قدّمت هذه الورقة العلمية في الملتقى الدولي الموسوم "جهود اللسانيين العرب في تعليمية اللغة العربية وحوسبتها"، يومي: 26-27 نوفمبر 2023م، جامعة الإخوة منتوري، قسنطينة1.

عنوان المداخلة: **جهود مركز الملك عبد الله بن عبد العزيز الدولي في حوسبة اللغة العربية**

Efforts of the king Abdullah bin Abdulaziz International Center for Arabic language

د/ نعيمة روابح أستاذ محاضر –أ-

جامعة الأمير عبد القادر للعلوم الإسلامية

naima.rouabah@univ-emir.dz

الملخص بالعربية:

تهدف هذه الورقة العلمية إلى التعريف ببعض جهود الباحثين العرب المحدثين في اشتغالهم على مسألة تطويع تقنيات الحاسوب لخدمة البحث اللساني العربي، ويعدّ مركز الملك عبد الله بن عبد العزيز الدولي أحد المراكز البحثية المهتمة بقضايا اللسانيات الحاسوبية وعلى رأسها قضية حوسبة اللغة العربية والذكاء الاصطناعي؛ وهو المجال الذي تتداخل فيه التصورات اللسانية مع البرامج الحاسوبية لتشكّل نظريات تهدف إلى معالجة الوقائع اللغوية وفق منهج آلي.

وعليه فإن البحث سيحاول الإجابة عن جملة من التساؤلات أهمها: ما أهم الإسهامات التي قدّمها المركز في مجال حوسبة اللغة العربية؟ وهل يمكن لهذه الجهود أن تكون كافية لتحقيق معالجة حاسوبية للغة العربية؟

Abstract :

This scientific paper aims to indroduce some of the efforts of modern Arab researchers in their work on the issue of adapting computer technologies to serve arabic linguistic research , the king Abdullah bin Abdulaziz International Center is one of the research centers concerned with issues of computational linguistics, most notably the issue of computerization of the Arabic language and artificial intelligence, which is the field in which linguistic concepts interact with computer programs to from theories aimed at treating linguistic facts according to an automated approach.

Accordingly, the research will attempt to answer a number of questions, including: what are the most important contributions made by the Center in the field of computerization of the Arabic language? Can these efforts be sufficient to achieve computer processing of the Arabic language?

تقديم:

تشهد ساحة البحث العلمي، في عصرنا هذا، عهدا جديدا تسعى فيه الآلة الصمّاء (الحاسوب) التي لا تعمل إلا وفق برامج صنعها الإنسان نفسه، أن تكون فيه بديلا للإنسان من خلال محاولة استكناه العمليات العقلية التي يقوم بها الدماغ البشري لإنتاج اللغة وفهمها وإدراكها.

إن من أهداف اللسانيات الحاسوبية أن تهيّء للحاسوب كفاية لغوية تماثل ما يكون عليه الإنسان حين يستقبل اللغة، ويفهمها، ليعيد إنتاجها وفق المطلوب، هذه القدرة التي يتمتع بها الحاسوب من قدرة على تمثيل الأفكار والحقائق، وتوصيف المعلومات ووضعها في التمثيلات الملائمة لها، يجعل الحاسوب أكثر الاختراعات تفوّقا وتميّزا.

1- اللغة العربية والحاسوب:

لقد أصبح الحاسوب أكثر التقنيات طلبا في البحث العلمي عموما واللساني خصوصا، إذ تكاد الاتجاهات اللسانية باختلاف مناهجها وأهدافها تتفق على ضرورة استخدام هذه التقنية، وبات ضرورة من ضروريات البحث اللساني في مختلف مستوياته: الصوتية، المورفولوجية، التركيبية، المعجمية والدلالية، مما جعل الباحثين العرب باختلاف اختصاصاتهم يلحّون على ضرورة اعتماد التقنية والإفادة من برامجها في خدمة اللغة العربية سواء في التعليم، أو الإحصاء، أو تمثيل البيانات اللغوية وتخزينها أو ترجمتها أو تعليمها، لما لهذه الآلة من دور في النهوض بالبحث العلمي بل في أسلوب الحياة الإنسانية عامة.

إن لحوسبة اللغة العربية فوائد كثيرة خاصة في تعليم اللغات؛ إما على مستوى اللغة الأم أو اللغة الأجنبية، إذ للحاسوب من المزايا المتعددة والطرق المنهجية التعليمية ما يساعد على تجسير الفجوة بين اللغة ومتعلمها، إذ هناك " أبحاث جادة من قبل اللغويين ومهندسي الحاسوب من أجل إدخال الحوسبة إلى الترجمة بما يعرف بالترجمة الآلية MT machine translation أو بمفهومها الآخر بشيء من الاختلاف في درجة استخدام الحاسوب في الترجمة".[1]

[1] - اللغة العربية ووسائل الاتصال الحديثة، وليد إبراهيم الحاج، دار البداية ناشرون وموزعون، عمان، 2011م، ص30.

يتساءل **لايكوف** في كتابه "**الفلسفة في الجسد**": إن كان هناك حاسوب في الدماغ مسؤول عن بناء التصورات والنماذج التصورية فينا؟ وإن كان الأمر كذلك فمن يقوم ببناء هذه التصورات؟ ثم يجيب قائلا:" إن ما يعدّ مؤثرا في هذه النماذج أنها حاسوبية. وحقل علم الأعصاب الحاسوبي له ارتباط بالموضوع، لا فيما يخص مكان حدوث الحوسبات العصبية فحسب، ولكن فيما يخص كيفية حدوثها، أي له ارتباط بالآليات العصبية الحاسوبية الدقيقة التي تنجز عمليات حسية حركية وتباشر بناء التصورات والمقْوَلة والتفكير وتعلم اللغة. كل نموذج يقوم بهذين العملين بتفصيل".[1]

