

تكامل علم اللغة والذكاء الاصطناعي

¹ هند إبراهيم محمد، ² عباس علاء مهدي، ³ شذى جاسم محمد¹ جامعة ديالى، كلية التربية المقداد، ² مديرية تربية ديالى، قسم تربية المقدادية، ³ الجامعة المستنصرية، كلية العلوم

(لعراق)

Integration of Linguistics and Artificial Intelligence

¹Hind Ibrahim Mohammed, ²Abbas Alaa Mahdi, ³Shatha Jassim Mohammed¹University of Diyala ,Al- Muqdad, College of Education,(Iraq), hindim@uodiyala.edu.iq²Directorate General of Education,Diyala, (Iraq), abbasatr@gmail.com³ Al-Mustansiriya University, College of Science, (Iraq), shonash77@uomustansiriyah.edu.iq

تاريخ الاستلام: 2025/ 10 / 3 تاريخ القبول: 2025 / 11 / 01 تاريخ النشر: 2026 / 04 / 01

المخلص:

يُعدّ علم اللغة أحد أهم الحقول الإنسانية التي تُعنى بدراسة بنية اللغة ووظائفها وأساليب استخدامها في التواصل البشري، في حين يمثل الذكاء الاصطناعي (AI) أحد أبرز إنجازات العصر الرقمي الذي يسعى لمحاكاة قدرات الإنسان الذهنية واللغوية. ومع التطور المتسارع للتقنيات الذكية. برزت الحاجة الملحة إلى دمج المعارف اللغوية بالنماذج والخوارزميات الحاسوبية من أجل بناء أنظمة قادرة على فهم اللغة الطبيعية ومعالجتها بصورة أقرب إلى إدراك الإنسان. إن هذا التكامل لا يقتصر على تطوير تطبيقات عملية مثل الترجمة الآلية، والمساعداات الذكية، وتحليل النصوص، بل يمتد ليشمل إعادة صياغة مفاهيم نظرية حول طبيعة اللغة والاتصال والمعنى. فبينما يمد علم اللغة الذكاء الاصطناعي بالأطر النظرية لفهم النحو والدلالة والتداولية، يوفر الذكاء الاصطناعي للغويين أدوات كمية وتجريبية تفتح آفاقاً جديدة للبحث والتطبيق. من هنا تأتي أهمية دراسة التكامل بين علم اللغة والذكاء الاصطناعي باعتباره مجالاً متعدد التخصصات يُساهم في إثراء المعرفة العلمية من جانب، ويعزز الابتكار التكنولوجي من جانب آخر، بما ينعكس إيجاباً على مجالات التعليم والإعلام، والصناعات الرقمية المتقدمة. يتناول هذا البحث العلاقة التكاملية بين علم اللغة (Linguistics) والذكاء الاصطناعي (AI)، من خلال استعراض النظريات اللغوية الأساسية مثل النحو التوليدي، الدلالات، التداولية، إضافة إلى استراتيجيات الذكاء الاصطناعي كالتعلم الآلي ومعالجة اللغة الطبيعية.

كلمات مفتاحية: علم اللغة، الذكاء الاصطناعي (AI)، معالجة اللغة الطبيعية (NLP)، التعلم الآلي (ML).

Abstract:

Linguistics is a vital field that explores the structure, functions, and use of language in human communication. Meanwhile, artificial intelligence (AI) represents a major technological advancement aimed at replicating human cognitive and linguistic abilities. With the rapid growth of intelligent systems, there is an increasing need to integrate linguistic knowledge with computational models and algorithms to create systems capable of understanding and processing natural language in a human-like manner. This integration extends beyond practical applications—such as machine translation, intelligent assistants, and text

analysis—to reshaping theoretical perspectives on language, communication, and meaning. Linguistics provides AI with frameworks for understanding syntax, semantics, and pragmatics, while AI offers linguists computational tools and empirical methods that open new avenues for research and innovation. Studying the intersection of linguistics and AI is important for advancing knowledge and fostering technological innovation, with positive implications for education, media, and advanced digital industries. This research examines the synergy between linguistic theories, including generative grammar, semantics, and pragmatics, and AI strategies such as machine learning and natural language processing.

Keywords: Linguistics, Artificial Intelligence (AI), Natural Language Processing (NLP), Machine Learning (ML).

المبحث الاول

الاطار العام

المقدمة

يبحث علم اللغويات في الخصائص التي تجعل اللغة البشرية فريدة، فضلاً عن مبادئ التواصل المشتركة مع الكائنات الحية الأخرى. سعى الذكاء الاصطناعي، الذي شهد تطوراً ملحوظاً منذ عام ١٩٥٠، إلى توصيف السلوك الذكي بشكل مستقل عن الوسيط؛ وبالتالي، يمكن تحليل كل من اللغة والذكاء كظاهرتين مترابطتين (Alberts, 2022). وقد ازداد الترابط بين المجالين بشكل متزايد، ويُعدّ فهم اللغويات أمراً حيوياً لإنتاج وتحليل أنظمة الذكاء الاصطناعي المعاصرة الموجهة نحو اللغة (Shormani, 2024). يُعدّ علم اللغويات مجالاً واسعاً يشمل نظريات النحو، والدلالات، والبراغماتية، والتوليد، والاكْتساب، والإملاء؛ ويتبع كل مجال فرعي تقنيات نظرية ومنهجية محددة، ويستخدم مصطلحات مميزة، ويوفر رؤى قيّمة للبحث في مجال الذكاء الاصطناعي (Zhang et al., 2023).

يشهد علم اللغويات الحاسوبية (CL) على العلاقة الوثيقة بين التقدم التكنولوجي والأنماط الفكرية، وهي صلة يُتوقع استمرارها في المستقبل غير المنظور. يتوازي تاريخ علم اللغويات الحاسوبية مع الوضع التاريخي والمعاصر لصناعة الذكاء الاصطناعي على نطاق أوسع؛ إذ يتأثر هذا المجال بشكل كبير بالابتكارات الخارجية والأفكار الجديدة من الداخل (Mandour, 2025). لطالما كان علم اللغويات الحاسوبية متعدد التخصصات، إذ يقع عند تقاطع اللغويات وعلوم الحاسوب، ومؤخراً التعلم الآلي، وبالتالي لم يكن له تعريف ثابت ومقبول عالمياً. في الماضي، كان المجال المعني هو اللغة والآلات، أو اللغويات الكلية والجزئية والحاسوبية، على سبيل المثال لا الحصر. بدأت الستينيات بالترجمة الآلية. وهكذا، يتألف تاريخ علم اللغويات الحاسوبية في جزء كبير منه من تصورات سابقة متعاقبة حول التحديات الرئيسية التي ينبغي مواجهتها، بعد أن أتاحت (ثم اختفت) (Khaleel et al., 2024).

مالت الروايات المبكرة للغة البشرية إلى التركيز على اكتساب المعاني والأفكار، إلا أن النظريات التوليدية نظرت إلى المعنى على أنه شيء يجب تفسيره بعد استنباط البنية العميقة. وقد شكّل هذا التناقض بين فلسفة العقل الداخلية والخارجية العقد الأول من أبحاث اللغة البشرية. ونظراً لأن نماذج GPT أصبحت أحدث التطورات في مجال الذكاء الاصطناعي، ولأن هذه النماذج تتعلم من خلال التنبؤ بالنص دون اللجوء إلى المعنى أو التأريض، فمن المناسب استحضار الخلافات التاريخية في

الماضي وتزويد قراء اليوم بما يلزم لاتخاذ قرارات أكثر استنارة حول كيفية التفكير في التأريض (Church & Liberman, 2021).

هدف البحث

- استعراض أهم التطبيقات العملية في معالجة اللغة الطبيعية (NLP) ، الترجمة، توليد النصوص، والتعرف على الكلام.
- اقتراح توجهات بحثية مستقبلية لدمج أعمق بين الذكاء الاصطناعي واللغويات.

أهمية البحث

- يساهم في سد الفجوة بين الدراسات النظرية في علم اللغة والتطبيقات العملية في الذكاء الاصطناعي.
- يعزز فهم الباحثين لكيفية تطوير أنظمة قادرة على فهم اللغة البشرية بصورة طبيعية وسياقية.
- يفتح المجال أمام دراسات جديدة تجمع بين علوم إنسانية وتقنية.
- يسلط الضوء على التحديات الأخلاقية كمسألة الانحياز والخصوصية في أنظمة الذكاء الاصطناعي.

مشكلة البحث

- على الرغم من التطور الكبير في الذكاء الاصطناعي، ما زالت الأنظمة تعاني من: صعوبة في تمثيل المعنى العميق للغة.
- محدودية في التعامل مع السياق الثقافي والاجتماعي.
- تحديات في بناء أنظمة متعددة اللغات بشكل فعال وعادل.

