

2023
فائز بتصدي
الربيع AR10 للتأليف السريع

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً



إبراهيم حبيبي

الفهم الآلي للغة العربية تحليل الأراء والمشاعر نموذجاً

تأليف

إبراهيم حبيبي

1445 هـ - 2024 م

الفهم الآلي للغة العربية: تحليل الآراء والمشاعر نموذجاً

جميع حقوق الطبع محفوظة ©

1445 هـ - 2024 م

الطبعة الأولى

اسم الكتاب: الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

ابراهيم حبيبي

مقاس الكتاب: 7" * 10"

عدد الصفحات: 172

المدقق اللغوي: فاطمة الزماحي

التنسيق الفني: عدنان قلعه جي

رقم التسجيل الدولي ISBN :



arid.my | info@arid.my

رعاة الكتاب



المحفل العلمي الدولي
www.almahfal.org



منصق أريد العلمي
www.arid.my



جامعة أريد الدولية
university.arid.my



أكاديمية أريد لعلوم الفضاء
falak.arid.my



وقف الذكاء الاصطناعي
www.waqf.ai



القرية العلمية
sci-village.com



أبصر فايزر للمخيمات الافتراضية
www.abservisior.com



منصق الفاتحة
www.fatiha.id



منصق أبناء العلماء
sos.arid.my



منصق قادرة
www.qadirah.com



أبصر للتعليم الإلكتروني
www.abser.org



بوابة فلسبي للدفع الإلكتروني
www.filspay.com

تقديم منصة أريد العلمية

الحمد لله الذي أنزل القرآن هدى للعالمين، وجعل العلم مفتاحاً للغيب، ووسيلةً لإدراك الكون وأسراره، والصلاة والسلام على خير الأنام محمد بن عبد الله، الذي بعثه الله بالهدى ودين الحق ليظهره على الدين كله، وعلى آله وصحبه ومن تبعهم بإحسان إلى يوم الدين.. أمّا بعد:

يُشرفنا في منصة أريد العلمية، الرائدة في تعزيز البحث العلمي ونشر المعرفة، أن نقدم كتاب "الفهم الآلي للغة العربية / تحليل الآراء والمشاعر نموذجاً" كإضافة قيمة إلى المكتبة العلمية العربية. هذا الكتاب، الذي نال شرف الفوز في "تحدي التأليف السريع خلال 24 ساعة" المنظم من قبل المنصة للفترة من 1 إلى 3 تموز 2023م، وتم الإعلان عن فوزه في حفل ختام المحفل العلمي الدولي الثالث عشر يوم 3 نوفمبر 2023م، يعكس التزامنا بتحفيز الإبداع والتميز العلمي.

الكتاب يُعالج بعمق موضوع اللغة العربية وتفاعلها مع النظم الحاسوبية، مُسلطاً الضوء على اللسانيات الحاسوبية والتحليل الآلي للمستويات اللسانية، وبذلك يُقدم مساهمة هامة في فهم كيفية تطبيق التكنولوجيا في تحليل اللغات الطبيعية وخاصة اللغة العربية. كما يستعرض الكتاب تطبيقات الفهم الآلي للغات الطبيعية، مع التركيز

على تحليل الآراء والمشاعر فى النصوص العربية، مما يُبرز أهمية الجمع بين العلوم التقنية والإنسانية فى معالجة التحديات المعاصرة.

نسأل الله تعالى أن يجعل هذا العمل خالصاً لوجهه الكريم، وأن يكون مصدر إلهام للباحثين والمهتمين بتوظيف التقنيات الحديثة فى خدمة اللغة العربية وتطوير مجالات البحث العلمى والتطبيقى. كما ندعو الله أن يُعيننا وإياكم على مواصلة العمل والإبداع فى سبيل تحقيق المزيد من التقدم والازدهار فى عالمنا العربى والإسلامى.

د. سيف السويدي

رئيس منصة أريد العلمىة

الملخص:

يعتبر الفهم الآلي للغات الطبيعية مجالاً من المجالات التي انبثقت عن تلاقي العديد من الحقول المعرفية التي جعلت من اللغات الطبيعية وإشكالياتها موضوعاً لها كاللسانيات الحاسوبية، وعلوم الحاسوب بالإضافة إلى الذكاء الاصطناعي، حيث يهتم مجال الفهم الآلي للغات الطبيعية بتطوير العديد من الأنظمة والتطبيقات التي تسعى إلى جعل الحواسيب ذات قدرة على فهم ومعالجة اللغة الطبيعية بطريقة مشابهة للإنسان. وكذا تمكين الحواسيب والهواتف الذكية من قراءة وتفسير مختلف النصوص ومحاولة التوصل إلى معانيها الدقيقة، كما يمكن من استخلاص الآراء التي يعبر عنها الإنسان وكذا والمشاعر التي تخالجه ويتقاسمها في وسائل التواصل الاجتماعي.

إن قراءتنا الاستقصائية في ما توفر من مؤلفات وإنتاجات في موضوع المعالجة الآلية والفهم الآلي بينت أن اللغة العربية استطاعت أن تلحق ركب اللغات الطبيعية الأخرى، حيث تمكنت العديد من المعاهد والمؤسسات الحكومية تطوير نماذج وأدوات تمكن الحواسيب من فهم النصوص المكتوبة باللغة العربية واستخلاص ما تضمنته من معلومات، وتحديد دلالات ومعاني الكلمات، بالإضافة إلى الكشف عن مستويات التفاعل الذي يتجلى في مختلف نظم الإجابة التلقائية على الأسئلة المطروحة، كما تم التمكن من بلورة نظم دقيقة تساعد على تحديد معاني ودلالات الكلمات المستعملة، وكذا تحليل واستخلاص الآراء والمشاعر المعبر عنها في فضاءات التواصل الاجتماعي،

وكذا تقديم تفسيرات للنصوص العربية بشكل تطبعه الدقة والفعالية، إلى غيرها من المجالات التي تتطلب الفهم الدقيق للغة العربية ومعالجتها بشكل آلي.

نقدم في هذا الكتاب بعض ما توصلت إليه الأبحاث والتجارب الخاصة باللغة العربية في إطار حلق اللسانيات الحاسوبية، منطلقين من تحديد مميزات وخصائص اللغة العربية، ثم الإشارة إلى المستويات اللسانية العربية التي يسرت سبل نمذجة اللغة العربية في الحاسوب، الشيء الذي حول بلورة العديد من التطبيقات الخاصة بالمعالجة الآلية للغة العربية، ثم التأسيس إلى الفهم الآلي للغات الطبيعية والإشارة إلى بعض النظم المرتبطة بفهم اللغة العربية آلياً، مركزين على مجال تحليل الآراء والمشاعر المكتوبة بها، حيث عرضنا الخطوات التي تقطعها نظم تحليل الآراء والمشاعر، انطلاقاً من أحد التجارب البحثية العربية التي اعتمدت على الموارد اللسانية العربية في بلورة نظامها.

الإهداء

أهدى ثمرة جهدى المتواضع

إلى روح جدتى الطاهرة، طيبة القلب، تغمدها الله برحمته؛

إلى منبع الحنان وملاذ الضيق فى كل حال أمى الغالية؛

إلى خيرة الرجال مكابد الصعاب أبى أطال الله عمره؛

إلى زوجتى؛ رفيقة دربى؛

إلى قررة عنبى محمد ياسين؛

إلى سندى على الزمان، إخوتى حفظهم الله؛

إلى من دعمنى للمضى قدما فى حياتى أخوالى وخالاتى؛

إلى من حفزنى وشجعنى للمضى قدما أساتذتى وزملاء عملى؛

إلى معلى وأستاذى وموجهى ومشرفى ومصباح طرىقى الدكتوران عبد الحميد الجهاد

والمحمد اسماعىلى علوى

إلى كل الأقارب والأصدقاء؛

إلى كل أبناء وطنى الغالى المغرب.

شكر وتقدير

أتوجه بخالص الشكر وعظيم الامتنان لمنصة أريد التي كانت مصدر تحفيز وتشجيع في الخوض في التحدي تأليف هذا الكتاب، فمن من خلال منصبها المبتكرة والمهمة، تمكنت من تجاوز حدودي الشخصية وتفجيري طاقتي وتحدي الصعاب من أجل إتمام هذا الكتاب.

كما لا يفوتني أن أتقدم بالشكر الخاص للدكتور بسام مصباح الأغبى على تحكيمه الدقيق لهذا العمل المتواضع الذي أضفى عليه قيمة لا تقدر، فأنا ممتن لتفانيه وعطائه، كما أقدر جهوده الكبيرة والفعالة في دعم وتعزيز مستوى العلم والتعليم والمعرفة.

كما أود التعبير عن شكري الجزيل وتقديري العميق للأستاذة الدكتورة أميرة سمبس على حسن تواصلها القيم ومتابعتها الدائمة لكل المشاركين في التحدي.

كما أتقدم بالشكر الجزيل إلى أفراد عائلتي الكريمة على دعمهم وتشجيعهم المستمر، حيث لم أكن لأتمكن من تحقيق هذا الإنجاز المتواضع دون دعمهم الثابت وحبهم اللا محدود.

وأعرب عن امتناني العميق لأساتذتي الأعزاء الذين ساهموا في بناء قدراتي وتوجيهي خلال رحلتي العلمية والأكاديمية. فبفضل علمهم ومتابعتهم وتوجيهاتهم القيمة، استطعت توسيع آفاتي والنمو في مجالي الأكاديمي.

فهرس المحتوى

6	الملمخص:
8	الإهداء
9	شكر وتقدير
10	فهرس المحتوى
12	فهرس الأشكال:
12	فهرس الصور:
12	فهرس الجداول:
13	مقدمة:
17	الفصل الأول:
17	اللغة العربية والنظم الحاسوبية.
18	المبحد الأول: اللغة العربية واللسانيات الحاسوبية.
18	مدخل:
42	المبحد الثاني: اللغة العربية والتحليل الآلي للمستويات اللسانية.
42	مدخل:

77	المبحث الثالث: بعض النظم الحاسوبية العربية
77	مدخل:
94	خلاصة:
94	خاتمة:
96	الفصل الثاني: الفهم الآلي للغات الطبيعية: تحليل الآراء والمشاعر العربية
97	المبحث الأول: تطبيقات الفهم الآلي للغات الطبيعية:
97	مدخل:
104	المبحث الثاني: تحليل وتقييم الآراء والمشاعر ضمن التجارب العربية
104	مدخل:
138	خلاصة:
139	المبحث الرابع: خطوات تحليل الآراء والمشاعر من خلال نظام (IERA) لتحليل الآراء والمشاعر:
139	مدخل:
148	الخاتمة:
151	المصادر والمراجع

فهرس الأشكال:

- 1 الشكل 1: بنية الكلمة حسب دافيد كوهين
- 2 الشكل 2: المحللات الصرفية العربية حسب الخرج
- 3 الشكل 3: خطوات مرحلة التحليل في المحللات الصرفية العربية
- 3 الشكل 4: مراحل إعداد المعجم العربي الحاسوبي

فهرس الصور:

- 7 الصورة 1: نموذج لتحليل جملة بواسطة نظام الخليل
- 9 الصورة 2: مثال لتصريف الفعل "رَجَحَ" في نظام تصريف الأفعال في اللغة العربية
- 2 الصورة 3: واجهة موقع شكلي
- 2 الصورة 4: اهتمامات مجال الفهم الآلي للغات الطبيعية

فهرس الجداول:

- 9 الجدول 1: عدد الدخلات المعجمية لمدونة ووردنت إصدار 3.0
- 3 الجدول 2: مثال تم تحليله بواسطة النظام
- 4 الجدول 3: التعليق بعد القيام بعملية المعالجة القبلية
- 4 الجدول 4: جذوع الكلمات التي تتركب منها التعليق
- 5 الجدول 5: تحديد الحقول المعجمية لكل مدخل معجمي
- 5 الجدول 6: تحديد المعلومات المتعلقة بالحروف التي تؤثر في توجيه الكلام العربي

مقدمة:

حظيت اللغة العربية كباقي اللغات الطبيعية باهتمام الباحثين اللغويين قدماء ومحدثين منذ بداية التفكير اللساني في الحضارة العربية إلى يومنا هذا؛ فقد عمل القدماء على وضع قواعدنا النحوية والصرفية والصواتية، وطريقة اشتقاق الكلمات وتقييد معانيها بوضع الحركات فوق وتحت حروفها. ولم تضع العرب علامات الشكل اعتباراً، بل لغرض التمييز الدلالي والنحوي للكلمات العربية الغنية، وهذا التمييز يحدده السياق التركيبي والدلالي للكلمة وهي مجتمعة مع كلمات أخرى في جملة مفيدة. ونظراً للتلاقح بين العلوم وحاجة علم إلى علم آخر لدراسة إشكالية بحثية معينة، فإننا نشهد تطورات العديد من الدراسات بما فيها الدراسات اللغوية التي تهتم باللغات الطبيعية والتي كانت في حاجة ماسة إلى الاستعانة بحقول معرفية أخرى، لدراسة بعض الظواهر اللغوية، سميت بالعلوم المعرفية التي جاءت نتاجاً للنظريات الحديثة، التي اتخذت اللغة موضوعاً أساساً لتطوير أبحاثها، باعتبارها وسيلة فعالة لتحقيق التواصل بين مختلف الشعوب.

تشكل العلوم المعرفية من مجموعة من التخصصات، من بينها، علم الاجتماع، علم النفس، علم الأعصاب، الرياضيات، الفلسفة، علم الحاسوب، اللسانيات، وتركز اهتمامها على دراسة الدماغ البشري والكيفية التي يتم بواسطتها تمثل المعرفة الإنسانية، ومن ثم تحليلها وفهمها، كما أنها اهتمت بالآليات التي يتم اعتمادها لتمثيل وتوليد هذه المعرفة.

إن الاهتمام الذي حظيت به اللسانيات باعتبارها علماً يهتم بدراسة اللغة الطبيعية مكنها من الانفتاح على العديد من العلوم، والاستفادة مما توصلت إليه هذه الأبحاث من نتائج إيجابية مثل علم النفس، وعلم الاجتماع وعلم الجغرافيا، وكذلك علم الحاسوب، هذا الأخير سهل على اللسانيات استثمار نظرياتها ونتائج أبحاثها بشكل عملي، فخلق برامج لسانية حاسوبية أضحت البشرية في حاجة ماسة إليها، ولقد زاد الاهتمام بهذا الفرع الجديد من اللسانيات نظراً للصعوبات التي تطرحها معالجة اللغة آلياً، ثم فهمها، وكذا إعادة توليدها، إلا أن الاهتمام قد انصب نحو المعالجة الآلية للغات الطبيعية منذ السنوات الخمسين والستين من القرن الماضي بما عرف

باللسانيات الحاسوبية، وبعدها جاء الاهتمام بكل من فهم وتوليد اللغة الطبيعية خاصة، وهذا ما نشهده في يومنا هذا بما يسمى ببرامج الدردشة أو برامج الذكاء الاصطناعي التي أحدثت ثورة في مجالي فهم وتوليد اللغات الطبيعية بما فيها اللغة العربية ومن بين هذه البرامج وأشهرها نجات جيبتي (Chat-GPT).

من هذا المنطلق تعد اللسانيات الحاسوبية علماً يجمع بين ما هو لساني وما هو حاسوبي؛ وتعنى بالبحث في الطريقة الأنسب لتحليل اللغة آلياً أساساً، ثم الفهم وبعده التوليد، هذه العملية المعقدة لا تقتصر على الحاسوب باعتباره جهازاً فقط، بل يتطلب إلى جانب ذلك، الدراية بعلوم أخرى مثل الرياضيات والمنطق وعلم النفس، وعلوم أخرى... وقد تشعبت مضامين هذا التخصص اللساني الحاسوبي، الدقيق وتعددت موضوعاته، ليشمل مجموعة من القضايا، الكبرى المتمثلة في: المعالجة آلياً، والفهم الآلي، والتوليد الآلي، ويندرج ضمن كل قضية مجموعة من التطبيقات الحاسوبية كالتحليل الصرفي والتركيبي، والترجمة الآلية، والتلخيص الآلي، والتشكيل الآلي، إلى غيرها من التطبيقات.

لقد حاول الدارسون العرب إدخال اللغة العربية إلى الحاسوبية، وإخضاعها للمعالجة الآلية كباقي اللغات الأخرى، إلا أن هذه المعالجة لم ترق في بداياتها إلى المستوى المطلوب، لأن اللغة العربية لغة غنية من ناحية اشتقاق جذورها وتشكيل كلماتها، وتركيب جملها، بالإضافة إلى علامات الشكل التي توضع فوق وتحت الكلمات، مما ينتج عنه تعدد المعاني الدلالية والنحوية للكلمة الواحدة أحياناً، وفي هذا الإطار تعد اللغة العربية لغة غنية المعاني، إلا أن هذا الغنى يطرح مجموعة من الإشكاليات في مجال اللسانيات الحاسوبية، خاصة فيما يتعلق بالمعالجة آلياً، أو الفهم الآلي، أو التوليد الآلي للغة العربية، وقد زادت هذه الإشكاليات بازدياد الإنتاج النصي الرقمي الذي نشهده في عصرنا هذا، خاصة بعدما تطورت تقنيات الويب، إذ أن الأبحاث اللسانية الحاسوبية قد مكنت من تحقيق الغنى المعرفي في ما يتعلق بالمعالجة الآلية للغات الطبيعية، بينما لا زال الباحثون يسعون إلى تطوير وبلورة نظم تعمل على فهم وتوليد اللغات الطبيعية، بما فيها اللغة العربية.

شهدت الشبكات التفاعلية بما فيها شبكات التواصل الاجتماعي، والمننديات والمواقع تجارية وغيرها، نموا ملحوظا في السنوات الأخيرة وأصبحت مكوناً مؤثراً في الحياة الاجتماعية، والثقافية، والاقتصادية، وحتى السياسية لما تحمل بياناتها الضخمة من معلومات قيمة يمكن أن تكون ذات فائدة كبيرة. وأصبح الكشف عما يحمله النص من بيانات وآراء ومشاعر أمراً بالغ الأهمية، وهو ما يستدعي الاعتماد على الأدوات التي تمت بلورتها بفضل الذكاء الاصطناعي لفهم تلك البيانات النصية واستثمارها في إعداد نظم توليد نصوص أخرى، حيث إن جمع البيانات واستخلاص الآراء والمشاعر تعتبر من بين أهم أولويات المعاهد والمؤسسات الحكومية والأحزاب السياسية لما تخدمها في تنمية مقترحاتها وتطوير أهدافها، وتحظى اليوم مواضيع مجال الفهم الآلي للغات الطبيعية، خاصة تحليل الآراء المشاعر بانتباه العديد من الباحثين والشركات، نظرا للخدمة القيمة التي يوفرها.

