ISSN: 2957-3874 (Print)
Journal of Al-Farabi for Humanity Sciences (JFHS)
https://www.iasj.net/iasj/journal/419/issues
مجلة الفارابي للعلوم الإنسانية تصدرها كلية الفارابي الجامعة

اللسانيات الحاسوبية وتحديات النمذجة اللغوية : مقاربات في بناء الخوارزميات وتطوير أنظمة التعلم الآلي

د شيماء عبد الكريم
الجامعة العراقية كلية التربية للبنات

المستخلص

يتناول هذا البحث اللسانيات الحاسوبية بوصفها حقلًا معرفيًا بينيًا يجمع بين علوم اللغة وعلوم الحاسوب، ويركّز على النمذجة اللغوية باعتبارها أحد الركائز الأساسية في أنظمة معالجة اللغة الطبيعية وتطبيقات الذكاء الاصطناعي. ويهدف البحث إلى تحليل التحديات التي تواجه النمذجة اللغوية، سواء كانت تحديات لغوية ناتجة عن تعقيد البنية اللغوية وتعدد مستوياتها، أو تحديات تقنية مرتبطة بوفرة البيانات وجودتها ومتطلبات الحوسبة والخوارزميات. كما يستعرض البحث تطور النماذج اللغوية، بدءًا من النماذج الإحصائية التقليدية، مرورًا بالنماذج العصبية، وصولًا إلى النماذج اللغوية الكبيرة القائمة على تقنيات التعلم العميق والمحولات. ويعالج كذلك مقاربات بناء الخوارزميات، من خلال المقارنة بين الأساليب القائمة على القواعد والأساليب المعتمدة على البيانات، مع إبراز دور البيانات الضخمة والتدريب المسبق في تحسين أداء النماذج.وقد خلص البحث إلى أن النمذجة اللغوية، رغم التطور الملحوظ الذي شهدته، لا تزال تواجه تحديات معقدة، خاصة في سياق اللغة العربية، مما يستدعي اعتماد مقاربات تكاملية تجمع بين البعد اللغوي والتقني، وتسهم في تطوير أنظمة أكثر دقة ومرونة في معالجة اللغة الطبيعية.الكلمات المفتاحية اللسانيات الحاسوبية، النمذجة اللغوية، معالجة اللغة الطبيعية، التعلم الآلي، التعلم العميق، البيانات الضخمة، الخوارزميات، اللغة العربية

Abstract

This study examines computational linguistics as an interdisciplinary field that integrates linguistics and computer science, with a particular focus on language modeling as a fundamental component of natural language processing and artificial intelligence systems. The research aims to analyze the challenges facing language modeling, including linguistic challenges related to the complexity and multi-level structure of language, as well as technical challenges associated with data availability, quality, computational resources, and algorithm design.The study also explores the evolution of language models, from traditional statistical models to neural approaches and large-scale language models based on deep learning and transformer architectures. Furthermore, it investigates different approaches to algorithm development, comparing rule-based methods with data-driven techniques, and highlighting the role of big data and pretraining in enhancing model performance.The study concludes that, despite significant advancements, language modeling still faces complex challenges—particularly in the context of the Arabic language—thus requiring integrated approaches that combine linguistic insights with advanced computational techniques to develop more accurate and efficient natural language processing systems.

KeywordsComputational Linguistics, Language Modeling, Natural Language Processing, Machine Learning, Deep Learning, Big Data, Algorithms, Arabic Language

مقدمة

يشهد العصر الراهن ثورة معرفية وتقنية غير مسبوقة، قوامها التطور المتسارع في تقنيات الذكاء الاصطناعي، وتحليل البيانات الضخمة، وأنظمة التعلم الآلي، الأمر الذي أفضى إلى إعادة تشكيل أنماط التفكير الإنساني، وأساليب إنتاج المعرفة وتداولها وفي خضم هذه التحولات، برزت اللغة بوصفها أحد أهم المفاتيح الأساسية لفهم هذا التحول؛ إذ تمثل الوسيط الرئيس الذي تُبنى من خلاله المعارف، وتُنقل به الخبرات، وتُصاغ عبره أنماط التواصل بين الأفراد والمجتمعات.ومن هذا المنطلق، لم تعد اللغة موضوعًا للدراسة الوصفية أو التأملية فحسب، كما كان سائدًا في الدرس اللساني

