جامعة الأزهر

كلية اللغة العربية بأسيوط

المجلة العلمية

اللغات الطبيعية واللسانيات الحاسوبية

Natural Languages And Computational Linguistics

إعداد

حنان ناصر الرشيدي

دكتوراه في اللغويات، المملكة العربية السعودية

(العدد الرابع والأربعون)

(الإصدار الثالث-أغسطس)

(الجزء الثالث (١٤٤٧هـ /٢٠٢٥م)

الترقيم الدولي للمجلة (ISSN) *2536- 9083*
رقم الإيداع بدار الكتب المصرية : ٢٠٢٥/٦٢٧١م

اللغات الطبيعية واللسانيات الحاسوبية

حنان ناصر الرشيدي

قسم اللغة والنحو والصرف، كلية اللغة العربية وآدابها، جامعة أم القرى، المملكة العربية السعودية

البريد الإلكتروني : *Al-7anan.1@hotmail.com*

الملخص:

تسعى هذه الدراسة إلى الكشف عن دور الحاسوب في معالجة اللغات الطبيعية، وإلقاء الضوء على أهم تطبيقات معالجة اللغة العربية، التي تسعى إلى تصميم برامج حاسوبية للنظريات اللغوية، ومعالجة الكلام البشري، وفهمه، وتحليله من خلالها، وتهدف بصفة عامة إلى تطبيق النماذج الحاسوبية على الملكة اللغوية، وتوظيف الحاسوب في مجال محاكاة اللغة والطرق التي تعمل بها.

واللسانيات الحاسوبية تعد إحدى علوم اللسانيات التطبيقية، فهي تُعنى بدراسة اللغات الطبيعية من منظور حاسوبي من خلال تصميم برامج للنظريات اللغوية والوصول إلى المعالجة الآلية للغات الطبيعية بالذكاء الاصطناعي.

واتبعت في هذه الدراسة المنهج الوصفي التحليلي، وقد توصلت الدراسة إلى العديد من النتائج والتوصيات جاء من أهمها أن اللسانيات الحاسوبية علم بيني يجمع بين علمي اللغة واللغات الطبيعية، والحاسوب ، وهندسة اللغة العربية وبرامجها الآلية تساعد في نقل المعارف والعلوم، من وإلى اللغة العربية، وإلى الأجيال القادمة، بالإضافة إلى حوسبة اللغة الطبيعية، تسهل الوصول في وقت قياسي إلى مفردتها، ومعرفة دلالاتها ومقابلاتها في اللغات الأخرى.

كلمات مفتاحية : اللغات الطبيعية، اللسانيات الحاسوبية، النظريات اللغوية.

Natural Languages and Computational Linguistics

Hanan Nasser Al-Rashidi

Depaartment of Language,Grammar, and Morphology, College of Arabic Language and Literature, umm Al-Qura university , Saudi Arabia

Email: *Al-7anan.1@hotmail.com*

Abstract:

This study seeks to reveal the role of the computer in natural language processing, and shed light on the most important applications of Arabic language processing, which seeks to design computer programs for linguistic theories, and the processing of human speech, understanding, and analysis through them, and aims in general to apply computer models to the linguistic queen, and the employment of computers in the field of language simulation and methods that Works out. Computational linguistics is an applied linguistics science that studies natural languages from a computational perspective by designing programs for linguistic theories and accessing the automated processing of natural languages with artificial intelligence.

In this study, the descriptive analytical approach was followed, and the study reached many results and recommendations, the most important of which was that computational linguistics is an interdisciplinary science that combines the sciences of language and natural languages, computer, and Arabic language engineering and its automated programs that help transfer knowledge and science, to and from the Arabic language, and to future generations, in addition to computerizing natural language, facilitating access in record time to its vocabulary, and knowing its semantics and interviews in other languages.

keywords: ***Natural Languages, Computational Linguistics, Linguistic Theories.***

مقدمة

الحمد لله رب العالمين، والصلاة والسلام على أشرف الأنبياء والمرسلين، ومن سار على نهجه واقتفى أثره إلى يوم الدين، أما بعد.

فتعد اللسانيات الحاسوبية إحدى علوم اللسانيات التطبيقية، وهي تُعنى بدراسة اللغات الطبيعية من منظور حاسوبي من خلال تصميم برامج للنظريات اللغوية والوصول إلى المعالجة الآلية للغات الطبيعية بالذكاء الاصطناعي.

وتسعى إلى تصميم برامج حاسوبية للنظريات اللغوية، ومعالجة الكلام البشري، وفهمه، وتحليله من خلالها، وتهدف بصفة عامة إلى تطبيق النماذج الحاسوبية على الملكة اللغوية، وتوظيف الحاسوب في مجال محاكاة اللغة والطرق التي تعمل بها.

ويهدف هذا البحث إلى الكشف عن دور الحاسوب في معالجة اللغات الطبيعية، وإلقاء الضوء على أهم تطبيقات معالجة اللغة العربية، منتهجة المنهج الوصفي التحليلي.

قابلية اللغة للتمثيل المنطقي والرياضي دفعت باتجاه معالجتها آليا، ويرجع الفضل في اكتشاف النسق الحاسوبي للغة إلى اللغوي هلمسليف، الذي رأى أن اللغة تتمتع بنظام يفرض نفسه على الأشياء، فقد صاغ نظريته التي تقول بوجود علاقات بين الوحدات اللغوية في صورة منطقية رياضية، رابطا بذلك بين التحليل اللغوي والمنطقي الرياضي بشكل عام.[1] وقد طور تشومسكي هذه النظرية، ويرى أن غاية النظرية اللسانية، تكمن في معرفة ما يجري في الدماغ البشري:" من مكانيكية وآلية تساعده في معرفة (المعرفة اللغوية) المتجلية من خلال هذه (المكنة) البيولوجية

(١) حلمي خليل، دراسات في اللسانيات التطبيقية، مصر،دار المعرفة، د.ط، ٢٩-٣٠.