إذن هذه النماذج كما يرى **لايكوف** تنتج عن عمل جهاز حاسوبي داخل الدماغ، لكن هذا التصور "ينافي واقع اللغة والتفكير، حقا توجد آليات عصبية بالدماغ تقوم بكل هذا العمل، لكنها ليست حاسوبية بالمعنى الدقيق، فهي قدرة فطرية في الدماغ (القدرة على التعميم والاستنتاج لدى المخ البشري وحده) تقوم بعملية الترتيب والتنظيم والتعميم والاستنتاج وتصدر الأوامر الحركية لأجهزة الجسم، وتبني التصورات، فهي قدرة فطرية بالمخ، يحاول الحاسوب البشري القيام ببعض مهامها، من هنا كان المخ يشبه الحاسوب".[2]

إن الغرض الذي يسعى إليه الترميز الحاسوبي من خلال الذكاء الاصطناعي هو إعادة إنتاج عمليات مماثلة للعمليات الذهنية الكامنة في أدمغة البشر، عن طريق بناء تصورات في داخل الحاسوب، وتحريكها، هذه التصورات يُفترض أنها تتماشى أو تتطابق مع نموذج ذهني معين، أي مع برامج ذهنية مخزنة سلفا في الذهن البشري، وهي "فرضية أساسية لفهم عمل الذكاء الاصطناعي في تقليده للذكاء البشري".[3]

2- استخدام الحاسوب في العملية التعليمية

يعدّ الحاسوب "وسيلة جد ناجحة حالة استخدامها في ميدان تعلّم اللغات أو استعمالها في ميادين المعرفة عامة"،[4] إذ يظهر دوره في تعليم اللغات في شكلين: كوسيلة في تركيب أو بناء المواد بغض النظر عن الكيفية التي يجب أن تقدم بها هذه المواد، وفي التقديم الفعلي لهذه المواد للمتعلم؛ فالحاسوب يقدّم خدمات

---

[1] - الفلسفة في الجسد، الذهن المتجسد وتحديه للفكر الغربي، جورج لايكوف، مارك جونسون، ترجمة وتقديم: عبد المجيد جحفة، دار الكتاب الجديدة المتحدة، ط1، بيروت، 2016م، ص 81.

[2] - اللسانيات العصبية اللغة في الدماغ (رمزية، عصبية، عرفانية)، عطية سليمان أحمد، الأكاديمية الحديثة للكتاب الجامعي، القاهرة، 2019م، ص 407.

[3] - المرجع نفسه، ص 435.

[4] - دروس في اللسانيات التطبيقية، صالح بلعيد، دار هومة، الجزائر، ص 107.

كثيرة للمتعلمين والمعلّمين على حدّ سواء، إذ يمكن له "أن يقدّم بيئة تعليمية متفاعلة، كما أن الحاسوب يوفر عنصر المنافسة أيضا".[1]

لقد حاول العديد من الباحثين في مجال الذكاء الاصطناعي "نمذجة برامجهم الحاسوبية متّبعين الذّكاء البشري، ويستلهمون ما يكتشف في معالجة المعلومات البشرية، وهي اكتشافات مصدرها موجود في علوم أخرى في العلم المعرفي، مثل علم النفس واللسانيات...وقد أدت محاولات الباحثين في الذّكاء الاصطناعي لبرمجة الأنسقة التي تستطيع فهم اللغة أو الإبصار أو حلّ المسائل".[2]

يتوجّه التعلّم العميق الآلي حديثا إلى صنع شبكة من الخلايا العصبية، وقد ازدادت كفاءة الترجمة الآلية في السنوات الأخيرة بفضل ما جدّ من تأثيرات أنتجتها الأبحاث العلمية في مجال الذّكاء الاصطناعي، فكيف تعمل هذه التقنية، وما فائدتها، وما نتائجها المحتملة بالنسبة لمنظومة التعليم في ظل تعدد اللغات؟

إن الإجابة عن هذه التساؤلات وغيرها نجدها في أبحاث علم الذّكاء الاصطناعي التي شهدت تقدّما كبيرا بفضل استخدام ما يسمى بالشبكات العصبية الاصطناعية في أربعينيات القرن الماضي (1940)، أين تم التوصل لفكرة محاكاة بنية المخ البشري بما يحتوي عليه من مليارات الخلايا العصبية المتشابكة (نويرونات) في شكل شبكات اصطناعية داخل الكومبيوتر، بيد أنه لم يتحقق شرْطيْ (وجود قدرة حاسوبية عالية بقدر كاف وكذا توفر كميات كبيرة من البيانات) لتحويل تلك الفكرة إلى واقع إلا في الآونة الأخيرة.

3- جهود مركز الملك عبد الله بن عبد العزيز الدولي في حوسبة اللغة العربية

يعدّ مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية أحد المراكز البحثية الهامة؛ ركّز اهتماماته على المجالات التي ما يزال فيها البحث بكرا، فسلّط الضوء عليها وحاول توجيه أنظار الباحثين والمؤسسات الأكاديمية إلى ضرورة الاهتمام بها، من هذه المجالات (التخطيط اللغوي، تعليم اللغة العربية للناطقين بها وللناطقين بغيرها، ...)، ويعدّ مجال (حوسبة اللغة العربية والذكاء الاصطناعي) أهمها، لارتباطها بالتقنيات الحديثة والعالم الافتراضي.

أصدر المركز في السنوات القليلة الماضية ما يناهز ستة عشرة مؤلفا مختّصا في (حوسبة اللغة العربية) و(المدونات اللغوية)، كما صدر له في عام 2019 م سبعة كتب هي (العربية والذكاء الاصطناعي، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، خوارزميات الذكاء الاصطناعي في تحليل النص العربي، مقدّمة

---

[1] - اللغة والحسابية، كريستوفر-س-بتلر، ضمن الموسوعة اللغوية، المجلد الثاني، ص 655.

[2] - عطية سليمان أحمد، اللسانيات العصبية، ص 440.

في حوسبة اللغة العربية، الموارد اللغوية الحاسوبية، المعالجة الآلية للنصوص العربية، تطبيقات أساسية في المعالجة الآلية للغة العربية).