تعتبر مهمة تصنيف الآراء والمشاعر المتواجدة في صفحات الويب من المهام الرئيسية في مجال الفهم الآلي للغات الطبيعية، حيث تمكن نظم تحليل الآراء والمشاعر من تحديد قطبيتها الإيجابية أو السلبية أو المحايدة، استنادا للألفاظ والتعابير المعبر بها عن الآراء والشاعر حول موضوع معين (كيان، منتج، شخصية، إلخ)، وذلك للتصرف واتخاذ القرارات المناسبة. وتتطلب مهمة تصنيف الآراء التوفر على موارد لسانية متخصصة، وبالتحديد معاجم متخصصة تشتمل على كلمات توظف للتعبير عن الآراء والمشاعر. أنجزت العديد من الأبحاث في هذا المجال باللغات اللاتينية وكذا اللغة العربية رغم طبيعتها وخصائصها الغنية وقواعدها المتشعبة ووجود لهجات متعددة في الدول العربية التي شكلت في البداية عقبة كبيرة في التمكن من بلورة نظم تمكن من استخلاص الآراء والمشاعر العربية.

يرتبط هذا الكتاب بمجال الفهم الآلي للغات الطبيعية بما فيها اللغة العربية؛ لذلك اقتضى الأمر تضمينه إطارا نظريا ودراسة تطبيقية؛ ضمن فصلين، تضمن الفصل الأول الموسوم بـ "اللغة العربية واللسانيات الحاسوبية" ثلاث مباحث أساسية؛ خصصنا المبحث الأول للحديث عن مميزات وخصائص اللغة العربية، وما تشتمل عليه من أقسام الكلام، كما قمنا بالتعريف باللسانيات الحاسوبية، وكذا المعالجة الآلية

لِللغات الطبيعية، وعرض أهم مواضيعها، وخصصنا المبحث الثاني للحديث عن المعالجة الآلية لبعض المستويات اللسانية العربية؛ من بينها المستوى الصرفي والمستوى المعجمي، والمستوى التركيبي، أما المبحث الثالث، فجسد أهم التطبيقات الحاسوبية العربية.

بينما تضمن الفصل الثاني " الفهم الآلي للغات الطبيعية: تحليل الآراء والمشاعر العربية" ثلاث مباحث؛ عُني الأول بتعريف مجال الفهم الآلي للغات الطبيعية، وتقديم بعض التطبيقات التي تندرج ضمن مجال الفهم الآلي للغات الطبيعية. أما المبحث الثاني، فخصصناه للحديث عن مجال تحليل وتقييم الآراء والمشاعر وما ارتبط بها من إنجازات على مستوى التجارب العربية؛ من حيث تحديد مجالات اهتمامها، وأهم مستويات التحليل، والاشارة إلى بعض الإشكالات التي تعترض تقدم نظم اللغة العربية، وأهم المقاربات المعتمدة في مجال تحليل وتقييم الآراء والمشاعر، وتقديم بعض الأعمال التي سعت إلى بلورة نظم عربية خاصة بتحليل الآراء والمشاعر. بينما تضمن المبحث الثالث والأخير عرضاً للخطوات التي تقطعها نظم تحليل الآراء والمشاعر أثناء عملية تحديد توجيه التعليقات العربية، وهي خطوات مقتبسة من نظام (IERA) الذي تمت بلورته من طرف معهد الدراسات والأبحاث للتعريب بالمغرب بشراكة مع المدرسة المحمدية للهندسة.

الفصل الأول: اللغة العربية والنظم الحاسوبية.

المبحث الأول: اللغة العربية واللسانيات الحاسوبية

مدخل:

اهتم العديد من الدارسين باللغة العربية لما تتبوؤه من مكانة معتبرة بين لغات العالم، إذ أصبحت مجموعة من الدول تعتمد اللغة العربية لغة أساسية في جامعاتها، كما أصبح العديد من الأجانب يقبلون على تعلم اللغة العربية سواء داخل بلدانهم أو من خلال التوجه إلى بلدان عربية لمخالطة أهلها الناطقين بها قصد التمكن من تحصيلها، وقد جاء هذا الاهتمام نظراً لما تزخر به اللغة العربية من خصائص لغوية ومعجمية وتركيبية إلى غيرها من الخصائص التي تميزها عن باقي لغات العالم.

لقد أدى بروز مجال المعالجة الآلية للغات الطبيعية، إلى الاهتمام بمعالجة اللغة الطبيعية حاسوبياً؛ فنتج عن ذلك حقل معرفي جديد، سمي باللسانيات الحاسوبية، الذي مكن مجموعة من الباحثين من الاهتمام بحوسبة اللغات الطبيعية، من بينها اللغة العربية التي استطاعت بفضل تآزر جهود مجموعة من الباحثين أن تحظى باهتمام كبير لحوسبة مستوياتها اللسانية وتُلقَى بركب اللغات الأخرى.

إن من بين ما تسعى إليه اللسانيات الحاسوبية من أهداف يتمثل في "فهم آلية عمل الدماغ البشري ومحاكاته في تلقي اللغة وإنتاجها"⁽¹⁾، وذلك بمعالجة أهم المستويات اللسانية الصرفية، والصوتية، والمعجمية، والتركيبية، والدلالية، إلى غيرها، كما تهدف إلى تهيئة كفاية لغوية للحاسوب مماثلة لتلك التي يوظفها الإنسان في استقباله للغة وإدراكه وفهمه لها ثم إعادة إنتاجها وفق المطلوب.

تنقسم اللسانيات الحاسوبية إلى قسمين أحدهما نظري وآخر تطبيقي؛ يتجلى القسم النظري في ما توصلت إليه النظرية اللسانية الحديثة في دراستها للغة

(1) فتيحة محمد الدبابسة، "نهادهاموسى وجهوده اللغوية" بحث الماجستير، إشراف: هاني صبري

البطاط، كلية الدراسات العليا، جامعة الخليل، 2011، ص: 146

الطبيعية، من بينها ما جاء به فرديناند ديوسوسير الذي قال بأن اللسانيات "هي الدراسة العلمية للغة"⁽¹⁾ بكل مستوياتها؛ الصرفية، والتركيبية، والدلالية، والصوتية، والتداولية.

أما القسم التطبيقي فيتمثل في استثمار كل ما حددته اللسانيات النظرية من آليات ومستويات الدرس اللساني في دراسة لغة من اللغات، حيث يتم برمجتها وفق نظم آلية ذكية اعتماداً على اللغة التي يتعامل بها الحاسوب، وذلك للعمل على إيجاد طريقة مثلى لتحسين التفاعل الآلي/ الإنساني، قصد فهم الكيفية التي يشتغل بها الذهن البشري.

كما اهتمت اللسانيات الحاسوبية كذلك بإحصاء الظواهر اللغوية على المستوى الصوتي، والمستوى التركيبي، والمستوى الصرفي، بالإضافة إلى عمليات التحليل والتركيب اللغويين والتي تمثل الخطوات الأولى للمعالجة الآلية للغة.⁽²⁾ لقد انطلقت البداية الفعلية للأبحاث المتعلقة بالمعالجة الآلية للغات الطبيعية بصفة عامة في الأربعينيات من القرن الماضي تزامناً مع انتشار الحواسيب، وقد تجلت تلك الأبحاث في المحاولات الأولى للترجمة الآلية، وقد ساهمت نتائجها الأولية والمشجعة في ظهور مجال المعالجة الآلية للغات الطبيعية وانتشاره أواسط المختبرات البحثية في بقاع العالم⁽³⁾، أما بالنسبة لمعالجة اللغة العربية، فقد ظهرت في الدول الغربية، حيث لم يكن دافعهم الاهتمام باللغة العربية هو النهوض والرقى بها، بل كان الدافع إلى برمجتها اقتصادياً؛ " إذ كانت البلاد العربية سوقاً تجارياً رائجة لتقنيات

(1) Saussure FERDINAND (1995) : "Cours De Linguistique Générale" P232.

(2) سامي عباد حنا، ونجيب جريس، وحسام زكي الدين كريم (د ت): "معجم اللسانيات الحديثة: انجليزي-عربي"، ص: 26، بتصرف

(3) Diksha Khurana et al.(2007), " Natural Language Processing: State of The Art, Current Trends and Challenges", *arXiv e-prints* 1708: arXiv:1708.05148.

الحاسوب بالعربية تتحرر شيئاً فشيئاً من الأيدي الأجنبية... فقد نجحت العديد من الشركات العربية بالتعاون مع شركات أجنبية في صناعة الحواسيب المصغرة الخاصة، وتعريب لغات البرمجة، وهكذا برزت لغة "نجلاء" وهي لغة "بازيك" عربية تشتغل على الحواسيب من نوع الفارابي من إنتاج الصناعات الحاسوبية السعودية، ولغة "الخوارزمي" وهي لغة مقتبسة عن "البازيك" أيضاً وتشتغل على الحواسيب من سلسلة "الرائد" الحاسوب "صخر" من إنتاج الشركة العالمية".⁽¹⁾

ومن أجل توضيح علاقة اللغة العربية باللسانيات الحاسوبية فإننا سنقدم في هذا المبحث تعريفاً للغة العربية، وتحديدًا لأقسام الكلام العربي، ثم سنعرض أهم الخصائص المميزة للغة العربية، لنعرج الحديث بعد ذلك عن اللسانيات الحاسوبية والاشارة إلى بعض مواضعها، ثم الحديث كذلك عن المعالجة الآلية للغات الطبيعية من بينها اللغة العربية.

1. تعريف باللغة العربية :

تميزت اللغة بكونها كائنًا بيولوجيًا⁽²⁾ لأنها تعد أداة أكثر استجابة للواقع الإنساني، كما تعد الميزة الأساس التي تميز بها الإنسان عن الحيوان حسب ما جاء على لسان ابن جني أبو الفتح عثمان في كتابه الخصائص الذي أقر ضمنه أن اللغة هي "أصوات يعبر بها كل قوم عن أغراضهم"⁽³⁾، وهو تعريف بين من خلاله أن اللغة وإن كانت وسيلة للتواصل، إلا أنها تتطور حسب رغبات الإنسان مثل التطور البيولوجي لباقي الكائنات الحية الموجودة على وجه الأرض.

(1) أحمد وليد العناتي (2006): "اللسانيات الحاسوبية العربية رؤية ثقافية"، مجلة فكر ونقد، ع: 82، ص: 33-34، بتصريف.

(2) <https://www.akhbarona.com/writers/31167.html>

(3) ينظر عثمان بن جني أبو الفتح، (د ت)، "الخصائص"، تحقيق محمد علي النجار، ج 1، الطبعة 11، ص: 33

لقت اللغة العربية اهتماما كبيرا من لدن اللغويين العرب، تمثل ذلك في العناية بها والتنقيب عنها وجمع مفرداتها وتبويب معانيها وتفسير ظواهرها والتفصيل في قوانين بناء كلماتها وضوابط قواعدها وتركيبها، ولهذا كُتِبَ لها البقاء منذ ولادتها، ويرجع ذلك إلى كونها لغة نزل بها القرآن الكريم من عند الله سبحانه وتعالى على نبيه سيدنا محمد ﷺ، وهي بذلك نالت حظها الأوفر من التطور والنماء - باعتبارها أداة للتفكير ونشر الثقافة في البلدان الناطقة بها وبغيرها - بفعل التطور التكنولوجي الذي شهده العصر في الآونة الأخيرة المتمثل في غزارة انتشار الصحافة الإلكترونية والكتابات المتوالية في ميادين شتى، كما ساهم الإقبال على تعلمها في انتشار المدارس وشيوع المعاهد المختصة في تعليمها، مما جعل التنافس قائما بين تلك المدارس والمعاهد متمثلا في تحديث المناهج ووضع المقررات وتأطير المدرسين وتكثيف تكوينهم للعناية بالدارس لهذه اللغة أشد عناية، كما لقيت اللغة حظا أوفر في الاهتمام بها والذي توج بالاحتفاء السنوي بها لتعريف الغافل عنها بأبرز خصائصها، وأهم مكوناتها، ودوافع صمودها، ومقومات دفاعها، وتصديها للغزو الثقافي، واستفحال معضلة العامية، وفكرة تكريس اللغة الثانية والعزوف عنها باعتبارها اللغة الثانوية التي لا تشبع جوف الجوعى، ولا تروي ظمأ العطشى.

تتميز اللغة العربية بالعديد من الخصائص لعل أبرزها الإيجاز الذي يوحى على الاقتصار في الكلام، والتعبير بأحسن الكلمات التي تومئ بإشارات معبرة، ومعاني خصبة تترك في نفوس المتلقي معاني غزيرة يملؤها ثراء ذوقى، ورقة دلالية، وبلاغة عربية، يصعب ترجمة معانيها للغة أخرى، ويكفيها ما حظيت به من مكانة راقية، ومنزلة عالية، في كونها تحتوي على أكثر من اثنتي عشرة ألف كلمة مضمنة للعديد من المرادفات التي تصل في الحب مثلا إلى أربع وعشرين كلمة مقارنة باللغات الأخرى كاللغة الإنجليزية التي لا تتعدى ست آلاف كلمة، وأقل منها الفرنسية التي لا تتعدى ثلاث مائة

كلمة، كما أن اللغة العربية تتميز بكونها تشتمل على أطول كلمة، أوردها الله سبحانه وتعالى في القرآن الكريم وهي: " فَأَسْقَيْنَاكُمُوهُ" ضمن سورة الحجر الآية 22، بلغت أحد عشر حرفاً، كما أنها تتميز باشمالها على كلمات تحوي أقل عدد حروف في كلمات لغات العالم، وعددها حرف واحد وحركة واحدة مثل "ع" وهي فعل أمر من "وعى"، ورغم كل ذلك فقد قسم العرب القدماء أقسام الكلام العربي إلى ثلاثة أقسام تتمثل في: "الاسم والفعل والحرف".

1. الكلام العربي:

يتفرع عن شجرة الكلام ثلاثة فروع وهي: الاسم والفعل والحرف، ينتج عن تلاقح هذه الفروع علاقات بين الكلام والكلمة واللفظ والقول والكلم، متشابهة من حيث مواد التكوين (أي أنها كلها تتكون من نفس الحروف والأصوات) ومختلفة من حيث الدلالات؛ مما مكن الحرف من أن يبرز مكانته في فلك الكلام وعلاقته بالكلمة واللفظ والقول والكلم، ويُعَيَّن لنفسه موضعاً في بيتها ومحيطها.

1.2. تعريف الكلام:

لخص ابن مالك في ألفيته شجرة الكلام بالإشارة إلى جل المصطلحات التي سبقت الإشارة إليها في قوله:

كَلَامُنَا لَفْظٌ مُفِيدٌ كَأَسْتَقِمُّ **** اسْمٌ وَفِعْلٌ ثُمَّ حَرْفٌ أَلْكَمُّ

وَإِحْدُهُ كَلِمَةٌ وَالْقَوْلُ عَمٌّ **** وَكَلِمَةٌ بِهَا كَلَامٌ قَدْ يُؤْمٌ¹

عرف ابن عقيل في شرحه لألفية ابن مالك الكلام بأنه "اللفظ المفيد فائدة يحسن السكوت عليها، والمقصود بالمفيد هو ما فهم من خلاله الكلام فهما يحسنه

(1) محمد بن عبد الله بن مالك الأندلسي، "ألفية ابن مالك في النحو والصرف"، المكتبة الشعبية، بيروت، لبنان، ص:2.

السكوت عليه ليخرج بذلك الكلمة والكلم" (1)، ونضيف تعريفاً آخر للزمخشري الذي قال فيه بأن "الكلام هو المركب من كلمتين أسندت إحداهما إلى الأخرى وذلك لا يتأتى إلا في اسمين كقولك: "زيد أخوك"، و"بشر صاحبك". أو في فعل واسم نحو قولك: "ضرب زيد"، و"انطلق بكر". وتسمى الجمل" (2).

يتألف الكلام كما سبق الإشارة إلى ذلك من ثلاثة عناصر: "الاسم والفعل والحرف"، إذ أن تلاقح هذه العناصر وإن كان ضمناً هو ما يُمكن من تحصيل الفائدة؛ فمن خلال تمثيله ب"استقم"، يتبين بأن الفعل باعتباره كلمة وإن كانت مفردة، إلا أنها تحيل عن كلام بعدها يتجسد في العنصرين المتبقيين (الاسم والحرف)؛ فأما الاسم فيجسده المخاطب الذي أمر بالاستقامة، وأما الحرف فيحيل على ما ينبغي الاستقامة عليه أو فيه؛ فنقول مثلاً: "استقم في مكانك"، أو "استقم على ما استقام الناس عليه".

2.2. أقسام الكلام:

يضم الكلام العربي ثلاث فئات وهي: الاسم والفعل والحرف، وفي هذا الصدد يقول الأنباري مفسراً للدور الذي تؤديه هذه الفئات بعد توظيفها في التعبير عن كل ما يتمثله الإنسان من آراء وأحاسيس وعواطف ومشاعر ورغبات، حيث قال: "وجدنا هذه الأقسام الثلاث يعبر بها عن جميع ما يخطر بالبال، ويتوهم في الخيال، ولو كان ههنا قسم رابع لبقى في النفس شيء لا يمكن التعبير عنه، ألا ترى أنه لو سقط آخر هذه الأقسام الثلاث - ويعني به الحرف- لبقى في النفس شيء لا يمكن التعبير عنه، إزاء ما

(1) عبد الله بن عقيل العقيلي بهاء الدين، (د.ت)، "شرح ابن عقيل على ألفية ابن مالك"، تحقيق محمد محي الدين عبد الحميد، المكتبة العصري صيدان - بيروت، الطبعة: الثانية، ج: 1، ص: 14 - 15، بتصرف.

(2) موفق الدين أبو البقاء بن يعيش الموصلي، (2001)، "شرح المفصل للزمخشري"، قدم له: الدكتور إميل بديع يعقوب، دار الكتب العلمية، بيروت - لبنان الطبعة: الأولى، ج: 1، ص: 70، بتصرف.

سقط، فلما عبر بهذه الأقسام عن جميع الأشياء دل على أنه ليس إلا هذه الأقسام الثلاث⁽¹⁾، ويمكن الإشارة في هذا الصدد إلى قول الزجاجي حينما عبر عن الكلام، باعتباره من الله سبحانه وتعالى، حيث قال: "ونحن نعلم أن الله إنما جعل الكلام ليعبر به العباد عمًا هجس في نفوسهم، وخاطب به بعضهم بعضا بما في ضمائرهم، مما لا يوقف عليه بإشارة ولا إيماء ولا رمز بحاجب ولا حيلة من الحياة، فإذا كان هذا معقولاً ظاهراً غير مدفوع، فيتبين أن المُخاطَبَ، والمُخاطَبَ، والمخبر عنه، والمخبر به، أجسام وأعراض تنوب في العبارة عنها اسماءً أو ما يعتبرونه معنى يدخله تحت هذا القسم من أمر، أو نهي، أو نداء، أو نعت، أو أشبه ذلك مما تختص به الأسماء، لأن الأمر والنهي إنما يقعان على الاسم النائب عن المسمى، فالخبر إذا هو غير المخبر والمخبر عنه وهما داخلان تحت قسم الاسم، والخبر هو الفعل وما اشتق منه أو تضمن معناه وهو الحديث الذي ذكرناه، ولا بد من رابط بينهما وهو الحرف ولن يوجد إلى معنى رابع سبيل فيكون للكلام القسم الرابع"⁽²⁾

1.2.2. تعريف الاسم:

لقد وضعت مجموعة من التعاريف اللغوية للاسم التي أجمعت على أن الاسم في اللغة مأخوذ من سما يسمو سموا، وفي هذا الصدد عمل الانباري على تفسير سبب جعل الاسم من سمي يسمو، حيث عرض اتجاهي المدرسة البصرية والكوفية إلا أنه رجح الاتجاه البصري الذي يحيل على أن الاسم إنما سمي بذلك "لأنه سمي على مسمى وعلا على ما تحته من معناه فسمي اسماً لذلك؛ فلما كان الاسم يخبر به ويخبر

(1) كمال الدين الانباري النحوي (1999)، "أسرار العربية"، تحقيق: يوسف بركات هبود، دار الأرقام بن ابي الأرقام - بيروت/ لبنان، ص: 35.