حدود البحث

- يركز البحث على الجانب اللغوي-التقني دون التعمق في البرمجة أو الخوارزميات الرياضية.
- يغطي بشكل رئيسي التطبيقات المتعلقة بمعالجة اللغة الطبيعية، الترجمة، والتوليد اللغوي.
- يعتمد على أمثلة من اللغة الإنجليزية بشكل أساسي مع الإشارة للغات أخرى.
- لا يتناول جميع التطبيقات الفرعية للذكاء الاصطناعي (مثل الرؤية الحاسوبية أو الروبوتات) إلا بقدر ارتباطها باللغة.

فرضية البحث

يمكن تعزيز فعالية أنظمة الذكاء الاصطناعي اللغوية بشكل ملموس إذا تم دمج الأطر والنظريات اللغوية (النحو، الدلالة، التداولية) مع تقنيات الذكاء الاصطناعي الحديثة (التعلم العميق، الشبكات العصبية، النماذج اللغوية الكبيرة)، مما يؤدي إلى أنظمة أكثر دقة في فهم اللغة البشرية ومعالجتها.

المحور الثاني

الاطار النظري

1. النظريات المبكرة في اللغويات

لطالما كانت دراسة اللغة البشرية، قبل ظهور تخصصين علميين يُطلق عليهما الذكاء الاصطناعي واللغويات، طموحةً للغاية، إذ ارتبطت بقدرة الإنسان وحكمته، وبالمناطق وطرق تنظيم المعرفة، وبالوعي والإدراك. في منتصف القرن العشرين، أشار أبراهام كابلان (١٩٥٢) إلى أن "أهداف اللغويات لم تُحدد منذ زمن طويل من خلال الاستبطان وعلم النفس فحسب، بل من خلال المنطق والفلسفة أيضًا". ميّز كابلان ثلاثة أنواع من الوصف: "النحوي، والدلالي، والبراغماتي" (انظر الأسس الفلسفية). ولا تزال هذه المفاهيم الأساسية محورية في الصياغات الحديثة للغويات، على الرغم من اختلاف تفاصيلها بين الأطر المعاصرة العديدة في اللغويات، والعلوم المعرفية، والفلسفة، والحوسبة (Haber et al., 2024). بدأت نظرية مجموعة القواعد التوليدية بالعمل الرائد الذي قام به نعوم تشومسكي في عام 1956، والذي قدم فكرة أن القواعد، وليس تطبيقها على أي نص معين، هي التي تحمل المعلومات المهمة (Khaleel et al., 2024). في أواخر الخمسينيات والستينيات من القرن الماضي، ارتقى العلماء الذين يدرسون اللغة البشرية إلى مستوى طموحات العقود الأولى. من عام 1958 إلى عام 1964، أطلق باحثو جورج تاون تجارب معالجة اللغة الطبيعية (NLP) الأولى من نوعها. أثارت هذه الجهود اهتمامًا واسع النطاق وأوجدت مفهومًا دائمًا للترجمة الآلية لم تنجح التطورات في اللغويات الحاسوبية ولا النمو الهائل للتعلم الآلي التجريبي في إزاحته (Ferryhough & Borghi, 2023).

2. تقنيات الذكاء الاصطناعي

مر تطوير الذكاء الاصطناعي (AI) بمراحل ملحوظة منذ الخمسينيات. تم استكشاف العديد من الأطر التقنية، بما في ذلك المناهج الرمزية (العميقة)؛ والشبكات العصبية الاصطناعية؛ والنمذجة القائمة على الوكيل؛ ومؤخرًا، التعلم الآلي (العميق) (Panchal et al., 2024). ويعود التركيز الحالي على أنظمة التعلم الآلي بشكل كبير إلى توافر مجموعات البيانات الكبيرة واقتراح أساليب تدريب فعالة. ومع ذلك، لا يزال التقدم بطيئًا لأن هذه التقنيات تتطلب كميات كبيرة من البيانات لاستخراج الميزات ذات الصلة. وتُعد بيانات اللغة، على وجه الخصوص، صعبة لأن النص لا يكشف بشفافية عن البنية الأساسية أو معنى العبارات. ولا تزال العديد من الأسئلة العملية دون حل، مما يضمن استمرار تطور الذكاء الاصطناعي لسنوات عديدة (Sridhar et al., 2025). ومن بين أقدم هذه النظريات القواعد النحوية التحويلية والتوليدية التي اقترحها تشومسكي (1957)، والتي عرّفت القواعد النحوية على أنها مجموعة من القواعد لتوليد جميع جمل اللغة دون الإفراط في التوليد. وارتبطت بسهولة بطريقة لتنفيذ فهم اللغة، وأثرت بشدة على البحث في مجال الذكاء الاصطناعي لعقود. وتوسيع نطاقه، طور فيلمور (1968) قواعد الحالة، التي تحصي الحالات التي تحدد أثر العبارات المتعلقة بمسند الجملة، ذات الصلة بتمثيل الدلالات والسياق. وفي تناوله للدلالات بشكل مباشر، طبق مونتاجو (1970) مناهج منطقية لتعريفها بطريقة مشابهة للتعبيرات الرياضية. كما تقترح اللغويات الرسمية أساليب لوصف اكتساب اللغة (ماركوس، 1993)، مما ألهم خوارزميات لتعلم القواعد تلقائيًا من بيانات الإدخال (Veeramani et al., 2025).

بالتوازي مع إنشاء اللغويات الرسمية، خضعت تقنية الذكاء الاصطناعي لتطورات رئيسية متعددة. في الأصل كان الهدف هو إنشاء ذكاء يشبه الإنسان، تحول المجال إلى مجموعة محدودة من القدرات في العقد الثاني. وبدءًا من العقد الثالث، كان التركيز على المناهج القائمة على البيانات والنماذج الإحصائية، وتقسيم الأنظمة إلى أنظمة تستمد البنية الأساسية للبيانات وأخرى تتعلم الأنظمة مع بنية المشكلة لحلها بكفاءة. وقد تم تطبيق المنهجيات المرتبطة لحل العديد من المشكلات المتعلقة باللغة. وفي الأثناء، ازدهرت الأساليب الرمزية مثل أنظمة الخبراء المتقدمة من حين لآخر (Shormani, 2024).

استندت نظريات تشومسكي الرائدة إلى الشكلية الرياضية، حيث اعتبرت اللغة محكومةً بقواعد نحوية غير مرتبطة بالسياق، ممتدة بقواعد تحويلية وصرفية، مع بناء نحوي مستقل في جوهره (Tasheva, 2024). تأثرت أفكاره بنشأة علوم الحاسوب، واللغويات الحاسوبية، والذكاء الاصطناعي، التي شاركت في الاعتقاد بإمكانية وصف الفكر بالمنطق الرسمي. ألهمت المدرسة التوليدية عقوداً من الفكر اللغوي، معتمدةً على رؤية اللغة كنظام يتلاعب بالرموز المجردة من خلال قواعد تركيبية مثل الدمج والتحرك. في البداية، هدفت معظم أبحاث الذكاء الاصطناعي إلى تحديد مبادئ شبيهة بالقوانين وراء السلوك البشري لنمذجة التفكير واللغة والرؤية رسمياً، والمعروفة باسم الذكاء الاصطناعي التقليدي الجيد (GOFAI). فشلت هذه المناهج القائمة على القواعد في تقديم حلول عملية للمهام التي يجدها البشر بسيطة. منذ الثورة الإحصائية في التسعينيات، فضّل الباحثون مناهج تستنتج الارتباطات من البيانات، مما يُمكن الأنظمة من محاكاة مهام محددة مثل التصنيف والتعرف على الأنماط دون فهم المنطق السببي. وقد تحوّل تركيز تطبيقات الذكاء الاصطناعي من نمذجة السلوك البشري العام إلى حل مشكلات محددة بكفاءة (Paradowski, 2011).

عادةً ما تكون تطبيقات الذكاء الاصطناعي في أنظمة الأتمتة أنظمة موزعة، ويتطلب تطويرها مشاركة العديد من الخبراء. يستخدم كل خبير لغة نمذجة خاصة بمجاله وأدوات ذات صلة. في الوقت الحالي، لا توجد لغة نمذجة رسومية متعددة التخصصات تُمكن من نمذجة تطبيقات الذكاء الاصطناعي كنظام شامل مفهوم لجميع التخصصات (Petit & De Cooman, 2021).