التقليدي، بل أضحت كيانًا معرفيًا قابلاً للتمثيل الصوري والمعالجة الخوارزمية، مما مهّد لظهور اللسانيات الحاسوبية بوصفها حقلًا علميًا بينيًا يتقاطع فيه اللساني بالحاسوبي، ويهدف إلى نمذجة اللغة الإنسانية تمثيلًا يمكن الآلة من فهمها ومعالجتها. وقد أسهم هذا التلاقي بين علوم اللغة وعلوم الحاسوب في نقل البحث اللغوي من مستوى الوصف والتحليل إلى مستوى المحاكاة والتوليد، حيث أصبحت اللغة تُدرس بوصفها نظامًا يمكن ترميزه، وتحويله إلى بنى رياضية وخوارزمية قابلة للتنفيذ.وفي إطار هذا التحول، برزت النمذجة اللغوية باعتبارها أحد الركائز الأساسية في بناء أنظمة معالجة اللغة الطبيعية، إذ تقوم على تمثيل اللغة في صورة نماذج احتمالية أو عصبية تسعى إلى محاكاة القدرة البشرية على إنتاج اللغة وفهمها. وقد شهدت هذه النماذج تطورًا ملحوظًا عبر مراحل متعددة، بدءًا من النماذج الإحصائية البسيطة القائمة على حساب التكرارات والاحتمالات، مرورًا بالنماذج العصبية التي تعتمد على الشبكات المتكررة، وصولًا إلى النماذج اللغوية الضخمة المعتمدة على بنية "المحوّلات"، والتي أحدثت نقلة نوعية في دقة الأداء وكفاءة المعالجة.غير أن هذا التقدم التقني، على الرغم من أهميته، لا يخلو من إشكالات معرفية ومنهجية عميقة، ترتبط بطبيعة اللغة الإنسانية ذاتها، بما تتسم به من تعقيد وتعدد في المستويات، وتداخل في البنى، وغنى في الدلالات. فاللغة ليست مجرد نسق شكلي يمكن تمثيله بسهولة، بل هي ظاهرة إنسانية مركبة تتداخل فيها الأبعاد الصوتية والصرفية والنحوية والدلالية والتداولية، فضلًا عن ارتباطها بالسياقات الثقافية والاجتماعية والنفسية.وتزداد هذه الإشكالية تعقيدًا في سياق اللغة العربية، التي تتميز بخصائص بنيوية فريدة، من أبرزها غناها الصرفي، وتنوع أنماط اشتقاقها، ومرونتها التركيبية، فضلًا عن ظاهرة الازدواجية اللغوية التي تتجلى في التعايش بين الفصحى والعامية، إلى جانب انتشار الثنائية اللغوية في كثير من المجتمعات العربية. وهذه الخصائص تجعل من عملية تمثيل اللغة العربية حاسوبيًا تحديًا مضاعفًا، يتطلب مقاربات أكثر دقة ومرونة.ومن جهة أخرى، تواجه النمذجة اللغوية تحديات تقنية لا تقل أهمية عن التحديات اللغوية، تتصل بالحاجة إلى كميات ضخمة من البيانات اللغوية المتنوعة، وتطوير خوارزميات قادرة على استيعاب هذا الكم الهائل من المعلومات، فضلًا عن متطلبات الحوسبة العالية التي تفرضها النماذج الحديثة، خاصة تلك القائمة على التعلم العميق. كما أن جودة البيانات وتنوعها تمثلان عاملًا حاسمًا في تحديد كفاءة النماذج، الأمر الذي يبرز أهمية بناء موارد لغوية رقمية دقيقة وشاملة.وفي ضوء ما تقدم، يتضح أن اللسانيات الحاسوبية والنمذجة اللغوية لا تمثلان مجرد أدوات تقنية لمعالجة اللغة، بل تشكلان إطارًا معرفيًا متكاملًا يسعى إلى إعادة بناء العلاقة بين اللغة والآلة، في أفق تحقيق فهم أعمق للظاهرة اللغوية، وتطوير أنظمة ذكية قادرة على محاكاتها بدرجة عالية من الكفاءة. ومن هنا تنبع أهمية هذا البحث، الذي يتناول التحديات التي تواجه النمذجة اللغوية، ويستعرض أبرز المقاربات المعتمدة في بناء الخوارزميات وتطوير أنظمة التعلم الآلي، سعيًا إلى الإسهام في تطوير هذا المجال الحيوي.

مشكلة البحث

على الرغم من التطور المتسارع الذي شهدته اللسانيات الحاسوبية، وما أفرزته من نماذج لغوية متقدمة قادرة على معالجة اللغة الطبيعية بدرجات عالية من الدقة، فإن النمذجة اللغوية لا تزال تواجه إشكاليات معقدة تحدّ من قدرتها على تحقيق تمثيل شامل ودقيق للغة الإنسانية. وتكمن جوهر هذه الإشكالية في الفجوة القائمة بين الطبيعة المركبة للغة بوصفها ظاهرة إنسانية غنية بالسياقات والدلالات، وبين الطابع الصوري والرياضي الذي تعتمد عليه النماذج الحاسوبية في تمثيلها.وتتجلى هذه المشكلة بشكل أوضح في سياق اللغة العربية، التي تتميز بخصائص لغوية معقدة، من حيث غناها الصرفي، وتنوع بنياتها التركيبية، وتعدد مستوياتها الاستعمالية، فضلًا عن ظاهرتي الازدواجية والثنائية اللغوية، وما تفرزانه من تداخل في الأنظمة اللغوية واختلاط في الدلالات، وهو ما يجعل من عملية نمذجة هذه اللغة تحديًا يتجاوز حدود المعالجة التقنية إلى إشكال معرفي ومنهجي يتعلق بمدى إمكانية تمثيل اللغة تمثيلًا حسابيًا دقيقًا.كما أن النماذج التقليدية القائمة على القواعد أثبتت محدوديتها في استيعاب التنوع اللغوي والسياقي، في حين أن النماذج الحديثة المعتمدة على التعلم الآلي والتعلم العميق – على الرغم من كفاءتها – تظل رهينة بوفرة البيانات وجودتها، وتعاني من صعوبات في تفسير مخرجاتها، فضلًا عن حاجتها إلى موارد حاسوبية ضخمة.وانطلاقًا من ذلك، تتمحور مشكلة البحث حول الإشكال الرئيس الآتي:

إلى أي مدى يمكن للنمذجة اللغوية، في ظل التحديات اللغوية والتقنية، أن تحقق تمثيلًا دقيقًا وفعالًا للغة الطبيعية – ولا سيما اللغة العربية – من خلال مقاربات بناء الخوارزميات وأنظمة التعلم الآلي؟

ويتفرع عن هذا الإشكال الرئيس مجموعة من التساؤلات الفرعية، منها:

ما أبرز التحديات اللغوية التي تعيق بناء نماذج لغوية دقيقة؟

١. كيف تؤثر الثنائية اللغوية والازدواجية في كفاءة النماذج الحاسوبية؟
٢. ما حدود المقاربات التقليدية في معالجة اللغة، وما أوجه القصور فيها؟
٣. كيف أسهمت تقنيات التعلم الآلي والتعلم العميق في تجاوز هذه الإشكالات؟
٤. ما دور البيانات الضخمة والتدريب المسبق في تحسين أداء النماذج اللغوية؟

أهمية البحث

تنبع أهمية هذا البحث من كونه يتناول موضوعًا يقع في صميم التفاعل المعاصر بين اللغة والتكنولوجيا، حيث تمثل اللسانيات الحاسوبية والنمذجة اللغوية ركيزة أساسية في تطوير تطبيقات الذكاء الاصطناعي ومعالجة اللغة الطبيعية. وتبرز أهمية البحث في عدة جوانب، من أهمها:

الأهمية العلمية: يسهم البحث في تعميق الفهم النظري لمجال النمذجة اللغوية، من خلال تحليل التحديات اللغوية والتقنية التي تواجهه، وربطها بالمقاربات الحديثة في بناء الخوارزميات.

الأهمية التطبيقية: يساعد في تطوير أنظمة لغوية أكثر كفاءة، خاصة في مجالات مثل الترجمة الآلية، والتعرف على الكلام، وتحليل النصوص، وهو ما ينعكس على تحسين التطبيقات الذكية.