وسيكون تركيزنا في هذه الورقة العلمية سيكون حول مصنّفين اثنين فحسب وهما: مقدّمة في حوسبة اللغة العربية، الموارد اللغوية الحاسوبية.

أولا/ كتاب مقدّمة في حوسبة اللغة العربية:

الباحثون المساهمون في العدد: محسن رشوان، أحمد راغب، محمد عطية، سامح الأنصاري، المعتز بالله السعيد.

الطبعة وسنة النشر: ط1، 2019م

محتوى العدد:

يضمّ هذا العدد مدخلا إلى حوسبة اللغة، يشتمل على تعريف هذا الميدان الحديث من البحث، وما يقابله من تسميات مثل (اللسانيات الحاسوبية Computational linguistics، معالجة اللغات الطبيعية Natural langague processing، وتقنيات اللغة الطبيعية Human langague technologies) وهي تسميات متعددة لمفهوم واحد وهو ذلك العلم الذي تُوَجَّهُ من خلاله أنظمة الحاسوب إلى فهم اللغة الإنسانية ومحاكاة الذكاء البشري.

انطلاقا من هذا التحديد، فإن حوسبة اللغة تقوم على ثلاثة محاور أساسية:[1]

-**تقنيات معالجة النصوص** (Text processing) ومن أمثلتها: الترجمة الآلية، والتلخيص الآلي (عن طريق الاستخلاص Extraction أو التلخيص Abstraction)، والتّنقيب في النصوص (تحليل المشاعر، تلخيص وتصنيف وتجميع النصوص، التصحيح الآلي للنصوص).

-**تقنيات معالجة الكلام المنطوق** (Speech processing) ومن أمثلتها: التعرّف الآلي على الكلام المنطوق، وتحويل النص المكتوب إلى الكلام المنطوق.

-**تقنيات معالجة الصور** (Image processing) ومن أمثلتها: التعرّف الآلي على الكتابة Optical character recognition.

[1] - ينظر: مدخل إلى حوسبة اللغة العربية، محسن رشوان، ضمن كتاب: مقدمة في حوسبة اللغة العربية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م، ص 17.

ثم إن عملية حوسبة اللغة تقتضي مراعاة التدرج في مستويات التحليل اللساني، وهي على التوالي:

1-1- المستوى الفونولوجي Phonological analysis :

يهتم بطريقة نطق الكلمة وحروفها التي تُنطق والتي تُنطق على غير أصلها؛ كالحروف المدغمة والحركات القصيرة مما يجعل إنجاز آلية النطق الآلي لكلمات اللغة العربية مهمة صعبة قياسا مع اللغات الأخرى.[1]

ولأن فونيمات اللغة العربية تخضع لقواعد فونولوجية تحكم ورودها في سياق كلمة أو جملة، مما قد يغيّر في صفته تأثّرا بما يسبقه أو يلحقه، لذلك فإن الوصف الفونيتيكي لأصوات العربية غير كافٍ ليمكّن الحاسوب من تمييز الفونيمات، وعليه يحتاج إلى توصيف آخر يهتم برصد التغييرات التي تلحق فونيمات العربية في سياقها الوظيفي؛ فالحرف المضعّف مثلا تزيد المدة الزمنية التي تستغرق نطقه، والدال في (أردتُ) تنطق تاء بسبب التأثير الرّجعي، وكلمتا (سوط/ وصوت) قد يتطابقان في سياق الأداء النّطقي،...[2]

2-1- التحليل الصّرفي Morphological analysis:

للمستوى الصرفي في اللغة العربية أهمية كبيرة؛ إذ "أن ميكنة الصرف في اللغة العربية يعدّ مدخلا أساسيا وقاسما مشتركا لمعظم نظُمها الآلية،...وتعدّ معالجة الصرف العربي آليا مطلبا أساسيا لميكنة عمليات تحليل النصوص المكتوبة والمنطوقة وفهمها، وتوليدها ذاتيا، علاوة على كونه أساسيا لا غنى عنه لميكنة المعاجم، واسترجاع المعلومات، وتحليل مضمون النصوص".[3]

---

[1] - للتوضيح أكثر، ينظر: التحليل الصوتي، الدراسة التحليلية للصوت اللغوي في العصر الحديث، أحمد راغب، ضمن كتاب: مقدمة في حوسبة اللغة العربية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م، ص65.

[2] - ينظر: المعالجة الآلية للكلام المنطوق، التعرف والآلية، سالم الغزالي، ضمن كتاب: استخدام اللغة العربية في المعلوماتية، المنظمة العربية للثقافة والتربية والعلوم، تونس، 1996م، ص 134.

[3] - اللغة العربية والحاسوب، نبيل علي، مؤسسة تعريب، الكويت، 1988م، ص 297.

يتم في هذا المستوى تحليل الكلمة إلى عناصرها الأساسية (السابق، الجذع، واللاحق) ونظرا للطبيعة الاشتقاقية للغة العربية يتم الاستعانة بالطرق الإحصائية التي تساعد في تحديد أقسم الكلم وما يتعلق بها من توصيفات.[1]

إن المفردة بالنسبة للمشتغلين بمعالجة اللغة[2] حاسوبيا على المستوى الصرفي هي وحدة أساسية للتعامل مع النّص، وتفاديا للوقوع في الالتباس (أيّ احتمال المفردة لأكثر من تأويل)، يتوجّب عليهم الانطلاق من بداية غير ملتبسة، يمكن تسمية نقطة البداية هذه (الكلمة الرقمية الخام crude digital word)؛ وهي عبارة عن تسلسل من الشّفرات يرمزُ كلّ منها إما إلى حرف هجائي عربي أو علامة ضبط صوتي عربي، على أن يحدّ هذا التسلسل الرقمي من كلا طرفيه شفرةٌ لا ترمز حرف هجائي عربيّ ولا علامة ضبط صوتي عربي، لتصبح بعدها الكلمات العربية جاهزة للمعالجة الحاسوبية التي يكون أولها غاليا التحليل الصرفيّ الآلي.[3]