(2) عبد الرحمن بن إسحاق أبي القاسم الزجاجي، (1979) "الإيضاح في علل النحو"، تحقيق: عثمان السيد محمد، دار الكتب العلمية، بيروت - لبنان، ص: 48.

عنه والفعل يخبر به ولا يخبر عنه والحرف لا يخبر به ولا يخبر عنه فقد سما الاسم على الفعل والحرف أي ارتفع".⁽¹⁾

وإذا نظرنا إلى الاسم في الاصطلاح فيمكن تقسيم التعاريف التي قدمت للاسم إلى قسمين؛ فالقسم الأول: عرف الاسم بعلاماته، كما سبقت الإشارة إلى ذلك في شرح ابن عقيل عن ألفية ابن مالك، ويمكن الإحالة على ما جاء به ابن هشام الانصاري في كتابه قطر الندى وبل الصدى الذي قيل فيه أن الاسم يعرف "بال كالرجل، والتنوين كرجل، وبالحدیث عنه كتاء ضربت"⁽²⁾، أما القسم الثاني: فهو الذي عرف الاسم بمعناه، ويمكن الإشارة في هذا الصدد إلى ما قاله الزجاجي بأن الاسم: "في كلام العرب ما كان فاعلاً أو مفعولاً أو واقعا في حيز الفاعل والمفعول به"⁽³⁾ كما عرفه مصطفى الغلاييني بأنه "ما دل على معنى في نفسه غير مقترن بزمان".⁽⁴⁾

وقد ميز النحويون الاسم بمجموعة من العلامات اختصرها ابن مالك في ألفيته بقوله:

بِالْجَرِّ وَالْتَّنْوِينِ وَالنَّدَى وَأَلْ *** وَمُسْنَدٌ لِلِاسْمِ تَمْيِيزٌ حَصَلْ⁽⁵⁾

فمن خلال هذا البيت يتبين بأن علامات الاسم تتضمن:

(1) كمال الدين الانباري النحوي (1999)، "أسرار العربية"، تحقيق: يوسف بركات هبود، مرجع سابق، ص: 4

(2) أبو محمد عبد الله جمال الدين بن هشام الأنصاري، (2012)، "شرح قطر الندى وبل الصدى"، تحقيق: محمد محي الدين عبد الحميد، شركة أبناء شريف الانصاري للطباعة والنشر والتوزيع، بيروت - لبنان، ص: 32.

(3) عبد الرحمن بن إسحاق أبي القاسم الزجاجي، (1979) "الإيضاح في علل النحو"، تحقيق: عثمان السيد محمد، مرجع سابق، ص: 48

(4) مصطفى الغلاييني، (2010)، "جامع الدروس العربية"، دار التوفيقية للتراث، القاهرة، مصر، ج: 1، ص: 9.

(5) محمد بن عبد الله بن مالك الأندلسي، "ألفية ابن مالك في النحو والصرف"، المكتبة الشعبية، بيروت، لبنان، ص: 2

- الجر: وليس المراد به حرف الجر، لأنه قد يدخل في اللفظ على ما ليس باسم كقولك: "أشرت إليه بأن يقم"، بل المراد كون الكلمة مجرورة نحو: "صلبت في المسجد الواسع" قال تعالى: ﴿بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ﴾.
- التنوين: وهو نون ساكنة زائدة لغير توكيد، تلحق آخر الأسماء لفظاً لا خطأً ولا وقفاً، نحو: كتب محمدٌ واجباً بخط جميلٍ، ويتضمن التنوين أربعة أقسام وهي: تنوين التمكين، تنوين التنكير، تنوين المقابلة، تنوين العوض.

- دخول (أل) على الكلمة نحو: قدم المسافر.

- النداء: وليس المراد به دخول حرف النداء، لأنه قد يدخل في اللفظ على ما ليس باسم، كقوله تعالى: ﴿قِيلَ ادْخُلِ الْجَنَّةَ قَالَ يَا لَيْتَ قَوْمِي يَعْلَمُونَ﴾⁽¹⁾ على أحد الإعرابين⁽²⁾ وإنما المراد كون الكلمة مناداة، نحو: يا خالد أكرم أباك. قال تعالى: ﴿وَنَدَيْتُهُ أَنْ يَا بُرْهِيمُ ۗ ۱٠٤ قَدْ صَدَّقْتَ الرُّءْيَا إِنَّا كَذَلِكَ نَجْزِي الْمُحْسِنِينَ ۗ ۱٠٥﴾⁽³⁾.

- الإسناد: وهو أن يُنسب إلى الكلمة حكم تحصل به الفائدة نفيًا أو إثباتًا، نحو: صلبت مع الجماعة.

2.2.2. تعريف الفعل

يمكن أن نعرف الفعل بأنه يمثل كل حركة يقوم بها الانسان أو الحيوان أو غيرها من المخلوقات على وجه الأرض، وإلى هذا أشار الفيروزآبادي في معجمه " فالفعل: بالكسر حركة الإنسان، أو كناية عن كل عمل متعمد أو غير متعمد، والفعل

(1) سورة يس، آية: 26.

(2) وهو أن تكون (يا) للنداء والمنادى محذوف، فإن كانت للتنبيه فلا شاهد فيها.

(3) سورة الصافات، آية: 105.

بِالْفَتْحِ مَصْدَرٌ فَعَلَ يَفْعَلُ"⁽¹⁾، كما أشار ابن سيدة الأندلسى فى المحكم والمحيط الأعظم إلى أن "الْفِعْلُ هُوَ كِنَايَةٌ عَن كُلِّ فِعْلٍ مُتَعَدٍّ أَوْ غَيْرِ مُتَعَدٍّ"⁽²⁾.

وإذا نظرنا إلى الفعل من وجهة نظر اصطلاحية، فيمكن أن نسرد بضع تعاريف من بينها " حدث حقيقة أو تشبها وتنزيلا قام به الفاعل فأوجده حقيقة أو تشبها وتنزيلا أو اتصف به نفيًا أو إثباتًا حقيقة أو تنزيلا ليعم ما وقع وما لم يقع"⁽³⁾، أما الزجاجى فقد عرف الفعل بأنه "ما دل على حدث، وزمان ماض أو مستقبل"⁽⁴⁾ إلا أنه نجده قد أغفل الزمن المضارع فى تعريفه هذا، ومن بين التعاريف كذلك ما تضمنته المقدمة الجزولية بأن الفعل "هو كل كلمة تدل على معنى فى نفسها وتعرض لزمن وجود ذلك المعنى"⁽⁵⁾.

لقد أشارت التعاريف إلى أن الفعل يتأسس وفق ركنين أساسيين هما: الحدث والزمان؛ فإن فقد الزمان فكان حدثًا فقط فهو مصدر مثل الضرب والأكل، وإن انتفى الحدث فأصبح زمانًا فقط فهو ظرف زمان.

يتميز الفعل بمجموعة من الخصائص كما أشار إليها ابن مالك فى ألفيته بقوله:

-
- (1) مجد الدين محمد بن يعقوب الفيروز آبادى، (2005)، "القاموس المحيط"، تحقيق محمد نعيم العرقسوسى، الطبعة الثامنة، مؤسسة الرسالة، بيروت - لبنان، ص: 1043.
 - (2) على بن إسماعيل بن سيدة (1985)، "المحكم والمحيط الأعظم"، تحقيق: عبد الستار أحمد فراج، مصطفى السقا، حسين نصار، الطبعة الأولى، معهد المخطوطات بجامعة الدول العربية، ج: 2، ص: 117.
 - (3) رؤوف جمال الدين، (د.ت)، "المعجب فى علم النحو يتضمن فلسفة علم النحو والحروف وفق المنهج العلمى الأصيل"، دار الهجرة، قم، إيران، ص: 26.
 - (4) عبد الرحمن بن إسحاق أبى القاسم الزجاجى، (1979) "الإيضاح فى علل النحو"، تحقيق: عثمان السيد محمد، مرجع سابق، ص: 52.
 - (5) عيسى بن عبد العزيز الجزولى (د.ت) "المقدمة الجزولية فى النحو" تحقيق: شعبان عبد الوهاب محمد، مطبعة أم القرى، ص: 4.

بَتَا فَعَلْتُ وَأَتَتْ وَيَا أَفْعَلِي *** وَنُونٌ أَقْبَلَنَّ فِعْلٌ يَنْجَلِي (1)

تتجسد علامات الفعل حسب قول ابن مالك في:

- تاء الفاعل: للمتكلم نحو: قمتُ بواجبي، أو للمخاطب نحو: أنتَ واسيت المريض، قال تعالى: ﴿إِنْ كُنْتُ قُلْتُهُ فَقَدْ عَلِمْتَهُ﴾ (١١٦) (2) أو للمخاطبة نحو: أنتِ أحببت أولادك، قال تعالى: ﴿فَإِذَا خِفتَ عَلَيْهِ فَأَلْقِيهِ فِي آيَمٍ وَلَا تَخَافِي وَلَا تَحْزَنِي إِنَّا رَادُّوهُ إِلَيْكَ وَجَاعِلُوهُ مِنَ الْمُرْسَلِينَ﴾ (3).
- تاء التأنيث الساكنة أصالة نحو: صامتُ هند يوم الخميس، قال تعالى: ﴿وَقَالَتْ لِأُخْتِهِ - فُصِيهَ فَبَصُرَتْ بِهِ - عَن جُنُبٍ وَهُمْ لَا يَشْعُرُونَ﴾ (4).

- ياء المخاطبة نحو: قول تعالى: ﴿فَكُلِّي وَأَشْرِبِي وَقَرِّي عَيْنًا فَإِمَّا تَرَيَنَّ مِنَ الْبَشْرِ أَحَدًا فَقُولِي إِنِّي نَذَرْتُ لِلرَّحْمَنِ صَوْمًا فَلَنْ أُكَلِّمَ آيَوْمَ إِنْشَاءً﴾ (5) وإنما قال ابن مالك: (ويا افعلي) ولم يقل: (ياء الضمير) لأن هذه تشمل ياء المتكلم، وليست خاصة بالفعل بل تكون فيه وفي الاسم والحرف، كقوله تعالى: ﴿رَبِّ آغْفِرْ لِي وَلِوَالِدَيَّ وَلِمَن دَخَلَ بَيْتِي مُؤْمِنًا وَلِلْمُؤْمِنِينَ وَالْمُؤْمِنَاتِ وَلَا تَزِدِ الظَّالِمِينَ إِلَّا تَبَارًا﴾ (6) وقوله تعالى: ﴿فَتَبَسَّمْ ضَاحِكًا مِّن قَوْلِهَا وَقَالَ رَبِّ أَوْزِعْنِي أَنْ أَشْكُرَ نِعْمَتَكَ الَّتِي أَنْعَمْتَ عَلَيَّ وَعَلَىٰ وُلْدِي وَأَنْ أَعْمَلَ صَالِحًا تَرْضَاهُ وَأَدْخِلْنِي بِرَحْمَتِكَ فِي عِبَادِكَ الصَّالِحِينَ﴾ (7).

(1) محمد بن عبد الله بن مالك الأندلسي، "ألفية ابن مالك في النحو والصرف"، المكتبة الشعبية،

بيروت، لبنان، ص:2

(2) سورة المائدة، آية: 116.

(3) سورة القصص، آية: 7.

(4) سورة القصص، آية: 11.

(5) سورة مريم، آية: 26،

(6) سورة نوح، آية: 28.

(7) سورة النمل، آية: 19.

- نون التوكيد ثقيلة كانت أو خفيفة؛ فالأولى نحو: والله لأفعلنَ الخير، والثانية نحو: احذرن قول السوء. قال تعالى: ﴿لَيْسُ جَنًّا وَلَيْكُونًا مِّنَ الصَّغِيرِينَ﴾ (1).

3.2.2. تعريف الحرف:

عرف النحاة العرب الحرف بكونه كلمة جاءت لتعبر عن معنى أو معاني بتعلقها بالاسم والفعل، وتنقسم الحروف في اللغة العربية إلى حروف مباني، وحروف معاني.

1. حروف المباني:

هي الحروف التي تتركب فيما بينها لبناء الكلام اسماً كان أو فعلاً، نحو كلمة "كتاب" فإنها تتكون من الحروف الآتية: "الكاف، والتاء، والألف، والباء" وقد عد النحاة من هذه الحروف ثمانية وعشرين حرفاً، رتبت ترتيباً هجائياً، وترتيباً أبجدياً، فالترتيب الأبجدي على الشكل الآتي: "أبجد هوز حطي كلمن سعفص قرشت ثخذ ضظغ" وهناك كذلك الترتيب الأصواتي الذي يعتمد على ترتيب الحروف حسب مخارجها، وبدأ هذا الترتيب بالحروف الخارجة من أقصى الحلق وصولاً إلى الحروف التي تخرج من الشفتين، وقد اعتمد الخليل بن أحمد الفراهيدي هذا الترتيب في ترتيب معجمه العين.

2. حروف المعاني:

تتمثل حروف المعاني في القسم الثالث من أقسام الكلام العربي التي تؤدي وظائف كثيرة في الكلام العربي؛ إذ إن هذه الحروف تقوم بربط الاسم بالاسم، والفعل بالفعل، والاسم بالفعل، هذا من الناحية التركيبية، كما تؤدي حروف المعاني أدواراً دلالية. تتمثل في قلب معنى الجملة، أو تغييره، أو تأكيده.

2. خصوصيات منظومة اللغة العربية من منظور معلوماتي:

(1) سورة يوسف، آية: 32.

تمتاز اللغة العربية بالعديد من الخصائص، وسنحاول في هذا المحور التركيز على الخصائص الأكثر دراسة في العديد من الحقول المعرفية أبرزها المعالجة الآلية للغة العربية، والمتمثلة أساساً في توسطها اللغوي، وحدة خاصيتها الصرفية، ومرونتها النحوية، وانتظاماتها الصوتية، وحساسيتها السياقية، وقوة تماسك عناصر منظومتها اللغوية.

1.3. الخصوصية الصوتية:

تشتمل اللغة العربية على جهاز صوتي لم يعرف له مثيل في اللغات الأخرى، تتوزع حروفها على مخارج صوتية متنوعة تتلخص في أربعة مخارج كبرى: الجوف، الحلق، اللسان والشفطان، وكل مخرج من هذه المخارج ينقسم إلى أقسام أخرى حسب الحروف التي تضمها، مما يعطي تنوعاً وانسجاماً بين الأصوات. ومن جهة ثانية فاللغة العربية لا تجمع بعض الحروف في الكلمة الواحدة مثل اجتماع حرف الجيم مع القاف، ومن بين الخصائص الصوتية التي تنفرد بها اللغة العربية عن غيرها من اللغات خاصية الإطباق التي قد لا نجدها في باقي اللغات الطبيعية، إلى غيرها من الخصائص الصوتية التي تعتبر ذات أهمية في توليد الكلام العربي، سواء على المستوى الطبيعي أو على مستوى المعالجة الآلية للكلام العربي.⁽¹⁾

2.3. الخصوصية الصرفية:

تعتبر اللغة العربية من بين اللغات التي تتسم بالاشتقاق كخاصية لها، إذ يتأسس على مجموعة من الأوزان أو الصيغ الصرفية باعتبارها قوالب يفرغ فيها الجذر للحصول على مجموعة من المشتقات، فينتج عن ذلك تكوين عائلة لغوية أصلها الجذر، وذلك مثل جذر "ك ت ب" الذي يتفرع منه جذع "كُتَب" باعتباره من الجذوع المستعملة، ويشتق منه العديد من الكلمات مثل: كُتِب - كاتب - مكتوب - كتابة -

(1) نبيل علي (1994)، "العرب وعصر المعلومات"، مجلة عالم المعرفة، ع: 184، المجلس الوطني للثقافة والفنون والآداب، مطابع السياسة الكويتية، الكويت، ص: 335، بتصرف.

كتاب، وهذا ما يساعد في تعلم اللغة العربية للناطقين بغيرها، حيث إن الطالب يتمكن من الوصول أو يقترب من معاني الكلمات إذا استوعب معنى الجذع، وبذلك فإن "انتظام الصرف العربي واطراده يزيد من قابليته للمعالجة الآلية، وتعد هذه المعالجة هي المدخل الطبيعي لمعالجة المنظومة الشاملة للغة العربية، حيث يعد المصرف الآلي مقوماً أساسياً في مَيْكَنَةِ المعجم العربي، وتطوير نظم آلية للإعراب الآلي والتشكيل التلقائي".⁽¹⁾

3.3. الخصوصية النحوية:

تتسم اللغة العربية بخاصية المرونة اللغوية التي تتجلى في تنوع تراكيها بتعدد ترتيب الكلمات داخل الجملة الواحدة، إذ أنه يمكن صياغة جملة واحدة باعتماد العديد من الصيغ التركيبية التي تساعد على فهم الكلام العربي ضمن معاني مختلفة، ومن أمثلة الخصائص النحوية نجد التقديم والتأخير، والزيادة والحذف، والحشو والإطناب، والفصل والوصل إلى غيرها من الخصائص النحوية التي شكلت تحدياً كبيراً في مجال المعالجة الآلية للغة العربية، إذ أن بلوغ نمذجة مثل هذه الخصائص سيمكن لا محال من بلوغ درجة الدقة في إنتاج النظم والتطبيقات المتعلقة بالتحليل النحوي الآلي للنصوص العربية.⁽²⁾

4.3. الخصوصية الدلالية:

تتميز اللغة العربية بمجموعة من الخصائص من بينها تنوع التراكيب اللغوية، والتي يسميها النحويون بالجملة، إلا أنه لا يعد في اللغة العربية تركيباً أو جملة أو كلاماً إلا إذا كانت له دلالة واضحة مفهومة، ولقد جاء في تعريف الكلام عند علماء

(1) المرجع نفسه، ص: 334، بتصرف.

(2) نبيل علي (1994)، "العرب وعصر المعلومات"، مرجع سابق، ص: 334 – 335، بتصرف

النحو بأنه: اللفظ المركب المفيد فائدة يحسن السكوت عليها⁽¹⁾، وإذا ما تأملنا كذلك في الجملة العربية فإن قواعدا تمكنا من تركيب جمل مفيدة يحسن السكوت عليها سواء كانت مطولة، أو موجزة بحذف بعض العناصر من الجملة المطولة، فتبقى مشتركة في نفس المعنى، أو قد تدل على معنى مناقض لها وذلك يفضي بنا إلى خاصية أخرى تمتاز بها اللغة العربية، وهي خاصية الحذف، ونمثل لذلك بما يلي:

- إن محمد أحرز نجاحاً باهراً.
- محمد أحرز نجاحاً باهراً.
- محمد أحرز نجاحاً.
- محمد لم ينجح.
- ربما محمد نجح.
- محمد نجح.