الأهمية اللغوية: يسلط الضوء على خصوصية اللغة العربية والتحديات المرتبطة بمعالجتها حاسوبيًا، مما يسهم في دعم الجهود الرامية إلى تطوير موارد لغوية رقمية عربية.

أهداف البحث

يسعى هذا البحث إلى تحقيق مجموعة من الأهداف العلمية والمنهجية، من أبرزها:

التعرف على مفهوم اللسانيات الحاسوبية والنمذجة اللغوية، وبيان تطورهما.

تحليل أبرز التحديات اللغوية التي تواجه النمذجة اللغوية، وخاصة في اللغة العربية.

الكشف عن التحديات التقنية المرتبطة ببناء النماذج اللغوية وتدريبها.

دراسة المقاربات المختلفة في بناء الخوارزميات، من النماذج التقليدية إلى التعلم الآلي والتعلم العميق.

بيان دور البيانات الضخمة والتدريب المسبق في تحسين أداء النماذج اللغوية.

الوصول إلى تصور علمي يسهم في تطوير نماذج لغوية أكثر دقة ومرونة.

منهجية البحث

يعتمد هذا البحث على منهج علمي تكاملي يجمع بين أكثر من مقاربة منهجية، وذلك على النحو الآتي:

المنهج الوصفي التحليلي: من خلال عرض المفاهيم الأساسية للسانيات الحاسوبية والنمذجة اللغوية، وتحليل التحديات التي تواجهها.

المنهج المقارن: عبر المقارنة بين المقاربات التقليدية في بناء الخوارزميات والمقاربات الحديثة القائمة على التعلم الآلي.

المنهج الاستقرائي: من خلال تتبع تطور النماذج اللغوية واستخلاص السمات العامة لكل مرحلة.

المنهج التطبيقي (ضمني): من خلال ربط الجوانب النظرية بالتطبيقات العملية في معالجة اللغة الطبيعية.

حدود البحث

الحدود الموضوعية: يقتصر البحث على دراسة النمذجة اللغوية في إطار اللسانيات الحاسوبية، مع التركيز على التحديات اللغوية والتقنية ومقاربات بناء الخوارزميات.

خطة البحث

ينتظم البحث في مبحثين رئيسين:

المبحث الأول: الإطار المفاهيمي للّسانيات الحاسوبية والنمذجة اللغوية

المطلب الأول: مفهوم اللسانيات الحاسوبية ونشأتها

المطلب الثاني: مفهوم النمذجة اللغوية وأنواعها

المبحث الثاني: تحديات النمذجة اللغوية ومقاربات بناء الخوارزميات

المطلب الأول: تحديات النمذجة اللغوية (اللغوية والتقنية)

المطلب الثاني: مقاربات بناء الخوارزميات وتطوير أنظمة التعلم الآلي

المبحث الأول: اللسانيات الحاسوبية والنمذجة اللغوية: المفاهيم والنشأة والتطور

المطلب الأول: مفهوم اللسانيات الحاسوبية ونشأتها

تُعدّ اللسانيات الحاسوبية من أبرز الحقول المعرفية البينية التي نشأت في سياق التفاعل المتزايد بين علوم اللغة وعلوم الحاسوب، حيث لم يعد النظر إلى اللغة مقتصراً على بعدها الوصفي أو التأملي، بل أضحت موضوعاً للتمثيل الصوري والمعالجة الخوارزمية. وفي هذا الإطار، يمكن تعريف اللسانيات الحاسوبية بأنها ذلك العلم الذي يُعنى بتوظيف الحاسوب وتقنيات المعالجة الآلية في دراسة اللغة الإنسانية، من خلال بناء نماذج قادرة على تمثيل بنيتها وتحليلها، بما يتيح للآلة أداء مهام لغوية معقدة، كالفهم، والتوليد، والترجمة، وتمييز الكلام.ولا يقتصر هذا الحقل على الجانب التطبيقي فحسب، بل يمتد ليشمل أبعاداً نظرية عميقة تتصل بطبيعة المعرفة اللغوية ذاتها. وفي هذا السياق، يذهب نهاد الموسى إلى أن اللسانيات الحاسوبية تمثل نظاماً معرفياً بينياً يتقاطع فيه اللساني مع الحاسوبي، ويُعنى بحوسبة جوانب الملكة اللغوية، من خلال مكونين رئيسيين: أحدهما تطبيقي يهدف إلى إنتاج برامج لغوية تحاكي الاستعمال الإنساني للغة، والآخر نظري يتصل ببناء نماذج صورية للمعرفة اللغوية التي تمكّن الإنسان من إنتاج اللغة وفهمها[١].ويكشف هذا التصور عن طبيعة مزدوجة لهذا الحقل، إذ يجمع بين السعي إلى تحقيق الكفاءة الإجرائية من جهة، ومحاولة تفسير البنية العميقة للغة من جهة أخرى، الأمر الذي يضعه في قلب الإشكال المعرفي المتعلق بإمكانية اختزال الظاهرة اللغوية ضمن نماذج حسابية.أما من حيث النشأة، فقد ارتبط ظهور اللسانيات الحاسوبية بتطور الحاسوب ذاته، إذ بدأت تطبيقاته الأولى في منتصف القرن العشرين ذات طابع رقمي صرف، حيث اقتصر استخدامه على معالجة البيانات العددية في المجالات التجارية. غير أن التحول النوعي الذي شهده الحاسوب خلال سبعينات القرن الماضي، حين أصبح قادراً على تخزين المعلومات واسترجاعها ومعالجتها، مهّد لظهور محاولات أولية للتعامل مع اللغة بوصفها معطى قابلاً للمعالجة الآلية وقد تعزز هذا الاتجاه بشكل لافت في ثمانينات القرن العشرين، مع الانتقال من معالجة المعلومات إلى معالجة المعرفة، حيث دخل الحاسوب في مواجهة مباشرة مع منظومة اللغة، بوصفها الأداة الأساسية لإنتاج المعرفة ونقلها. وقد فرضت هذه المواجهة تطويراً ملحوظاً في قدرات الحاسوب، سواء من حيث السرعة أو سعة التخزين أو أساليب البرمجة، بما مكّنه من التعامل مع البنى اللغوية الأكثر تعقيدا[٢].وفي سياق تطور هذا الحقل، اعتمدت اللسانيات الحاسوبية في بداياتها على التحليل الإحصائي للمفردات والتراكيب، قبل أن تنتقل إلى بناء معاجم إلكترونية وقواعد بيانات لغوية، شكّلت الأساس لتطوير تطبيقات أكثر تقدماً، كالترجمة الآلية وأنظمة التعرف على الكلام. وقد أسهم هذا التطور في تقليص الزمن اللازم لمعالجة اللغة، وفتح آفاق جديدة أمام توظيف الحاسوب في الدراسات اللغوية[٣].أما على مستوى اللغة العربية، فقد شهدت اللسانيات الحاسوبية تطوراً تدريجياً، بدأ بمبادرات فردية، ثم اتجه نحو الطابع المؤسسي، مع بروز الحاجة إلى تضافر جهود اللغويين والحاسوبيين لتطوير أنظمة قادرة على معالجة الخصائص البنيوية المعقدة للعربية. وقد أكدت دراسات عدة على ضرورة بناء معاجم حاسوبية، وتوحيد المصطلحات، وترجمة الأعمال العلمية، فضلاً عن إدماج هذا الحقل ضمن المناهج الأكاديمية[٤]ومع ذلك، لا تزال معالجة اللغة العربية حاسوبياً تواجه تحديات متعددة، تتصل بطبيعتها الصرفية الغنية، وتعدد مستوياتها، واتساع الفجوة بين الفصحى واللهجات، فضلاً عن الإشكالات المرتبطة بتمثيل الأصوات والرموز الكتابية. الأمر الذي يجعل من اللسانيات الحاسوبية مجالاً مفتوحاً على مزيد من البحث والتطوير، خاصة في ظل التحولات المتسارعة التي يشهدها عصر المعلومات.وبذلك، يتضح أن اللسانيات الحاسوبية لم تعد مجرد تطبيق تقني على اللغة، بل أصبحت إطاراً نظرياً وتطبيقياً يسعى إلى إعادة صياغة العلاقة بين الإنسان واللغة والآلة، في أفق تحقيق فهم أعمق للظاهرة اللغوية، وتطوير أنظمة قادرة على محاكاتها بدرجة عالية من الكفاءة.المطلب الثاني: مفهوم النمذجة اللغوية وأنواعها