العاملة في الدماغ البشري، ألا وهي اللغة ومن ثم استغلال هذه المعرفة اللغوية لتصب في المعارف الإنسانية الأخرى".[١]

وقد فتح هذا الكشف العلمي الطريق أمام رواد الذكاء الاصطناعي لوصف اللغات البشرية بطريقة صورية، تجعلها خاضعة للبرمجة على الحاسوب، وذلك من الناحية التركيبية، كون اللغات الطبيعية، تشكل لغات معقدة ضمن اللغات الصورية المعروفة. إلا أن هناك فرقا بين النظام الشكلي للغة ونظام الرياضيات، فالأخير يعتمد على الوضوح والتحديد الدقيق والموضوعي، بينما اللغات الطبيعية يشوبها الغموض والتعقيد، واللسانيات الحاسوبية تحاول أن" تجمع بين هذين النظامين الرياضي واللغوي، بين العلاقات الرياضية وبين العلاقات اللغوية، إن هذا الجمع يجعلك تقف وتفكر بشكل عميق".[٢]

وبذلك تظهر الحاجة لفهم البنية اللغوية القائمة في الدماغ البشري، ولا يتحقق هذا الفهم إلا بالتجريد الرياضي، إذ توصف اللغات الطبيعية على شكل صيغ رياضية تمثل مادة للاستعمال في الحاسبات الالكترونية، ولا يُراد من ذلك أن يغير الحاسوب اللغة بل عليه "أن يستغل خصائصها الكامنة وعلاقتها الدفينة لتيّسير أمور معالجتها آليا، واكتشاف طرائق جديدة لإكساب الآلة خاصية الذكاء الاصطناعي".[٣]

---

(١) مازن الوعر، دراسات لسانية تطبيقة، دمشق، ط١، د.ت، ٦٣.

(٢) مازن الوعر، مرجع سابق، ٧٦.

(٣) نبيل علي، اللغة والحاسوب، الكويت،مؤسسة تعريب، د.ط، د.ت، ١٥.

المبحث الأول

مفهوم اللسانيات الحاسوبية

ظهر هذا العلم في أمريكا من خلال جامعة جورج تاون سنة ١٩٥٤م، في حقل الترجمة الآلية من اللغات الأخرى إلى اللغة الإنجليزية، وتعد أقدم محاولة لدراسة اللغة بواسطة الحاسوب عند الأوربيين سنة ١٩٦١م، بجامعة (قوتبرغ) السويدية، إلا أن هذه المحاولة ذات بعد محلي لم تلقَ مستوى الرواج والانتشار في المحيط الأوربي، لكن البداية الحقيقية بمركز التحليل الآلي للغة بمدينة (قالارات) الإيطالية، الذي كان يشرف عليه (روبارتو بوزا) حيث وضع سنة ١٩٦٢م، الدعائم الأولى لاستخدام الحاسوب في دراسة اللغة، وشهد بعد ذلك تزايدا في افتتاح المراكز الحاسوبية للغة منها، المركز الحسابي لدراسة الأدب واللغة في جامعة كامبردج سنة ١٩٦٤م، والمركز المعجمي بمجمع (دالاكروسكا) بإطاليا، ومعهد الألسنية التاسع لمجمع العلوم ب(كييف) في أوكرانيا(الإتحاد السوفياتي)سنة ١٩٦٤م.[١]

أما عند العرب فقد كانت بداية الاستفادة من هذا العلم في السبعينيات وذلك في مجال العلوم الشرعية، إذ اقتصرت بادئ الأمر على إدخال أجزاء معينة من القرآن الكريم في الحاسوب، ثم أتت محاولة لتعريبه وذلك باستبدال الحروف اللاتينية بالحروف العربية.

أما في مجال الدرس اللغوي فكانت البدايات إحصائية بالدرجة الأولى وبتوجيه من إبراهيم أنيس سنة ١٩٧١م، الذي التقى بالدكتور علي حلمي موسى أستاذ الفيزياء النظرية في جامعة الكويت، وطرح علية فكرة الاستفادة من الحاسبات الآلية في

---

(١) ينظر:عبد الرحمن حسن عارف، توظيف اللسانيات في خدمة الدراسات اللغوية العربية، مجلة مجمع اللغة العربية، ٢٠٠٧م، العدد٤٩،٧٣.

إحصاءات الحروف الأصلية لمواد اللغة العربية بهدف الوقوف على نسج الكلمة العربية، ثم اتفقا على البدء بدراسة إحصائية لجذور اللغة كما جاءت في معجم الصحاح للجوهري.[١]

وتعد هذه البداية الأولى لحوسبة التراث العربي من خلال الدراسة الإحصائية، ثم توالت جهود الأفراد والمؤسسات العلمية في هذا المجال من طريق نشر مؤلفات ومقالات، وإقامة ندوات وملتقيات علمية.

الجدير بالذكر أن المؤتمر الذي عقد بالرباط سنة ١٩٨٣م، من أوائل الجهود في مجال اللسانيات الحاسوبية، وقد أسهم في عقده المركز القومي للتنسيق والتخطيط للبحث العلمي والتقني، ومعهد الدراسات والأبحاث للتعريب في المغرب، وقد نشر أعماله في كتاب (اللسانيات العربية التطبيقية والمعالجة الإشارية والمعلوماتية).

تعريفها:

اللسانيات الحاسوبية وتسمى اللغويات المعلوماتية، وهي علوم حديثة تستعمل الحواسيب في تحويل النصوص، والمعلومات اللغوية إلى لغات الحاسب الرقمية لتحليلها، وترجمتها للغات أخرى، وهي "دراسة علمية للغة الطبيعة من منظور حاسوبي، وهذه الدراسة لا يمكن أن تتم إلا ببناء برامج حاسوبية لأنظمة اللغات البشرية من خلال تقييس ومحاكاة نظام عمل الدماغ البشري لنظم عمل الحاسب الآلي".[٢]وعرفها نهاد الموسى بأنها: "الدراسة العلمية للنظام اللغوي في سائر

(١) ينظر:عبد الرحمن حسن، المرجع نفسه،٤٩-٥٠.

(٢) مهديوي عمر، توليد الأسماء من الجذور الثلاثية الصحيحة في اللغة العربية، مقاربة لسانية، كلية الآداب والعلوم الإنسانية، الدار البيضاء،١٧.

مستوياته بمنظار حاسوبي، ويتجلى هدفها في تطبيق النماذج الحاسوبية على الملكة اللغوية"[١].

نستنتج من هذه التعريفات أن اللسانيات الحاسوبية علم بيني يجمع بين علمي اللغة ومجاله الرئيس اللغات الطبيعية، والحاسوب ومجاله الرموز من أجل تحويل اللغة إلى رموز رياضية، أي لإعداد اللغة الطبيعية لتصبح لغة تخاطب وتحاور مع الحاسوب.

تسهم اللسانيات في المساعدة على فهم خصائص المعطيات اللغوية، وكذلك تقديم نظريات تفيد في كيفية بناء اللغة واستعمالها، ومن أهم الخدمات التي يقدمها علم اللغة الحاسوبي المشاركة في تقديم نظريات وتقنيات تمكن من وضع برامج حاسوبية تساعد على فهم اللغة الطبيعية.