إن دلالة هذه التراكيب يتم اعتمادها حسب الحاجة التي يريد المتكلم وهي بيان طبيعة تأكيد نجاح محمد، أو الاكتفاء بالإخبار عن نجاح محمد، أو نفي نجاح أو الشك في نجاحه، تؤكد مدى تأثير خاصية الحذف في اللغة العربية التي ترتبط بدلالة الإخبار، وقد فصل البلاغيون في ذلك بعد استفادتهم من قواعد الحذف في الجملة العربية، وجعلوا لكل تركيب في الجملة دلالة بعد الحذف، فبذلك تعتبر التراكيب اللغوية العربية أو الجمل العربية، وإجادة ترتيبها وسيلة أساسية في تحقيق دلالة سليمة أو صحيحة، وإلى ذلك أشار أبو هلال العسكري في كتابه الصناعتين بقوله في فصل (تمييز الكلام): "الكلام أيدك الله يحسن بسلاسته وسهولته وفصاحة لفظه وإصابة معناه، وجودة مطالعه، وأنس مقاطيعه، واستواء تقاسيمه، وتعادل أطرافه، وتشبيهه

(1) عبد الله بن عقيل العقيلي بهاء الدين، (د.ت)، "شرح ابن عقيل على ألفية ابن مالك"، تحقيق محمد محي الدين عبد الحميد، مرجع سابق، ج: 1، ص: 19، بتصرف.

أعمدته بهواديه مع قلة ضروراته، بل عدمها أمثل حتى لا يكون لها في الألفاظ أثر ...
فإذا كان الكلام كذلك كان بالقول ضليعا، وبال حفظ خليقاً⁽¹⁾.

5.3. الخصوصية اللغوية :

يلعب التوسط اللغوي دورا بارزا بالنسبة للغة العربية، ويتجلى ذلك في كونها لغة توازن يمكنها من التموقع في الوسط بين أطراف كثيرة من المحاور التي تحدد مجالات التنوع اللغوي، التي ترتكز أساسا على انحياز اللغة العربية إلى الاستعمالات الشائعة وترك الاستعمالات الشاذة التي تعثر بنطقها اللسان، ويستصعب السمع بتوظيفها الإفهام، وهو الأمر الذي سيزداد أهمية إثر ضرورة الاهتمام بحوسبتها، إذ أصبح من الضرورة العناية بالدراسات اللغوية المقارنة والتقابلية التي أضحت مصدرا غنيا لإفهام اللغة العربية في إطار أشمل وأعمق لإغناء مناعتها ضد ما ستعرض له من قضايا بفعل المعالجة الآلية، التي يهدف من خلالها تمكين الخبراء من الاستفادة من العدد الهائل مما وصلت إليه نظم المعالجة الآلية للغات الطبيعية وانتقاء الأجود منها والأنسب لخدمة اللغة العربية⁽²⁾.

6.3. الخصوصية السياقية:

تتسم اللغة العربية بحساسية سياقية متمثلة في تأخي العناصر اللغوية في ما بينها من أجل تقديم عبارة سليمة من كل لحن، ومن بين أهم مظاهر الحساسية السياقية نجد شكل كتابة الحروف العربية حسب موقعها ضمن الكلمة؛ فشكل الحرف في بداية الكلمة ليس هو نفسه في وسط الكلمة وليس هو نفسه كذلك في آخر الكلمة، فإذا أخذنا على سبيل المثال حرف الغين؛ فإن صور كتابته تتعد بتعدد وروده

(1) أبو هلال العسكري (1984)، "كتاب الصناعتين"، تحقيق وضبط مفيد قميصه، دار الكتب

العلمية بيروت - لبنان، الطبعة: 2، ص: 14.

(2) محمد الهليلي (1998)، "اللغة العربية والحاسوب"، بحث لنيل شهادة الإجازة، كلية الآداب والعلوم الانسانية مكناس، جامعة مولاي إسماعيل، ص: 14، بتصرف.

في الكلمة مثل: غلام، لغز، مبالغ، فراغ، وهذا ينطبق على كل الحروف العربية التي تتعدد أشكال كتابتها بتعدد مواقعها في الكلمة، مما يجعل المعالجة الآلية للغة العربية تختلف عن المعالجة الآلية للغات الطبيعية كاللغة الإنجليزية في العديد من المواضع، مثل "تعامل الكمبيوتر مع نسق الكتابة الإنجليزية ذات الحروف المنفصلة يعد بسيطاً إذا ما قورن بذلك للكتابة العربية ذات التشبيك والتشكيل فيكفي أن الأخيرة تحتاج- خلافاً عن الإنجليزية- إلى وسيلة برمجية آلية لانتقاء شكل الحرف المناسب تلقائياً وفقاً للحرف المحيط به"⁽¹⁾، وهذا ما يجعل تقنية التعرف الضوئي على الحروف العربية (OCR) في تأخر ملحوظ.

7.3. خصوصية تماسك المستويات اللغوية:

يعتبر التماسك خاصية أساسية لفهم المستويات اللغوية، واللغة العربية تكاد تشكل بواسطة مستوياتها عنصراً واحداً، حيث يصعب امتثال عنصر بمعزل عن العناصر الأخرى، فعلى سبيل المثال لا يمكن تصور المستوى الصرفي بمعزل عن المستوى المعجمي في بناء الكلام العربي، ولا المستوى الصرفي عن المستوى النحوي في بناء الجمل، إلى غيره من التماسك اللغوي، وهذا يجعل من الصعب تصور برمجة مستوى من المستويات اللسانية بمعزل عن الآخر، إذ أن هذا التماسك هو الذي يقلل من عدم الفهم الآلي للغة العربية وإبعاد اللبس والاحتمال في تأويل ما تم تحليله.

لقد شكلت الخصائص المذكورة أعلاه تحدياً بالنسبة لأعمال فرق البحث التي سعت جاهدة إلى تجاوزها أثناء المعالجة الآلية للغة العربية؛ فبفضل الميزة الرياضية التي تتميز بها اللغة العربية فقد تمكنت العديد من فرق البحث من قطع أشواط بارزة ضمن مجال المعالجة الآلية⁽²⁾ للغة العربية، توجت بإنتاج مجموعة من التطبيقات

(1) نبيل علي، "العرب وعصر المعلومات"، مرجع سابق، ص: 336.

(2) نبيل علي، "العرب وعصر المعلومات"، مرجع سابق، ص: 336، بتصرف.

التي سيأتي الحديث عنها في المبحث الخاص بالتطبيقات الحاسوبية المتعلقة باللغة العربية.

3. اللسانيات الحاسوبية:

لقد قدمت مجموعة من التعاريف للسانيات الحاسوبية من أبرزها:
" اللسانيات الحاسوبية: هي الدراسة العلمية للغة من وجهة نظر حاسوبية، حيث يعمل بموجها اللغويون والحاسوبيون على تقديم نماذج تعمل على معالجة اللغة الطبيعية (على مستوى الإنتاج والفهم) بالإضافة إلى أنواع مختلفة من الظواهر اللغوية. فالعمليات الحاسوبية يتم برمجتها للحصول على مجموعة من التطبيقات، مثل نظم التعرف على الكلام، نظم توليد الكلام، نظم الرد الصوتي الآلي، ونظم محركات البحث على شبكة الإنترنت، ونظم تحرير النصوص، ونظم تعليم اللغة"⁽¹⁾
"اللسانيات الحاسوبية: هي تداخل بين اللسانيات والتطبيقات اللغوية المحوسبة، حيث يتم تزويد الحاسوب بالآليات التي تعمل على معالجة اللغة الطبيعية بجميع مستويات الدرس اللساني"⁽²⁾.
يتبين من خلال هذين التعريفين أن اللسانيات الحاسوبية تهتم بدراسة اللغة الطبيعية دراسة علمية حاسوبية تعتمد على المستوى النظري للسانيات بكل مستوياتها؛ التركيبية، والدلالية، والصرفية، والصوتية، والسعي إلى نمذجتها على مستوى الحاسوب.

(1) Jack C. Richards and Richard Schmidt (2002): "Langman Dictionary Of Language Teaching And Applied Linguistics", p: 100.

(2) Hadumod Dussmann (1998) : "Routledge Dictionary of Language and linguistics", p:224

ويضيف كل من إغور بلوشكوف (Igor A. Bolshakov) وألكسندر غلبخ (Alexander Gelbukh) أن اللسانيات الحاسوبية هي مقابل للمعالجة الآلية للغة الطبيعية، إن اللسانيات الحاسوبية ليست إلا بنية للنظم الحاسوبية التي تعمل على معالجة كلمات، ونصوص اللغة الطبيعية⁽¹⁾.

فاللسانيات الحاسوبية إذن تقوم بدراسة نصوص اللغات الطبيعية بالكيفية التي تدرسها اللسانيات النظرية، وذلك ببرمجة مستويات الدرس اللساني، فهي بذلك تنشطر إلى شطرين: جانب نظري، وجانب تطبيقي:

يتمثل الجانب النظري في النظرية اللسانية الحديثة وما توصلت إليه في دراستها للغة الطبيعية، باعتبارها "الدراسة العلمية للغة"⁽²⁾ بكل مستوياتها؛ الصرفية، والتركيبية، والدلالية، والصوتية، والتداولية.

أما الجانب التطبيقي فيتمثل في استثمار كل ما حددته اللسانيات النظرية من آليات ومستويات الدرس اللساني في دراسة لغة من اللغات، حيث يتم برمجتها وفق نظم آلية ذكية اعتماداً على اللغة التي يتعامل بها الحاسوب، بالإضافة إلى الاهتمام "بالنتائج العلمي لنمذجة الاستعمال الإنساني للغة، وهو يهدف إلى إنتاج نظم ذات معرفة باللغة الإنسانية"⁽³⁾. أي أنها تسعى إلى العمل على إيجاد طريقة مثلى لتحسين التفاعل الآلي/الإنساني، قصد فهم الكيفية التي يشتغل بها الذهن البشري.

(1) Igor A. Bolshakov and Alexander Gelbukh (2004): "Computational Linguistics: Models, Resources, Applications", p:25

(2) Ferdinand Saussure (1995): "Course De Linguistique Générale "P :232.

(3) وليد أحمد العناتي (2006): "اللسانيات الحاسوبية العربية رؤية ثقافية"، مجلة فكر ونقد، ع:82 ص:27.

كما اهتمت اللسانيات الحاسوبية بإحصاء الظواهر اللغوية على المستوى الصوتي، والمستوى التركيبي، والمستوى الصرفي، بالإضافة إلى عمليات التحليل والتركيب اللغويين والتي تمثل الخطوات الأولى للمعالجة الآلية للغة.⁽¹⁾

ويبقى الهدف الذي تسعى إليه اللسانيات الحاسوبية وراء أبحاثها هو تهيئة وتوفير كفايات لغوية للحاسوب تشبه العمليات الذهنية للإنسان حين يستقبل اللغة ويدركها ويفهمها ثم يعيد إنتاجها وفق المطلوب، والكفاية المقصودة إنما تعني:

- استذخال قواعد اللغة، في نظامها الصوتي، وأنساقها الصرفية، وأنماط نظمها الجملي، وأنحاء أعاريها، ودلالات ألفاظها، ووجوه استعمالها وأساليب البيان، وأحكام رسمها الإملائي.

- إنتاج ما لا ينتهي من الأداءات اللغوية الصحيحة.
- مرجع في تمييز الخطأ من الصواب.
- ومن تمام هذه الكفاية كفاية تواصلية، تتعلق بالعناصر الخارجية التي تتدخل في الموقف الكلامي".⁽²⁾

4. موضوع اللسانيات الحاسوبية:

امتازت اللسانيات الحاسوبية منذ ظهورها بالانفتاح على مجموعة من القضايا المتعلقة بالمعالجة الآلية للغات الطبيعية، بل استطاعت أن تواكب كل التطورات، خاصة بعدما أصبح استعمال الحواسيب أمراً ضرورياً في الحياة اليومية لمجموعة من الشرائح الإنسانية صغيرها وكبيرها، بالإضافة إلى انتشار الشبكة العنكبوتية التي لم يعد تصور الحياة اليومية بدونها، مما أدى إلى تزايد هائل في

(1) سامي عياد حنا، ونجيب جريس، وكريم زكي حسام الدين (د.ت): "معجم اللسانيات الحديثة: انجليزي-عربي"، ص: 26، بتصرف.

(2) وليد العناتي، (2012): "العربية في اللسانيات التطبيقية"، ص: 278، بتصرف.

البيانات الرقمية التي لم يعد استيعابها ممكناً إلا باستعمال الحاسوب، وقد أفضت المعالجة الآلية لتلك البيانات إلى ظهور العديد من المواضيع نلخص البعض منها في ما يلي:

- إحصاء الظواهر اللغوية على المستوى الصوتي والصرفي والتركيبية.
- المعالجة الآلية للمستويات الصوتية والصرفية والتركيبية والدلالية.
- صناعة المعاجم، سواء تعلق الأمر بتخزين المعاجم إلكترونياً، أو تحليل العلاقات التي تربط بين مفردات المعجم وتحديد السمات الدلالية لها.
- بناء نظم آلية لفهم السياق اللغوي في صورته المنطوقة والمكتوبة.
- بناء نظم لتعلم اللغات.
- التلخيص الآلي.
- التنقيب في النصوص.
- تحليل الآراء والمشاعر وتصنيفها.

5. معالجة اللغات الطبيعية:

اهتم الباحثون في مجال الذكاء الاصطناعي بمعالجة اللغة البشرية، ومحاولة الوصول إلى المعاني التي يحملها النص اعتماداً على النمذجة اللغوية التي تساعد الحاسوبي في برمجة المستويات اللسانية على مستوى الحاسوب، وقد جاء حقل معالجة اللغة الطبيعية ضمن مجال الذكاء الاصطناعي تزامناً مع التيار اللساني الذي تقدم به اللساني نعوم تشومسكي، والذي جعل من توليد اللغة الطبيعية (Natural

Language Generation) تحديا كبيرا، سعى إلى تحقيقه اعتمادا على مجموعة من القواعد الرياضية والمنطقية، لتكون بذلك اللسانيات أهم الحقول المعرفية التي ساهمت في التأسيس لمجال معالجة اللغات الطبيعية ضمن مجال الذكاء الاصطناعي إلى جانب مجموعة من الحقول المعرفية الأخرى كعلم النفس وعلم والاجتماع والمنطق إلى غيرها... مما مكن مجموعة من الباحثين منذ البداية من بناء أنظمة حاسوبية؛ تعمل على إنتاج نص ذي لغة جيدة ومعبرة وخالية من التعقيد، وتمثيلها بشكل متطور داخل الحاسوب،⁽¹⁾ كما سعى الباحثون إلى بناء أنظمة تمكن من تقديم إجابات لأسئلة حول موضوع موجود مسبقا ضمن قاعدة المعارف⁽²⁾ ...، فيكون بذلك نظام توليد اللغة الطبيعية أسهل من نظام فهمها لأنه ينطلق من المعنى إلى النص عكس نظام فهم اللغة الطبيعية الذي يصعب التحكم فيه نظرا لتعقيد بنية اللغة التي يتلقاها كمدخلات.⁽³⁾

ومن أبرز الاهتمامات العامة التي ساهمت في بروز مجال معالجة اللغات الطبيعية، نجد الترجمة الآلية واستخراج المعلومات، وكذا التصنيف والتجميع، وينطوي تحت هذه الاهتمامات العامة اهتمامات فرعية أخرى سنشير إلى البعض منها ضمن مباحث هذا البحث، والإشارة إلى أهم المستويات التي حققت في برمجتها دقة عالية، والتي لا زالت في طور التدريبات من أجل الحصول على دقة يمكن الوثوق بنتائجها، ثم تقديم أهم التطبيقات المتعلقة بالمستويات اللسانية التي تمكن مجموعة من فرق البحث من إنتاجها.

(1) Paul Semaan, (2012) " Natural Language Generation: An Overview ", Journal of Computer Science 1, no 3: 8. بتصرف .

(2) Ehud Reiter et Robert Dale, "Building Applied Natural Language Generation Systems", Natural Language Engineering 3, no 1 (mars 1997): 57-87, <https://doi.org/10.1017/S1351324997001502>.

(3) Paul Semaan, (2012) "Natural Language Generation: An Overview ", Journal of Computer Science 1, no 3: 8. بتصرف .

6. فهم اللغة الطبيعية:

بدأت المعالجة الآلية للغات الطبيعية يبرز في الخمسينيات من القرن الماضي، فشرع مبلورو البرامج المعلوماتية في اختبار مدخلات لغوية بسيطة، ثم بادرت مجموعة من الباحثين مع بداية الستينيات في تطوير مجال جديد سمي بالفهم الآلي للغات الطبيعية (Natural Language Understanding) بدافع الرغبة في الحصول على أجهزة حواسيب تستطيع فهم إدخال اللغة الأكثر تعقيداً باعتبار هذا المجال فرعاً من فروع معالجة اللغات الطبيعية، ويُعرف مجال الفهم الآلي للغات الطبيعية كذلك بالنظم القائمة على المعرفة التي تستخدم كمية كبيرة من المعلومات حول مجال معين قصد فهم التعابير الإنسانية فهما صحيحاً باعتماد مجموعة من الخوارزميات المهيئة والتي يتم تخزينها داخل النظام بالإضافة إلى استخدام بعض مخططات تمثيل المعرفة مثل: المنطق أو الدلالات الإجرائية أو الشبكات الدلالية،⁽¹⁾ ليكون بذلك مجالاً ينحصر اهتمامه باللغة بشكل أساسي من أجل جعل الحواسيب قادرة على فهم ما تتضمنه النصوص من معارف ومعلومات.

يستعين الفهم الآلي للغات الطبيعية بمجموعة من خوارزميات الذكاء الاصطناعي لتمكين النظم من الفهم الآلي للقدرات الإنسانية مثل: تحديد المواقف والمشاعر، وإنتاج النصوص أو تلخيصها النصوص. خلاصة.

تميزت اللغة العربية كغيرها من اللغات الطبيعية بمجموعة من المميزات التي جعلتها من اللغات التي تحظى باهتمام الباحثين من بينها؛ الانتظام الصوتي والتماسك اللغوي والتنوع الدلالي إلى غيرها من المميزات، ويرجع الفضل في ذلك إلى اللغويين

(1) Avron Barr et Edward A. Feigenbaum, éd., « Chapter IV - Understanding Natural Language », in The Handbook of Artificial Intelligence (Butterworth-Heinemann, 1981), 223-321, <https://doi.org/10.1016/B978-0-86576-089-9.50009-0>, بتصرف.