3. معالجة اللغة الطبيعية

معالجة اللغة الطبيعية (NLP) هي مجال ضمن اللغويات والذكاء الاصطناعي يُنَبِّت نماذج وأساليب معالجة اللغة (Sung & Hopper, 2023). وتبحث في مفاهيم أساسية مثل علم الأصوات، وعلم الصرف، والنحو، والدلالات، وتطبيقاتها في الأنظمة والخدمات العملية. وقد تناولت الدراسات المبكرة تسلسلات الصوتيات، وتكوين الكلمات، والأنماط النحوية، والعلاقات الأساسية بين الكلمات كأساس للنصوص. واستكشفت الدلالات القائمة على المتر كآداة بسيطة لحل المشكلات (Alberts, 2022). لا يهدف تطوير اللغويات الرسمية إلى بناء نماذج وصفية فحسب، بل يهدف أيضاً إلى توفير أوصاف واضحة يمكن للحاسوب استخدامها كمرجع، بهدف نهائي يتمثل في توليد نماذج ذات توجه سياسي تأخذ في الاعتبار نوايا المتحدث والوظيفة البراغماتية للألفاظ (Torfi et al., 2020). غالباً ما تُشكل مواصفات النظريات اللغوية أساساً للأساليب المطبقة في العديد من أنظمة معالجة اللغة الطبيعية (NLP). إذا كانت المواصفات ذات سلامة لغوية كافية، ووفرت لغة برمجة تمثيلاً إجرائياً دقيقاً للصيغ، فإن النظام الناتج يكون إلى حد كبير تطبيقاً حرفياً للنظرية الأصلية. تُعد مواصفات نظام القواعد وأنظمة إعادة الكتابة النوعين الرئيسيين للأوصاف اللغوية الحاسوبية المستخدمة بشكل متكرر في مناهج الذكاء الاصطناعي لمعالجة اللغة الطبيعية (Chrupała, 2023).

في معالجة اللغة الطبيعية (NLP)، تُمثل بيانات اللغة تسلسلات تُستخدم غالباً كمدخلات لأنظمة الذكاء الاصطناعي. يسمح جمع الأمثلة الطبيعية لهذه البيانات بقياس أداء النظام في مهمة ما أو طلب تعليقات توضيحية تُوجه عملية التعلم. تتضمن الأساليب البديلة هندسة الميزات يدوياً لكل مهمة، ولكن العديد من الأساليب المعاصرة، وخاصةً تلك التي تستخدم الشبكات العصبية، تتعلم الميزات بشكل غير مباشر من خلال مجموعات بيانات واسعة (Mullis et al., 2024).

تُظهر الأسس النظرية للصيغ اللغوية ونماذج معالجة اللغة الطبيعية تداخلاً كبيراً. تُوفر اللغات الرسمية والأتمتة الحاسوبية المصاحبة لها، والتي تدرسها اللغويات النظرية، أساساً لتصنيف وتحليل وفهم مختلف اللغات الطبيعية

والنماذج الحاسوبية، بما في ذلك تلك المستخدمة في معالجة اللغة الطبيعية. نتيجةً لذلك، ابتكر هذا المجال فئات متنوعةً للنماذج الحاسوبية، مثل آلات الحالة المحدودة والقواعد النحوية، وأثبتت نتائج نظريةً عديدةً تتعلق بأساليب متنوعة، بدءًا من الصياغات التوليدية والاحتمالية وصولاً إلى الشبكات العصبية (Başarslan & Kayaalp, 2023).

1.3 تطبيقات معالجة اللغة الطبيعية في الذكاء الاصطناعي

أحدثت تقدم تقنيات الذكاء الاصطناعي الحديثة تحولاً جذرياً في مهام معالجة اللغة الطبيعية (NLP) مثل التلخيص التلقائي والترجمة الآلية والإجابة على الأسئلة. وقد أدى الانتقال من الأساليب القائمة على القواعد إلى الأساليب الإحصائية والتعلم الآلي والتعلم العميق إلى تحسين قدرة الآلات الذكية على فهم اللغة البشرية وإنشاء نصوصٍ تُشبه نصوص الإنسان (Khurana et al., 2023).

في التطبيقات العملية، مهّدت معالجة اللغة الطبيعية (NLP) الطريق للعديد من تقنيات الذكاء الاصطناعي، بما في ذلك المساعدات الاصطناعية، والتلخيص التلقائي، والشروحات الذكية. تُظهر هذه التطورات دمج النماذج اللغوية والأدوات التحليلية في أنظمة الذكاء الاصطناعي (Hatta et al., 2024).

يُعد التعلم المُشرف وغير المُشرف مفهوميْن أساسيين في الذكاء الاصطناعي الحديث. تتضمن المهام المُشرفة تصنيف بيانات التدريب بشكل صحيح. على سبيل المثال، قد تُربط مجموعة من الصور بأسماء مثل "كلب" أو "قط" أو "جبل". يفترض التعلم المُشرف أن مجموعة البيانات المُستخدمة لتدريب النموذج تُوفر معلومات كافية لتنفيذ المهمة المطلوبة. في المقابل، لا تفترض مهام التعلم غير المُشرف توفر الأسماء بالضرورة. بدلاً من ذلك، تحاول النماذج غير المُشرفة تحديد البنية الضمنية في بيانات الإدخال. في سياق متصل، يُنظر إلى التعلم ذاتي الإشراف بشكل مُختلف كفتحة من التعلم المُشرف أو غير المُشرف، مع استمرار العمل بوجود بنية ضمن بيانات الإدخال. ومن أمثلة المهام ذاتية الإشراف التنبؤ المُشترك في اللغة، حيث تُحجب مجموعة فرعية من سلاسل الرموز من عينة من لغة ما، ويُطلب من النموذج التنبؤ بالرموز المفقودة (Pantano et al., 2025). كلا فئتي التعلم مهمتان لنجاح دمج اللغويات والذكاء الاصطناعي. تُعدّ دراسة اللغة مفيدة بشكل خاص لتصميم وفهم التعلم غير الخاضع للإشراف؛ وقد أسفرت أساليب التعلم الآلي نفسها عن حسابات نظرية المعلومات لاكتساب اللغة. وفي الوقت نفسه، تُعدّ بنية اللغة غنية بالمعلومات في التعلم الخاضع للإشراف، حيث غالباً ما يكون توليد سمات النموذج من البيانات أفضل باستخدام سمات AIS مقارنةً بالسمات العامة. بالنسبة لمهام اللغة الطبيعية، قد يكون توافر البيانات الخاضعة للإشراف نادراً جداً، لذلك تستخدم جميع نماذج معالجة اللغة الطبيعية تقريباً التدريب غير الخاضع للإشراف أو التدريب الذاتي للإشراف لمعلماتها العامة، مع الاحتفاظ بالإشراف لنسبة صغيرة نسبياً من معلمات النموذج (Laurer et al., 2024).

2.3 استخراج السمات من البيانات اللغوية

يُحدد اختيار السمات بشكل جوهري نجاح تطبيقات التعلم الآلي لمعالجة اللغة الطبيعية. خلال مشروع نظام الخبراء بلغة الكري في إمبيرال، اكتُشف مبدأ "الاستقرار البناء" (Hellmann et al., 2013): تحدث إخفاقات كارثية إذا لم تُقدّم السمة اللغوية المطلوبة لإجراء التحليل المستهدف؛ ولن تُنقد إضافة المزيد من البيانات الموقف. لذلك، ركّز تطوير القواعد النحوية على تحديد كيفية استخلاص المعلومات اللغوية لتكون سمات تحليلية فعّالة.

شكّل تطوير مجموعة صغيرة من السمات المصنّعة يدوياً أرضيةً للعمل. وقد أخذت هذه السمات في الاعتبار الظواهر السياقية التي تؤثر على المكونات النحوية، مثل ارتباط الجملة بحرف الجر، وتعدد الأفعال، وحل الإحالة إلى المجاز. واستُخدمت أنطولوجيا قائمة على الاستيعاب للكيانات الدلالية، مُهيكلت حول عناصر البيئة الصوتية. وقد قبلت الفلسفة

المُستخدمة عدم تحديد عدد كبير من المعايير اللغوية. ولم تُدمج عملية استخلاصها إلا ما يتطلبه الهدف طويل المدى لأنطولوجيا نمذجة المجال، وما تستطيع القواعد النحوية الحالية توفيره (Zafar, 2025).

3.3 بناء الجملة والتحليل

تركز النظريات النحوية على المكونات والتبعية. تُحدد تحليلات المكونات مجموعات الكلمات، أو المكونات، التي تعمل بشكل مشابه ضمن عبارة أكبر، مما يُساعد في تمثيل الحلول والاستدلال. تُركز تحليلات التبعية على العلاقات الهرمية، مع وجود روابط ثنائية مهيمنة بين الكلمات الفردية. تختلف العلاقات وهياكل شجرة التحليل عبر الأطر النظرية. تشمل المناهج الشائعة قواعد بنية العبارة، وقواعد الزاوية اليسرى، والقواعد المعجمية والمعجمية، وقواعد الاستبدال الشجرية، وقواعد التبعية، وقواعد الارتباط، والقواعد التجميعية الفئوية (Ge, 2010). في مهام معالجة اللغة الطبيعية، تُفيد التحليلات النحوية في اتخاذ القرارات وتُعزز أساليب التعلم الآلي بمعلومات مفيدة من البيانات غير المُسمّاة. تُعالج العديد من الخوارزميات الإحصائية وخوارزميات التعلم الآلي التحليل النحوي، بما في ذلك القواعد الاحتمالية الخالية من السياق، والمناهج المعجمية المتنوعة، وخوارزميات شجرة التمييز القابلة للتعديل، وتحليلات التبعية، وهياكل البيانات القوية الثابتة (Jaf & Calder, 2019).