أولًا: تعريف النمذجة اللغوية

في اللغة:

يرتبط مفهوم النمذجة اللغوية بمصطلح "النموذج" الذي يدل في أصل وضعه اللغوي على معنى المثال والتمثيل فقد ورد في القاموس المحيط للفيروزآبادي (ت ٨١٧هـ) أن «النموذج: مثال الشيء»[٥]، وهو ما يشير إلى كونه صورة يُحتذى بها ويُقاس عليها.

كما جاء في المعجم الوسيط، من تأليف نخبة من علماء مجمع اللغة العربية بالقاهرة، أن «الأنموذج (النموذج): مثال الشيء، وهو معرّب من الفارسية»[٦]، مما يعزز دلالته على التمثيل والتقريب.

وفي معجم اللغة العربية المعاصرة لأحمد مختار عمر (ت ١٤٢٤هـ) عُرِّف النموذج بأنه «ما يُتخذ مثالًا يُحتذى به، أو وصف أو نهج يُستعان به في توضيح ما لا يمكن إدراكه مباشرة»[٧]، وهو تعريف يبرز البعد الوظيفي للنموذج في التفسير والتوضيح.

ومن خلال هذه التعريفات يتبين أن مفهوم "النموذج" يقوم على محاكاة الواقع أو تمثيله، وهو ما يُعد أساسًا علميًا لظهور النمذجة اللغوية التي تهدف إلى تمثيل اللغة وتحليلها باستخدام صيغ وقواعد منظمة.

في الاصطلاح

يتسع مفهوم "النموذج" في الاصطلاح ليأخذ بعدًا منهجيًا يتجاوز دلالته اللغوية، إذ يُقصد به الصورة المعيارية أو المثال الذي يُحتذى به، ويُرجع إليه في التقويم والقياس. فالنموذج يمثل إطارًا مرجعيًا تُقاس عليه الظواهر أو الأشياء التي تنتمي إلى جنسه، بما يجعله أداة منظمة للفهم والتوجيه وعلى هذا الأساس، يمكن النظر إلى النموذج بوصفه تمثيلًا مثاليًا لشيء معين، يُسترشد به في إنتاج ما يماثله؛ فـ"نموذج الكتاب" – على سبيل المثال – لا يقتصر على كونه صورة مادية، بل قد يكون تصورًا ذهنيًا يُستحضر عند التأليف، ويُقاس عليه في بناء العمل وتقويمه، سواء تم الرجوع إليه بصورة مباشرة أم على نحو اعتباري قائم في الذهن[٨].

انطلاقًا من مفهوم النموذج، تُعدّ النمذجة عملية معرفية ومنهجية تقوم على تمثيل الظواهر أو نقل الخبرات من خلال بناء نموذج يُحتذى به فهي استراتيجية تُستخدم لتبسيط المعارف وتوضيحها، عبر تقديم صورة منظمة يمكن محاكاتها أو القياس عليها تعتمد النمذجة في جوهرها على آليتي التقليد والمحاكاة، حيث يُقدَّم النموذج بوصفه بنية مرجعية تُنقل من خلالها الخبرة من فرد إلى آخر أو من إطار نظري إلى تطبيق عملي وقد أدى اتساع استخدام هذا المفهوم في مجالات متعددة – كالتربية واللسانيات وعلوم الحاسوب – إلى تعدد تعريفاته وتنوع مقارباته، الأمر الذي يعكس ثراءه الاصطلاحي ومرونته في التوظيف داخل السياقات العلمية المختلفة، ولا سيما في الدراسات اللسانية التي تتخذ من النمذجة وسيلة لتحليل اللغة وتمثيل بنياتها[٩].

النماذج اللغوية

في سياق اللسانيات الحاسوبية، يُقصد بالنموذج اللغوي ذلك النموذج الذي يعمل على تمثيل اللغة تمثيلًا رياضيًا أو إحصائيًا، حيث يُسند احتمالات مختلفة لتتابعات الكلمات داخل اللغة. وقد عُرِّف النموذج اللغوي بأنه نموذج إحصائي يُعطي قيمة احتمالية لكل سلسلة من الكلمات اعتمادًا على التوزيع الاحتمالي[١٠].