القدماء الذين سعوا جاهدين إلى التأصيل للغة العربية، والحرص على تدوين كل قواعدها، والتعريف بأقسام كلامها، وتعيين مميزات ووظائف كل أقسامها، والتأكيد على أن تكامل أجزاء الكلام العربي، هو ما يستطيع من خلاله العربي تحقيق كلام عربي يتسم بمعاني فصيحة، وبفضل تطور الدراسات وتقدم الأبحاث اللغوية فقد ظهر علم اللسانيات الذي جاء ليدرس اللغات الطبيعية دراسة علمية؛ دراسة مهدت الطريق إلى تلاقح المجالات المعرفية في ما بينها، الأمر الذي ساهم في ظهور اللسانيات الحاسوبية التي وجهت إلى الاهتمام باللغات الطبيعية وجهة حاسوبية هدفها معالجة اللغة الإنسانية معالجة آلية لفهمها، وإدراك آليات إنتاجها، مما سيحفز العديد من المؤسسات البحثية إلى تعميق البحث وتمعين النظر في المستويات اللسانية، وفي المبحث الموالي سنحاول الإحاطة بجهود الباحثين الذين استثمروا طاقاتهم، وكرسوا جهودهم في حوسبة المستويات اللسانية العربية.

المبحث الثاني: اللغة العربية والتحليل الآلي للمستويات اللسانية.

مدخل:

باتت الجهود الفردية لا تجدي في مجال البحث العلمي الرصين؛ لعدم قدرتها للتصدي للقضايا الراهنة كيفما كان نوعها، بل أصبح من الضروري التكامل المنهجي والتداخل المعرفي بين مجموعة من العلوم أمراً ضرورياً من أجل الماضي قدما بتطوير الأبحاث، وتدقيق النتائج في شتى مجالات الثورة المعرفية التي أضحت تغزو أواسط كل التخصصات العلمية، وما تحمله من تحديات العصر، وما ستأخذه على عاتقها مستقبلاً في إطار تضافر الجهود الجماعية، وهو ما يعرف في عصرنا الحديث بالعلوم البينية (Interdisciplinary Systems)؛ القاضية إلى ضرورة الامتثال إلى التداخل المعرفي، وتشارك التخصصات بين الباحثين، من أجل تكوين عائلة علمية تتصف بالتعالق المعرفي على حد تعبير الدكتور عبده الراجحي، ونظراً لما تؤديه اللغة من دور باعتبارها حاضنة لمجموعة من الجوانب؛ فقد تم وضعها كعنصر أساس في العديد من الحقول المعرفية الإنسانية؛ كحقل علم الاجتماع، وعلم النفس، والفلسفة وغيرها من الحقول.

وفي هذا الإطار أصبح لزاماً وضع اللغة الطبيعية في دائرة تضم مجموعة من العلوم، لتتولى اللغة بذلك التأليف بين هذه العلوم المتمثلة في العلوم الإنسانية (علم النفس، وعلم الاجتماع، وعلم المعرفة، ...) والعلوم الطبيعية (البيولوجيا والكيمياء) والعلوم الصورية (الإحصاء، والمنطق، والحاسوب، ...) ⁽¹⁾ ثم تكوين عائلة معرفية تمكن من إفهام اللغة للمضي بها قدما والحفاظ عليها من الزوال، خاصة أنها أساس التعايش بين الإنسان ومفتاح التواصل بين الشعوب في بقاع العالم الذي أصبح قرية صغيرة

(1) نبيل علي، ونادية حجازي (2005)، "الفجوة الرقمية"، عالم المعرفة، ع: 318، بدون طبعة، مطابع السياسة - الكويت: المجلس الوطني للثقافة والفنون والأدب، الكويت، ص: 316 بتصرف.

نتيجة التقدم العلمي السريع، فلوحظ أنه لتكوين عائلة تحمل على عاتقها هذا الهدف الأسمى، لا بد لها من التزاوج بين علوم الحاسوب وعلوم اللغة، خاصة بعدما بينت القدرة الحاسوبية مكانتها في تعزيز مناهج اللغة، وتيسير وسائلها وتبسيط مستوياتها وتسريع نموها واسترجاع قيمتها وسط أخواتها، ففرغت عائلة معرفية جديدة سميت بالمعالجة الآلية للغات الطبيعية تناولت اللغة بمختلف مستوياتها الصوتية والصرفية والمعجمية والتركيبية والدلالية والتداولية من أجل تحقيق النمو المعرفي في ظل الانفجار المعلوماتي واللاحق بركب الثورة التكنولوجية.

لقد تأسس حقل المعالجة الآلية للغات الطبيعية بناء على إطار نظري دقيق من أجل محاولة إفهام اللغة الطبيعية ألياً مستعينا بالشراكة بين مجموعة من العلوم من بينها اللسانيات والرياضيات والمنطق، باعتبارها هندسة لتنزيل الإطار النظري على شكل نظم مختلفة باختلاف المستويات اللسانية، سعياً لتحقيق الفهم الآلي للغة من خلال توصيف المستويات اللسانية توصيفاً منطقياً ورياضياً يسهل برمجتها على مستوى الحاسوب. وسنعرض ملخصاً مقتضباً لمجموعة من المستويات اللسانية، ونخص بالذكر: (المستوى الصوتي، والصرفي، المعجمي، والتركيب، والدلالي)، باعتبارها مستويات يستحيل الاستغناء عنها في إنتاج تطبيقات المعالجة الآلية للغات الطبيعية.

1. التحليل الصرفي للنصوص العربية:

تتميز اللغة العربية بسمّة الاشتقاق الصرفي الذي يعتمد على مجموعة من الصيغ الصرفية التي تسهل التعرف على أصل الكلمة، لهذا الأساس نجد الدراسات الحديثة قد أولت اهتماماً كبيراً لهذا المستوى باعتباره نقطة الالتقاء مع باقي المستويات؛ بل يعتبر المستوى الصرفي ركيزة للمستوى الصوتي ومدخلاً للمستوى النحوي، كما يمثل المستوى الصرفي أداة أساسية لنمذجة المستويين المعجمي والدلالي، كل ذلك وجه البحث اللساني الحاسوبي إلى إعطاء الأهمية الكبرى لهذا المستوى سعياً لتحقيق هدفين رئيسيين نلخصهما في ما يلي:

● محاكاة التفكير الإنساني العربي في مدى قدرته على التعرف على أصل الكلمة واشتقاقاتها.

● تسهيل عملية تحليل النصوص التي أصبحت متزايدة بكثرة في مواقع الإنترنت.

ومن خلال هذين الهدفين سعى الباحثون إلى بلورة محلات صرفية قادرة على استخلاص عناصر بنية الكلمة العربية، وتحديد مجموعة من السمات المتمثلة أساساً في:

● السمات الصرفية: المتعلقة بتحديد جذر الكلمة العربية ووزنها الصرفي.

● السمات الصرف صوتية: والتي تهتم بالتغيرات الصوتية التي تطرأ على الكلمة مثل الإعلال، والإبدال، والقلب...

● السمات الصرف نحوية: وتهتم بتحديد ما لحق الكلمة من سوابق ولواحق وزوائد، كما تهتم بتقسيم الكلمة إلى اسم، أو فعل، أو حرف.⁽¹⁾

وقد انطلقت أول فكرة لبناء محلل صرفي سنة 1961 مع ديفيد كوهين (David Cohen) الذي خلص فيها إلى تقديم نموذج لمعالجة صرفية أدنوية (Minimaliste) منطلقاً من أن الكلمة العربية مركبة من مجموعة من العناصر اللغوية، قد يصل تركيبها إلى الجملة من قبيل "أنالهماه"، المركبة من جملة فعلية، فعلها متعدي إلى مفعولين جاء ضميرين متصلين، وفاعله ضمير مستتر، وقد تأسست عملية التحليل حسب ديفيد كوهين على صياغة كل الأشكال اللسانية وفق أوزان وجدوع.

(1) Abdelhamid El Jihad, (2013) « نظام تصريف الأفعال في اللغة العربية »، Revue de l'Information Scientifique et Technique 20، no 2: 27-48. بتصرف.

بين نموذج كوهين أن الكلمة العربية يمكن أن تنقسم إلى خمسة أجزاء
نوضحها على الشكل (1) الآتي:⁽¹⁾

Poste fixe	Suffixe	Coprs schématique	Préfixe	Antéfixe
الضمائر	اللواحق النحوية	الكلمة	السوابق	الحروف (حروف التعريف - حروف الجر)
Enclitique		Forme fléchie	Proclitique	

الشكل 1: بنية الكلمة حسب دافيد كوهين

وقد تطورت الأبحاث بعد دفيد كوهين والهادفة إلى بلورة المحللات الصرفية العربية من أجل التعرف على الكلمة وما ارتبط بها من سوابق ولواحق ومقدمات، وعلى هذا الأساس فقد اعتمدت المحللات الصرفية في تحليل الكلام العربي على ثلاث مقاربات وهي:

- المقاربة المؤسسة على المعجم (Morphological Analysis)
(Based On Lexicon).
- مقارنة التجديع الخفيف (Light Stemming).
- مقارنة التحليل الإحصائي⁽²⁾ (Statistical Analysis).

(1) Nebia BENZATER (2014) « Analyse Morphologique du Texte Arabe pour Son Indexation Sémantique. » (Magistère en Informatique، Oran، Université des Sciences et de la Technologie - Mohamed Boudiaf، p: 16، بتصرف .

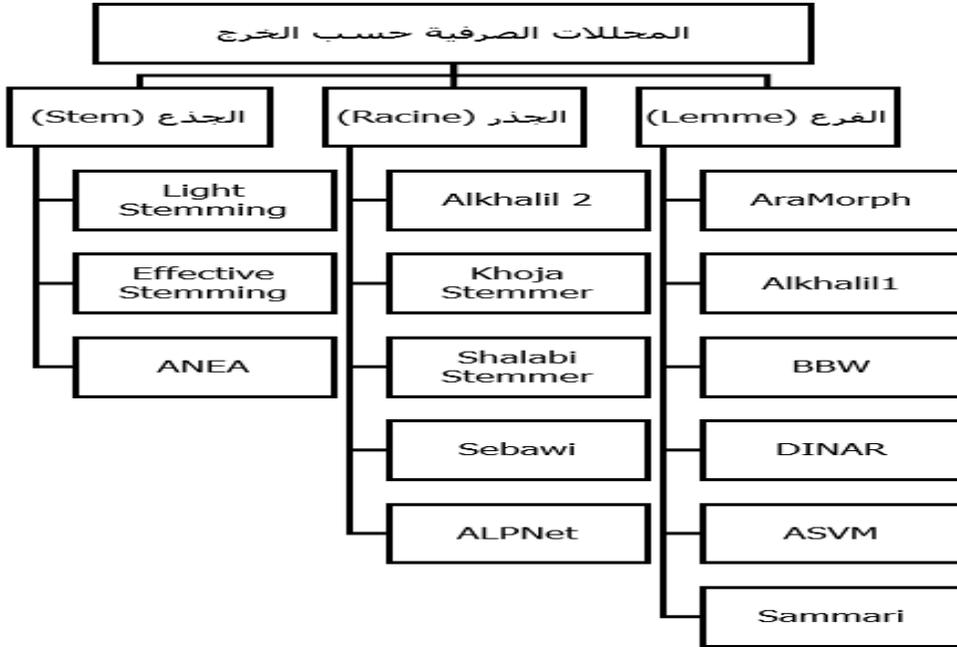
(2) Leah S. Larkey، Lisa Ballesteros، et Margaret E. Connell (2007) « Light Stemming for Arabic Information Retrieval » ، in Arabic Computational Morphology : Knowledge-Based and Empirical Methods، éd. Par Abdelhadi Souidi، Antal van den Bosch، et Günter Neumann، Text، Speech and Language Technology (Dordrecht : Springer Netherlands، 221-43، https://doi.org/10.1007/978-1-4020-6046-5_12، بتصرف.

كما تتفاوت المقاربات المتعلقة بخرج (output) النظام بعد عملية التحليل، وهو ما بينته الدراسات في كون خرج المحللات الصرفية العربية لا يخرج على واحدة من مخرجات ثلاثة وهي:

- محللات صرفية تعتمد في خرجها على الفرع (Lemma).
- محللات صرفية تعتمد في خرجها على الجذر (Racine).
- محللات صرفية تعتمد في خرجها على الجذع (Stem).

ونقدم فيما يلي مجموعة من المحللات التي اعتمدت أصناف المخرجات

السابقة نلخصها في الشكل (2) الآتي:



الشكل 2: المحللات الصرفية العربية حسب الخرج⁽¹⁾؛

(1) BENZATER Nebia, « Analyse Morphologique du Texte Arabe pour Son Indexation Sémantique. » (Magistère, Oran, Université des Sciences et de la Technologie - Mohamed Boudiaf, 2014), 26.

1.1. مراحل التحليل الصرفي الآلي:

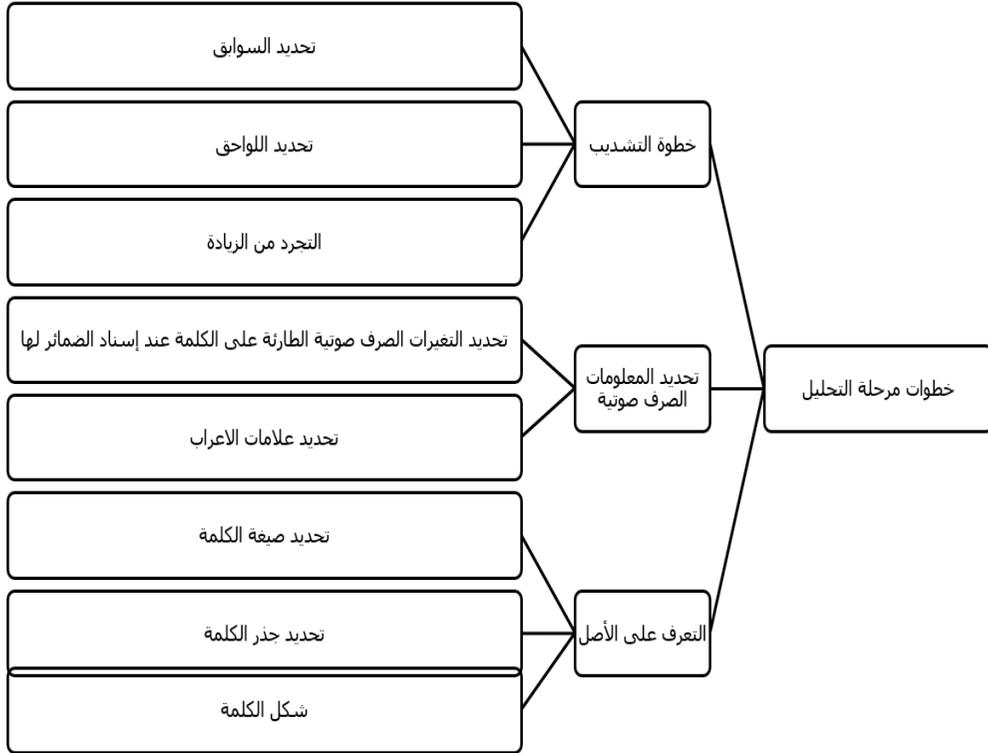
يهتم علم الصرف بدراسة الكلمة سواء كانت فعلا أو اسما أو حرفا، فيقدم المعلومات الصوتية، والصرفية، والنحوية لكل نوع من أنواع الكلام العربي، تتمثل هذه المعلومات عموما في تحديد: جذر الكلمة إذا كانت اسما أو فعلا على اعتبار أن الحرف لا أصل له، كما يحدد نوع الكلمة، والسوابق، واللواحق، والوزن... إلى غيرها من المعلومات، وعلى هذا الأساس بنيت المحللات الصرفية الآلية العربية، ونظرا لغياب شكل الكلمات العربية الذي ينجم عن تعدد احتمالات الكلمة، وكذا غنى اللغة العربية من حيث تعدد ألفاظها، فإن التحليل الصرفي الآلي يعتبر الكلمة العربية أثناء تحليلها على أنها اسما ثم فعلا ثم حرفا. إن عملية التحليل في غالبية نظم التحليل الصرفي تمر بثلاث مراحل أساسية هي:

- مرحلة الدخل.
 - مرحلة التحليل.
 - مرحلة الخرج.
- وفيما يلي تفصيل لكل مرحلة.
- مرحلة الدخل:

تتمثل مرحلة الإدخال بإدراج النص أو الجملة المراد تحليلها، وينبغي الإشارة إلى أن المحلل الصرفي يركز في عملية التحليل على الكلمة بمفردها بدون اعتبار السياق الذي جاءت فيه، وبما أن اللغة العربية تتميز بخاصية الشكل، فإن المحللات الصرفية تقبل الكلمة أو الجملة أو النص مشكولا كاملا، أو مشكولا جزئيا أو غير مشكول، إلا أن الشكل يقلل من عدد الاحتمالات في مرحلة الخرج أو قد يقدم تحليلا واحدا، عكس نظيرتها غير المشكولة.

● مرحلة التحليل:

تتميز مرحلة التحليل بتفكيك النص إلى كلمات اعتماداً على الفراغات الموجودة بين الكلمة والكلمة الأخرى، ثم يتم تفكيك كل كلمة إلى عدة أجزاء يتم فيها تحديد السوابق، واللواحق، والمقدمات، والأوزان الصرفية، والفئات النحوية، وغيرها من المعلومات الصرفية التي تختلف باختلاف نوع الكلمة، وتتلخص خطوات مرحلة التحليل في ثلاث خطوات أساسية، يتفرع عن كل خطوة منها خطوات فرعية نلخصها في الشكل (3) الآتي:



الشكل 3: خطوات مرحلة التحليل في المحللات الصرفية العربية

وللتمكن من فهم أكثر للخطوات السابقة الذكر نورد أمثلة تجسد كيفية تحليل الفعل والاسم والحرف تحليلاً آلياً تم تلخيصها من مقال "مواصفات نظام التحليل الصرفي في اللغة العربية" لمروان البواب، الذي نشر عقب اجتماع خبراء

المحللات الصرفية الحاسوبية للغة العربية بدمشق، حيث تمكن مروان البواب في هذا المقال من تدقيق تصور التحليل الصرفي للكلام العربي مقدما خطوات تحليل الاسم والفعل والحرف.⁽¹⁾

1. خطوات تحليل الفعل:

□ الخطوة الأولى: تشذيب الفعل:

يعمل المحلل في هذه المرحلة على إزالة كل سوابق ولواحق الفعل للحصول على الجذر، أو الجذع، أو الفرع حسب الخرج المعتمد في المحلل الصرفي، وتتم عملية التشذيب وفق ثلاث خطوات فرعية وهي:

● تحديد سوابق الفعل:

حدد العرب القدماء في تعريفهم للكلام العربي مميزات يتميز بها كل قسم من أقسام الكلام، ففي هذه المرحلة يقوم المحلل الصرفي بتحديد سوابق الفعل، التي قد تكون سابقة واحدة أو سابقتان؛ وسوابق الفعل هي:

○ حروف المضارعة الأربع المجتمعة في "أنيت، أو نأتي، أو نأيت".

○ حرف الاستقبال "السين".

○ الحروف الناصبة وهي: لام الجحود، لام التعليل، فاء السببية، وواو

المعية.

○ حروف العطف التي تلتصق بالكلمة وهي: الفاء، والواو.

○ همزة الاستفهام.

(1) <http://www.arabacademy.gov.sy/uploads/nadawat/10/13.pdf> consulter le 23 juin

إلا أن هذه الحروف قد تكون أصلية في الفعل ونمثل لذلك بالفعل

الآتي:

▪ وَلَسَعَتْ: و(العطف)+ل(القسم)+سعت

▪ وَلَسَعَتْ: و(العطف)+لسعت.