تقسم جوانب اللسانيات الحاسوبية إلى قسمين:

-**الجانب النظري**: يبحث في الإطار النظري العميق الذي به يمكننا أن نفترض كيف يعمل الدماغ الإلكتروني لحل المشكلات اللغوية.

-**الجانب التطبيقي**: يُعنى بالناتج العملي لنمذجة الاستعمال الإنساني للغة، وإنتاج برامج ذات معرفة باللغة الإنسانية.[٢]

---

(١) نهاد الموسى، العربية نحو توصيف جديد في ضوء اللسانيات الحديثة، مؤسسة الدراسات العربية للنشر ،ط١، ٢٠٠٠م،٤٤.

(٢) ينظر: عبد الرحمن حسن العارف، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية،٥٢-٥٣.

المبحث الثاني

معالجة اللغات الطبيعية

تتفرع عن الذكاء الاصطناعي مجالات معرفية عديدة، ومنها المجال الذي بصدد التحدث عنه، أعني معالجة اللغات الطبيعية (Natural Language Processing-NLP)، وهو مجال معرفي يُسعى من خلاله إلى توجيه الآلة إلى فهم اللغة الطبيعية عبر مستوياتها المتعددة، ومعالجة وحداتها في هذه المستويات تحليلا وتوليدا، وخلق بيئة تفاعلية قادرة على تحقيق التواصل بين الإنسان والآلة.

ويطلق اسم اللغات الطبيعية للإشارة إلى اللغات الإنسانية، الإنجليزية، والفرنسية، والصينية وغيرها، لتمييزها عن اللغات الاصطناعية مثل لغات البرمجة.

وتكمن أهمية برمجيات معالجة اللغات الطبيعية في جعل الحوار بين الإنسان والآلة ممكنا بلغة أقرب ما تكون إلى اللغات الطبيعية التي يستخدمها في الحوار مع أقرانه، وذلك باستخدام الذكاء الصناعي. فاللغة هي منظومة من الرمز والمعاني والأصوات والقواعد التي يؤدي اندماجها إلى التخاطب مع الإنسان.

المعالجة الآلية للغات الطبيعية تعني معالجة اللغة الطبيعية في أشكالها كافة؛ مكتوبة (نصوص)، أو مطبوعة (وثائق)، أو مسموعة، أو مقروءة (توليد الكلام).

تنقسم معالجة اللغات الطبيعية إلى قسمين: معالجة النصوص المكتوبة؛ باستخدام المعطيات المعجمية والنحوية والدلالية، ومعالجة النصوص المنطوقة؛ باستخدام المعطيات المعجمية والنحوية والدلالية والصوتية.[١]

---

(١) ينظر: ديانا ماينادر، وآخرون، معالجة اللغات الطبيعية،ترجمة خالد الميمان، مركز الملك عبدالله بن عبدالعزيز لخدمة اللغة العربية، ٢٠١٩م، ٤٤.

المطلب الأول

المنهجيات المتبعة في المعالجة اللغوية

هناك نوعان من المنهجيات المتبعة في مهام المعالجة اللغوية، الأول المنهجية القائمة على المعرفة والثاني المنهجية المبنية على التعلم، علما أنه يمكن دمجهما معا.

أولا: المنهجية القائمة على المعرفة أو القائمة على القواعد:

تعد من الأساليب التقليدية بصفة عامة، وتستند على قواعد مكتوبة يدويا، وتجري كتابة هذه القواعد على يد متخصصين في مجال اللغات الطبيعية، وتتطلب معرفة قواعد اللغة والمهارات اللغوية، فضلا عن امتلاك ملكة البديهة.

ومن المزايا لهذه المنهجية سهولة فهم النتائج، عندما يتعرف النظام على شيء ما بشكل غير صحيح، يكون بوسع المطور التحقق من القواعد وتحديد سبب حدوث الخطأ، إذ يكون بمقدوره تصحيح القواعد أو كتابة قواعد إضافية لحل المشكلة.

ومع ذلك تستهلك كتابة القواعد الكثير من الوقت، وفي حال حدوث تغيير في المهمة، فقد يضطر المطور إلى إعادة كتابة العديد من القواعد.

ثانيا: منهجيات تعلم الآلة:

هو أسلوب تحليلي متعلق بالذكاء الصناعي، يحظى باهتمام في الآونة الأخيرة مع ظهور الأجهزة المتطورة، وبسبب عدم وجود ضرورة لامتلاك خبرة في المجال المعني أو امتلاك معرفة لغوية، ولذلك أصبح بالإمكان إنشاء نظام خاضع للإشراف بسرعة

كبيرة، إذا توفرت بيانات تدريبية كافية، ويمكن الحصول على نتائج معقولة بعد تدريب محدود جدا.(١)

ويعنى بتطوير خوارزميات قادرة على تحليل البيانات والتعلم منها لتحسين أدائها في مهمة محددة، كاتخاذ قرار معين أو تصنيف شيء ما، وبعد بناء برامج تعلم الآلة، فإنها تمر بمرحلة تدريب على بيانات كثيرة مصنفة بشريا لتكسب خوارزمية تعلم الآلة القدرة على تعلم تنفيذ نفس المهمة لاحقا على بيانات جديدة غير مصنفة، ويبدأ الباحثون باختبار أداء الأنظمة بعرض بعض المدخلات على الخوارزمية المدربة ومقارنة النتيجة التي تعطيها هذه الخوارزمية بالتصنيف الصحيح لها.

أوضح ذلك بمثال لخوارزمية لديها القدرة على تحديد نوع الفاكهة التي تظهر بصورة ما. في كل مرحلة تأخذ هذه الخوارزمية صورة لإحدى الفواكه كمدخل، في المرحلة الأولى يتم بناء الخوارزمية بحيث تكون قادرة على استقبال صور وإعطاء أوسمة محددة كمخرج. ثم تبدأ مرحلة التدريب بحيث تعطى هذه الخوارزمية عددا كبيرا من صور الفواكه ومع كل صورة تعطي اسم الفاكهة التي تظهر في تلك الصورة، فتقوم الخوارزمية بتحليل كل صورة من أجل إيجاد علاقة بين الصورة ونوع الفاكهة المرفق معها (كالشكل، واللون، والحجم) حتى تتمكن الخوارزمية من إيجاد علاقة مطردة بين الصور وأوسمتها. ثم تبدأ مرحلة الاختبار للخوارزمية بأن تعطى بعض الصور الجديدة (أي صور لم تستخدم في مرحلة التدريب)، لفواكه نفس الأنواع التي تم تدريب الخوارزمية عليها، ثم تقييم الخوارزمية وحساب دقتها بتحديد نسبة التصنيفات الصحيحة.(٢)

---

(١) ينظر: المرجع السابق، ٤٣.