ولأن المعالجة الحاسوبية للكلمة العربية لا تتوقف عند الكلمة الرقمية الخام فحسب، إذ قد لا يتوصل المحلل الصرفي الآلي إلى أي كلمة عربية حقيقية تناظر الكلمة العربية الرقمية الخام لاحتمالين:

الأول: تمييز الكلمات الأجنبية المكتوبة بحروف عربية وما يلحق بها من مشكلات منها التغير الذي يلحق بهذه الكلمات، إذ نجد باستمرار كلمات تَجدُّ وأخرى تختفي وبمعدل عالٍ يزيد من تسارعه الانفتاح الهائل على الاتصالات والعولمة، إذ من الصعوبة إنشاء جداول أو تحديثها بقائمة هذه الكلمات. إضافة إلى هذا هناك مشكلة ترتبط بالهجاء الذي تكتب به كل كلمة من هذا النوع لغياب المرجعية العربية، وعليه فالأمر يبقى مجرد اجتهاد قد يتغير من مترجم إلى آخر، ومن كاتب إلى آخر.[4]

وقد يزداد الأمر صعوبة في حال حصر هذه الكلمات في جداول بسبب دخول اللواصق العربية على الكلمات الأجنبية المكتوبة بحروف عربية، لما لهذا من تغيير في هجائها كقولنا: أَرْمِينْيَا، الأَزْمَنْ، أَرْمَنِيّ/ فرنسا، الفرنسيون، الفرنساويّ.....

ولتجاوز هذه المشكلات، فإن الأنظمة الأكثر نضجا لمعالجة اللغة العربية حاسوبيا تقدم جملة من المقاربات؛ قسم منها يقوم بإجراء تحليلات لغوية لتعيين أقسام الكلم، أو الدلالة، أو النحو، في حين

---

[1] - ينظر: مدخل إلى حوسبة اللغة العربية، محسن رشوان، ص 19.

[2] - ينظر في تعريف مصطلح المعالجة اللغوية: المعالجة الآلية للغة العربية (إنشاء نموذج لساني صرفي إعرابي للفعل العربي)، فارس شاشة، ماجستير، كلية العلوم الإنسانية والاجتماعية، جامعة الجزائر، 2008م، ص13.

[3] - ينظر: التحليل الصرفي الآلي للمفردات العربية، محمد عطية، ضمن كتاب: مقدمة في حوسبة اللغة العربية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م، ص 94.

[4] - من أمثلة الهجاءات المتداولة: Birmingham: بِيرْمِينْجِهامْ، بِرْمِنْجِهَامْ/ Dayton: دَايْتُونْ، دَاتُونْ، دِيتُونْ. ينظر: المرجع نفسه، ص 96.

يؤجل مسألة تمييز الكلمات الأجنبية المكتوبة بحروف عربية، أما القسم الثاني فيخرج المسألة من طبقة التحليل الصرفي إلى طبقة أعلى وهي التطبيق القائم على هذه المعالجة اللغوية، لأنه أحيانا في التطبيق لا يتطلب هذا التمييز، وربما إذا وُجد كان تمييزا جزئيا فقط.[1]

ولتوضيح طريقة المعالجة اللغوية مع هذا النوع من الكلمات، نورد الأمثلة التالية:[2]

1- كلمة عربية رقمية خام:

1-1- عربية سليمة، مصرّفة، منتظِمة: للباحثينَ، فالبلدان، يقرؤونهُ،....

1-2- غير منتظِمة (استثناء صرفي): الله، تقوى،...

1-3- غير مصرّفة (جامدة): هو، الذي، مِنْ،....

1-4- غير مصرّفة (معرّبة): آمين، إبراهيم، استبرق،...

2- غير عربية، أجنبية مكتوبة بحروف عربية: واشنطن، برلمانية،.....

2-1- خاطئة: لنبسمتثص، كشهحصو،.....

آخر ما نشير إليه في هذا العنصر، أن حوسبة البنية الصرفية للكلمة العربية يجب أن تراعى فيها أربع مسائل جوهرية:

1-اللواصق المركبة؛ وهي التي تتركب من أكثر من عنصر واحد، مثل (بالـ...في كلمة بالأدوات)و(فسيـ... في كلمة (فسيعلمون) أو (.....تان في كلمة كاتبتان) و(هوم في كلمة ليساعدوهم).

2-عدم التنافر بين أي جزأين من الأجزاء الثلاثة الأساسية للكلمة (السابقة، الجذع، اللاحقة)؛ كأن تتلاحم السابقة (ال) مع الجذع (صالح) مع اللاحقة (ة) مشكلة الكلمة (الصالحة)، أو التحام السابقة (يـ) مع الجذع (تنافس) مع اللاحقة (ون) لإسناد الفعل المضارع إلى جمع المذكر ورفعه بثبوت النون مشكلة الكلمة (يتنافسون)، في حين لا يمكن أن تلتحم السابقة (يـ) المضارعة مثلا مع جذع الكلمة الاسمي (مصابيح) لحصول تنافر بين الأجزاء.[3]

3-التوافق مع الفونولوجيا العربية؛ لأن البناء الصرفي العربي محكوم بالقوانين الفونولوجية العربية وهي قوانين لا تمنع التعويض المباشر بين الحروف المتغيرة للصيغة الصرفية كالإعلال، والإدغام، والإبدال والحذف.[4]

---

[1] - ينظر: التحليل الصرفي الآلي للمفردات العربية، محمد عطية، ص 97.

[2] - ينظر: المرجع نفسه، ص 99، 100.

[3] - ينظر: المرجع السابق، ص 102.

[4] - ينظر: المرجع نفسه، ص 106.

بناء على هذه العمليات، إذا توافر لدينا جذرٌ، وصيغة صرفية، وسابقة، ولاحقة، نستطيع عبر الصيغة البنائية العامة للكلمة العربية أن نقوم حسابيا بتركيب الكلمة العربية المناظرة لتلك المكوّنات (بشرط عدم تنافرها).