وتُستخدم النماذج اللغوية في العديد من تطبيقات معالجة اللغة الطبيعية، مثل التعرف على الكلام، والترجمة الآلية، وتحليل النصوص، واسترجاع المعلومات، وهو ما يعكس أهميتها في تطوير الأنظمة الذكية كما يُنظر إلى نموذج اللغة بوصفه محاولة لمحاكاة القدرة البشرية على إنتاج اللغة الطبيعية وفهمها، حيث يسعى إلى تمثيل الأنماط اللغوية التي يستخدمها الإنسان في التواصل، مما يجعله أحد الركائز الأساسية في أنظمة الذكاء الاصطناعي الحديثة[١١].

أهمية النماذج اللغوية وتطورها

تتجلى أهمية النماذج اللغوية في كونها أداة أساسية تُوظَّف في طيف واسع من مهام معالجة اللغة الطبيعية، حيث تُستخدم في التعرف على الكلام، والترجمة الآلية، وتوليد اللغة الطبيعية، بما يُسهم في إنتاج نصوص تحاكي النصوص البشرية من حيث البنية والأسلوب[١٢]. كما تدخل في تطبيقات أخرى مثل التعرف الضوئي على الحروف، والتعرف على الكتابة اليدوية، فضلًا عن استقراء القواعد النحوية واسترجاع المعلومات[١٣].

وقد شهدت النماذج اللغوية تطورًا ملحوظًا عبر مراحل متعددة؛ إذ بدأت بالنماذج الإحصائية البحتة، مثل نماذج السلاسل (N–gram) التي تعتمد على حساب احتمالات تتابع الكلمات، ثم تطورت إلى النماذج العصبية المعتمدة على الشبكات العصبية المتكررة(RNNs) ، والتي أتاحت تمثيلًا أعمق للسياق اللغوي[١٤].

وفي المرحلة الأحدث، برزت النماذج اللغوية الكبيرة (LLMs) بوصفها أكثر النماذج تقدمًا، حيث تعتمد على بنية "المحوّلات (Transformers) " المدربة على كميات ضخمة من البيانات النصية، غالبًا ما تُستمد من مصادر متنوعة مثل شبكة الإنترنت. وقد أسهم هذا التطور في تحقيق نقلة نوعية في أداء النماذج، متجاوزة بذلك قدرات النماذج السابقة، سواء الإحصائية أو العصبية التقليدية[١٥].

ويُلاحظ أن هذا التطور يعكس مسارًا تصاعديًا في القدرة على تمثيل اللغة البشرية، والاقتراب من محاكاة خصائصها المعقدة، وهو ما جعل النماذج اللغوية ركيزة أساسية في تطبيقات الذكاء الاصطناعي الحديثة.

المبحث الثاني: إشكاليات النمذجة اللغوية وآليات تطوير الخوارزميات في اللسانيات الحاسوبية

المطلب الأول: تحديات النمذجة اللغوية

أولًا: التحديات اللغوية تُعدّ الثنائية اللغوية من أبرز الظواهر التي تُسهم في تعقيد عملية النمذجة اللغوية، لما تفرضه من تداخل بين الأنظمة اللغوية المختلفة داخل الفرد أو المجتمع. فاللغة – كما عرّفها ابن جني – «أصوات يُعبّر بها كل قوم عن أغراضهم»[١٦]، وهو تعريف يبرز بعدها الوظيفي والاجتماعي.أما الثنائية اللغوية، فهي ظاهرة تشير إلى استخدام لغتين مختلفتين في سياق واحد، سواء على مستوى الفرد أو المجتمع، حيث يتعايش نظامان لغويان أو أكثر داخل البيئة اللغوية الواحدة[١٧]، كما تُفهم أيضًا بوصفها قدرة الفرد على استعمال لغتين بدرجات متفاوتة من الكفاءة[١٨].وقد ميّز الباحثون بين الثنائية اللغوية والازدواجية (Diglossia)، حيث تشير الأولى – في الاصطلاح الشائع – إلى استخدام لغتين مختلفتين، بينما تدل الثانية على وجود مستويين لغويين داخل اللغة الواحدة، كالفصحى والعامية[١٩].

وتظهر آثار الثنائية اللغوية في مستويات متعددة، منها المستوى الصوتي، حيث يحدث تداخل في النطق والتنغيم، والمستوى النحوي الذي يتمثل في اضطراب ترتيب الجملة واستخدام الضمائر، والمستوى المفرداتي والدلالي الذي يتجلى في اقتراض الألفاظ واختلاط المعاني بين اللغتين[٢٠].ومن ثم، تُشكّل هذه الظاهرة تحديًا رئيسًا أمام النمذجة اللغوية، لما تسببه من غموض دلالي، وتعدد في المعاني، وصعوبة في بناء نماذج دقيقة قادرة على تمثيل اللغة تمثيلًا صحيحًا.

ثانيًا: التحديات التقنية

إلى جانب التحديات اللغوية، تواجه النمذجة اللغوية مجموعة من التحديات التقنية، أبرزها الحاجة إلى كميات ضخمة من البيانات اللغوية المتنوعة، فضلًا عن متطلبات الحوسبة العالية اللازمة لمعالجة هذه البيانات وتدريب النماذج عليها،كما أن التداخل اللغوي الناتج عن الثنائية اللغوية يزيد من صعوبة تدريب النماذج، إذ يصعب على الأنظمة التمييز بين الأنماط اللغوية المختلفة، مما يؤثر في دقة النتائج وكفاءة الأداء(٢١). ومن ثم، تتطلب النمذجة اللغوية تطوير خوارزميات أكثر قدرة على التعامل مع هذا التعقيد اللغوي والتنوع في البيانات.

المطلب الثاني: مقاربات بناء الخوارزميات وتطوير أنظمة التعلم الآلي تُعدّ مقاربات بناء الخوارزميات في اللسانيات الحاسوبية من الركائز الأساسية التي يقوم عليها تطوير أنظمة المعالجة الآلية للغة، إذ شهد هذا المجال تحولًا نوعيًا من الاعتماد على القواعد اللغوية الصارمة إلى توظيف النماذج الإحصائية وأساليب التعلم الآلي، مدعومة بتطور الموارد الرقمية والبيانات الضخمة، وقد أسهم هذا التحول في تعزيز قدرة الأنظمة على تحليل اللغة الطبيعية وفهمها بشكل أكثر دقة ومرونة.