(٢) بتصرف يسير: يوسف سالم العريان، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، مركز الملك عبدالله بن عبدالعزيز، ٢٠١٩م، ١٧٧.

وتصنف خوارزميات التعلم الآلي إلى:[1]

-التعلم المراقب (التعلم بإشراف): يقوم الإنسان في التعلم المراقب بتزويد عدد كبير من الإدخال للتنبؤ بقيمة المخرجات المطلوبة، بالإضافة إلى تزويد مدى دقة التنبؤات أثناء تدريب الخوارزمية، وحالما تنتهي الخوارزمية من التعلم، سوف تطبق ما تعلمته على بيانات جديدة.

-التعلم غير المراقب: لا حاجة لتدريب الخوارزمية مع المخرجات المطلوبة، وتستخدم خوارزميات التعلم غير المراقب في مهام معالجة أكثر تعقيدا من أنظمة التعلم المراقب.[2]

أولا: الشبكات العصبية الاصطناعية:

الشبكات العصبية الاصطناعية (ANN) منهجية من منهجيات تعلم الآلة مستوحاة من الخلايا العصبية.

تتكون الشبكات العصبية الاصطناعية من طبقات لنشر البيانات، أو أوزان مدخلات تحسب في مرحلة التدريب ثم يتم تحديد الناتج أو التصنيف عبرها أثناء الاستعمال.

يمكن تصنيف طبقات الشبكات العصبية إلى الأنواع التالية:

-طبقة المدخلات: وهي المسؤولة عن إدخال البيانات إلى الشبكة العصبية، وعدد العصبونات في هذه الطبقة يساوي عدد ميزات البيانات المدخلة.

---

(١) ينظر: ندى بدر جراح، تقنيات الذكاء الصناعي لتطوير التعلم الآلي الاحصائي،٢٠١٩م، ٤٦

(٢) ينظر: ندى بدر جراح، المرجع السابق، ٤٦.

-الطبقات الخفية: تقع مجموعة الطبقات هذه ما بين طبقة المدخلات وطبقة المخرجات، ووظيفتها الأساسية تحويل البيانات المدخلة إلى المخرجات المطلوبة.

طبقة المخرجات: وهي المسؤولة عن استقبال نتائج الطبقات الخفية وإصدار النتيجة النهائية للشبكة العصبية.(١)

ثانيا: التعلم العميق:

هو أحد فروع علم تعلم الآلة، والذي يهتم بتطوير خوارزميات تُمكن الحاسب الآلي من تعلم أداء المهام الصعبة(٢)، وهي شبكات عصبية اصطناعية، ولكنها تحتوي على عدد كبير من الطبقات الخفية، تؤدي هذه الزيادة إلى زيادة تعقيد عملية التدريب وتتطلب قدرا أكبر من البيانات لتدريبها.

ومع التعلم العميق، أصبح من الممكن للحاسوب إمكانية الكشف عن النقاط التي يجب ملاحظتها بين البيانات التي تسمى الميزات حيث يكتسب تلقائيا ميزات يصعب شرحها من حيث المنطق واللغات.(٣)

ثالثا: أهم تطبيقات التعلم العميق في خدمة اللغة العربية:

-تطبيقات التعلم العميق في مجال تحليل اللغة العربية الطبيعية:

اقترح الباحثون نموذجا لغويا (language model) على مستوى الحرف يقوم بتعيين قيمة محتملة لكل سلسلة من الحروف عن طريق التوزيع الاحتمالي، يطبق

---

(١) ينظر: يوسف سالم العريان، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية،١٨٨.

(٢) ينظر: ديانا ماينادر، المرجع السابق ،٥٠.

(٣) ينظر: ندى بدر جراح، المرجع السابق، ٤٦-٤٧.

البحث الشبكات العصبية الالتفاتية على أحرف الإدخال قبل إدخالها إلى الشبكات ذات الذاكرة القصيرة، وطبقت هذه الخوارزمية على لغات من ضمنها اللغة العربية.

-تطبيقات التعلم العميق في مجال التعرف على الكلام العربي المنطوق:

استخدم الباحثون الشبكة العصبية المتكررة للتعرف على الأرقام العربية المنطوقة. تكونت شبكتهم العصبية من طبقتين خفيتين وكان أداؤها جيدا لبيانات عدة متكلمين.(١)

وتمكن باحثون سنة ٢٠١٦م في جامعة لومان بفرنسا من تسخير تقنية التعلم العميق للتعرف على النماذج الصوتية العربية، وتحقق تحسين للدقة بنسبة ١٥,٧%

-تطبيقات التعلم العميق في مجال التعرف على الحروف العربية المكتوبة:

يعتبر استخدام تقنية التعلم العميق في مجال التعرف الضوئي الآلي على النصوص العربية من أكثر مجالات خدمة اللغة العربية انتشارا.

واللغة العربية تختلف عن اللغات الأخرى بخصائص منها:

- اتجاه الكتابة في العربية من اليمين لليسار على عكس اللغات اللاتينية.
- شكل الحرف العربي يعتمد على اتصاله بما حوله.
- طبيعة اللغة العربية مختلفة في الاشتقاق والصرف والنحو والتشكيل.
- بعض الحروف متشابهة لحد كبير، بحيث تختلف بعدد وموضع النقاط.

ويبدو مما سبق أن هذه الخصائص لها تأثير عند تصميم خوارزميات التعرف الآلي على النصوص العربية المكتوبة.

---

(١) ينظر: أحمد الحايك، التعلم العميق وتطبيقاته المرتبطة باللغة العربية، ٢٠١٩م، ١٥٦-١٥٧.

وفي عام ٢٠١٧م قدم شوقي بوفنار وزملاؤه عملا استخدم فيه الشبكة العصبية الالتفاتية العميقة للتعرف على صور الأحرف العربية المكتوبة بخط اليد، وسجلت نتائج البحث دقة تصل إلى ٩٧,٣٢%.

وكذلك قام أحمد الصاوي وزملاؤه ببناء شبكة عصبية التفاتية وتطبيقها للتعرف على الحروف العربية المكتوبة باليد، وسجلت نتائج وصلت دقتها إلى ٩٤%.[١]

(١) ينظر: المرجع السابق، ١٥٦.

