *arXiv preprint*. https://doi.org/10.48550/arXiv.2312.08400

47. Kwon, S., Bhatia, G., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). Beyond English: Evaluating LLMs for Arabic grammatical error correction. In *Proceedings of ArabicNLP 2023* (pp. 101–119). Singapore: Association for Computational Linguistics.

48. Lemberg, I., Schröder, B., & Storrer, A. (2001). Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen, Germany: Max Niemeyer Verlag.

49. Maged, S., et al. (2024). HistoryQuest: Arabic question answering in Egyptian history with LLM fine-tuning and transformer models. In *2024 Intelligent Methods, Systems, and Applications (IMSA)* (pp. 135–140). Giza, Egypt: IEEE. https://doi.org/10.1109/IMSA61967.2024.10652824

50. Manjunath, S. H., & McCrae, J. P. (2021). Encoder-attention-based automatic term recognition (EA-ATR). In *3rd Conference on Language, Data and Knowledge (LDK 2021)* (pp. 23:1–23:13). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik. https://doi.org/10.4230/OASIcs.LDK.2021.23

51. Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing* (1st ed.). Cambridge, USA: MIT Press.

52. Mellah, Y., Touahri, I., Kaddari, Z., Haja, Z., Berrich, J., & Bouchentouf, T. (2022). LARSA22 at Qur'an QA 2022: Text-to-text transformer for finding answers to questions from Qur'an. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT)* (pp. 89–94). Marseille, France: ELRA.

53. Mohan, V., & Senthilkumar, S. (2023). A brief review of the development path of artificial intelligence and its subfields. *International Journal of Engineering Technologies and Management Research, 10*(6), 1–12. https://doi.org/10.29121/ijetmr.v10.i6.2023.1331

54. Mozannar, H., Maamary, E., El Hajal, K., & Hajj, H. (2019). Neural Arabic question answering. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop (WANLP)* (pp. 108–118). Florence, Italy: ACL.

55. Müller-Spitzer, C. (2014). Textual structures in electronic dictionaries compared with printed dictionaries: A short general survey. In R. Gouws, U. Heid, W. Schweickard, & H. Wiegand (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (Vol. 5, pp. 623–635). Berlin, Germany: De Gruyter Mouton.

56. Murel, J., & Kavlakoglu, E. (n.d.). Topic modeling. *IBM*. Retrieved April 14, 2024, from https://www.ibm.com/sa-ar/think/topics/topic-modeling

57. Nazar, R., & Acosta, N. (2023). Termout: A tool for the semi-automatic creation of term databases. *Terminology and Artificial Intelligence (TAI)*, Paris, France.

58. Nesi, H. (2008). Dictionaries in electronic form. In A. P. Cowie (Ed.), The Oxford history of English lexicography (1st ed., pp. 458–478). Oxford, UK: Oxford University Press.

59. Newell, A. (1983). Intellectual issues in the history of artificial intelligence. In F. Machlup & U. Mansfield (Eds.), *The study of information: Interdisciplinary messages* (pp. 187–294). New York, USA: John Wiley & Sons.

60. Omar, C. A. M. B. C., & Dahan, H. B. A. M. (2011). The development of E-dictionary for the use with maharah al-Qiraah textbook at a matriculation centre in a university in Malaysia. *Journal of Language Studies, 10*, 255–264.

61. Pasha, A., Al-Badrashiny, M., Diab, M. T., El Kholy, A., Eskander, R., Habash, N., et al. (2014). MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)* (pp. 1094–1101). Reykjavik, Iceland: ELRA.

62. Polzin, T., & Rieser, H. (1991). Parsing with situation semantics. *Nordic Journal of Linguistics, 14*(2), 141–189. https://doi.org/10.1017/S0332586500002432

63. Pustejovsky, J., & Boguraev, B. (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence, 63*(1–2), 193–223.

64. Radford, A., et al. (2019). Language models are few-shot learners. *arXiv preprint*. https://arxiv.org/abs/2005.14165

65. Ranta, A. (2004). Grammatical framework. *Journal of Functional Programming, 14*(2), 145–189. https://doi.org/10.1017/S0956796803004738

66. Regragui, Y., Abouenour, L., Krieche, F., Bouzoubaa, K., & Rosso, P. (2016). Arabic WordNet: New content and new applications. In *Proceedings of the 8th Global WordNet Conference (GWC)* (pp. 333–341). Bucharest, Romania: Global Wordnet Association

67. Reiter, R. M. (1975). Formal reasoning and language understanding systems. In *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing (TINLAP '75)* (pp. 175–179). Stroudsburg, PA, USA: ACL. https://doi.org/10.3115/980190.980238

68. Saini, F., Sharma, T., & Madan, S. (2021). A comparative analysis of expert opinions on artificial intelligence: Evolution, applications, and its future. *Advanced Journal of Graduate Research, 11*(1), 10–22.