تتكون اللغة من سلاسل، كل سلسلة منها عبارة عن تسلسل خطي من صفر أو أكثر من العناصر الأساسية تُسمى الرموز. على سبيل المثال، في اللغة الإنجليزية، الرمز هي كلمات أو علامات ترقيم، وفي اللغة اليابانية المكتوبة، الرمز هي أحرف، وفي التعرف على الكلام، غالبًا ما تكون الرموز أصواتًا كلامية فردية. يُقال إن السلسلة نحوية إذا كانت جملة في اللغة. بما أن الرموز لا تكون مستقلة تمامًا عن بعضها البعض، فإن اللغات عبارة عن مجموعات لا نهائية من سلاسل الرموز، تتفق على بنية خاصة بكل لغة تربط رمزًا برمز. في اللغة الإنجليزية، تتكون البنية من سلسلة من العلاقات التأسيسية والتبعية بين الرموز. تجمع العلاقة التأسيسية الرموز مع رموز أخرى، ولكنها لا تحدد أدوار أو علاقات الرموز داخل المجموعة بشكل دقيق. تربط علاقة التبعية كل رمز داخل المجموعة برمز آخر فيها - في اللغة الإنجليزية، يكون هذا الرمز عادةً أبعد إلى اليسار (Borenstein et al., 2024).

4.3 خوارزميات التحليل في الذكاء الاصطناعي

يكشف التحليل النحوي عن العلاقات النحوية والبنوية بين الكلمات في الجملة، ويربط تسلسلات الكلمات بالمعاني، ويُمكن من إزالة الغموض عن المدخلات اللغوية بشكل صحيح. بخلاف البيانات النصية، يُعرض الكلام في جمل غالبًا ما تكون عفوية وغير منظمة، وتدمج عدة محادثات، مما يتطلب معلومات سياقية إضافية وتحليلًا معقدًا. غالبًا ما يكون تحويل الكلام إلى كلمات غير مكتمل أو غامض أو خاطئ. تُعد التقنيات التي تعالج هذه المشكلات وتعزز الفهم الدلالي هدفًا بحثيًا رئيسيًا. تتطلب تطبيقات اللغة المنطوقة في سياقات الحوار هياكل تحليل للكلام تدريجيًا وفي تزامن زمني. لذلك، يجب على المحلل جدول الموارد عند ظهور المتطلبات، بناءً على شروط التعرف واختيار القيود التي تحد من مساحة البحث (Menzel, 1994).

5.3 النظريات الدلالية

تقدم العديد من النظريات الدلالية تفسيرات منهجية لمعنى التعبيرات اللغوية، والتي يمكن استخدامها لتفسير النصوص أو بناء أنظمة اصطناعية (Sukaton, 2019). في مجال معالجة اللغة الطبيعية (NLP) والمجالات المماثلة، تشرح هذه النظريات كيفية فهم التواصل البشري عبر اللغة على عدة مستويات: علم النحو، وعلم الصرف، وعلم الدلالة، وعلم التداولية¹، وتقتح كيفية نقل هذه المستويات إلى الحاسوب للتفاعل بشكل سليم مع البشر. ووفقًا لبويسيو، فإن فهم

الكلام يتضمن تحليله، والاستعانة بالمعلومات المعجمية، وفهم سياقه، وممارسة التفكير المنطقي السليم (Tian et al., 2024).

6.3 تمثيل المعنى في الذكاء الاصطناعي

ميلت المناهج المبكرة لفهم الحاسوب إما إلى مساواة الفهم بالقدرة على الإجابة على الأسئلة أو إلى استخدام المنطق كوسيلة للتمثيل. أما الرأي المعاصر، فيشير إلى أن الأمر يتطلب أكثر من ذلك بكثير. تُظهر تعريفات القاموس المذكورة أعلاه أن القدرة على التفسير تعتمد على جميع جوانب اللغة وتساهم فيها. وللتفسير، يجب أن يكون الحاسوب الذكي قادرًا على التعبير عما يسمعه أو يقرأه بلغته الخاصة، وأن يسأل نفسه عن مدى معقولية هذا التعبير واتساقه وملاءمته. ويجب أن تكون طريقة التعبير عن المعنى بحيث تتمكن الخوارزميات من العمل على هذه التعبيرات لمراعاة معلومات المعقولة والاتساق والملاءمة التي يوفرها المُعطي، واتخاذ خيارات مدروسة - لا عشوائية - بين التفسيرات ومسارات العمل (Cui et al., 2023).

تميل النظريات اللغوية الحالية للبراغماتية إلى التركيز على أفعال الكلام كما أشار إليها أوستن (1962) وسيرل (1969)، والبنية التخاطبية وفقًا لجريس (1989)، ومبادئ الصلة التي وضعها سيربر وويلسون (1986). ولا تزال تحليلات جريس مؤثرة بشكل خاص بين اللغويين (Hirst, 1997). ومع ذلك، لم يتوصل باحثو الذكاء الاصطناعي بعد إلى نظرية شاملة تنقلهم من معنى الجملة وسياقها المحلي إلى نوايا المتكلم. ورغم استخدام أنظمة التخطيط والاستدلال للأغراض العامة لاستخلاص استنتاجات براغماتية، إلا أن هذه الأنظمة لا تزال هامشية بالنسبة للذكاء الاصطناعي السائد. لذلك، غالبًا ما تُعتبر البراغماتية تخصصًا يتجاوز حدود الذكاء الاصطناعي نفسه. ومع ذلك، يجب على التطبيقات العملية للذكاء الاصطناعي تفسير العبارات تفسيرًا براغماتيًا، مع التمييز بين الطلبات والأوامر والتهديدات، على سبيل المثال، حتى عندما تتشابه الجمل في هيكلها. نتيجة لذلك، سعت المناهج الحديثة إلى تطوير نماذج براغماتية يمكن دمجها بسهولة في الأنظمة العملية. لا تحتاج هذه النماذج إلى تقديم تحليل شامل للنية؛ بل تُستغل أطر التخطيط لأنظمة الذكاء الاصطناعي المبكرة لإصلاح التفسيرات الأولية غير الكافية. ينشأ تحدٍ كبير لأن جميع نظريات اللغة محدودة. فاللغة، أكثر من أي نشاط آخر، تعتمد بشكل كبير على السياق، وتمتد هذه التبعيات إلى ما هو أبعد من مجرد النطق المباشر أو السياق الاجتماعي الثنائي. يثبت النهج النظري للموقف في السياق فائدته هنا، إذ يسمح بدمج فئات كبيرة من التبعيات في مفهوم شكلي واحد (Deroy, 2023).

7.3 النظريات البراغماتية

يتطلب اكتساب اللغة واستخدامها القدرة على صياغة العبارات في سياق براغماتي غني. يتضمن هذا عادةً التفكير في معرفة ونوايا كل من المتحدث والمستمع. على الرغم من أن مجموعة كبيرة من الأبحاث قد ركزت على نماذج اكتساب اللغة أو التواصل الناشئ، إلا أن القليل منها يدمج التفكير البراغماتي. غالبًا ما تأخذ الأعمال الحالية بعين الاعتبار متعلمي اللغة فقط، متجاهلةً أثر المُحاور. بناءً على أبحاث سابقة تُنمذج الفاعلين الفرديين باستخدام التفكير البراغماتي الصريح، نقدم نموذجًا متماثلًا يُمكن فاعلين من إنتاج وتفسير لفظة بشكل مشترك. يتمتع كل فاعل بحالات معتقدات وأهداف وقدرات عقلية خاصة بالآخر. نقدم مجموعة متنوعة من التحليلات لفهم سلوك النموذج الناتج، ونثبت أن التفكير البراغماتي يُمكن الفاعلين من حل الألعاب المرجعية بنجاح أكبر. تُظهر مجموعة من التجارب، فضلًا عن مجموعة بيانات الأشكال الرمزية واسعة الاستخدام، أن نموذجنا قادر على تحقيق التواصل البراغماتي دون معرفة مسبقة. يُمكن النهج المُقدم من ظهور بنية تركيبية قابلة للتفسير، قادرة على استيعاب رموز جديدة. وتتمثل الخطوة التالية الطبيعية في استكشاف آثار أنظمة المعلومات المختلفة. وقد تتبنى الدراسات المستقبلية منظورًا تطوريًا وتستخدم خوارزميات جينية لمحاكاة الظهور التدريجي للعوامل البراغماتية. (Kang et al., 2020).