أولًا: الخوارزميات التقليدية مقابل التعلم الآلي

اعتمدت المقاربات التقليدية في معالجة اللغة الطبيعية على بناء خوارزميات قائمة على القواعد اللغوية، حيث يتم تمثيل الظواهر اللغوية في صورة قواعد نحوية وصرفية دقيقة تُطبَّق على النصوص باستخدام موارد ثابتة، مثل المعاجم والقواميس الحاسوبية. ويُعد هذا النهج قائمًا على تحليل هيكلي للغة يعتمد على المعرفة اللغوية المسبقة، ويتطلب تدخلًا بشريًا مباشرًا من الخبراء في صياغة القواعد وضبطها[٢٢].

غير أن هذه المقاربة واجهت صعوبات عديدة، أبرزها محدودية قدرتها على التعامل مع التنوع اللغوي والتغيرات السياقية، فضلًا عن تعقيد بناء قواعد شاملة تغطي جميع الظواهر اللغوية. ومع تطور اللسانيات الحاسوبية، ظهر اتجاه جديد يعتمد على المناهج الكمية والإحصائية، حيث يتم تدريب الخوارزميات على كميات كبيرة من البيانات لاستخلاص الأنماط اللغوية بشكل آلي، دون الحاجة إلى تحديد القواعد بشكل صريح[٢٣].وقد أدى هذا التحول إلى تقليل الاعتماد على القواعد اللغوية الصارمة، واستبدالها بنماذج تعتمد على الاحتمالات والتكرارات المستخلصة من المدونات اللغوية، مما أتاح للأنظمة قدرة أكبر على التكيف مع اللغة في سياقاتها المختلفة.

ثانيًا: التعلم العميق والشبكات العصبية

مع تطور التعلم الآلي، برز التعلم العميق بوصفه أحد أهم المقاربات الحديثة في بناء الخوارزميات اللغوية، حيث يعتمد على الشبكات العصبية الاصطناعية متعددة الطبقات، التي تحاكي في بنيتها طريقة عمل الدماغ البشري. ومن أبرز هذه النماذج: الشبكات العصبية المتكررة (RNN)، التي تُستخدم لمعالجة البيانات المتسلسلة، ونماذج المحولات (Transformers) التي أحدثت نقلة نوعية في فهم السياق اللغوي[٢٤].

وتتميز هذه النماذج بقدرتها على استخراج السمات اللغوية تلقائيًا من البيانات، دون الحاجة إلى تدخل بشري في تحديد الخصائص أو القواعد، كما أنها قادرة على التعامل مع النصوص الطويلة وفهم العلاقات السياقية المعقدة. وقد أدى ذلك إلى تحسين أداء العديد من تطبيقات معالجة اللغة الطبيعية، مثل الترجمة الآلية، والتلخيص، وتحليل المشاعر.

كما أن التعلم العميق يعتمد على بنية طبقية (Layers)، حيث تُعالج البيانات عبر مستويات متعددة من التمثيل، مما يسمح باستخلاص خصائص لغوية دقيقة، ويُسهم في زيادة دقة النماذج كلما ازداد عمقها وعدد طبقاتها[٢٥].

ثالثًا: دور البيانات الضخمة (Big Data)

تمثل البيانات الضخمة عنصرًا حاسمًا في بناء النماذج اللغوية الحديثة، إذ تعتمد الخوارزميات المتقدمة على كميات هائلة من البيانات النصية التي يتم جمعها من مصادر متعددة، مثل الكتب، والمقالات، ومواقع الويب، ووسائل التواصل الاجتماعي.

وتُعرف البيانات الضخمة بأنها مجموعة من البيانات التي يصعب معالجتها باستخدام الأساليب التقليدية، نظرًا لضخامتها وتنوعها وسرعة تدفقها، مما استدعى تطوير أنظمة وتقنيات جديدة لمعالجتها وتحليلها[٢٦].

وقد أسهمت هذه البيانات في تحسين أداء النماذج اللغوية، حيث تعتمد الخوارزميات الحديثة على تحليل التكرارات والأنماط الإحصائية داخل هذه البيانات، مما يساعد في بناء نماذج أكثر دقة في التنبؤ وفهم السياق. كما أن توافر المدونات اللغوية الحاسوبية (Corpora) شكّل قاعدة أساسية لتدريب هذه النماذج، حيث توفر تمثيلًا واقعيًا للغة في استعمالاتها المختلفة[٢٧].

رابعًا: التدريب المسبق (Pretraining) وتحسين الأداء

يُعد التدريب المسبق من أهم التقنيات التي أحدثت تطورًا كبيرًا في مجال النمذجة اللغوية، حيث يتم تدريب النماذج على كميات ضخمة من البيانات العامة قبل توجيهها لأداء مهام محددة. ويتيح هذا الأسلوب للنماذج اكتساب معرفة لغوية شاملة، تشمل القواعد والسياقات والعلاقات الدلالية بين الكلمات.ومن خلال هذه المرحلة، تصبح النماذج قادرة على التكيف مع مهام مختلفة، مثل الترجمة الآلية، وتحليل النصوص، واسترجاع المعلومات، دون الحاجة إلى تدريب كامل من الصفر. كما يُسهم التدريب المسبق في تحسين دقة النتائج وتقليل الحاجة إلى بيانات تدريبية خاصة بكل مهمة[٢٨].

الخاتمة

في ختام هذا البحث، يتبين أن اللسانيات الحاسوبية تمثل أحد أبرز الحقول المعرفية التي تجسد التفاعل العميق بين اللغة والتكنولوجيا، حيث أسهمت في نقل دراسة اللغة من إطارها الوصفي التقليدي إلى مجال المعالجة الآلية والنمذجة الصورية. وقد شكّلت النمذجة اللغوية ركيزة أساسية في هذا التحول، لما توفره من أدوات قادرة على تمثيل اللغة وتحليلها والتنبؤ بأنماطها في سياقات متعددة.