المطلب الثاني

مسارات مهام معالجة اللغات الطبيعية:

تتكون مسارات مهام ما قبل معالجة اللغات الطبيعية إجمالا من المكونات الآتية:

تقطيع الكلمات Tokenization.

تقسيم الجمل Sentence splitting

تصنيف أقسام الكلام part-of-speech tagging.

التحليل الصرفي Morphological analysis.

التحليل النحوي وتجزئة النص Parsing and chunking.

فالمهمة الأولى تجزئة كلمات النص إلى قطع، تليها مهمة تقسيم الجمل بهدف تقطيع النص إلى وحدات لغوية (تكون في العادة كلمات وأرقام وعلامات ترقيم والمسافات بين الكلمات) وجمل على التوالي، وتضع مهمة تصنيف أقسام الكلام كل جزء من أجزاء الجملة في فئة نحوية، والتحليل الصرفي يقوم بإيجاد جذر كل كلمة، وفيما يلي تفصيل لها:

أولا: تقطيع كلمات النصTokenization :

تجزئة كلمات النص إلى قطع أي تقسيم النص المدخل إلى وحدات بسيطة، وهذه الوحدات اللغوية تشير عموما إلى الكلمات والأرقام والرموز، وتعد خطوة مطلوبة في جميع تطبيقات المعالجة اللغوية، لأن الخوارزميات الأكثر تعقيدا مثل خوارزميات

تصنيف أقسام الكلام، تتطلب وجود هذه الوحدات كمدخلات لها، بدلا من النص الخام.[١]

ثانيا: تقسيم الجمل Sentence splitting:

تمييز الجمل هي مهمة تقسيم النص إلى الجمل المكونة له، وتشتمل هذه المهمة على تحديد ما إذا كانت علامات الترقيم، مثل نقطة النهاية، والفواصل وعلامات التعجب وعلامات الاستفهام، تدل على نهاية الجملة أو على شيء آخر (الكلام المقتبس، الاختصارات).

وتستخدم معظم مقطعات الجمل قوائم الاختصارات للمساعدة في تحديد المطلوب، مثلا: تدل نقطة النهاية على نهاية الجملة ما لم تأت بعد اختصار مثل السيد.(.Mr).

ثالثا: تصنيف أقسام الكلام part-of-speech tagging:

يعنى تصنيف أقسام الكلام بوضع علامات على الكلمات تشير إلى التصنيف الذي تنتمي إليه، مثلا الأسماء، والأفعال، والصفات، وتقسم هذه الفئات اللغوية الأساسية إلى أصناف دقيقة، مثلا: المفرد، والجمع، وزمن الفعل (الماضي، الحاضر، المستقبل)، وتكمن صعوبة هذه العملية في تصنيف أقسام الكلام بناءً على السياق، مثلا يمكن أن تصنف كلمة(سعيد) على كونها اسم أو صفة حسب سياق الكلام.

رابعا: التحليل الصرفي Morphological analysis:

يتعلق التحليل الصرفي بشكل أساسي بالتعرف على الوحدات اللغوية داخل الكلمة وتصنيفها، عن طريق تجزئة الكلمة إلى الجذر مع السوابق واللواحق، مثلا: الفعل في اللغة الإنجليزية (walked من الجذر walk واللاحقة ed)، وفي اللغات الأخرى

---

(١) ينظر: ديانا ماينادر، وآخرون، معالجة اللغات الطبيعية، ٤٥.

يمكن استخدام السوابق (إضافة مقطع في بداية الكلمة)، أو المتوسطات (إضافة مقطع في وسط الكلمة)، إلى جانب تغييرات أخرى.[١]

خامسا: التحليل النحوي وتجزئة النص Parsing and chunking:

يقصد بها تحليل الجمل، وعملية التحليل تشرح بشكل أساسي كيف ترتبط العناصر المختلفة في الجملة بعضها ببعض، وهناك الكثير من النظريات النحوية المختلفة في علم اللغويات الحاسوبية، حيث تطرح هذه النظريات أنواعا مختلفة من البنى النحوية، ولهذا السبب تختلف أدوات التحليل بعضها عن بعض، ليس من حيث الأداء، بل أيضا من حيث نوع التمثيل الشكلي الذي تنتجه، وعلى النظرية النحوية التي تستخدمها.[٢]

تجزئة النص: تميل هذه الخوارزميات للعمل في أحسن صورها عندما يكون النص الذي تعالجه مشابها للنص الذي سبق تدريبها عليه، ومهمة تجزئة النص أكثر تعقيدا من مهام المعالجة الأخرى، وتقسم أدوات تجزئة النص إلى مجزئات العبارات الاسمية ومجزئات العبارة الفعلية.

(١) ينظر: المرجع السابق،٤٦

(٢) ينظر: المرجع السابق، ٦٠-٦١.

المبحث الثالث

تطبيقات المعالجة الآلية للغات:

في هذا المبحث نتطرق إلى أهم تطبيقات معالجة اللغات الطبيعية، وهي الترجمة الآلية، والمعجم الإلكتروني، والتلخيص الآلي، تحليل الآراء والمشاعر.

أولا: الترجمة الآلية:

هي استخدام برمجيات الحاسب في ترجمة النصوص أو الكلام من لغةٍ إنسانية لأخرى، وتعمل برامج الترجمة الآلية على استبدال الكلمات باللغة المترجم منها بالكلمات المقابلة لها في اللغة المترجم إليها(١)، ويكمن موضوع الترجمة في "تحليل النص الأصلي ونقل عناصره من اللغة التي سيترجم إليها، ثم توليد هذا النَّص اعتمادًا على التَّحليل والنقل"(٢)، وتتيحُ برمجيات الترجمة الآلية الحاليَّة تخصيصَ الترجمة حسب المجال أو المهنة، إذ تُحسَّن الترجمة النهائية من خلال حصر نطاق الاستبدال المسموح به، وهذا الأسلوب فعَّالٌ للغاية خاصة في المجالات التي تستخدم فيها اللغة الرسميَّة أو الاصطلاحيَّة، وعامةً تكون نتائج الترجمة الآليَّة أفضل في النُّصوص الحكوميَّة والقانونيَّة؛ لأنَّها تعتمد على قوالب من الجمل والعبارات بعكسِ النصوص العامة والمحادثات(٣).

---

(١) ينظر: الحجاج محمد البشير، المعالجة الآلية للغة العربية جهود الحاضر وتحديات المستقبل، مجلة لغة العصر،٢٠٠٩م، ص ٦.