1.7.3 نماذج الذكاء الاصطناعي السياقية

في الذكاء الاصطناعي واللغويات، يتضمن "السياق" معلومات تتجاوز المدخلات اللغوية اللازمة للتفسير، ولا يُمثّل الرسالة نفسها. سعت أبحاث الذكاء الاصطناعي الرائدة المبكرة إلى مبادئ شبيهة بالقانون، تُشكّل أساس السلوك البشري، لنمذجة التفكير واللغة والرؤية، والمعروفة باسم الذكاء الاصطناعي التقليدي الجيد (GOFAI) (Alberts, 2022). ومع ذلك، غالبًا ما أثبتت هذه الأساليب القائمة على قواعد واستثناءات مكتوبة بخط اليد أنها غير عملية. منذ الثورة الإحصائية في التسعينيات، سمحت أساليب استنتاج الارتباطات الإحصائية من أمثلة وفيرة للأنظمة بمحاكاة السلوك البشري في تطبيقات ضيقة ومحددة دون فهم المنطق السببي، مما حوّل التركيز من الذكاء الاصطناعي العام إلى حلول عملية خاصة بمهام محددة. ونتيجة لذلك، تكثف التركيز على تقنيات فهم اللغة الطبيعية القوية (NLU) لتسهيل التواصل بين الإنسان والآلة. وأعدت معالجة اللغة الإحصائية القائمة على مجموعة النصوص إحياء التفاؤل بشأن محاكاة القدرة اللغوية البشرية واجتياز اختبار تورينج، حيث استخرجت الأساليب الإحصائية أنماطاً سطحية وحلت حالات عدم اليقين بشكل احتمالي. وتؤكد النظريات المجسدة للإدراك على أثر البيئات المادية والاجتماعية في تشكيل الإدراك، مما يسلط الضوء على تحدي ترسيخ فهم لغة الذكاء الاصطناعي في تجربة العالم الحقيقي والتفاعل. وبالتالي، يظل دعم التفسير السياقي في أنظمة الذكاء الاصطناعي أمراً بالغ الأهمية لاستخدام اللغة في العالم الحقيقي (Kumar et al., 2024). تتفوق نماذج اللغة الكبيرة (LLMs) مثل GPT في إنتاج وفهم اللغة والمعرفة والرمز وغيرها من الوسائط. يتيح تقديمها الحوسبة الواعية للسياق، حيث يمكن إجراء نمذجة السياق والاستدلال - وهي تحديات تاريخية - من خلال التفاعل، كما هو موضح في الدراسات (Xiong et al., 2023). يعتبر النموذج طلبات المستخدمين وأجهزة الاستشعار والأوامر كمدخلات نصية؛ يُوجّه وكيل تلقائي برنامج ماجستير القانون (LLM) باستخدام بيانات سياقية لوضع خطط عمل، والتي تُعزز، عند تنفيذها، الوعي بالسياق. تشمل الأمثلة تشغيل ذراع متنقلة لتوصيل الأدوية. وقد تم تحليل العوامل المؤثرة على أداء الوعي بالسياق المُدار من قبل ماجستير القانون، مع تحديد أسئلة بحثية مفتوحة. تتخطى نماذج الذكاء الاصطناعي السياقية حدود النحو والدلالات، حيث تلتقط وتُطبّق البرجماتية؛ فالمعنى المقصود للكلمة لا يعتمد فقط على النحو والدلالات والاستدلالات، بل يعتمد أيضاً على الموقف التواصلية والمعرفة السابقة والافتراضات المسبقة. تُعزز أنظمة الذكاء الاصطناعي التي تستوعب هذه العناصر التكامل اللغوي (Zou et al., 2023).

8.3 توليد اللغة

يُمكن توليد اللغة الأنظمة من التواصل بفعالية مع المستخدمين من خلال الكلام أو النص، وهي قدرة بالغة الأهمية عندما يفتقر المستخدمون إلى الإلمام باللغات الرسمية (McKeown, 2005). كما يُمكن الحواسيب من تقديم شروحات وتوضيحات ووصف للإجراءات والنتائج، مما يُسهّل التفاعل مع قواعد البيانات ومستودعات المعلومات وأنظمة الخبراء. يتضمن تطوير أنظمة توليد اللغة معالجة العديد من القضايا، لا سيما تحديد المعلومات التي يجب إيصالها، وتحديد وقت إيصالها، والتعبير عن النية بوضوح ومباشرة. على سبيل المثال، تستخدم أنظمة الإجابة على الأسئلة مثل TEXT وTAILOR توليد اللغة للرد بشكل تفاعلي على استفسارات المستخدمين. تفتقر العديد من هياكل الذكاء الاصطناعي إلى الذكاء التخاطبي العام، وهي إما لا تتعامل مع اللغة أو تعتمد على أساليب يدوية قائمة على القوالب (Ramesh & Kolonin, 2021). في المقابل، يُؤدّد النهج القائم على قواعد الربط تلقائياً جملاً صحيحة نحوياً ومتناسكة تتجاوز الخطوط الأساسية الحالية. تُنتج أنظمة معالجة اللغة القابلة للتفسير نماذج ونتائج مفهومة، على عكس هياكل الذكاء الاصطناعي المغلقة،

وتؤد لغة قابلة للتفسير وصحيحة نحويًا مبنية على قواعد نحوية مفهومة. تتطلب أساليب تعلم القواعد النحوية الحالية، بما في ذلك الشبكات العصبية العميقة، عادةً كميات كبيرة من بيانات التدريب المُشرف عليها (Basavatia et al., 2024).

1.8.3 تقنيات توليد النصوص

شغل تطوير تقنيات توليد النصوص الاصطناعية الباحثين في اللغويات والذكاء الاصطناعي والحوسبة منذ الأيام الأولى للمعالجة الآلية للغات. وقد أدت هذه الجهود إلى مجموعة واسعة من الأساليب، بدءًا من تلخيص المستندات ووصولًا إلى السرد الآلي بالكامل مثل مناخات المدن الخيالية. يقدم هذا القسم لمحة عامة موجزة عن توليد النصوص ومكانته في المشهد المتكامل للذكاء الاصطناعي واللغويات (Zhang et al., 2023).

ونظرًا لأن انتشار الذكاء الاصطناعي يُبرز الطلب المستمر على الحلول التقنية لتحديات معالجة اللغة الطبيعية، فإن إنتاج نص متماسك وحساس للسياق ومتصل لا يزال يتطلب الكثير من منظور اللغويات الحاسوبية، حيث يركز على سمائيات وسيلة التعبير بدلاً من الرسالة أو المعنى (Santhanam & Shaikh, 2019). نشأ توليد اللغة في الأصل كعكس للتحليل، ويتطلب استراتيجيات أو آليات لتحويل الرسائل إلى سلاسل نصية خطية بالعمق المناسب من التفصيل أو مستوى التجريد. وبينما قد يكون تقدير أهمية أحداث كلامية معينة مجالًا للاستدلال أو لبنية الذكاء الاصطناعي الأوسع وبيئة التشغيل، فإن إنتاج "البصمة" اللغوية لأي حدث واحد يُعد مهمة توليد نص. وبالتالي، فإنه يشمل عادةً مجموعة من المهام الفرعية، بما في ذلك الاختيار المعجمي، والتجميع والاختزال، وتعيين المرجع المشترك، وهيكل الخطاب، والتكوين البلاغي، والإدراك اللغوي، إلى جانب التجزئة السطحية المناسبة (Cole et al., 1995). ويعتمد نطاق هذه العمليات وتعقيدها على القوة التعبيرية ومرونة النهج المختار، أو مجموعة الافتراضات، أو البنية العامة. وقد تقع بعض المهام أيضًا، بطبيعة الحال، خارج النطاق اللغوي الصرف، وتندرج بدلاً من ذلك تحت عنوان التعيين الأوسع لحالة المعلومات، أو التركيز، أو أهمية الخطاب. مع ذلك، يُمكن رسم تقسيم ثلاثي تقريبي بين الأنظمة الموجهة نحو السطح (والتي عادةً ما تعتمد على نمذجة النصوص، أو القوالب، أو النصوص الجاهزة)، والبيانات المنظمة أو الأنظمة المعتمدة على قواعد البيانات (بما في ذلك مناهج الاستدلال أو الاستدلال)، والمناهج "الهيكليّة" الأكثر تطورًا والقائمة على تحديد دقيق للقصد التواصلي أو الإجراءات المهمة ضمن نطاق المشكلة. وتُعدّ التكهنات المتعلقة بالعلاقات الأوسع بين توليد النصوص ونظريات أفعال الكلام، والتخطيط، والفعل والفاعلية، والحوار أمرًا شائعًا في معظم الأدبيات، ويُشير عمومًا إلى تقارب مستقبلي بين هذه المجالات (Triantafyllopoulos & Schuller, 2024).