69. Sammoudi, M., Habaybeh, A., Ashqar, H. I., & Elhenawy, M. (2024). Question-answering (QA) model for a personalized learning assistant for Arabic language. In *Proceedings of the International Conference on Intelligent Systems, Blockchain, and Communication Technologies* (pp. 356–367). Cham, Switzerland: Springer Nature.

70. Schaefer, J. F., & Kohlhase, M. (2020). GLIF: A declarative framework for symbolic natural language understanding. In *Proceedings of the Workshop on Formal and Cognitive Reasoning (FCR@KI)*. Saarbrücken, Germany: CEUR-WS.

71. Schwarcz, R. M. (1969). Towards a computational formalization of natural language semantics. In *Proceedings of the 1969 Conference on Computational Linguistics (COLING '69)* (pp. 1–53). Stroudsburg, PA, USA: ACL. https://doi.org/10.3115/990403.990432

72. Siyam, F., Hidayat, R., Rochmat, C., Maulaya, R., Avilya, A., & Maulidi, M. (2024). Accuracy analysis of artificial intelligence in Arabic language translation and grammatical rule mapping. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab, 16*(2), 558–573. https://doi.org/10.24042/albayan.v16i2.24588

73. Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 427–448). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

74. Tamer, A., Hassan, A.-A., Ali, A., Salah, N., & Medhat, W. (2023). Fine tuning of large language models for Arabic language. In *2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–4). Giza, Egypt: IEEE. https://doi.org/10.1109/AICCSA59173.2023.10479346

75. Wang, J.-T. (1973). On logical formulation of the computation process in semantical systems. In *Proceedings of the 5th International Conference on Computational Linguistics (COLING)* (pp. 87–100). Pisa, Italy: ACL. https://doi.org/10.3115/992532.992540

76. Zeroual, I., & Lakhouaja, A. (2018). Arabic diacritization using deep learning. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 3148–3153). Miyazaki, Japan: ELRA.

77. Zettlemoyer, L., & Collins, M. (2007). Online learning of relaxed CCG grammars for parsing to logical form. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 678–687). Prague, Czech Republic: ACL.

روابط الانترنت:

https://2u.pw/Od4RBxYH
https://2u.pw/y6VD7wp
https://2u.pw/Od4RBxYH
https://goo.su/dXz9ML
https://ai.gov.ae/ar/ai-dictionary/
https://babelnet.org/
https://lexicala.com/
https://2u.pw/riUTf
https://2u.pw/Tgynp
https://goo.su/HOltzUJ

Artificial Intelligence and Cognitive Integration Modeling Conceptual Content in Arabic Digital Lexicons: Between the Horizons of Development and the Challenges of Conceptual Delimitation

Abstract

The paper presents a theoretical analysis of the process of modeling conceptual content in Arabic digital lexicons, focusing on the role of artificial intelligence and mechanisms of cognitive integration. It explores how contemporary technologies—particularly AI—can contribute to developing the structure of lexical entries, improving retrieval methods, and enhancing the accuracy of definitions and their contextualization. Through a critical examination of a sample of relevant studies, the research reveals existing gaps in representing conceptual content and proposes methodological frameworks for more effective development. Furthermore, the study envisions the horizons of building smarter and semantically aware Arabic digital lexicons capable of meeting user needs in rapidly evolving digital environments. The research adopts a descriptive-analytical approach, aiming to contribute to advancing the digital lexicographic industry within the context of cognitive integration and technological progress."

Keywords: Artificial Intelligence, Arabic Digital Lexicons, Linguistic Challenges, Technological Opportunities, Conceptual Content Modeling, Cognitive Integration.