1-3- التحليل النّحوي Syntactic analysis :

يتم فيه تعيين وظيفة الكلمة في الجملة أو التركيب بناء على موقعها الإعرابي، وصولا للمعنى أو الدلالة، والتحليل النحويّ في اللغة العربية يحتاج إلى قرائن كثيرة (مثل بناء الكلمة، ترتيب عناصر الجملة...) للوصول إلى الدلالة المطلوبة نظرا لمرونة العربية في ترتيب عناصر الجملة (تقديم المفعول به على الفاعل، تقديم الخبر على المبتدأ....).[1] فالنحو في اللغة العربية مرتبط ارتباطا وثيقا بمنظومتي الصرف والمعجم.

ويعدّ التحليل النحوي والمعالجة الآلية مقوّما لا غنى عنه، إذ تعتمد عليه العديد من تطبيقات معالجة اللغة مثل التطبيقات الخاصة بتحليل النصوص وفهمها آليا، وعمليات الاستخلاص والتلخيص والفهرسة الآلية، نظم الترجمة الآلية، اكتشاف الأخطاء الهجائية والنحوية آليا.[2]

1-4- التحليل الدّلالي Semantic analysis :

للتحليل الدلالي درجات عدّة؛ أدناها تحديد دلالة الكلمة في سياقها، وأما الدرجات الفرعية الأخرى نكر منها:

-فكّ الالتباس الدّلالي للكلمة مثل: كلمة (عين) فلها من المعاني ما يأتي بمعنى: الجاسوس، وعين الماء، الجارحة،....

-فكّ الالتباس البلاغي ومثاله: رأيت أسودا في المعركة، والقصد: رأيت فرسانا شجعانا في المعركة.

-الفصل بين الموضوعات: ففي أحايين كثيرة نجد في المقالة الواحدة موضوعات عدّة، ولا نجد لها عناوين فرعية تُبين عنها، والفصل هنا يكون ضروريا لتسهيل عملية استرجاع المعلومات وقت الحاجة إليها.[3]

1-5- **التحليل المعجمي Lexical analysis :**

---

[1] - ينظر: مدخل إلى حوسبة اللغة، محسن رشوان، ص 19.

[2] - ينظر: التحليل التركيبي، سامح الأنصاري، ضمن كتاب: مقدمة في حوسبة اللغة العربية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م، ص 137.

[3] - ينظر: مدخل في حوسبة اللغة العربية، محسن رشوان، ص 21.

يساعد على تعيين مكوّنات المعجم اللغوي وتوصيفها على النحو الذي يُمكّن من تمييز مباني اللغة ومعانيها، وهو يجمع المستويات اللسانية السابقة؛ فيهتم بنطق الكلمات ومبانيها، كما يبحث في سياقات المفردات ومعانيها الوظيفية وصولا إلى معانيها المعجمية.

إنه على الرغم مما توفّره التقنيات الحديثة (الحاسوب) من خدمات للغة العربية ولغيرها من اللغات، إلا أن إخضاع اللغة العربية للمعالجة الحاسوبية ما يزال يواجه جملة من التحدّيات، أهمها:[1]

●الكتابة العربية: إذ الكلمة فيها ترد حروفها متّصلة بعضها ببعض، لذلك التعرّف على حروف هذه الكلمات مسألة في غاية الصعوبة إذا قورنت بالنصوص المكتوبة باللغة اللاتينية (تكتب حروف كلماتها منفصلة بعضها عن بعض)، وهذا يعدّ تحديا آخر في تحويل الكلام المنطوق في العربية إلى نص مقروء.

●أسماء الأعلام والكائنات: إن وجود علامة إضافية في كتابة أسماء الأعلام أو الكائنات يساعد في التعرف عليها آليا وهو ما تفتقد إليه العربية قياسا مع غيرها من اللغات التي تميّزها بحروف كبيرة؛ ففي قولنا (زرت القاهرة) لا توجد أي علامة تميّز رسم حروف الاسم (القاهرة) في حين هذا التمييز نجده في الإنجليزية (I have visited Cairo) وهو أمر تستدعيه محركات البحث.

●غياب التشكيل عند بنية الكلمة أو نهايتها: لتمييز النّطق بالكلمة الواحدة؛ كالتمييز في نطق (عين) إن كانت (عَينْ، عَيَّنَ، عُيِّنَ،...) إذ تغيير حركة الحرف يصحب معه تغييرا في المعنى، هذا الأمر يجعل التعامل مع الكلمة العربية حاسوبيا أمرا معقّدا لأن وجود الحركات في الكلمة يمثّل جزء ضروريا من حقيقتها،

وإذا كان الأمر سهلا على القارئ العربي لدرايته بلغته فإن الأمر بالنسبة للحاسوب ليس بالسهولة نفسها مما يزيد الالتباس لديه ويُصعّب المعالجة الحاسوبية.

●كثرة الأخطاء الشائعة كالتمييز بين كتابة الهمزات (ا، أ،إ).

●الهاء والتاء المربوطة كما في كتابة الكلمة (القاهرة) إذ كثيرا ما تكتب (القاهره).

●الأخطاء الإملائية: (شيء) تكتب (شئ، شيئ)

●من الناحية التركيبية فاللغة العربية تجيز التقديم والتأخير بين كلماتها ولهذه الظاهرة قيم دلالية وبلاغية، كذلك الوجود التقديري للكلمة (الضمير المستتر مثلا)، وغيرها، كل هذا يزيد من صعوبة المعالجة النحوية والبلاغية للعربية.

---

[1] - ينظر: المرجع نفسه، ص 24، 25.

ثانيا/ كتاب الموارد اللغوية الحاسوبية:

الباحثون المساهمون في العدد: عبد العاطي هوّاري، سامح الأنصاري، المعتز بالله السعيد، محسن رشوان.

محتوى العدد:

تناول العدد مسائل هامة توزّعت على خمسة فصول يمكن إجمال محتواها في الآتي:

أولا: الموارد المعجمية[1] العربية الحاسوبية: هذه الموارد ناتجة عن دخول الحاسوب إلى مجال العمل المعجمي، وما صحبه من تقنيات جديدة في الصناعة المعجمية.