2.8.3 التعرف على الكلام وتوليده

يُعد التعرف التلقائي على الكلام وتوليده أحد أهم سبل التكامل بين اللغويات والذكاء الاصطناعي. يسعى كلا التطبيقين إلى إعادة إنتاج العمليات التي يفهمها المتحدثون البشر الكلام وينتجونه في برامج الحاسوب (Mou, 2002). يُظهر توليف الكلام اختلافات في الأسلوب واللهجة وسرعة الكلام، والتي يجب تحليلها جميعًا وتوليدها بشكل مُرضٍ. ومع ذلك، فهي تُوفر وسيلة لتقييم نظريات التواصل الكلامي أقل تكلفة بكثير من بناء أنظمة فهم كاملة للغة المنطوقة. تطورت المناهج النظرية من الأطر المبكرة القائمة على القواعد إلى تقنيات التعلم العميق الحديثة، حيث يُساهم كل منها في تحسين الدقة والطبيعية (Wallis, 2021). كما أن التطورات في تكنولوجيا الأجهزة ومعالجة الإشارات تجعل من الواقعي التفكير في تطوير تطبيقات متعددة اللغات.

يستمر تطوير تقنيات وأنظمة مُختلفة للتعرف على الكلام وتوليده، بهدف محاكاة القدرات البشرية بدقة في هذا المجال. تحليل طيف الكلام، الذي يدرس خصائص تردد إشارات الكلام، يوفر معلومات مثل مواقع المكونات، والتردد الأساسي، وطاقة الإشارة. كما تستخلص معالجة إشارات الكلام سمات الكلام الأساسية، بما في ذلك درجة الصوت، وتوزيع الطاقة

الطيفية للمسالك الصوتية، وأنماط التردد الزمني. تستخلص الأنظمة المستقلة عن المتحدث سمات تبقى ثابتة بغض النظر عن المتحدث. يُسهّل التشوه الزمني الديناميكي (DTW) محاذاة زمنية فعّالة للأنماط المتشابهة من تسلسلات الكلام، ويُستخدم لتعلم نماذج أولية لكلمات معزولة من نماذج متعددة للتعرف على الكلام. أما التعرف على الأنماط، القائم على تجميع متجهات السمات المشتقة من سمات الكلام، فيربط كل نمط إدخال بالنموذج الأولي الأكثر تشابهاً، وهو مناسب لمراقبة الكلمات المعزولة. يستخدم تكميم المتجهات المحسّن (VQ) رسماً بيانياً شبيهاً بالكلمات لمحاذاة كل عنصر من نمط الإدخال مع أقرب عنصر من نمط مرجعي، مع تعديل يُحدد بواسطة سمة محددة مُعدّلة للكلام. يحاكي كل من أساليب المعالجة هذه جانباً واحداً على الأقل من قدرات معالجة الكلام البشرية (Amaro, 2016).

4. أنظمة الذكاء الاصطناعي متعددة اللغات

تحظى أنظمة الذكاء الاصطناعي الحديثة بدعم متزايد لعدد متزايد من اللغات الطبيعية - فعلى سبيل المثال، يمكن لخدمة Google Translate الترجمة بين أكثر من 100 لغة، بينما تدعم Siri أكثر من 20 لغة. ومع ذلك، لا تزال العديد من مهام الذكاء الاصطناعي المهمة، مثل التعرف على الكلام، والإجابة على الأسئلة، واسترجاع المعلومات بلغات متعددة، تُمثل تحدياً. السبب الأساسي هو التنوع اللغوي وندرة الموارد اللغوية للغات غير الممثلة تمثيلاً كافياً - كحالة متطرفة، لا تحتوي مئات اللغات في العالم على بيانات رقمية على الإطلاق. الويب، هذا المصدر الرئيسي لبيانات اللغة، هو في الغالب باللغة الإنجليزية. حتى بين صفحات الويب غير الإنجليزية، فإن غالبية المحتوى النصي موجود بلغات شائعة مثل الإسبانية واليابانية والبرتغالية (Rehm et al., 2020).

ستظل التعددية اللغوية تحدياً محورياً لمجتمع المعلومات بشكل عام، ولتقنيات اللغات المتقدمة بشكل خاص، لفترة طويلة قادمة. يُعدّ دعم اللغة الإنجليزية وحدها مقبولاً للشؤون الداخلية للعديد من المنظمات أو للأنشطة ذات التأثير التكنولوجي المنخفض، ولكنه غير مقبول على الإطلاق للتطبيقات المدنية في الكيانات متعددة اللغات مثل الاتحاد الأوروبي (Mohamed et al., 2024). وفيما يتعلق بتطبيقات تقنيات اللغات المتقدمة، لا يقتصر التعدد اللغوي على توافر الموارد اللغوية وتغطية تقنيات اللغات، على الرغم من أهميتها في حد ذاتها؛ بل يتطلب أيضاً توافر أدوات دعم دقيقة تغطي سلسلة قيمة تقنيات اللغات بأكملها، مما يوفر إمكانية الوصول والاستغلال الفعال والإدارة الفعالة لموارد اللغات (Buddemeyer et al., 2021).

يتعين على العديد من أنظمة الذكاء الاصطناعي اليوم معالجة المدخلات والمخرجات عبر مجموعة متنوعة من اللغات لتكون قابلة للتطبيق عالمياً. لذلك، يتعاون عدد كبير من المؤسسات الأوروبية في مجموعة متنوعة من مشاريع البحث والتطوير في مجال تقنيات اللغات لتطوير أدوات وموارد لدعم فرق أبحاث تقنيات اللغات في جميع أنحاء أوروبا. يقدم المجلس الأوروبي لتكنولوجيا اللغات (ELTEC) مساهمات في أنشطة تكنولوجيا اللغات متعددة اللغات واسعة النطاق إلى المجلس الأوروبي لتكنولوجيا اللغات التابع للمفوضية الأوروبية، وهو مصمم لتمثيل مجال تكنولوجيا اللغات الأوروبية بأكمله. يدعم المجلس المهام الاستراتيجية ويعزز التعددية اللغوية المدعومة بالتكنولوجيا لتلبية الاحتياجات الأوروبية بنجاح. تُعد القدرات متعددة اللغات والوسائط، والمعايير، وقابلية التشغيل البيئي، واستخراج المعلومات الموحد متعدد اللغات من مونت-ديوس، وأفضل ممارسات تكنولوجيا اللغات المستلمة عبر الإنترنت، من الأولويات الاستراتيجية في هذا الصدد. يُسهّل المجلس التواصل العلمي والتقني، فضلاً عن مهام التنسيق في جميع الأنشطة والبرامج الدولية المتعلقة بأبحاث تكنولوجيا اللغات وتطويرها وابتكارها. هناك حاجة ماسة لتطوير وتطبيقات تكنولوجيا اللغات متعددة التخصصات واسعة النطاق بما يعود بالنفع الكامل على الاقتصاد والمجتمع، مع توفير توافر أوسع ووصول أوسع للجمهور في

المقام الأول (Zhang et al., 2023). على الرغم من أن تقنيات اللغة البشرية، المتجذرة أصلاً في الذكاء الاصطناعي واللغويات الحاسوبية، وخاصةً لفهم اللغة الطبيعية، كانت على أعتاب مرحلة النضج منذ عقود، إلا أن الحدود بينهما أصبحت ضبابية بشكل متزايد، حيث تعتمد العديد من المناهج المتعلقة باللغة على أطر التعلم العميق متعددة الأغراض، المرتبطة عادةً بالذكاء الاصطناعي. وبالتالي، أصبح الذكاء الاصطناعي المرتكز على اللغة واقعاً علمياً. تُطوّر تقنيات وتقنيات الذكاء الاصطناعي الحالية نماذج عصبية عميقة شديدة التنافسية، وتستفيد منها، وتتضمن وسائط متعددة، وتتعلم باستمرار وبشكل عام من بيانات محدودة، وتستغل تمثيلات سياقية أكثر ثراءً ووضوحاً. تتطلب هذه الأنظمة جهوداً بحثية مختلفة ومتعددة التخصصات، تعتمد على خبرات إضافية تتجاوز البحث والتطوير التقليدي لتكنولوجيا اللغة (Bayer et al., 2023).