● تحديد لواحق الفعل:

ترتبط بالفعل لاحقتان هما:

○ ضمائر الرفع المتصلة وهي: تاء المفرد، وألف الاثنين، وواو الجماعة.

○ ضمائر النصب المتصلة وهي: "الياء، للمتكمم الواحد، و"نا" للمتكمم المعظم نفسه أو معه غيره، و"ك" للمخاطب المفرد المذكر، و"ك" للمخاطبة المفردة المؤنثة، و"كُما" للمثنى المخاطب، و"كُم" لجماعة الذكور المخاطبين، و"كُنَّ" لجماعة الإناث المخاطبات، "هُ" للغائب المفرد المذكر، و"ها" للغائبة المفردة المؤنثة، و"هُما" للمثنى الغائب، و"هُم" لجماعة الذكور الغائبين، و"هُنَّ" لجماعة الإناث الغائبات

وللتعرف عن كيفية تحليل الفعل مع اللاحقة نمثل بالفعل الآتي:

▪ دَلَّكَ: دَلَّ+ك (المخاطب المفرد المذكر).

▪ دَلَّكَ: دَلَّكَ باعتباره فعلاً مضارعاً يُدَلِّكُ

● التجريد من الزيادة:

يستعين المحلل الصرفي بمجموعة من الجداول تضم الأوزان الصرفية العربية الممكنة وفق الأفعال المجردة والمزيدة، بأقسامها الثلاثية والرباعية، ونظراً لغياب الشكل في الكلام العربي فإن المحلل الصرفي الآلي يقدم مجموعة من الاحتمالات

فيما يتعلق بالمجرد والمزيد، وكمثال على ذلك نأخذ فعل: "علم"; حيث يقدم المحلل الصرفي باعتبار التجرد والزيادة احتمالين هما:

- علم: عَلِمَ.
- علم: عَلَّمَ.

□ الخطوة الثانية: تحديد المعلومات الصرف صوتية

تتمثل الخطوة الثانية في كشف المحلل الصرفي للعمليات الصرف صوتية الممكن حدوثها للفعل عند إسناده للضمائر، أو في حالة بناء الفعل للمجهول أو بنائه للمعلوم، وكذلك إعراب الفعل وما قد ينتج عنه من تغييرات خاصة إذا كان الإعراب بالحروف، وتشمل هذه التغييرات بكثرة الأفعال المعتلة وتندرج ضمن هذه الخطوة خطوات فرعية يقوم بها المحلل الصرفي الآلي وهي:

● إسناد الفعل:

يعمل المحلل الصرفي الآلي على تحديد ما إذا كان الفعل قد اسند إليه ضمير أم لا، باعتماد مجموعة من القرائن، فإذا أخذنا على سبيل المثال فعل "رَمَيْتُ"، مشكولاً فالمحلل الصرفي يتمكن من التعرف على الضمير المسند إلى الفعل "رمى" باعتماد قواعد صوتية، كما أن المحلل الصرفي يقدم مجموعة من الاحتمالات إذا كان الفعل "رمى" غير مشكول، وتتمثل هذه الاحتمالات في:

- رَمَيْتِ.
- رَمَيْتُ.
- رَمَيْتَ.
- رَمَيْتَ.

● تحديد صيغة البناء للمعلوم والبناء للمجهول:

إن غياب شكل الكلمات العربية بصفة عامة والأفعال على وجه الخصوص يؤثر تأثيراً كبيراً؛ إذ يصعب على المحلل الصرفي الآلي في كثير من الأحيان التعرف على ما إذا كان الفعل مبنيًا للمعلوم أو مبنيًا للمجهول؛ ولهذا فإن المحلل الصرفي يقدم احتماليين:

- البناء للمعلوم "رَمَيْتُ".

- البناء للمجهول "رُمِيْتُ".

● تحديد العلامة الإعرابية:

يقدم المحلل الصرفي في غياب الشكل جميع الاحتمالات الممكنة للحركات الإعرابية، سواء كانت حركات أم حروفاً مع الأخذ بعين الاعتبار التغييرات الصرف صوتية وما لها من تأثير في حذف حروف العلة؛ وكمثال على ذلك نقدم فعل "رمى"، فإذا جاء في صيغة الأمر فإنه يبنى بحذف حرف العلة فنقول: "إِزْم".

□ الخطوة الثالثة: التعرف على الأصل

يتمكن المحلل الصرفي في كثير من الحالات من التعرف على أصل الفعل، فيقدم المحلل حسب الخرج المعتمد (إما الجذر، وإما الجذع، وإما الفرع)، ويتفرع عن هذه الخطوة بعض الخطوات على الشكل الآتي:

● تحديد صيغة الفعل:

بعد حذف اللواحق يعمل المحلل الصرفي على تقديم صيغة الفعل المعالج، والتي قد تحيل على أكثر من صيغة إذا كان الفعل غير مشكول، وكمثال على ذلك نقدم فعل "أكرم":

- في الزمن الماضي: "أَكْرَمَ" على وزن "أَفْعَلْ".

- في الزمن المضارع: "أُكْرِمُ" على وزن "أُفْعِلْ".

- فى زمن الأمر: "أَكْرِمُ" على وزن "أَفْعِلْ".

● تحديد الجذر:

يقوم المحلل الصرفى بتقديم جذر الفعل المعالج، فى هذه المرحلة تكثر الاحتمالات إذ قد يكون للفعل أكثر من جذر، لذلك فإن المحلل الصرفى فى هذه الحالة يقدم جميع الجذور الممكنة المرتبطة بالفعل الخاضع لعملية التحليل، وكمثال على ذلك نقدم فعل "يعد"، فإن احتمال جذر هذا الفعل يُرد إلى ثلاثة احتمالات وهى:

▪ يعد: يَعِدُ من جذر "وع د".

▪ يعد: يَعُدُّ من جذر "ع د د".

▪ يعد: يَعُدُّ من جذر "ع و د".

● شكل الفعل:

يقدم المحلل الصرفى للفعل غير المشكول جميع الاحتمالات الممكنة لشكله.

ب. خطوات تحليل الاسم:

□ الخطوة الأولى: تشذيب الاسم:

يعمل المحلل فى هذه المرحلة على إزالة كل سوابق ولواحق الاسم لإنتاج الجذر، أو الجذع، أو الفرع حسب خرج المحلل الصرفى، وتجدر الإشارة إلى أن الاسم ينفرد بسابقة "ال" التعريف، كما يختلف عن الفعل فى السوابق المتعلقة بالمضارعة، "أنيت"، بالإضافة إلى أن الفعل والاسم يشتركان فى حروف العطف التى تلتصق بالكلمة، وتتم عملية تشذيب الاسم بنفس الخطوات الفرعية التى يمر بها الفعل وهى:

● تحديد سوابق الاسم:

تتمثل السوابق الخاصة بالاسم فى:

○ حروف النداء (أ - وا - ...)

○ حروف القسم (ت - و - ب)

○ (ال) للتعريف.

○ وحروف الجر (ل - ب - ك - ت - ...).

يقوم المحلل الصرفي في هذه المرحلة بتحديد سوابق الاسم، التي قد تكون سابقة واحدة أو سابقتين أو ثلاث سوابق، أو أربعة سوابق، ونمثل لذلك بالأمثلة الآتية:

▪ الكتاب.

▪ وكتاب.

▪ والكتاب.

▪ وبالكتاب.

▪ أقبالكتاب.

● تحديد لواحق الاسم:

تضاف للاسم الضمائر المتصلة وتأتي غالباً لأداء وظيفة التعريف وهذه الضمائر هي: "الياء"، للمتكلم الواحد، و"نا" للمتكلم المعظم نفسه أو معه غيره، و"ك" للمخاطب المفرد المذكر، و"ك" للمخاطبة المفردة المؤنثة، و"كُما" للمثنى المخاطب، و"كُم" لجماعة الذكور المخاطبين، و"كُنن" لجماعة الإناث المخاطبات، و"هُ" للغائب المفرد المذكر، و"ها" للغائبة المفردة المؤنثة، و"هُما" للمثنى الغائب، و"هُم" لجماعة الذكور الغائبين، و"هُنن" لجماعة الإناث الغائبات، إلا أنه قد تكون إحدى هذه الضمائر أصلية في الكلمة، وكمثال على ذلك:

فكه: فَكُّ + هُ.

▪ فكه: فَكَّ + هُ.

▪ فكه: فَكِّ + هِ.

▪ فكه: فَكِّه.

□ الخطوة الثانية: تحديد المعلومات الصرفية

كما سبقت الإشارة إلى أن هذه الخطوة تختص بمعالجة القضايا الصرف صوتية، ويدخل ضمن هذه الخطوة تحديد عنصرين أساسيين:

● تحديد نوع الاسم:

يقوم المحلل الصرفى بتحديد الصيغ الصرفية للكلمة (مفردة أو مثنى أو جمع).

▪ المفرد: مسلم.

▪ المثنى: مسلمان.

▪ الجمع: مسلمون.

● تحديد العلامة الإعرابية:

يقدم المحلل الصرفى الحركات الإعرابية سواء كانت حركات أم حروفاً، وكمثال

على ذلك:

▪ المفرد المرفوع: مُسَلِّمٌ.

▪ المفرد المنصوب: مُسَلِّمًا.

▪ المفرد المجرور: مُسَلِّمٍ.

▪ المثنى المرفوع: مُسَلِّمَانِ.

▪ المثنى المنصوب: مُسَلِّمَيْنِ.

▪ المثنى المجرور: مُسَلِّمَيْنِ.

▪ الجمع المرفوع: مُسَلِّمُونَ.

▪ الجمع المنصوب: مُسَلِّمِينَ.

▪ الجمع المجرور: مُسَلِّمِينَ.

الخطوة الثالثة: التعرف على الأصل

يتمكن المحلل الصرفي في كثير من الحالات من التعرف على أصل الاسم، فيقدم المحلل حسب الخرج المعتمد (إما الجذر، أو الجذع، أو الفرع)، وتتفرع عن هذه الخطوة بعض الخطوات على الشكل الآتي:

● تحديد وزن الاسم:

يتمكن المحلل الصرفي في هذه الخطوة من تقديم وزن الاسم باعتماد قاعدة معطيات تضم كل أوزان الاسم العربي، فيقدم المحلل وزن الاسم المعالج.

● تحديد الجذر:

يقوم المحلل الصرفي بتقديم جذر الاسم المعالج مثل:

▪ اسم: "س م و".

▪ دية: "و دي".

● تحديد نوع الاسم حسب الفئة النحوية:

يعمل المحلل على إعطاء الفئة النحوية للاسم باعتماد وزنه، فيستطيع التمييز بين اسم الفاعل واسم المفعول، والصفة المشبهة، إلى غيرها من الفئات النحوية العربية.

● تحديد الاسم المتصرف وغير المتصرف:

أغلب الأسماء العربية تعتبر متصرفة، لكن يوجد بعض الأسماء غير المتصرفة التي حدد لها العرب قرائن تعرف بها.

● تحديد نوع الاسم من حيث التذكير والتأنيث:

يحدد المحلل الصرفي الاسم المعالج ما إذ كان مؤنثاً أو مذكراً، فيتمكن من معرفة أن الأسماء الآتية: "شمس - أرض" كونها أسماء مؤنثة، كما يتمكن من معرفة أن الأسماء الآتية: "قمر - رجل" أسماء مذكرة، كما أن هناك بعض الأسماء التي

يستوي فيها التذكير والتأنيث، ونخص بالذكر الصفة مثل: "صبور"، فهي صفة يستوي فيها المذكر والمؤنث فنقول:

▪ رجل صبور.

▪ امرأة صبور.

● تحديد نوع الاسم من حيث النسبة والتصغير:

يتمكن المحلل الصرفي من التعرف على الكلمة سواء دلت على النسبة أو

التصغير ونمثل لذلك بالأمثلة الآتية:

▪ خلوي: نسبة إلى خلية.

▪ طُبِّيُّ تصغير طبي.

● شكل الاسم:

تتم في هذه الخطوة إعطاء جميع احتمالات شكل الاسم المعالج غير

مشكول، ونمثل لذلك بالكلمة الآتية:

▪ المدرسة: المُدْرِسَةُ

▪ المدرسة: المُدْرِسَةَ

▪ المدرسة: المُدْرِسَةِ

▪ المدرسة: المُدْرِسَةُ

▪ المدرسة: المُدْرِسَةَ

▪ المدرسة: المُدْرِسَةِ

إلى غيرها من الاحتمالات الممكنة لنفس الكلمة.

5. خطوات تحليل الحرف:

يقتصر تحليل الحرف على خطوتين أساسيتين هما: خطوة التشذيب،

وخطوة تحديد علامة البناء.

□ الخطوة الأولى: التشذيب

● تحديد السوابق:

قد تكون السابقة حرفاً واحداً، أو حرفين، أو ثلاثة أحرف ونمثل لذلك

بالحرف الآتى:

▪ غير: لغير.

▪ غير: فلغير.

▪ غير: أفلغير.

● تحديد اللواحق:

تتمثل لاحقة الحرف فى ضمائر الجر المتصلة مثل: غيرهما.

الخطوة الثانية: تحديد علامة البناء:

تتمثل هذه الخطوة فى تقديم علامات بناء الحروف المختلفة؛ فمنها ما يبنى

على السكون، ومنها ما يبنى على الضم ومنها ما يبنى على الفتح ومنها ما يبنى على الكسر ومنها ما يبنى على السكون.

● مرحلة الخرج:

تقوم نظم التحليل بتقديم نتائج الخرج الخاصة بكل قسم من أقسام الكلام،

ومن أمثلة ذلك ما يلى:

1. خرج الفعل:

نأخذ على سبيل المثال فعل "فكتبت" فخرج عملية تحليل الفعل تقدم

مجموعة من الاحتمالات، وسنكتفى بتقديم البعض منها مع المعلومات الخاصة بكل

احتمال على الشكل الآتى:

▪ فكتبت: فَكَتَبْتُ = فَ + كَتَبَ + ت

○ السابقة: ف.

- الجذر "ك ت ب".
- الجذع "كتب".
- الفرع "كَتَبَ".
- فعل ماض مبني للمعلوم.
- مجرد على وزن فَعَلَ يَفْعُلُ.
- مبني على الفتح.
- مسند إلى الغائب المفرد.
- متصل بتاء التأنيث.
- فكتبت: فَكَتَبْتُ = فَ + كَتَبْتُ + تُ
- السابقة: ف.
- الجذر "ك ت ب".
- الجذع "كتب".
- الفرع "كَتَبَ".
- فعل ماض مبني للمعلوم.
- مجرد على وزن فَعَلَ يَفْعُلُ جذر "ك ت ب".
- مبني على السكون.
- مسند إلى المتكلم المفرد.
- متصل بضمير الرفع (ت).

ب. خرج الاسم:

نأخذ على سبيل المثال الاسم "والدرس" فخرج عملية تحليل الاسم تقدم مجموعة من الاحتمالات، وسنكتفي بتقديم البعض منها مع المعلومات الخاصة بكل احتمال على الشكل الآتي:

- والدرس: وَالْدَرْسُ = وَ + ال + دَرْسٌ.
 - السابقة و (العطف) + ال (للتعريف).
 - اسم مذكر على وزن فَعْلٌ.
 - الجذر (د ر س).
 - الجذع (درس).
 - الفرع (دَرْسٌ)
 - مرفوع وعلامة رفعه الضمة الظاهرة.
 - والدرس: وَالْدَرْسَ = وَ + ال + دَرْسَ.
 - السابقة و (العطف) + ال (للتعريف).
 - اسم مذكر على وزن فَعْلٌ.
 - الجذر (د ر س).
 - الجذع (درس).
 - الفرع (دَرْسَ)
 - منصوب وعلامة نصبه الفتحة الظاهرة.
5. خرج الحرف:

نأخذ على سبيل المثال الحرف "فمنكم" فخرج عملية تحليل الحرف تقدم المعلومات الآتية:

- فمنكم: فَمِنْكُمْ = فَ + مِنْ + كُمْ
 - السابقة (فَ)

○ من: حرف جر مبني على السكون.

○ كم: ضمير متصل بضمير الجر (من)

ويستمر المحلل الصرفي بعرض جميع الاحتمالات الممكنة للكلمة المحللة إلى أن يقدم كل الاحتمالات خاصة إذا كانت الكلمة غير مشكولة، أما إذا كانت الكلمة مشكولة فإنه لا يقدم إلا نتيجة واحدة وفقاً لعلامات الشكل التي تحملها الكلمة المحللة، وقد كانت الخطوات التي سبق ذكرها بمثابة مواصفات شاملة تجمع بين جميع المحللات الصرفية العربية.⁽¹⁾

2. التحليل التركيبي للنصوص العربية:

اهتم النحويون العرب منذ القدم بتراكيب اللغة العربية، فقعدوا لها في مجموعة من المؤلفات معتمدين في جمعها على ما تلفظت به أفواه القبائل العربية والبوادي المجاورة لها خوفاً عليها من الضياع، فتمكنوا أشد تمكن من الإحاطة بكل القواعد المنظمة للكلام العربي، لغرض تسهيل اكتسابها وتيسير تعلمها وتبسيط تعليمها اعتماداً على مناهج بسيطة واصفة، ومع تطور الدراسات اللغوية في القرن الماضي، وهو ما اصطُح عليه باللسانيات فقد عرج بدراسة اللغة دراسة حديثة موظفة مجموعة من النظريات، أهمها النظرية التوليدية المتأثرة بالجانب الرياضي التي قعد لها نعوم تشومسكي في أعماله منذ كتابه الأول المسمى بالبنى التركيبية (Synthetics Structures) الصادر سنة 1957، وتلته مجموعة من الأعمال انتهت بكتابه الأخير المسمى بالبرنامج الأدنى (The Minimalist Programs) الصادر عام 1995.

لقد تناول تشومسكي اللغة ضمن أعماله من وجهة صورية متأثرة بالجانب الرياضي، ومرد ذلك إلى طبيعة انتمائه البحثي لمعهد ماساتشوستس التكنولوجي

(1) مروان البواب، (2009)، "مواصفات نظام التحليل الصرفي في اللغة العربية"، اجتماع خبراء المحللات الصرفية الحاسوبية للغة العربية (26-28/4/2009) - دمشق، بتصرف.

(Massachusetts Institute Of Technology) الشيء الذي سيحدث ثورة لسانية غيرت منحى الدراسات اللغوية إلى ما هو منطقي صوري رياضي، بعدها سيساهم في إحداث مجال معرفي يمزج بين اللسانيات وعلم الحاسوب وهو اللسانيات الحاسوبية التي بسطت نمذجة قواعد اللغة وساعدت في تناول اللغة ومعالجتها معالجة آلية.

لقد سعى مجموعة من اللسانيين والباحثين في لغات مختلفة إلى محاولة التفكير في صورنة المستويات اللسانية من بينها النحو العربي وتسهيل برمجته على مستوى الحاسوب، باعتباره "قنطرة الوصول التي تعبر خلالها مسارات الافتراض المتبادل بين علوم اللغة وعلم الحاسوب، فضلاً عن كونه المسؤول عن توفير المعطيات اللازمة لتحليل اللغوي الأعمق ألا وهو الفهم الآلي للنصوص اللغوية"⁽¹⁾.