وقد سعى هذا البحث إلى استكشاف التحديات التي تواجه النمذجة اللغوية، سواء على المستوى اللغوي أو التقني، حيث تبيّن أن تعقيد البنية اللغوية، وتعدد مستوياتها، ووجود ظواهر كالثنائية اللغوية والازدواجية، تمثل عوائق حقيقية أمام بناء نماذج دقيقة. كما كشفت الدراسة عن تحديات تقنية تتعلق بوفرة البيانات وجودتها، ومتطلبات الحوسبة العالية، وتعقيد تصميم الخوارزميات القادرة على استيعاب هذا التنوع اللغوي.

وفي المقابل، أبرز البحث التطور الكبير الذي شهدته مقاربات بناء الخوارزميات، بدءًا من النماذج التقليدية القائمة على القواعد، وصولًا إلى النماذج الحديثة المعتمدة على التعلم الآلي والتعلم العميق، والتي أسهمت في تحسين أداء أنظمة معالجة اللغة الطبيعية بشكل ملحوظ. كما أظهر الدور المحوري للبيانات الضخمة وتقنيات التدريب المسبق في تعزيز كفاءة النماذج اللغوية، وتمكينها من التعامل مع السياقات اللغوية المعقدة.ومع ذلك، يظل تحقيق تمثيل دقيق وشامل للغة – ولا سيما اللغة العربية – هدفًا يتطلب مزيدًا من البحث والتطوير، من خلال تبني مقاربات تكاملية تجمع بين المعرفة اللغوية العميقة والتقنيات الحاسوبية المتقدمة. كما تبرز الحاجة إلى تطوير موارد لغوية رقمية عربية، وتعزيز التعاون بين اللغويين والحاسوبيين، بما يسهم في بناء أنظمة أكثر كفاءة ومرونة.وبناءً على ذلك، يوصي البحث بضرورة توجيه الجهود المستقبلية نحو:تطوير مدونات لغوية عربية غنية ومتنوعة.

تحسين جودة البيانات المستخدمة في تدريب النماذج.

دعم الأبحاث البينية التي تجمع بين اللسانيات وعلوم الحاسوب.

توظيف تقنيات الذكاء الاصطناعي الحديثة في معالجة خصوصيات اللغة العربية.

وفي ضوء هذه النتائج، يمكن القول إن اللسانيات الحاسوبية والنمذجة اللغوية تمثلان مجالًا واعدًا، يحمل إمكانات كبيرة لتطوير أنظمة ذكية قادرة على فهم اللغة الإنسانية ومحاكاتها، بما يعزز من دور اللغة في عصر المعرفة الرقمية.

المراجع

الأدب المقارن (مرحلة الماجستير)، مناهج جامعة المدينة العالمية، إعداد: جامعة المدينة العالمية، جامعة المدينة العالمية.

التفاعل بين الإنسان والآلة في الترجمة الحاسوبية، سلمان الواسطي، دار الفكر.

توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية: جهود ونتائج، عبد الرحمن العارف، ملحق الثقافية (جريدة الجزيرة)، العدد ٣٥٣، ١٤٣٢هـ.

الثنائية اللغوية وأثرها في المجتمع، محمود، دار الفكر، الطبعة الأولى، ٢٠٠٢م.

الخصائص، أبو الفتح عثمان بن جني، تحقيق: محمد علي النجار، الهيئة المصرية العامة للكتاب، القاهرة، الطبعة الرابعة، ٢٠٠٦م.

حصاد القرن في اللسانيات، نهاد الموسى، دار الشروق، عمّان، ٢٠٠٠م.

اللغة العربية والحاسوب أو حوسبة اللغة العربية، نعيمة بوزيدي، دار الكتاب الحديث، القاهرة.

المعجم الوسيط، نخبة من علماء مجمع اللغة العربية بالقاهرة، مجمع اللغة العربية، القاهرة، الطبعة الثانية، ١٩٧٢م.

معجم البيانات والذكاء الاصطناعي (عربي–إنجليزي)، الهيئة السعودية للبيانات والذكاء الاصطناعي، ٢٠٢٢م.

معجم اللغة العربية المعاصرة، أحمد مختار عمر، عالم الكتب، القاهرة، الطبعة الأولى، ٢٠٠٨م.

الموارد اللغوية الحاسوبية، محسن رشوان وآخرون، مركز الملك عبد الله بن عبد العزيز، الرياض، الطبعة الأولى، ٢٠١٩م.

برنامج قائم على إستراتيجية النمذجة لعلاج مشكلات التحدث الناتجة عن الثنائية اللغوية لدى المرحلة الإعدادية بمدارس اللغات، صابر علام عثمان علام، المجلة العلمية لكلية التربية، جامعة أسيوط، ٢٠٢١م.

القاموس المحيط، مجد الدين محمد بن يعقوب الفيروزآبادي، تحقيق: مؤسسة الرسالة، بيروت، الطبعة الثامنة، ٢٠٠٥م.

مدخل إلى التربية الخاصة، قحطان أحمد الظاهر، دار وائل للنشر، عمّان، الطبعة الثالثة، ٢٠١٥م.

مدخل إلى اللسانيات الحاسوبية، عبد الله بن يحيى الفيفي، مركز الملك عبد الله بن عبد العزيز الدولي، الرياض، الطبعة الأولى، ٢٠١٧م.

Blank, I. A. ،What Are Language Models Supposed to Model?، مجلة Trends in Cognitive Sciences، المجلد ٢٧، العدد ١١، ٢٠٢٣م.

Benoît Habert ،Outiller la linguistique : de l’emprunt de techniques aux rencontres de savoirs ،Revue française de linguistique appliquée ،vol. IX ،n°١، ٢٠٠٤.

Corpus Linguistics at Work ،Elena Tognini-Bonelli ،John Benjamins ،Amsterdam، ٢٠٠١.

Jurafsky, D. & Martin, J. H. ،Speech and Language Processing (٣rd ed.)، ٢٠٢١م.

Language Processing with Perl and Prolog ،Pierre M. Nugues ،Springer ،London، ٢٠١٤.

Ludovic Tanguy & Cécile Fabre ،Évolutions de la linguistique outillée : méfaits et bienfaits du TAL ،Revue l’Information Grammaticale، ٢٠١٤.

Ponte, J. M. & Croft, W. B. ،A Language Modeling Approach to Information Retrieval، مؤتمر ACM SIGIR، ١٩٩٨م.