5. أمثلة تطبيقية في مجال معالجة اللغة العربية بالذكاء الاصطناعي

تمثل تطبيقات معالجة اللغة العربية مجالاً خصباً لتطبيق خوارزميات الذكاء الاصطناعي في تحليل النصوص وفهم المعاني. وتُعد أنظمة تحليل المشاعر مثالاً بارزاً على ذلك، إذ تعتمد على خوارزميات *التعلم العميق* لتحديد النغمة العاطفية للنصوص المستخرجة من وسائل التواصل الاجتماعي، كتمييز ما إذا كانت إيجابية أو سلبية أو محايدة (Cui et al., 2023). وقد تم تطوير نماذج لغوية مثل AraBERT وCAMELBERT لتجاوز تحديات اللهجات العربية، مما حسن دقة تحليل النصوص العربية بشكل كبير (Khurana et al., 2023).

أما في مجال التعرف على الكيانات الاسمية (Named Entity Recognition)، فقد ساهمت النماذج القائمة على الشبكات العصبية العميقة في استخلاص أسماء الأشخاص والأماكن والمؤسسات من النصوص العربية بدقة أعلى مقارنةً بالأساليب الإحصائية التقليدية (Hellmann et al., 2013).

كذلك، شهدت تطبيقات الترجمة الآلية تطوراً ملحوظاً بفضل دمج المعرفة اللغوية من النحو والدلالة العربية في النماذج الذكية، مما ساعد على إنتاج ترجمات أكثر سلاسة ودقة تعكس المعنى المقصود في السياق الأصلي (Shormani, 2024). تؤكد هذه الدراسات على أهمية التعاون بين المتخصصين في علم اللغة وخبراء الذكاء الاصطناعي من أجل بناء أنظمة قادرة على فهم اللغة العربية فهماً إدراكياً وليس حسابياً فحسب، مع مراعاة قضايا التحيز الأخلاقي والشفافية في مخرجات الأنظمة الذكية (Mandour, 2025).

6. عرض النتائج ونفسيرها

أظهرت نتائج الدراسة التحليلية للأدبيات أن التكامل بين علم اللغة والذكاء الاصطناعي لم يعد يقتصر على الجوانب النظرية، بل أصبح يشكل أساساً لتطوير تطبيقات عملية متقدمة في معالجة اللغة الطبيعية. فقد بيّنت الدراسات أن اعتماد النماذج اللغوية الكبيرة وتقنيات التعلم العميق أسهم بشكل واضح في تحسين أداء أنظمة الترجمة الآلية، وتوليد النصوص، والتعرف على الكلام، مقارنةً بالمقاربات التقليدية القائمة على القواعد فقط.

كما كشفت النتائج أن دمج المستويات اللغوية المختلفة، مثل النحو والدلالة والتداولية، داخل النماذج الحاسوبية يؤدي إلى زيادة قدرة الأنظمة الذكية على فهم السياق والتعامل مع الغموض اللغوي بصورة أكثر دقة. وأظهرت التطبيقات الحديثة في تحليل النصوص العربية ونماذج اللغة متعددة اللغات أن الاستفادة من المعرفة اللغوية تساعد على تجاوز العديد من التحديات المرتبطة بندرة البيانات أو اختلاف اللهجات.

من جهة أخرى، بيّنت النتائج استمرار وجود فجوة بين القدرة الشكلية للنماذج اللغوية الحديثة وبين الفهم العميق للمعنى والسياق الثقافي، إذ ما زالت الأنظمة تعتمد بدرجة كبيرة على الأنماط الإحصائية أكثر من اعتمادها على تمثيل معرفي شامل للغة.

تشير هذه النتائج إلى أن التطور السريع في تقنيات الذكاء الاصطناعي لم يكن منفصلاً عن التقدم في الدراسات اللغوية، بل إن العلاقة بينهما أصبحت علاقة تكاملية تقوم على تبادل الأدوار؛ إذ يوفر علم اللغة الأسس النظرية اللازمة لتنظيم المعرفة اللغوية، بينما يقدم الذكاء الاصطناعي أدوات تحليلية قادرة على معالجة كميات ضخمة من البيانات. ويُفسّر التحسن الملحوظ في أداء أنظمة معالجة اللغة الطبيعية بانتقال المجال من النماذج الرمزية المحدودة إلى النماذج الإحصائية والشبكات العصبية العميقة، التي تستطيع تعلّم السمات اللغوية ضمنياً من البيانات. ومع ذلك، فإن استمرار مشكلات تمثيل المعنى والبراغماتية يوضح أن الذكاء الاصطناعي ما زال بحاجة إلى دمج أعمق للنظريات اللغوية المعرفية والسياقية.

كما تعكس النتائج أن التعددية اللغوية تمثل تحدياً رئيسياً، حيث تميل النماذج الحديثة إلى تحقيق أداء أعلى في اللغات الغنية بالبيانات مقارنة باللغات ذات الموارد المحدودة، مما يبرز أهمية تطوير نماذج تراعي الخصوصيات الثقافية واللغوية. بناءً على ذلك، يمكن القول إن مستقبل أنظمة الذكاء الاصطناعي اللغوية يعتمد على بناء نماذج هجينة تجمع بين التحليل اللغوي العميق والتعلم الآلي، بما يسمح بالانتقال من مجرد معالجة سطحية للنصوص إلى فهم أكثر قرباً من الإدراك البشري.

7. التوصيات

استناداً إلى ما تم عرضه من نتائج وتحليل لأبعاد التكامل بين علم اللغة والذكاء الاصطناعي، يمكن اقتراح مجموعة من التوصيات العلمية التي تساهم في تطوير هذا المجال وتعزيز تطبيقاته المستقبلية:

1. تعزيز التكامل بين النظريات اللغوية والنماذج الحاسوبية الحديثة من خلال تصميم أطر بحثية تجمع بين النحو والدلالة والتداولية مع تقنيات التعلم العميق، بما يساعد على بناء أنظمة قادرة على فهم المعنى والسياق بصورة أعمق.
2. الاهتمام بتطوير نماذج ذكاء اصطناعي متعددة اللغات تراعي الخصوصيات اللغوية والثقافية، مع التركيز على اللغات ذات الموارد المحدودة مثل اللغة العربية، وذلك للحد من التحيز اللغوي وتحقيق عدالة أكبر في تطبيقات الذكاء الاصطناعي.
3. توسيع الدراسات التطبيقية في معالجة اللغة العربية عبر تطوير مجموعات بيانات معيارية، وتشجيع التعاون بين اللغويين وخبراء الذكاء الاصطناعي لتصميم نماذج قادرة على التعامل مع اللهجات والسياقات الثقافية المختلفة.
4. تعزيز البعد الأخلاقي في تصميم الأنظمة الذكية من خلال تضمين مبادئ الشفافية والمسؤولية وتقليل الانحياز في مراحل التدريب والتقييم، خاصة في التطبيقات التعليمية والإعلامية.
5. تشجيع الأبحاث متعددة التخصصات التي تجمع بين علوم اللغة والعلوم المعرفية وعلوم البيانات، بهدف الوصول إلى نماذج أكثر قدرة على تفسير النوايا البشرية والتفاعل الطبيعي مع المستخدمين.
6. التركيز على تطوير نماذج تفسيرية (Explainable AI) في مجال معالجة اللغة الطبيعية، بحيث تكون مخرجات الأنظمة مفهومة وقابلة للتحليل، مما يساهم في زيادة الثقة في التطبيقات الذكية.

بيانات الإفصاح:

- الموافقة الأخلاقية والموافقة على المشاركة: تم الاتفاق على المشاركة في البحث وفقاً للإرشادات الخاصة بالمجلة.
 - توافر البيانات والمواد: كافة البيانات والمواد متاحة عند الطلب.
 - مساهمة المؤلفين: يتحمل المؤلفين مسؤولية كافة محتويات البحث والتحليل والمنهجية والمراجعة الكاملة.
 - تضارب المصالح: لا يوجد تضارب في المصالح لأي طرف من خلال تصميم البحث وتقديمه وتقييمه.
 - التمويل: لا يوجد أي تمويل مخصص لهذا البحث.
 - شكر وتقدير: الشكر الجزيل لأكاديمية التطوير العلمي ومجلة المؤتمرات العلمية (JSC) على الدعم والإرشادات
- [\(/https://sdasmart.org/jsconf/\)](https://sdasmart.org/jsconf/)

8. Reference

- [1] Alberts, L. (2022). Not cheating on the Turing Test: towards grounded language learning in Artificial Intelligence. arXiv preprint arXiv:2206.14672.
- [2] Amaro, R. (2016). Teaching computational linguistics: challenges and target audiences. In II World Congress on Computer Science. Engineering and Technology Education. Portugal: Castelo Branco.
- [3] Başarslan, M. S., & Kayaalp, F. (2023). MBI-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis. Journal of Cloud Computing, 12(1), 5.
- [4] Basavatia, S., Murugesan, K., & Ratnakar, S. (2024). Starling: Self-supervised training of text-based reinforcement learning agent with large language models. arXiv preprint arXiv:2406.05872.
- [5] Bayer, M., Kaufhold, M. A., Buchhold, B., Keller, M., Dallmeyer, J., & Reuter, C. (2023). Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers. International journal of machine learning and cybernetics, 14(1), 135-150.
- [6] Borenstein, N., Svete, A., Chan, R., Valvoda, J., Nowak, F., Augenstein, I., ... & Cotterell, R. (2024). What languages are easy to language-model? a perspective from learning probabilistic regular languages. arXiv preprint arXiv:2406.04289.
- [7] Buddemeyer, A., Walker, E., & Alikhani, M. (2021). Words of wisdom: Representational harms in learning from AI communication. arXiv preprint arXiv:2111.08581.
- [8] Chrupała, G. (2023). Putting natural in natural language processing. arXiv preprint arXiv:2305.04572.
- [9] Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. Frontiers in Artificial Intelligence, 4, 625341.
- [10] Cole, R., Hirschman, L., Atlas, L., Beckman, M., Biermann, A., Bush, M., ... & Zue, V. (1995). The challenge of spoken language systems: Research directions for the nineties. IEEE transactions on Speech and Audio processing, 3(1), 1-21.