إن الفهم الآلي للنصوص اللغوية العربية من منظور تركيبى ليس بالأمر اليسير نظراً لتعدد جوانبه؛ فالوصول إلى معالجة المستوى التركيبى العربي آلياً يفرض المرور بمجموعة من المحطات ليتمكن الحاسوبى من برمجة القواعد الضابطة لهذا المستوى، وتتمثل هذه المحطات في استعراض الشق النظري للنحو العربي اعتماداً على الموروث، وما قُعد له من مفاهيم مؤسسة لمنظومة النحو ككل، ويمكنها أن تتوافق مع أدوات المعالجة الآلية، ثم يأتي بعد ذلك البحث عن اتجاهات النحو الحديثة والتي يمكنها تفسير النحو العربي مع العلاقات(*) التي تحكمه وتُطبعه ليكون قابلاً للبرمجة على مستوى الحاسوب، ثم بعد ذلك يأتي دور الحاسوبى المتمثل في اقتراح المنطقات، وتعيين المتطلبات الأساسية، وتحديد المقاربة أو المقاربات المناسبة، لتمكين الآلة من معالجة النحو العربي آلياً، ثم يعرض الحاسوبى في النهاية وصفاً للتطبيقات العلمية

(1) أحمد روبي محمد عبد الرحمان (2017)، "التحليل النحوي"، مجلة مباحث لغوية، "مدخل إلى اللسانيات الحاسوبية" ع: 30 تحرير: عبد الله بن يحيى الفيقي، الطبعة: الأولى، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض - المملكة العربية السعودية، ص: 77، بتصرف.

التي تدور في فلك المعالجة الآلية للتركيب العربي وفق ما اعتمد عليه النظام من قواعد النحو التي صيغت من قبل.⁽¹⁾

بدأت إرهابسات التحليل التركيبي للغات الطبيعية منذ بداية النصف الثاني من القرن العشرين وفق النظرية الإحصائية التي جاء بها العالم الأمريكي كلود شانون (Claude Shannon) والقائمة على "الاحتمال الاحصائي لوقوع الظاهرة اللغوية، والتي مكنت من قياس كمية المعلومات التي تتضمنها المقولات اللغوية المختلفة على أساس أن كمية المعلومات تتناسب عكسياً مع احتمال وقوع الظاهرة التي تشير إليها، فكلما نُدُر حدوثها زادت كمية معلوماتها والعكس بالعكس"⁽²⁾، وقد كان هذا الصرح العلمي بمثابة بداية للتقدم البحثي في هذا المجال، فجاء بعده ماركوف شين (Markov Chain) العالم الرياضي الروسي الذي درس الظواهر اللغوية المتغيرة زمنياً، كما درس علاقة الارتباط الزمني بين الأحداث باعتماد السلاسل الزمنية (Time Series)، فبواسطة هذه السلاسل تمكن من إقامة نماذج إحصائية للغة في هيئة شبكة كثيفة من علاقات التلازم والترابط والتوارد وما شابهه.⁽³⁾

(1) نبيل علي (1988)، "اللغة العربية والحاسوب (دراسة بحثية)"، تقديم: أسامة الخولي، المجلد الثامن، التعريب، دار الفكر العربي، القاهرة، ص: 353 بتصرف.

(*) أورد نبيل علي مجموعة من العلاقات التي تحكم المقولات النحوية من بينها العلاقات الدلالية والوظيفية والتركيبية التي ستمكن الحاسوبية من برمجة النحو العربية، ويندرج ضمن هذه العلاقات علاقة الرتبة، وعلاقة التضام، والعلاقة الوظيفية النحوية، والعلاقة الوظيفية الدلالية. اللغة العربية والحاسوب لنبيل علي ص: 340

(2) نبيل علي، و نادية حجازي (2005)، "الفجوة الرقمية"، عالم المعرفة، ع: 318، مرجع سابق، ص: 318. بتصرف.

(3) أحمد روبي محمد عبد الرحمان، "التحليل النحوي"، مجلة مباحث لغوية، "مدخل إلى اللسانيات الحاسوبية" ع: 30 تحرير: عبد الله بن يحيى الفيقي، مرجع سابق، ص: 84، بتصرف.

ومنذ ذلك الحين والأبحاث في تطور بشكل سريع، حيث لم يعد العمل مقتصرًا على الجهود الفردية كما بدأ في الأول، بل أخذت على عاتقها مجموعة من المؤسسات الجامعية والمعاهد البحثية تتبنى مثل هذه المشاريع العلمية؛ لما لهما من أهمية في تطوير البحث اللغوي الحديث، فحاولت في البداية إعداد مجموعة من المدونات اللغوية، صدر أولها سنة 1970 واصفة النحو السويدي في جامعة "لند" (Lund) وتوظيفها في نماذج إحصائية كتلك التي جاء بها ماركوف تحت إشراف "أولف تلمان" (Ulf Telman)، وقد ساعد كل ذلك على التفكير في النمذجة للبنوك الشجرية النحوية باعتبارها مواردًا لسانية ضرورية لتجريب مجموعة من التطبيقات الحاسوبية من بينها نظم الترجمة الآلية، المحللات التركيبية، فك اللبس الدلالي للكلمات ... ، و"البنك الشجري" (Treebank) في مجمله هو عبارة عن مدونة لغوية توصيفية، تتضمن بعض التحليلات النحوية التي تتجاوز بيان مستوى الأجزاء الكلامية إلى الكشف عن العلاقات التركيبية بين هذه الأجزاء"⁽¹⁾. لقد شهد البحث في مجال البنك الشجري النحوي تقدماً منذ بدايته في منتصف القرن الماضي إلى يومنا هذا، ونعرض فيما يلي جزءاً من المسار التاريخي لبناء البنك الشجري الخاص بمجموعة من اللغات.

- 1978 وسم مدونة ضمت عنونة النحو الإنجليزي من الباحث السويدي "ألكرد" (Ellegard) بمعية طلبته المهندسين في جامعة "غوتنبورغ" (Göteborgs).

- 1980 وضع منهج متكامل لعنونة المدونات سمي بـ "توسكا" (TOSCA) تحت إشراف فريق بحث بجامعة

(1) أحمد محمد روبي (2017)، "البنك الشجري النحوي"، سلسلة الرسائل الجامعية، عدد: 4، الطبعة: الأولى، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض - المملكة العربية السعودية، ص: 11، بتصرف.

"نلجمجن" (Nijmegen) الهولندية وتسخيرها للاستفادة منها في الدراسات النحوية.

- عنونة مدونة "لانكاستر" (Lancaster) للغة الإنجليزية البريطانية في جامع لانكاستر، لكنه لم يتم إتمامها نظراً لأسباب تقنية.

- من 1980 إلى 1990 بداية استثمار المدونات في بناء تطبيقات حاسوبية للغات الطبيعية مثل المحللات النحوية وأنظمة الترجمة الآلية.

- 1995 قدمت شبكة البيانات اللغوية ببندسلفانيا خدمة جليلة متمثلة

في بناء بنك شجري بقيادة ميتشل ماركوس (Mitchell Marcus) وأن تايلور (Ann Taylor) كان سببا في تطوير العديد من تطبيقات المعالجة الآلية للغة العربية، وقد لقي دعماً كبيراً من طرف مجموعة من الشركات وحماساً فائق النظير من طرف العديد من الباحثين الذين انخرطوا في هذا العمل، فكان نقطة انطلاق للمؤسسات العلمية والتجارية في بناء بنوك نحوية على غرار هذا البنك، رغم وصفه بعدم الدقة والاكتمال والغموض على مستوى التمثيل النحوي للجمل.⁽¹⁾

- 1996 الإعلان عن إطلاق مجموعة من البنوك الشجرية النحوية لمجموعة من اللغات من بينها بنك "بندسلفانيا الشجري" (Penn Treebank) الخاص باللغة الإنجليزية.⁽²⁾

(1) أحمد محمد روبي (2017)، "البنك الشجري النحوي"، سلسلة الرسائل الجامعية، عدد: 4، الطبعة: الأولى، مرجع سابق ص: 83-86 بتصرف.

(2) Ann Taylor, Mitchell Marcus, and Beatrice Santorini, (2003) «The Penn Treebank: An Overview», in *Treebanks*, éd. par Anne Abeillé, vol. 20 (Dordrecht: Springer Netherlands, 5-22, http://www.springerlink.com/index/10.1007/978-94-010-0201-1_1, بتصرف.

- 2002 إصدار بنك بنسلفانيا الشجري للغة الكورية (Korean Treebank) ضم أربعة وخمسين ألف كلمة⁽¹⁾.
- 2003 إصدار بنك بنسلفانيا الشجري للغة الصينية (Penn Chinese Treebank) ضم مائتين وخمسين ألف كلمة⁽²⁾.
- 2004 إصدار أول بنك خاص باللغة العربية مستمد من بنك بنسلفانيا (Arabic Treebank) ضم ما يقارب "542,543" كلمة موسومة⁽³⁾، كما صدر في نفس السنة بنك براغ الخاص باللغة العربية (Prague Arabic Dependency Treebank) من طرف معهد اللغويات التطبيقية والصورية بجامعة كارلوفيا ببراغ⁽⁴⁾.
- 2006 إصدار البنك الشجري للغة التشيكية من طرف معهد براغ (Prague Dependency Treebank).
- 2009 الإعلان عن البنك الشجري الكولومبي الخاص باللغة العربية (The Columbia Arabic Treebank) (Catib) الذي احتضنته وزارة الدفاع الأمريكية.

(1) Chung-hye Han et al.(2001), « Penn Korean Treebank : Development and Evaluation », 69-78, <http://hdl.handle.net/2065/12256>, بتصرف.

(2) Naiwen Xue et al. (juin 2005) « The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus », Natural Language Engineering 11, no 2: 207-38, <https://doi.org/10.1017/S135132490400364X>, بتصرف.

(3) Mohamed Maamouri et al. (1 janvier 2004) « The penn arabic treebank: Building a large-scale annotated arabic corpus », NEMLAR Conference on Arabic Language Resources and Tools, بتصرف.

(4) Jan Hajic et al., (2004), « Prague Arabic dependency treebank: Development in data and tools », in Proc. of the NEMLAR Intern. Conf. on Arabic Language Resources and Tools, 110-117, بتصرف.

- 2009 الإعلان عن البنك الاعتمادي الخاص بالقرآن الكريم (QADT)

(Quranic Arabic Dependency Treebank)) بتمويل من جامعة ليدز.⁽¹⁾

لقد شكلت كل الجهود السابقة وما عقها من أعمال أنجزتها فرق بحث مختلفة بدعم من المؤسسات الجامعية والشركات دفعة قوية، تُمننت بظهور مجموعة من المحللات التركيبية كان الغرض منها تحديد أجزاء الكلام والتعرف على سماته واستنتاج مختلف معانيه باختلاف سياقات وروده، ومن جهة أخرى يمكن القول بأن الغرض من المحللات التركيبية هو استنتاج مدى طواعية اللغة للبرمجة على مستوى الحاسوب، واستنتاج مدى قدرة هذه الأنظمة على "تناظر الأداء الإنساني".⁽²⁾

3. التحليل الدلالي للنصوص العربية:

عُرِّفت الدلالة بتعاريف مختلفة يمكن إجمالها في كونها: "علم يُعنى بدراسة معاني الألفاظ ضمن حقلها الدلالي ودراسة دلالاتها حسب سياق ورودها"، وهو نفس الهدف الذي أقر به الباحثون في مجال المعالجة الآلية للغات الطبيعية، باعتباره مجالاً يعنى باستخلاص معاني الألفاظ استناداً إلى مجموعة من الموارد اللسانية المنظمة وفق إطار لساني قادر على صورنة الدلالة وتسهيل نمذجتها على مستوى الحاسوب، بالإضافة إلى قواعد معطيات معجمية تشمل ألفاظ اللغة مقرونة بمعانيها، وتوظيف مجموعة من الأدوات الحاسوبية اعتماداً على مقارنة من المقاربات المعتمدة في مجال المعالجة الآلية للغات الطبيعية، ولهذه المتطلبات، تعتبر المعالجة الآلية للدلالة من بين الرهانات الكبيرة في مجال المعالجة الآلية للغات الطبيعية.

(1) أحمد محمد روبي (2017)، "البنك الشجري النحوي"، ص: 13، بتصرف.

(2) أحمد روبي محمد عبد الرحمان (2017)، "التحليل النحوي"، مجلة مباحث لغوية، "مدخل إلى

اللسانيات الحاسوبية" ع: 30 تحرير: عبد الله بن يحيى، مرجع سابق، ص: 87، بتصرف.

(*) سيتم الحديث بتفصيل عن المقصود بالمدونة اللغوية في المبحث الخاص بالمدونات العربية

يُحدّد الإطار اللساني لدراسة الدلالة في:

- الترادف (Synonyms).
- الالتباس (Ambiguity).
- التضاد (Antonyms).
- التباين (Monosemy).
- المشترك اللفظي أو الجنس (Homonyms).
- التضمن⁽¹⁾ (Hyponym).
- التعميم (Huponymy).

- دراسة الكلمات المفردة (Individual Words): أو ما عرف

بالدلالة المعجمية للكلمة، ويندرج ضمنها دراسة مجموعة من القضايا
الدلالية من بينها:

- دراسة الكلمة ضمن سياقها (Combined Into The Meaning

.(Of Sentences

لقد اهتم البحث اللساني بنمذجة المستوى الدلالي لما له من أهمية كبرى في
محاولة برمجة أنظمة تستطيع تقديم معلمات دلالية كاستخلاص المعاني المعبر عنها في

(1) Christopher D Manning and Hinrich Schiitze (2000), «Foundations of Statistical Natural Language Processing», Second printing, Cambridge États-Unis: The MIT Press Cambridge, Massachusetts London, England, p: 110. بتصرف.

النصوص المعالجة، فكما أن المستوى التركيبي وفرت له مجموعة من البنوك الشجرية النحوية لما لها من تأثير في تحقيق نتائج جيدة على مستوى التحليل الآلي للتركيب، فإنه من الواجب توفير مدونات خامدة للمستوى الدلالي من أجل تسهيل نمذجته مع ضرورة مراعات ما سبقت الإشارة إليه على مستوى دراسة الكلمة، إما منفردة أو مضمنة في سياق، وهذا ما سيساعد على دراسة النصوص واستخلاص مجموعة من القضايا من بينها:

- التحليل الآلي - وضع الخصائص والسمات للكلمات.
للنصوص.

- فهم النصوص. - الترجمة الآلية.

- تلخيص النصوص. - تحليل النصوص وتقييم الأراء الواردة فيها والمشاعر المعبر عنها.....⁽¹⁾

لقد شكلت مدونة وردنيت (Wordnet) الانطلاقة الفعلية لتناول الجانب الدلالي من منظور حاسوبي والتي أصدرها "كريستيان فيلبوم" (Christiane FELLBAUM) سنة 1998، حيث ضمت عددا كبيرا من الكلمات الإنجليزية في صيغة قابلة للقراءة على مستوى الحاسوب، حيث حاول الباحث جمع كلمات اللغة الإنجليزية اعتمادا على ما اصطلح عليه بالترادف (Synsets)، ونمثل له بكلمة «بارد» التي قد توجد لها مرادفات لها في نفس المجموعة مثل: «شتاء»، و«قارس»، و«تجمد»، كما ترتبط كل مجموعة مترادفة بمجموعة أخرى باستخدام مجموعة من العلاقات الدلالية مثل

(1) عمر مهديوي وسلوى السيد حمادة (2008)، "المعالجة الدلالية الآلية للغة العربية: نحو بناء قاعدة بيانات معجمية للعلاقات الدلالية بين الكلمات"، مجلة فكر ونقد، ص: 42، بتصرف.

علاقة التضمن؛ فمثلا كلمة «رجل» تندرج ضمن مجال أوسع وهو كلمة «إنسان»⁽¹⁾. وقد ضمن أنواعا مختلفة من أقسام الكلام من بينها "الاسم (Nouns)، والفعل (Verbs)، والصفة (Adjectives)، والظروف (Adverbs)⁽²⁾، وقد أعدت هذه المدونة خلال مجموعة من الإصدارات آخرها إصدار هو "وردنت 3.0" (3.0Wordnet) ضم ما يقارب 118,431 مترادفا (Synsets) تشكلت هذه المترادفات من عدد كبير من الدخلات المعجمية بلغت (207995) موزعة بين الأفعال، والأسماء، والصفات، والظروف، نلخصها في الجدول (1) الآتي:

الجدول 1: عدد الدخلات المعجمية لمدونة ووردنت إصدار 3.0

نوع الكلام	الاسم	الفعل	الصفة	الظرف	المجموع
العدد	147360	25051	30004	5580	207995

وقد ترجمت هذه المدونة إلى ما يقارب مائة وخمسين لغة من بينها: "اللغة الفرنسية، والإسبانية، والأسترالية، والفارسية، والعربية،..."⁽³⁾ ونظرا لما توصلت إليه الأبحاث من إنجاز لمجموعة من المحللات كالمحللات الصرفية والتركيبية، فقد أصبح الخوض في مجال المعالجة الآلية للدلالة أمرا متاحا يتمثل في بناء العديد من التطبيقات من بينها:

(1) إشراق علي أحمد الرفاعي (2017)، "التحليل الدلالي"، مجلة مباحث لغوية، "مدخل إلى اللسانيات الحاسوبية" ع: 30 تحرير: عبد الله بن يحيى الفيضي، الطبعة: الأولى، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض - المملكة العربية السعودية ص: 110 بتصرف.

(2) Christiane Fellbaum. « La représentation des verbes dans le réseau sémantique WordNet », Langages 33, no 136 (1999): 27-40. بتصرف <https://doi.org/10.3406/lg.1999.2210>

(3) <http://compling.hss.ntu.edu.sg/omw/summx.html> consulter le 03/07/2019 à 19:17

- تطبيقات فك اللبس الدلالي.
- تطبيقات تحليل الآراء والمشاعر.
- تطبيقات الانطولوجيا.⁽¹⁾

4. المعالجة الآلية للمستوى المعجمي:

لقد كان الهدف من المعاجم الورقية هو الحفاظ على الموروث المعجمي من الضياع، فتضافرت جهود مجموعة من اللغويين القدماء، وثمرت بإنتاج مجموعة من المعاجم المختلفة الترتيب والمتنوعة الموضوعات والمتعددة المصادر، فشملت بذلك ألفاظا عربية وغير عربية، ومع تطور الدراسات اللغوية الحديثة فقد تم تناول المعاجم وفق النظريات الحديثة، ولما أصبح الوردُ البحثي الواحد غير كاف ليُروى به العطش العلمي، فقد أضحي لزاما الاستعانة بمنابع معرفية أخرى، فظلت اللغة معطا أساسيا في محور مجموعة من الدراسات جعل المعجم محركا لها، ففرض التطور التكنولوجي الذي غزا العالم وتوغل في العلم حوسبة المعجم لهدف تخزين المادة وترتيبها طبقا للنظام المطلوب، وتعديل وحذف بعض المعطيات، وتحديد المعجمات بسهولة، إضافة إلى النقل المباشر إلى المطبعة والحصول على أجزاء محدودة من داخل المادة المخزونة⁽²⁾، وتهيئها لدراسة المستويات اللسانية الأخرى في إطار المعالجة الآلية للغة.