(٢) اليوبي بلقاسم، اللسانيات الحاسوبية مفهومها وتطورها ومجالات تطبيقاتها، مجلو مكناسة، العدد١٢، ص ٥٢.

(٣) ينظر: الحجاج محمد بشير، المرجع نفسه، ص٦.

وتبقى الترجمة الآلية "وسيلة فعالة من وسائل توظيف المعرفة العلمية والتقنية في المجتمع العربي؛ لأن معظم المعرفة قد أُنتجت ونُشرت وحُفظت باللغة الإنجليزية، وللوصول إليها لا بد من تفعيل دور الترجمة ومؤسساتها، وبهذا فإن الترجمة الآلية من اللغات الأخرى إلى العربية أو العكس تعتبر سبيلًا لسد الفجوة العلمية الناتجة عن تضخم الإنتاج العالمي الثقافي بالقياس إلى نظيره العربي"[١].

والترجمة الآلية بصفة عامة وصلت إلى مستوى متقدم جدا، ونجحت في مساعدة المترجم البشري على تحسين عمله بل، إلا أنها لم تستطع أن تتغلب على المترجم البشري، خاصة في ترجمة الحوارات والمحادثات ويزيد الأمر صعوبة لو كانت هذه الحوارات باللغة العامية أو اللغة غير الرسمية.[٢]

يمكن تقسيم طرق الترجمة الآلية إلى: ترجمة مباشرة وغير مباشرة، والطرق غير المباشرة تنقسم إلى قسمين طرق تعتمد على التحويل وأخرى على التمثيل الموحد للغات.

ومن أهم محاولات الترجمة الآلية من اللغة العربية وإليها:

ظهرت عدة محاولات للترجمة الآلية من اللغة العربية وإليها، وقد أثمر بعضها بتكوين أنظمة ترجمة آلية، وهي كالآتي:

- محاولة بشاي الأستاذ بجامعة هارفارد، في أوائل السبعينات.

- نظام المترجم العربي الذي طورته شركة (ATA) في لندن، وقد طورت الشركة برنامجا مصغرا أسمته(الوافي).

---

(١) مهديوي عمر، المرجع نفسه،٣٦.

(٢) ينظر: الحجاج محمد البشير، المرجع نفسه،٦

– نظام (عربترانز)، وقد طورته شركة عربية في لندن.[١]

– نظام (الناقل العربي) الذي طورته شركة سيموس العربية في باريس، ولدى الشركة أربعة برامج للترجمة بين الإنجليزية والعربية وبين الفرنسية والعربية، وبرنامج لكل اتجاه.

–وظفت شركة صخر لبرامج الحاسب، محركا للترجمة الآلية الخاص بها، وهي خدمة ترجمة فورية تقوم بترجمة أي صفحة، في بدايتها مجانية، بينما الآن يمكن إرسال النص المطلوب ترجمة والحصول على الترجمة بأجور.

–برنامج شركة جوجل، وهو برنامج مجاني يستند إلى الترجمة الإحصائية من ذخيرة لغوية مأخوذة من الأنترنت.[٢]

أولا: المعجم الإلكتروني:

يعد المعجم الإلكتروني من مخرجات المعالجة الآلية للغات الطبيعية، وهو نتيجة الاستفادة من علوم الحاسب في مجال الصناعة المعجمية، ويعرف بأنه قاعدة بيانات آلية تقنية للوحدات اللغوية، وما تعلق بها من معلومات من قبيل كيفيات النطق بها، وأصولها الصرفية، ومعانيها الدلالية، وكيفيات استخدامها ومفاهيمها المخصوصة التي تحفظ بنظام معين في ذاكرة تخزين ذات سعة كبيرة، ويقوم جهاز آلي بإرادة المعطيات الفنية والمضمونية التي يتضمنها المعجم الإلكتروني وفق برنامج محدد.[٣]

---

(١) ينظر: محمد زكي خضر، اللغة العربية والترجمة الآلية المشاكل والحلول، مؤتمر التعريب الحادي عشر، ٢٠٠٨م، ٢٧.

(٢) ينظر: محمد زكي خضر، المرجع السابق، ٢٨

(٣) أنوار الجمعاوي، المعجم الإلكتروني العربي المختص قراءة نقدية في نماذج مختاره،العدد ٢٠١٣-٢٠١٤، بحث بمؤتمر العربي للترجمة ،المغرب، ٤

ويُبنى المورد المعجمي على أسس من المفاهيم الحاسوبية، وتمثيل المعلومات المعجمية، بالاعتماد على تصور نظري، يحقق متطلبات المقاربة الحاسوبية للمادة المعجمية، ومن أهم المشاريع المعجمية العربية، المعجم العربي التفاعلي، ومشروع شبكة الكلمات العربية.(١)

أولا: أنواع المعاجم الإلكترونية: (٢)

– من حيث معيار اللغة: معاجم إلكترونية أحادية اللغة، وهي المعاجم التي تقتصر على في عملها على لغة معينة، ومعاجم إلكترونية ثنائية اللغة، وأخرى متعددة اللغات، حيث تورد ألفاظ لغة ما وترتبها بحسب المنهج المأخوذ به في هذه اللغة، وتجعلها مداخل المادة، ثم تذكر ما يقابلها من كلمة في لغة أخرى أو أكثر وذلك كأن تذكر الكلمة العربية متبوعةً بما يقابلها في الإنجليزية أو الفرنسية أو كلتيهما.

- من حيث المحتوى المعرفي:

معجم إلكتروني عام، حيث يشتمل على وحدات لغوية تنتمي إلى مجالات معرفية مختلفة، ومعجم إلكتروني خاص، ويشتمل على وحدات لغوية تنتمي إلى مجال معرفي محدد.

ثانيا: مزايا المعجم الإلكتروني(٣):

–حوسبة اللغة الطبيعية، تسهل الوصول في وقت قياسي إلى مفردتها، ومعرفة دلالاتها ومقابلاتها في اللغات الأخرى.

---

(١) ينظر: مجموعة من الباحثين، الموارد اللغوية الحاسوبية، ط١، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، ٢٠١٩م، ٣٨.

(٢) ينظر: أنوار الجمعاوي، المعجم الإلكتروني العربي المختص قراءة نقدية في نماذج مختاره،٥.

(٣) ينظر: المرجع السابق،٥–٦–٧.

– تتميز هذه المعاجم بالشمولية وبطاقة تخزينية كبيرة بحيث تشتمل على المصطلحات القديمة والحديثة، وأيضًا تتضمن لغات مختلفة.