- [11] Cui, J., Wang, Z., Ho, S. B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 56(8), 8469-8510.
- [12] Deroy, O. (2023). The ethics of terminology: Can we use human terms to describe AI?. *Topoi*, 42(3), 881-889.
- [13] Fernyhough, C., & Borghi, A. M. (2023). Inner speech as language process and cognitive tool. *Trends in cognitive sciences*, 27(12), 1180-1193.
- [14] Ge, R. (2010). Learning for semantic parsing using statistical syntactic parsing techniques.
- [15] Haber, Y., Levkovich, I., Hadar-Shoval, D., & Elyoseph, Z. (2024). The artificial third: a broad view of the effects of introducing generative artificial intelligence on psychotherapy. *JMIR Mental Health*, 11, e54781.
- [16] Hatta, M., Wahid, W. N., Yusuf, F., Hidayat, F., Santoso, N. A., & Aini, Q. (2024). Enhancing predictive models in system development using machine learning algorithms. *International Journal of Cyber and IT Service Management*, 4(2), 80-87.
- [17] Hellmann, S., Lehmann, J., Auer, S., & Brümmer, M. (2013, October). Integrating NLP using linked data. In *International semantic web conference* (pp. 98-113). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [18] Hirst, G. (1997). Context as a spurious concept. arXiv preprint [arXiv:1907.09712](https://arxiv.org/abs/1907.09712).
- [19] Jaf, S., & Calder, C. (2019). Deep learning for natural language parsing. *IEEE Access*, 7, 131363-131373.
- [20] Kang, Y., Wang, T., & de Melo, G. (2020). Incorporating pragmatic reasoning communication into emergent language. *Advances in neural information processing systems*, 33, 10348-10359.
- [21] Khaleel, M., Jebrel, A., & Shwehdy, D. M. (2024). Artificial intelligence in computer science: <https://doi.org/10.5281/zenodo.10937515>. *Int. J. Electr. Eng. and Sustain.*, 01-21.
- [22] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, 82(3), 3713-3744.
- [23] Kumar, K., Syed, M. H., Bhargava, S., Mohakud, L. L., Serajuddin, M., & Lourens, M. (2024, March). Natural language processing: Bridging the gap between human language and machine understanding. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-6). IEEE.
- [24] Laurer, M., Van Atteveldt, W., Casas, A., & Welbers, K. (2024). Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and bert-nli. *Political Analysis*, 32(1), 84-100.
- [25] Mandour, W. S. (2025). AI-Powered Tools for Linguistic and Sentiment Analysis in Language Learning: Applications Using Google Sheets and Large Language Models. In *Using AI Tools in Text Analysis, Simplification, Classification, and Synthesis* (pp. 67-112). IGI Global Scientific Publishing.

- [26] McKeown, K. R. (2005). Language generation: Applications, issues, and approaches. *Proceedings of the IEEE*, 74(7), 961-968.
- [27] Menzel, W. (1994). Parsing of spoken language under time constraints. *arXiv preprint cmp-lg/9409008*.
- [28] Mohamed, Y. A., Khanan, A., Bashir, M., Mohamed, A. H. H., Adiel, M. A., & Elsadig, M. A. (2024). The impact of artificial intelligence on language translation: a review. *Ieee Access*, 12, 25553-25579.
- [29] Mou, X. (2002). Towards a unified framework for sub-lexical and supra-lexical linguistic modeling (Doctoral dissertation, Massachusetts Institute of Technology).
- [30] Mullis, J., Chen, C., Morkos, B., & Ferguson, S. (2024). Deep neural networks in natural language processing for classifying requirements by origin and functionality: An application of bert in system requirements. *Journal of Mechanical Design*, 146(4), 041401.
- [31] Panchal, B. Y., & Shah, A. (2024, December). NLP Research: A Historical Survey and Current Trends in Global, Indic, and Gujarati Languages. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 1263-1272). IEEE.
- [32] Pantanowitz, L., Pearce, T., Abukhiran, I., Hanna, M., Wheeler, S., Soong, T. R., ... & Rashidi, H. H. (2025). Nongenerative artificial intelligence in medicine: advancements and applications in supervised and unsupervised machine learning. *Modern Pathology*, 38(3), 100680.
- [33] Paradowski, M. B. (2011). Developing embodied multisensory dialogue agents. *arXiv preprint arXiv:1111.7190*.
- [34] Petit, N., & De Cooman, J. (2021). Models of Law and Regulation for AI. *The routledge social science handbook of AI*, 199-221.
- [35] Ramesh, V., & Kolonin, A. (2021). Natural language generation using Link Grammar for general conversational intelligence. *arXiv preprint arXiv:2105.00830*.
- [36] Rehm, G., Marheinecke, K., Hegele, S., Piperidis, S., Bontcheva, K., Hajič, J., ... & Yvon, F. (2020). The European language technology landscape in 2020: Language-centric and human-centric AI for cross-cultural communication in multilingual Europe. *arXiv preprint arXiv:2003.13833*.
- [37] Santhanam, S., & Shaikh, S. (2019). A survey of natural language generation techniques with a focus on dialogue systems-past, present and future directions. *arXiv preprint arXiv:1906.00500*.
- [38] Shormani, M. Q. (2024). Generative linguistics contribution to artificial intelligence: Where this contribution lies? *arXiv preprint, arXiv:2410.20221*.
- [39] Sridhar, A., Pesala, B., Radhakrishnan, G., Niezgoda, J., & Gopalakrishnan, S. (2025). Evolution of Artificial Intelligence. In *Artificial Intelligence and Biological Sciences* (pp. 26-37). CRC Press.
- [40] Sukaton, O. (2019). Semantics in natural language processing and language teaching. *ELS Journal on Interdisciplinary Studies in Humanities*, 2(1), 58-65.

- [41] Sung, J., & Hopper, J. L. (2023). Co-evolution of epidemiology and artificial intelligence: challenges and opportunities. *International journal of epidemiology*, 52(4), 969-973.
- [42] Tasheva, N. (2024). The Evolution of Modern Linguistics: Key Concepts and Trends. *Medicine, pedagogy and technology: theory and practice*, 2(11), 31-39.
- [43] Tian, Y., Cui, W., Deng, D., Yi, X., Yang, Y., Zhang, H., & Wu, Y. (2024). Chartgpt: Leveraging llms to generate charts from abstract natural language. *IEEE Transactions on Visualization and Computer Graphics*, 31(3), 1731-1745.
- [44] Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2020). Natural language processing advancements by deep learning: A survey. *arXiv preprint arXiv:2003.01200*.
- [45] Triantafyllopoulos, A., & Schuller, B. W. (2024). Expressivity and speech synthesis. *arXiv preprint arXiv:2404.19363*.
- [46] Veeramani, M., Karthick, P., Venkateswaran, S., Sriman, B., Bhanu, S. T., & Devi, V. S. (2025). Transparency in Text: Unraveling Explainability in Healthcare Natural Language Processing. *Explainable Artificial Intelligence in the Healthcare Industry*, 131-160.
- [47] Wallis, P. (2021). Introducing the Talk Markup Language (TalkML): Adding a little social intelligence to industrial speech interfaces. *arXiv preprint arXiv:2105.11294*.
- [48] Xiong, H., Bian, J., Yang, S., Zhang, X., Kong, L., & Zhang, D. (2023). Natural Language based Context Modeling and Reasoning for Ubiquitous Computing with Large Language Models: A Tutorial. *arXiv preprint arXiv:2309.15074*.
- [49] Zafar, H. (2025). Examining the OpenAlex Concepts: A Detailed Case Study of Machine-Derived Classification (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- [50] Zhang, H., Song, H., Li, S., Zhou, M., & Song, D. (2023). A survey of controllable text generation using transformer-based pre-trained language models. *ACM Computing Surveys*, 56(3), 1–37.
- [51] Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, 15(4), 2872.