Simplifying the Difference: Machine Learning vs Deep Learning ،SCS، ٢٠٢٠

---

(١) حصاد القرن في اللسانيات، نهاد الموسى، دار الشروق، عمّان، ٢٠٠٠، ص ٤٧.

(٢) اللغة العربية والحاسوب أو حوسبة اللغة العربية، نعيمة بوزيدي، دار الكتاب الحديث، القاهرة، ص ١٩٦.

(٣) التفاعل بين الإنسان والآلة في الترجمة الحاسوبية، سلمان الواسطي، دار الفكر، ص ٧.

(٤) . توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية: جهود ونتائج، عبد الرحمن العارف، ملحق الثقافية (جريدة الجزيرة)، العدد ٣٥٣، ١٤٣٢هـ.

(٥) القاموس المحيط، مجد الدين محمد بن يعقوب الفيروزآبادي (ت ٨١٧هـ)، تحقيق: مكتب تحقيق التراث، مؤسسة الرسالة، بيروت، ط٨، ١٤٢٦هـ/٢٠٠٥م، ص ٢٠٨.

(٦) المعجم الوسيط، نخبة من علماء مجمع اللغة العربية بالقاهرة، مجمع اللغة العربية، القاهرة، ط٢، ١٩٧٢م، ج٢، ص ٩٥٦.

(٧) معجم اللغة العربية المعاصرة، أحمد مختار عمر (ت ١٤٢٤هـ)، عالم الكتب، القاهرة، ط١، ١٤٢٩هـ/٢٠٠٨م، ج٣، ص ٢٢٨٤.

(٨) الأدب المقارن ، مناهج جامعة المدينة العالمية، إعداد: جامعة المدينة العالمية، الناشر: جامعة المدينة العالمية، ص ٤٥٣.

(٩) برنامج قائم على إستراتيجية النمذجة لعلاج مشكلات التحدث الناتجة عن الثنائية اللغوية لدى المرحلة الإعدادية بمدارس اللغات، صابر علام عثمان علام، المجلة العلمية لكلية التربية، جامعة أسيوط، المجلد ٣٧، العدد ١٠، أكتوبر ٢٠٢١م، ص ١٠.

(١٠) معجم البيانات والذكاء الاصطناعي (عربي–إنجليزي)، الهيئة السعودية للبيانات والذكاء الاصطناعي، ٢٠٢٢م، ص ٨١.

(١١) Blank, I. A ، What Are Language Models Supposed to Model?، مجلة Trends in Cognitive Sciences، المجلد ٢٧، العدد ١١، ٢٠٢٣م، ص ٩٨٧.

(١٢) Blank, I. A. ، . What Are Language Models Supposed to Model?، مجلة Trends in Cognitive Sciences، المجلد ٢٧، العدد ١١، نوفمبر ٢٠٢٣م، ص ٩٨٧.

(١٣) Ponte, J. M. & Croft, W. B. ، . A Language Modeling Approach to Information Retrieval، مؤتمر ACM SIGIR ، ١٩٩٨م، ص ٢٧٥.

(١٤) Jurafsky, D. & Martin, J. H ،rd ed.)٣ . Speech and Language Processing (.، فصل: N–gram Language Models، ٢٠٢١م، ص ٥٧٢

(١٥) Jurafsky, D. & Martin, J. H. ، . Speech and Language Processing (3rd ed.)، فصل N–gram : Language Models، 2021م، ص ٥٧١.

(١٦) الخصائص، أبو الفتح عثمان بن جني (ت ٣٩٢هـ)، تحقيق: محمد علي النجار، الهيئة المصرية العامة للكتاب، القاهرة، الطبعة الرابعة، ٢٠٠٦م، ج١، ص ٦٧.

(١٧) برنامج قائم على إستراتيجية النمذجة لعلاج مشكلات التحدث الناتجة عن الثنائية اللغوية لدى المرحلة الإعدادية بمدارس اللغات، صابر علام عثمان علام، المجلة العلمية لكلية التربية، جامعة أسيوط، المجلد ٣٧، العدد ١٠، أكتوبر ٢٠٢١م، ص ١٠٥.

(١٨) معجم اللغة العربية المعاصرة، أحمد مختار عمر (ت ١٤٢٤هـ)، عالم الكتب، القاهرة، الطبعة الأولى، ١٤٢٩هـ/٢٠٠٨م، ج١، ص ١٠.

(١٩) الثنائية اللغوية وأثرها في المجتمع، محمود، دار الفكر، الطبعة الأولى، ٢٠٠٢م، ص ٥٩.

(٢٠) مدخل إلى التربية الخاصة، قحطان أحمد الظاهر، دار وائل للنشر، عمّان، الطبعة الثالثة، ٢٠١٥م، ص ١٠٩–١١٠.

(٢١) مدخل إلى التربية الخاصة، قحطان أحمد الظاهر، مرجع سابق،، ص ١٠٩–١١٠.

(٢٢) Benoît، Outiller la linguistique : de l’emprunt de techniques aux rencontres de savoirs، Habert، Revue française de linguistique appliquée، vol. IX، n°١، ٢٠٠٤، ص ١٢.

(٢٣) Évolutions de la linguistique outillée : méfaits et bienfaits du TAL، Ludovic Tanguy & Cécile Fabre، Revue l'Information Grammaticale، ٢٠١٤، ص ١٦

(٢٤) Language Processing with Perl and Prolog، Pierre M. Nugues، Springer، London، ٢٠١٤، ص ٤.

(٢٥) Simplifying the Difference: Machine Learning vs Deep Learning، SCS، ٢٠٢٠، متاح على: https://www.scs.org.sg/articles/machine-learning-vs-deep-learning

(٢٦) مدخل إلى اللسانيات الحاسوبية، عبد الله بن يحيى الفيفي، مركز الملك عبد الله بن عبد العزيز الدولي، الرياض، الطبعة الأولى، ٢٠١٧م، ص ٥.

(٢٧) الموارد اللغوية الحاسوبية، محسن رشوان وآخرون، مركز الملك عبد الله بن عبد العزيز، الرياض، الطبعة الأولى، ٢٠١٩م، ص ٥٧.

(٢٨) Corpus Linguistics at Work، Elena Tognini-Bonelli، John Benjamins، Amsterdam، ٢٠٠١، ص ٦.