– توفر بعض المعاجم ميزة التدقيق الإملائي للكلمة المدخلة، وتوضح احتمالات الخطأ عند إدخال المستخدم للكلمة، وتقدم كلمات بديلة قريبة للكلمة المراد البحث عنها.

– يوفر المعجم الإلكتروني عدة تطبيقات لغوية، بحيث يتمكن المستخدم من معرفة تصريف الكلمة، والبحث في أصولها الصوتية، وسياقاتها الدلالية والمفهومية.

– يسهل المعجم الإلكتروني تعلم اللغات، بما يوفره من مدونات لغوية محسوبة.

ثالثا: التلخيص الآلي:

يعد من تطبيقات المعالجة الآلية للغة، حيث يقوم بإنشاء نص مختصر من ملف أو مستند بواسطة برنامج حاسب آلي، على أن يحتوي النص المختصر على أهم الأفكار في النص الأصلي، وتكمن أهمية التلخيص الآلي في كثرة المعلومات وتزايدها عن قدرة المرء على الملاحقة والمتابعة، وينبغي على البرمجيات تقديم خلاصات متماسكة، وأن تأخذ في الاعتبار عدة متغيرات مثل الطول وأسلوب الكتابة والبناء من أجل إنشاء ملخص مفيد.[١]

ويمكن إجراء التلخيص الآلي بإحدى الطريقتين:

–التلخيص الاقتباسي: يعمل على تلخيص النص من خلال اقتباس أهم العبارات والمفاهيم الواردة فيه بدون توليد أي نصوص جديدة أو اختزال معان غير مهمة،

(١) ينظر: الحجاج محمد البشير،المعالجة الآلية للغة العربية جهود الحاضر وتحديات المستقبل،مجلة لغة العصر،٢٠٠٩م ٨.

وجميع الجمل الملخصة هي جمل وتعابير موجودة في النص الأصلي تم تصنيفها من قبل الخوارزمية على أنها مهمة وتلخص الموضوع بقدر كافٍ.

-التلخيص الخلاصي: يعمل على توليد نصوص تختصر محتوى ومعنى النص الإجمالي باستخدام نص جديد صحيح لغويا وإملائيا، ويحتاج إلى خوارزميات متقدمة تستطيع فهم النص أولا، ومن ثم توليد نص صحيح يلخص النص الأساسي.[١]

نماذج من أنظمة التلخيص للنصوص العربية:

-نظام لخص(lakhas): هو نظام للتلخيص الآلي تم تطويرة في جامعة مونتريال الكندية، يستخدم النظام الأسلوب الإحصائي في عمليات التلخيص على المراحل التالية: تجزئة النص إلى مجموعة من الجمل، تجزئة الجمل إلى كلمات، وضع حروف الكلمات في صور موحدة مثل:(هـ، ة) ومثل(أ،ا).

نظام (Acbtss): من جامعة إسكس البريطانية، ويعتمد على تقنية بايز الإحصائية ويحتاج هذا النظام مدونة لغوية مرمزة ومزودة بالحواشي، يدرب النظام على استخراج خصائص الجمل التي يحتفظ بها في ناتج التلخيص، وهذه الخصائص يتم تحديدها من خلال المعالجة اللغوية للنص (تحليل صرفي، ترميز أجزاء الكلام)، من خلال موقع الجمل داخل النص.

نظام (The summar of aramedia): يتميز هذا النظام من شركة صخر بوجود وظيفة تصحيح الأخطاء اللغوية الشائعة، يستخدم النظام في تلخيص النصوص

(١) ينظر: أحمد الحايك، التعلم العميق وتطبيقاته المرتبطة باللغة العربية، مركز الملك عبدالله لخدمة اللغة العربية، ٢٠١٩م، ١٨٨.

الإنجليزية والعربية.[١]

ثالثا: تحليل الآراء والمشاعر:

تحليل المشاعر هو أحد مجالات اللسانيات الحاسوبية، ويعد من أنشط فروع هذه العلوم بحثا نظرًا لأهمية تطبيقاته وكثرة المحتوى النصي لإجراء البحوث فيه، لاسيما ما تقدمه شبكات التواصل الاجتماعي اليوم من كميات هائلة من النصوص المحملة بآراء أصحابها تجاه كل أنواع القضايا.

تعمل خوارزميات تحليل المشاعر بتحليل النص اللغوي من أجل الكشف عن المشاعر التي يعبر عنها الكلام تجاه موضوع النص، وتصنيفها إلى مشاعر إيجابية أو سلبية أو محايدة، فإن بعض الخوارزميات تذهب إلى تصانيف أشمل يتضمن حالات شعورية أكثر تفصيلا كالسعادة والغضب..إلخ.

وترجع أصول مجال تحليل الآراء والمشاعر إلى علم الفلسفة، وتستند الكثير من الدراسات الأولى في الموضوع إلى أفكار فريدريك نيتشه ونظرياته حول تعدد الآراء التي تتلخص في أن الحقيقة ممكن أن تكون ذات أوجه متعددة، وأن كثير من القضايا التي يتجادل حولها الناس ليس لها حقيقة مطلقة بالضرورة، والفيلسوف الأمريكي ريتشارد سكاشت درس أفكار نيتشه "وأعاد صياغتها بحيث فرق بين نوعين من الأفكار: الأفكار المرتبطة بحقائق، والأفكار التي تعبر عن رأي. وتعتبر هذه الدراسات هي الجهود الأولى التي ارتكزت عليها الكثير من الدراسات الحديثة في مجال تحليل الآراء". [٢]

---

(١) ينظر: محسن رشوان، والمعتز بالله، المعالجة الآلية للنصوص العربية، الطبعة الأولى، ٢٠١٩م، ١٥٥.

(٢) ينظر: يوسف أبو جبارة، تحليل الآراء العربية إلكترونيا، بحث منشور، ٢٠١٩م، ١٠٥-١٠٦-١٠٨.

ولما كان الكلام المكتوب والمنطوق هو الوسيلة للتعبير عن الأفكار ومشاركتها مع الآخرين،" فقد انصب كثير من اهتمام الباحثين في هذا المجال على دراسة العلاقة بين طبيعة الكلام المستخدم في الحديث والآراء التي يحملها المتحدث، حتى ظهر مجال في علم اللغويات متخصص بدراسة اللغويات النفسية". (١)

أما جهود البحث في تحليل الآراء العربية جاءت متأخرة، بعد أن وفر انتشار وسائل التواصل الاجتماعي حافزا لدى كثير من الباحثين العرب للبحث في هذا المجال.