وتتنوع هذه الموارد من حيث الشّكل وطبيعة المحتوى؛ فمن ناحية الشّكل تُصنّف إلى:

1-1- **المعجم المقروء آليا:** وهو تمثيل حاسوبيّ للمعجم الورقي التقليدي، يظهر في هيئة إلكترونية تسمح للحاسوب بالقيام بمعالجات عليها، هذه المعاجم المقروءة تعدّ موردا قيّما للمعلومات اللغوية المستخدمة في مجال معالجة اللغات الطبيعية.[2]

2-1- **قاعدة البيانات المعجمية:** وهي صياغة للمادة المعجمية في شكل قاعدة بيانات، لما توفره هذه الأخيرة من إمكانات في التخزين، والضبط، والربط العلائقي، وإمكانات البحث، والفهرسة، والاستخلاص وغيرها، ويمكن اعتبار قاعدة البيانات المعجمية هذه مخزنا هائلا للثروة اللفظية يمكن توظيف محتواها في بناء موارد معجمية أخرى، أو برمجيات حاسوبية فيما يخص معالجة اللغات الطبيعية.[3]

أما من ناحية طبيعة المحتوى فهي مصنّفة إلى:

1-1- **الموارد المعجمية:** وهي موارد تقوم على تمثيل المعنى وفق منهجيتين: الأولى تعريفية Dictionary تعتمد التعريف المعجمي في تمثيل المعلومات والمعارف.

---

[1] - يمكن التمييز بين ثلاث مستويات بحثية معجمية: الأول يسمى المعجم الذهني Mental lexicon يدرس الجانب الذهني من المعجم مثل اكتساب الثروة اللفظية وطريقة انتظامها في الذهن وآليات توليدها واستعمالها. الثاني المعجم اللغوي lexicon وهي الثروة اللفظية الموجودة عند جميع المتحدثين باللغة الواحدة. أما الثالث المعجم المُدوَّن Dictionary وهو محاولة لتمثيل المعجم اللغوي للغة ما في صورة مورد معجمي فهو ينتمي إلى الصناعة المعجمية. ينظر: المعجم العربي، نماذج تحليلية جديدة، عبد القادر الفاسي الفهري، توبقال للنشر، الدار البيضاء،1999م.

والمعجمة والتوسيط، عبد القادر الفاسي الفهري، المركز الثقافي العربي، الدار البيضاء، 1997م.

[2] - ينظر: الموارد المعجمية العربية الحاسوبية، عبد العاطي هواري، ضمن كتاب: الموارد اللغوية الحاسوبية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م، ص 14، 15.

[3] - ينظر: المرجع نفسه، ص 16.

والثانية علائقية شبكية تعتمد في بنائها وتمثيلها للمعنى على مقاربة العلاقات الدلالية القارة في المعجم اللغوي، كما أنها توظّف تقنيات العمل الحاسوب في توثيق العلاقات بين مفردات وعبارات المعجم.[1]

2-1- **موارد معجمية دلالية:** وهي موارد تقوم على تمثيل دلالي تركيبي للوحدات المعجمية، تهدف إلى تمثيل الأبنية التركيبية والمحتوى الدلالي للوحدات المعجمية خصوصا الأفعال والمشتقات فتقوم بتصنيفها تصنيفا تركيبيا ودلاليا في آن. وتكمن أهميتها الكبيرة في تطبيقات معالجة اللغات الطبيعية لارتباطها بالسياقات التركيبية المحتملة للكلمة.[2]

لقد أسهمت الحاسوبية في صناعة معاجم تقوم على أسس مفاهيمية حاسوبية تتعدى مجرد استخدام الأدوات الحاسوبية في التخزين أو التحليل أو الاستدعاء؛ إذ ساعدت البرامج الحاسوبية في إجراء عمليات الجمع والإحصاء والفهرسة والتحليل والتصنيف، ليقوم المعجمي بعدها بعمليات تحليل النصوص، والسياقات التي وردت فيها الكلمة وصولا للدلالات، ثم تصنيفها إلى دلالات مركزية وأخرى هامشية، كما أن هذه البرامج تسهّل كثيرا عملية تنظيم المادة اللغوية وتخزينها حاسوبيا في قاعدة بيانات معجمية.

ثانيا: المدوّنات اللغوية: وتسمى أيضا لسانيات المدوّنة Corpus linguistics تهتم بالبحث في الظواهر اللغوية وتفسيرها من خلال مجموعة من النصوص التي تمثّل الواقع اللغوي، وهي منهجٌ حديث نسبيا، أما أداته فهي المدونة اللسانية مكتوبة أو منطوقة التي يمكن التعامل معها آليا والتحكم في بياناتها ومدخلاتها.[3]

ويخضع اختيار نصوص المدونات اللغوية عند بنائها إلى إحدى الطرائق الثلاثة الآتية:[4]

1-1- **الاستبانة** Questionnaire يعتمد فيها صُنّاع المدونة على توجيه مجموعة من الأسئلة على أشخاص يمثّلون المجتمع اللغوي لهذه النصوص، وتركز الاستبانة على حقول معرفية معينة (عناوين الكتب المختارة، أسماء الكتاب والمصنفين في الاختصاص، ...)، وتعد هذه الطريقة أكثر الطرائق شيوعا في المدونات الموضوعة لأغراض تعليمية وتربوية.

---

[1] - ينظر: المرجع نفسه، ص 17.

[2] - ينظر: المرجع نفسه، ص 17، 18.

[3] - ينظر: المدونات اللغوية العربية بناؤها وطرائق الإفادة منها، سلطان ناصر المجهول وآخرون، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م، ص 19. والمدونات اللغوية ودورها في معالجة النصوص العربية، أيمن الدكروري، دار وجوه للنشر والتوزيع، ط1، المملكة العربية السعودية، 2018م، ص 38، 39.

[4] - ينظر: الموارد المعجمية العربية الحاسوبية، عبد العاطي هواري، ص 55

2-1- **الحصر الشامل للنصوص** Comprehensive Inventory هي أشمل من الاستبانة لأنها تُلزم صنّاع المدونة بحصر نصوص المجتمع اللغوي جميعها دون استثناء شيء، فهي تقوم على عملية مسحٍ شاملة في بناء نصوص المدونات.

3-1- **العيّنات الإحصائية** Statistical Sampling أكثر الطرائق شيوعا، يقوم صُنّاع المدونة اللغوية باختيار عينة من النصوص التي تتفق واهدافهم البحثية، عشوائية كانت أم غير عشوائية، وتعتمد الطريقة أساليب التحليل الإحصائي في بناء المدونات اللغوية، ويشيع استخدامها في صناعة المعجمات عموما (عامة، تاريخية،..)، وفي ميادين حوسبة اللغة التي تعنى باسترجاع المعلومات أو الترجمة الآلية.

وتُصنّف المدونات اللغوية العربية باعتبار هيئة النصوص إلى:[1]

1-1- مدونات لغوية نصية: Texte corpora وتستمد مادتها من مصادر مكتوبة كالصحف، والمجلات، والكتب المطبوعة، والوثائق المخطوطة وغيرها، يغلب عليها أن تكون تمثيلا لمستوى اللغة الفصحى، تستخدم في الدراسات اللغوية التاريخية وبناء المعجمات، والتعرّف الآلي على الكلام المكتوب.

2-1- مدونات لغوية منطوقة: Spoken Corpora /Speech تستمد مادتها من مصادر منطوقة، كالأفلام، والمسلسلات الإذاعية، نشرات الأخبار والمحادثات الهاتفية وغيرها، يغلب عليها أن تكون تمثيلا لمستوى اللغة الدّارجة، تستخدم في الدراسات اللغوية الوصفية وصناعة الأطالس اللغوية، والتعرّف الآلي على الكلام المنطوق.

وتُصنّف المدونات اللغوية باعتبار تعدّد اللغة إلى:[2]

1-1- **مدونات أحادية اللغة:** Monolingual Corpora تستمد مادة نصوصها من لغة واحدة، ويغلب عليها أنها تمثّل مستوى لغويا واحدا (اللغة الفصحى أو الدّارجة)، تستخدم في بناء المعاجم أحادية اللغة.

2-1- **مدونات ثنائية اللغة:** Bilingual Corpora تستمد مادة نصوصها من لغتين تنتميان إلى فصيلة لغوية واحدة او فصيلتين، تستخدم في بناء المعاجم ثنائية اللغة وتطبيقات الترجمة الآلية وتعليم اللغات.

3-1- **مدونات متعددة اللغات:** Miltilingual Corpora تستمد نصوصها من لغات عدة.

---

[1] - ينظر: المرجع نفسه، ص 61.

[2] - ينظر: المرجع السابق، ص 62.

وتُصنّف المدونات اللغوية بحسب توافق النصوص إلى:[1]

1-1- **مدونات لغوية متوازية:** Parallel Corpora تستمد نصوصها من لغتين أو أكثر، وتكون النصوص أصلا في إحدى هذه اللغات وتسمى (اللغة المصدر)، وترجمة في اللغة أو اللغات وتسمى (اللغة أو اللغات الهدف)، يلتزم فيها صناعها ببعض الضوابط لتيسير معالجتها وفقا للغرض الذي وُضعت من أجله المدونة.

2-1- **مدونات لغوية غير متوازية:** Non-parallel Corpora وهي مدونات متقاربة أو متقابلة، تستمد نصوصها من لهجات اللغة الواحدة، أو لغتين، أو مجموعة لغات، تكون النصوص أصلا غير مُترجم في أيّ من اللغات التي تضمها المدونة، وهي نادرة الاستخدام لندرة النصوص المتقاربة بين اللغات.

وتُصنّف المدونات اللغوية باعتبار طبيعة النصوص إلى:[2]

1-1- **مدونات مخصصة:** Specialized Corpora تستمد مادة نصوصها من حقل معيّن أو مجموعة معينة من الحقول، كالحقول المعرفية (علمية، قانونية، ...)، أو تاريخية (العصور الأدبية مثلا)،وغيرها، وقد يقتصر جمع النصوص فيها على اللغة المكتوبة أو المنطوقة، كما قد تُجمع من نصوص أديب أو كاتب معين (النصوص المسرحية لشكسبير مثلا).

2-1- **مدونات عامة:** General Corpora لا تتقيّد بنوع معيّن من النّصوص بل تتنوع مادتها لأنها مستمدة من حقول مختلفة (معرفية وتاريخية وجغرافية)، كما أنها قد تجمع بين نصوص اللغة المكتوبة والمنطوقة، أو مستويات لغوية متعددة.

ولفهرسة نصوص المدوّنات اللغوية التي سبق الحديث فيها، يمكن اتباع آليتين؛ **الأولى** تُعرف ب**الفهرسة الألفبائية**، تقوم على ترتيب المفردات ألفبائيا دون النظر في السوابق واللواحق، ودون مراعاة الطبيعة الاشتقاقية للغات وعلى رأسها اللغة العربية.

أما **الثانية** فتُعرف ب**الفهرسة الجذعية**، والجذع جزء من الكلمة، يكون مشتقا أو جامدا، وهو ناتج عن اتحاد المورفيمات المكوّنة لبنية الكلمة الأساسية (عاود ← فعاودتُهُم/ مكتب ← مكتبات). تتيح هذه الآلية البحث عن جذوع الكلمات بعد تجريدها من السوابق واللواحق.[3]

خاتمة:

---

[1] - ينظر: المرجع نفسه، ص 63.

[2] - ينظر: المدوّنات اللغوية، المعتز بالله السعيد، ضمن كتاب: الموارد اللغوية الحاسوبية، ص 65.

[3] - ينظر: المرجع السابق، ص 77.

-للمخ البشري قدرة فائقة على معالجة المعلومة تفوق قدرة الحاسوب، لأن بداخله برامج يعجز الحاسوب بكل برامجه أن يجمع بينها، ويُنسّق ويخلق أنساقا جديدة وغير متوقعة .