ركزت هذه الجهود على مواءمة المقاربات المستخدمة لتحليل الآراء في اللغة الإنجليزية واللغات الأخرى للغة العربية، وتضمن هذا بناء موارد لغوية تخدم تحليل الآراء العربية كمعاجم آراء ومدونات لغوية، ومكتبات برمجية لتحليل الآراء.

ولقد وضفت هذه الجهود البحثية للتعامل مع التحديات الخاصة باللغة العربية كتعدد اللهجات العربية، ودراسة أثر المعالجة المسبقة للنص العربي (كالتحليل الصرفي والتجذير والتجذيع) على دقة تحليل الآراء.(٢)

---

(١) المرجع السابق، ١٠٨.

(٢) ينظر: المرجع السابق، ١٠٨.

الخاتمة

تم بحمد الله هذا البحث الذي تناول معالجة اللغات الطبيعية، واستعرضت أهم تطبيقاتها، كما عرفت بأهم الوظائف التي ينبغي الإلمام بها للمهتم في هذا المجال، وفيما يلي **عرضٌ لأهم النتائج**:

– اللسانيات الحاسوبية علم بيني يجمع بين علمي اللغة واللغات الطبيعية، والحاسوب ومجاله الرموز من أجل تحويل اللغة إلى رموز رياضية، أي لإعداد اللغة الطبيعية لتصبح لغة تخاطب وتحاور مع الحاسوب.

–اللغة العربية غنية بالدلالات اللفظية والقواعد الصرفية والنحوية، وبذلك تصبح من أهم اللغات الطبيعية، ولابد من الاهتمام في مجال معالجة اللغة العربية خصوصا مع تطور خوارزميات التعلم العميق التي تساعد في معالجة اللغات الطبيعية وتطبيقاتها.

–هندسة اللغة العربية وبرامجها الآلية تساعد في نقل المعارف والعلوم، من وإلى اللغة العربية، وإلى الأجيال القادمة.

–تتميز اللغة العربية بمجموعة من الخصائص التي تساعد في تسهيل أو تعقيد معالجتها آليا بشكل يتفاوت من خاصية لأخرى.

– حوسبة اللغة الطبيعية، تسهل الوصول في وقت قياسي إلى مفردتها، ومعرفة دلالاتها ومقابلاتها في اللغات الأخرى.

المراجع والمصادر:

١-أحمد الحايك، التعلم العميق وتطبيقاته المرتبطة باللغة العربية، مركز الملك عبد الله لخدمة اللغة العربية، ٢٠١٩م.

٢-أحمد جنات، توصيف جديد في ضوء اللسانيات الحاسوبية، الأردن، ٢٠٠٧م.

٣-أحمد روبي، البنك الشجري النحوي، مركز الملك عبد العزيز الدولي لخدمة اللغة العربية، ط١، ٢٠١٨م.

٤-أحمد روبي، نحو بناء بنك شجري نحوي للغة العربية الفصحى المعاصرة، جامعة الفيوم، ٢٠١٥م

٥-أنوار الجمعاوي، المعجم الالكتروني العربي المختص قراءة نقدية في نماذج مختاره، بحث بمؤتمر العربي للترجمة، المغرب، ٢٠١٤م.

٦-الحجاج محمد البشير، المعالجة الآلية للغة العربية جهود الحاضر وتحديات المستقبل، مجلة لغة العصر،٢٠٠٩م.

٧-حلمي خليل، دراسات في اللسانيات التطبيقية، مصر،دار المعرفة، د.ط.

٨-حمد محمد الحناش، المعاجم الألكترونية للغة العربية، مؤسسة دار العرفان، بحث منشور.

٩-ديانا ماينادر، وآخرون، معالجة اللغات الطبيعية،ترجمة خالد الميمان، مركز الملك عبدالله بن عبدالعزيز لخدمة اللغة العربية، ٢٠١٩م.

١٠-راضية بن عربية، حوسبة النظام اللغوي: المعجم الآلي عند عبد الرحمن الحاج صالح، المجلس الأعلى للغة العربية، ٢٠١٧م.

١١-سعيد فاهم، قراءة في الإسهمات اللسانية الحاسوبية العربية آفاق ورهانات، مجلة دراسات جامعة الاغواط، عدد٣٦، ٢٠١٥.

١٢-صالح بلعيد، دروس في اللسانيات التطبيقية، الجزائر، دار هومه، د.ت.

١٣-عبد الرحمن حسن عارف، توظيف اللسانيات في خدمة الدراسات اللغوية العربية، مجلة مجمع اللغة العربية، ٢٠٠٧م، العدد٧٣.

١٤-مازن الوعر، دراسات لسانية تطبيقة، دمشق، ط١، د.ت.

١٥-محمد زكي خضر، اللغة العربية والترجمة الآلية المشاكل والحلول، مؤتمر التعريب الحادي عشر، ٢٠٠٨م.

١٦-مهديوي عمر، توليد الأسماء من الجذور الثلاثية الصحيحة في اللغة العربية، مقاربة لسانية، كلية الآداب والعلوم الإنسانية، الدار البيضاء.

١٧-نبيل علي، اللغة والحاسوب، الكويت، مؤسسة تعريب، د.ط، د.ت.

١٨-ندى بدر جراح، تقنيات الذكاء الصناعي لتطوير التعلم الآلي الاحصائي،٢٠١٩م.

١٩-نزار حبش، مقدمة في المعالجة الطبيعية للغة العربية، ترجمة: هند خليفة، جامعة الملك سعود، ط١، ٢٠١٤م.

٢٠-نهاد الموسى، اللغة العربية نحو توصيف جديد في ضوء اللسانيات الحديثة، مؤسسة الدراسات العربية للنشر، ط١، ٢٠٠٠م

٢٢-وليد إبراهيم الحاج، اللغة العربية ووسائل الاتصال الحديثة، ط١، دار البداية، عمان،٢٠١١م.

٢٣-اليوبي بلقاسم، اللسانيات الحاسوبية مفهومها وتطورها ومجالات تطبيقاتها، مجلة مكناسة، العدد١٢.

٢٤-يوسف سالم العريان، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، مركز الملك عبد الله بن عبد العزيز، ٢٠١٩م.

٢٥- يوسف أبو جبارة، تحليل الآراء العربية إلكترونيا، بحث منشور، ٢٠١٩م.

٢٦- مجموعة من الباحثين، الموارد اللغوية الحاسوبية، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض، ٢٠١٩م، ط١.