-إن اقتراح علماء النفس الاتصالي صنع نموذج لوظائف الدماغ هو محاولة منهم لتطبيق آلية صناعة الأنماط الحاسوبية على النماذج الدماغية، وهو أمر بيّن ضيّق أفقهم وقِلة عِلمهم بكيفية عمل الدماغ الذي يختلف تماما عن عمل الحاسوب، إذ تعمل الحاسوبية وفق برامج محددة وضعها فيها صانعوها مسبقا قبل أن تعمل، أما النماذج العصبية الموجودة في الدماغ فتعمل وفق برامج غير محددة ولا نعرف بدقة كل شيء عنها وعمّا يؤثر في عملها أو تعمل وفقه.

-الإمكانيات الخاصة بالنماذج الدماغية تختلف عن الحاسوب الذي تُسجل فيه برامجه على أُلوحٍ (رامات) فإذا حدث تلفٌ في جزء من البرنامج فإن البرنامج يسقط كليّا ولا يعمل، أما الدماغ إذا أصيب جزء منه أو حدث مرضٌ ما (الزهايمر مثلا) فإن البرامج المسجلة في الدماغ لا تسقط كليا، بل تعمل بكفاءة أقل حسب نوع المرض أو الإصابة.

-تحليل للفونيمات العربية لا بد أن يتجاوز التوصيف الفونيتيكي إلى التوصيف الفونولوجي الذي يرصد التغيرات التي تلحق بالفونيمات في سياقها الوظيفي مما يساعد على معالجتها حاسوبيا.

-من التطبيقات الحاسوبية التي تنتمي إلى التحليل الصرفي: الربط المعجمي العربي الآلي بإضافة مداخل أخرى لشرح الصيغ الصرفية للمفردة المراد شرحها.

-يساعد التحليل الصرفي أيضا على التحليل الدلالي والمعجمي بالكشف عن العلاقات الدلالية بين المفردات.

-يوفر التحليل الصرفي البرمجيات المعتمدة في تعليم اللغات (الهجاء، الإملاء،...).

-تتعدد روافد تطوير العمل المعجمي الحاسوبي بين الروافد اللسانية والروافد الحاسوبية، فالأول يشكّل التأسيس النظري في مجال بناء هذه الروافد، والثاني تطبيقي يقوم بالمعالجة الحاسوبية لهذه الروافد.

-يتم بناء معجم لساني حاسوبي باعتماد خطوات هي: التصميم، التحليل المعجمي الدلالي، التحليل الصرفي، التحليل التركيبي.

-من الملاحظات المسجلة على العمل المعجمي العربي الحاسوبي مشكلات في التأسيس النّظري، مركزية الصرفي في المعالجات الحاسوبية للمعجم وطغيان فكرة (الجذر، الجذع، المجرد المزيد،...)، التركيز على الجوانب القارة للمعجم العربي بالتركيز على سرد المعاني أو الدلالات الخاصة بالكلمات وإهمال الجوانب

الديناميكية التوليدية التي تهدف إلى توليد دلالات جديدة في سياقات جديدة، إضافة إلى تقليدية المقاربة الحاسوبية للمعجم التي تظهر في مقولات التصنيف الصرفية والتركيبية والدلالية المعتمدة.

-إن تحديث المفاهيم اللغوية الحاسوبية يجب أن يسبق التحديث التقني، وليس العكس، ثم إن حل المشكلات النظرية حلاّ نظريا في المعجم التقليدي ستظهر آثاره في المعجم الحاسوبي؛ فالمعالجة الحاسوبية للمعجم مرآة تعكس المنجز التنظيري.

-تستخدم المدونات اللغوية في معالجة اللغات الطبيعية عن طريق جملة من التطبيقات منها: الترجمة الآلية، آلية التطبيق الإملائي، آلية تشكيل النصوص، آلية فكّ الالتباس الدلالي.

بيبليوغرافيا المراجع المعتمدة:

1- دروس في اللسانيات التطبيقية، صالح بلعيد، دار هومة، الجزائر.

2- الفلسفة في الجسد، الذهن المتجسد وتحديه للفكر الغربي، جورج لايكوف، مارك جونسون، ترجمة وتقديم: عبد المجيد جحفة، دار الكتاب الجديدة المتحدة، ط1، بيروت، 2016م.

3- اللغة والحسابية، كريستوفر-س-بتلر، ضمن الموسوعة اللغوية، المجلد الثاني.

4- اللسانيات العصبية اللغة في الدماغ (رمزية، عصبية، عرفانية)، عطية سليمان أحمد، الأكاديمية الحديثة للكتاب الجامعي، القاهرة، 2019م.

5- اللغة العربية والحاسوب، نبيل علي، مؤسسة تعريب، الكويت، 1988م.

6- اللغة العربية ووسائل الاتصال الحديثة، وليد إبراهيم الحاج، دار البداية ناشرون وموزعون، عمان، 2011م.

7- المدونات اللغوية ودورها في معالجة النصوص العربية، أيمن الدكروري، دار وجوه للنشر والتوزيع، ط1، المملكة العربية السعودية، 2018م.

8- مقدمة في حوسبة اللغة العربية، مجموعة باحثين، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م.

9- المعالجة الآلية للغة العربية (إنشاء نموذج لساني صرفي إعرابي للفعل العربي)، فارس شاشة، ماجستير، كلية العلوم الإنسانية والاجتماعية، جامعة الجزائر، 2008م.

10- المعالجة الآلية للكلام المنطوق، التعرف والآلية، سالم الغزالي، ضمن كتاب: استخدام اللغة العربية في المعلوماتية، المنظمة العربية للثقافة والتربية والعلوم، تونس، 1996م.

11- المعجم العربي، نماذج تحليلية جديدة، عبد القادر الفاسي الفهري، توبقال للنشر، الدار البيضاء،1999م.

12- المعجمة والتوسيط، عبد القادر الفاسي الفهري، المركز الثقافي العربي، الدار البيضاء، 1997م.

13- الموارد اللغوية الحاسوبية، مجموعة باحثين، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، ط1، المملكة العربية السعودية، 2019م.