وقد كانت البدايات الأولى لحوسبة المعجم في منتصف القرن الماضي، حيث لقي اهتماما كبيرا من لدن فرق البحث وسط مجموعة من المؤسسات الجامعية والمعاهد العليا سعيا في بدايتها إلى "استعمال هذه المعاجم كموارد لغوية للتحليل الدلالي للغات الطبيعية على المستوى الصرفي والنحوي والدلالي، فكانت المعاجم بمثابة

(1) إشراق علي أحمد الرفاعي (2017)، "التحليل الدلالي"، مجلة مباحث لغوية، "مدخل إلى اللسانيات الحاسوبية" ع:30، ص: 112-121، بتصرف.

(2) جيلالي بن يشو (2019)، "حوسبة المعجم العربي الواقع والآفاق"، شبكة الصورة العربي،

<https://www.voiceofarabic.net/ar/articles/2662> consulter le 03/07/2019 a 23:43.

قواعد معطيات معجمية واصطلاحية تحتوي على معلومات في مختلف المستويات اللسانية ... لمواكبة تطور تطبيقات المعالجة الآلية للغات الطبيعية ... لتقديم خدمات في البحث عن معاني المفردة أو العبارة المراد البحث عنها"⁽¹⁾، ولم يتأت ذلك إلا بفضل مجموعة من الأبحاث المنجزة ضمن مجال اللسانيات الحاسوبية من أجل، "إعادة البنية المعجمية، في مختلف مستويات المعجم وارتباطه، الذي يعتبر الحل الأمثل لتبرئ قواعد معطيات المعالجة الآلية للغة العربية، ونحن لا نختلف معها في ذلك، لأن صناعة المعجم، أو بالأحرى معاجم للمستويات اللسانية بإمكانها تيسير سبل المعالجة الآلية للغة العربية، ويعتبر المعجم المدخل الأساس لولوج الحاسوب عموماً"⁽²⁾.

وقد دعت الحاجة الماسة للمعاجم الرقمية الباحثين إلى توفير مجموعة من المعاجم الآلية التي تهتم بمفردات اللغة في مستواها الإفرادي والتركيبى؛ فأما على المستوى الإفرادى للكلمة، فيتطلب من المعجمي الحاسوبي أن يُضمّن في المعجم الكلمات مرفوقة بوصف صوري لكل المفردات اللغوية بدءاً من كيفية كتابة كل لفظة مروراً بكيفية النطق بها، ثم إرفاقها بالمعلومات الصوتية والصرفية والنحوية ومحاولة جمع المعلومات الدلالية لكل دخلة معجمية.

(1) عبد الكريم بوزويغ وعبد الفتاح حمدني (2016)، "المعاجم الرقمية العربية: من المصطلح إلى الدلالة"، مجلة اللسان العربي، ع: 76، المنظمة العربية للتربية والثقافة والعلوم - مكتب تنسيق العريب، ص: 79-93، بتصرف.

(2) عبد الواحد دكيكي (2015)، "التخطيط المعجمي للجملة في العربية، السمات والتوزيع في التعابير المسكوكة نموذجاً" ضمن، البحث المعجمي ورهانات تأهيل اللغة العربية قضايا في النظرية والتخطيط والهندسة، تحرير: عبد الواحد دكيكي وآخرون، طبعة جديدة ومنقحة، وقائع الندوة العلمية الدولية الأولى للسانيات الحاسوبية والمعجم (الكلية المتعددة التخصصات بالرشيدية - جامعة مولاي إسماعيل المغرب: فريق البحث هندسة اللغات الطبيعية وتكنولوجيا الحوسبة GELINFO، ص: 388، بتصرف.

أما جمع المفردات باعتبار المستوى التركيبى، فيتضمن كل المعطيات التى يمكن أن تكتشف المفردة من خلالها، وما يقتضيه المستوى التركيبى الذى يحدد الوظيفة اللغوية لكل مفردة حسب سياق ورودها.⁽¹⁾

لقد أصبح المعجم فى ظل الحوسبة قاعدة بيانات تضم ألفاظ اللغة مرتبة فى أكثر من جدول، يشتمل كل جدول على مجموعة من الحقول، يشكل كل حقل المعلومات الصرفية، والصوتية، والتركيبية والدلالية لىتم الرجوع إليها عند كل عملية تحليلية آلية طبقت على النصوص، وهو ما أصبحنا نراه فى العديد من التطبيقات الحاسوبية ذات الاستعمال اليومى، وسيأتى الحديث عن بعض هذه التطبيقات فى المحور الذى سنتحدث فيه عما أنجز من تطبيقات حاسوبية.

ومن بين الاهتمامات كذلك التى أخذها الباحثون الحاسوبيون على عاتقهم هي حوسبة الجمل العربية على مستوى الحاسوب، كما أشار إلى ذلك الدكتور عبد الواحد دكيكي؛ إذ دعا إلى توفير معجم يضم نماذج مختلفة للجمل العربية، الشىء الذى سيمكن لا محال نظم "المعالجة الآلية من التعرف على مختلف البنيات الجمالية العربية، ورصد مختلف التفاعلات التضامية التى تنتجها مكونات الجملة تواردياً فى الماتريس، تفاعلات تحصل عادة بين المستويات اللسانية؛ النحوية، والصرفية، والمعجمية والدلالية والسياقية فى إنتاج جمل سليمة التكوين مبنى ومعنى".⁽²⁾

ولتحقيق ذلك كان لزاماً الانطلاق من عتاد لسانى تضبط به طرق جمع المعلومات، واختيار المداخل وترتيبها للوصول إلى عملية التحرير. أما العتاد الحاسوبى

(1) أمين قدرأوى (2009)، "نحو بناء معجم إلكترونى للمعالجة الآلية للغة العربية مفردات ألفاظ القرآن الكريم أنموذجاً مقارنة لسانية حاسوبية"، بحث لنيل درجة الماجستير، كلية الآداب واللغات - قسم اللغة العربية وآدابها، جامعة أبى بكر بلقايد - تلمسان، ص: 203 بتصرف.

(2) عبد الواحد دكيكي (2015)، "التخطيط المعجمى للجمل فى العربية، السمات والتوزيع فى التعابير المسكوكة نموذجاً" ص: 392.

فيتطلب إعداء قاعدة بيانات نصية إلكترونية منظمة طبقاً لقواعد معينة، ثم تنظيم المعطيات على شكل علاقات، وجعل كل ما سبق ذكره قابلاً للتحديث.⁽¹⁾

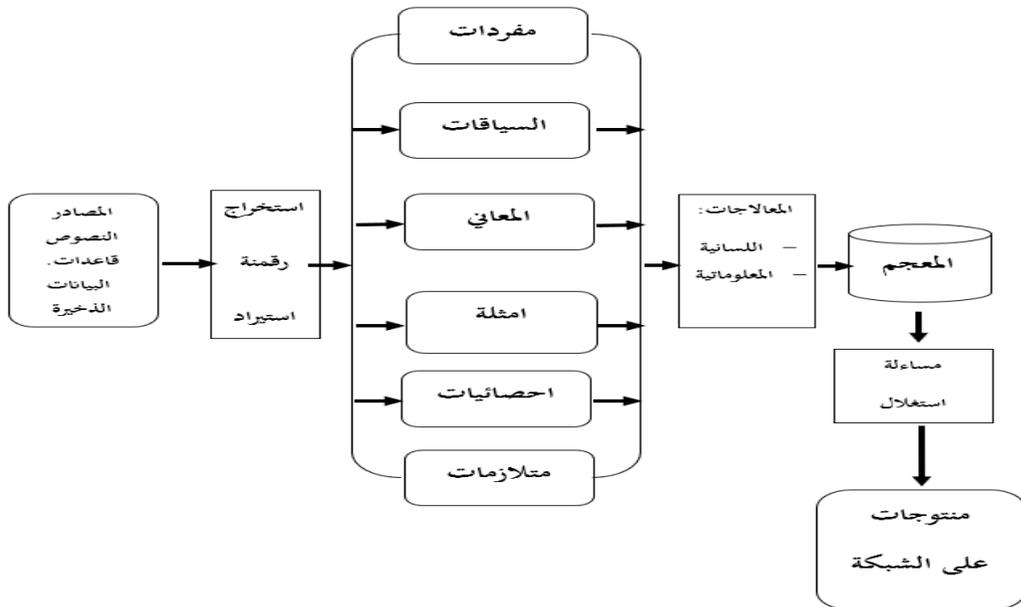
ويظهر دور اللسانيات من خلال قيام اللسانيين بإعطاء صورة تامة لكل كلمة، وقد سعى البعض هذا المجال باللسانيات الإحصائية، باعتباره مجالاً يقدم " نماذج نظرية وتطبيقية لصناعة المعاجم الإلكترونية "⁽²⁾. ومن أمثلة المعاجم التي أصبحت إلكترونية بالكامل نجد معجم لاروس (LA ROUSSE) الفرنسي، رغم صدوره ورقياً كل سنة، إلا أنه انسجم مع التقدم الإلكتروني ليصبح معجماً إلكترونياً بفضل جهود أبناء اللغة، ويعد هذا المعجم من المعاجم الموسوعية، حيث صمم بطريقة ممنهجة وفق الكيفية التي تُصمم بها الموسوعات " لاستيعاب كل ما يمكن جمعه من مواد معجمية، فقد نسج له فضاء من أحد عشر باباً يظهر مرسوماً على الشاشة مع بداية التشغيل، كما يظهر لك أحد عشر باباً. والباب الرئيسي منها جميعاً هو الباب المسى " قاموس " (Dictionary). وهو محور إلكتروني يمكن عن طريق المدخل " ابحث " من الحصول على تعريف كاف لكل كلمة يراد معرفة معناها والوقوف على مختلف وجوه استعمالها بصورة تراتبية قوامها الترتيب الألف بائي."⁽³⁾ فالملحظ أن هذا القاموس قد وُضع بمنهج علمي متميز لغرض تعليمي بالدرجة الأولى، ومن بين القواميس التي صممت على نفس الطريقة: قاموس أكسفورد (Oxford South African School)، الذي يشرح الكلمة

(1) فاتن الخولي (2014)، " المعجم بين اللغة والحاسوب بناء المعجم الإسلامي من منظور لساني حاسوبي"، ضمن المعجمية العربية قضايا وآفاق، تقديم حافظ إسماعيلي وآخرون، الطبعة الأولى، سلسلة المعرفة اللسانية، ج: الثاني، دار كنوز المعرفة، الأردن، عمان، ص: 177-98، بتصرف.

(2) بلقاسم اليوبي (1999)، "اللسانيات الحاسوبية: مفهوماً وتطورها و مجال تطبيقاتها"، مجلة مكناسة، عدد 12 (مكناس: كلية الآداب والعلوم الإنسانية).

(3) محمد شندول (2009)، "القاموس الإلكتروني: قاموس « Le petit Larouss Multimédia » الإلكتروني نموذجاً"، الدراسات المعجمية، العدد السابع والثامن، منشورات الجمعية المغربية للدراسات المعجمية، الرباط - المغرب، ص: 122.

الإجليزية ويقدم الاستعمالات الممكنة للكلمة حسب السياقات المختلفة التي وردت فيها، وقد وضع رهن إشارة المتعلم بالطريقتين، الورقية، والإلكترونية. وفي إطار حوسبة المعجم العربي، ظهرت مجموعة من المشاريع العلمية، والأفكار البحثية العربية المختلفة التي تسعى إلى حوسبة معجم عربي حاسوبي في مجموعة من المجالات يشتمل على كل الخصائص التي تميز الكلمة العربية، ومن هذه الأعمال فكرة الأستاذ عبد الغني أبو العزم مع السعدية آيت الطالب وفوزية بن جلون، في إطار "المعجم الحاسوبي العربي: التصور والمنهجية" حيث قدموا مراحل إعداد مشروع المعجم العربي الحاسوبي تلخص في الشكل (4) الآتي:



الشكل 4: مراحل إعداد المعجم العربي الحاسوبي (1)

(1) عبد الغني أبو العزم، السعدية آيت الطالب، وفوزية بن جلون (2009)، "المعجم الحاسوبي العربي التصور، والمنهجية"، الدراسات المعجمية، العدد السابع والثامن، منشورات الجمعية المغربية للدراسات المعجمية، الرباط - المغرب، ص: 14.

من خلال الشكل (4) أعلاه يتبين أن حوسبة المعجم ليس بالأمر السهل، بل يتطلب فريق عمل كبير ذي تخصصات متنوعة، كما تحتاج حوسبة المعجم جهداً مضاعفاً من أجل النهوض بهذا العمل، والرقى باللغة العربية وإعطائها المكانة التي تستحقها، بين اللغات الأخرى.

خلاصة:

أصبحت المعالجة الآلية للغات الطبيعية بصفة عامة واللغة العربية على وجه الخصوص، من أهم الاهتمامات البحثية التي كشفت عن نتائج تجاربها إلى تيسير الإمساك بالنصوص الضخمة في ظل الانفجار المعلوماتي الذي نعيشه اليوم، مما استوجب على مجموعة من الباحثين تبني المنهج العابر للتخصصات، فساهم الانطلاق من اللسانيات وتلاقحها مع علم الحاسوب في المسك بزمام المستويات اللسانية، والتمكن من نمذجتها، وهذا ما مكن مجموعة من الفرق البحثية من إنتاج مجموعة من النظم الحاسوبية.

إن المعالجة الآلية للغة العربية وغيرها من اللغات الطبيعية، لم تتوقف في الصورنة والتبسيط من أجل إدخال اللغة العربية في الحاسوب، بل ستنقل إلى وضع مجموعة من النظم التي تعكس الفعل البشري في تعامله مع اللغة في مختلف مستوياتها الصرفية، والمعجمية، والتركيبية، والدلالية، وصولاً إلى الأداءات الإنسانية للغة، وسنسى في المبحث الموالي إلى التعريف ببعض النظم الحاسوبية العربية التي تمت بلورتها بعد تحقيق نسب دقة عالية.

المبحث الثالث: بعض النظم الحاسوبية العربية

مدخل:

تمكن الباحثون في مجال اللسانيات الحاسوبية والمعالجة الآلية للغات الطبيعية من إنتاج مجموعة من النظم والتطبيقات المختلفة من بينها: نظم التحليل الصرفي والتحليل التركيبي وكذا التحليل الدلالي، وصولاً إلى أحدث النظم التي تمكن من تحليل الكلام قصد استخلاص مضمونه، وتقييم الآراء الواردة فيه، واكتشاف طبيعة المشاعر والعواطف المستترة ضمنه.

لقد تمكنت اللغة العربية بفضل جهود المؤسسات والمعاهد البحثية من اللحاق بركب اللغات الطبيعية الأخرى رغم تعقيد خصائصها التركيبية والغنى المفرداتي الذي يؤثر في تحقيق نسبة دقة أداء جيدة بعد المراحل التدريبية، وفيما يلي تقديم لأهم وأبرز التطبيقات التي بنيت من أجل خدمة اللغة العربية في مختلف مستوياتها الصرفية، والتركيبية، والدلالية، وغيرها من المستويات.

1. المحللات الصرفية:

1.1. المحلل الصرفي للغة العربية:

عمل فريق المعالجة الآلية للغة العربية التابع لمعهد الدراسات والأبحاث

للتعريب بالمغرب على بناء محلل صرفي للغة العربية مروراً بمرحلتين:

المرحلة الأولى: الاهتمام بالتحليل الصرفي للأسماء العربية.

المرحلة الثانية: الاهتمام بالتحليل الصرفي للأفعال العربية.

وقد تبين أن الفريق سعى - خلال المرحلتين - إلى بناء برنامج يتميز بالدقة موثوق بها في تحليل الأسماء والأفعال في اللغة العربية تحليلاً دقيقاً، حيث اعتمد الفريق على مجموعة من الموارد اللسانية الخاصة بالأفعال والأسماء العربية؛ وقد ساعدت هذه الموارد أساساً في تحديد:

- طبقات الاسم في ثلاث طبقات.⁽¹⁾

- طبقات الفعل في مائتين وستة وثلاثين طبقة.⁽²⁾

- القواعد المعجمية، والصرفية والتركيبية والصوتية الخاصة بالأسماء والأفعال.⁽³⁾

- أنواع الأسماء العربية التي تشمل: "اسم الجنس، واسم العلم، واسم الآلة، واسم المكان، واسم الزمان، واسم الحدث، واسم الفاعل، واسم المفعول، والتصغير، وصيغة المبالغة.

- الأبواب التصريفية العربية المجردة والمزيدة.⁽⁴⁾

○ أوزان الثلاثي المجرد:

(1) فريق المعالجة الآلية للغة العربية (2007)، "التحليل الصرفي للأسماء العربية"، ضمن وقائع الندوة الدولية الثانية للمعالجة الآلية للغة العربية إشراف: عبد الفتاح حمداني، معهد الدراسات والأبحاث للتعريب، الرباط - المغرب، ص: 51-63، بتصرف.

(2) عبد الحميد الجهاد وآخرون، (2009)، "محلل صرفي للأفعال العربية"، ضمن وقائع الندوة الدولية الثالثة حول المعالجة الآلية للغة العربية، المدرسة المحمدية للمهندسين ومعهد الدراسات والأبحاث للتعريب، الرباط - المغرب، ص: 1-3، بتصرف.

(3) فريق المعالجة الآلية للغة العربية (2007)، "التحليل الصرفي للأسماء العربية"، مرجع سابق، بتصرف.

(4) نفس المرجع.

1. فَعَلٌ - يَفْعَلُ.

3. فَعَلٌ - يَفْعَلُ.

5. فَعِلٌ - يَفْعِلُ.

2. فَعَلٌ - يَفْعَلُ.

4. فَعِلٌ - يَفْعِلُ.

6. فَعُلٌ - يَفْعُلُ.

○ أوزان الثلاثي المزيد بحرف:

7. فَعَلٌ - يُفَعِّلُ 8. فَاعَلٌ - يُفَاعِلُ 9. أَفَعَلٌ - يُفَعِّلُ

○ أوزان الثلاثي المزيد بحرفين:

10. تَفَعَّلٌ - تَفَاعَلٌ - 11. تَفَاعَلٌ - 12. انْفَعَلٌ - 13. اِفْتَعَلٌ - 14. اَفْعَلٌ

يَتَفَعَّلُ. يَتَفَاعَلُ. يَنْفَعِلُ. يَفْتَعِلُ. لٌ - يَفْعَلُ

○ أوزان الثلاثي المزيد بثلاثة أحرف

16. اَفْعُوْعَلْ - يَفْعُوْعِلُ.

15. اسْتَفْعَلْ - يَسْتَفْعِلُ.

18. اَفْعُوْلْ - يَفْعُوْلُ.

17. اَفْعَالٌ - يَفْعَالُ.

- أوزان الرباعي المجرد.
19. فَعَلَلَّ - يُفَعَّلِلَ
- أوزان الرباعي المزيد بحرف.
20. تَفَعَّلَلَّ - يَتَفَعَّلَلُّ
- أوزان الرباعي المزيد بحرفين.
21. أَفَعَّلَلَّ - يَفَعَّلَلُّ
22. أَفَعَّلَلَّ - يَفَعَّلَلُّ⁽¹⁾

- جرد السوابق واللواحق الخاصة بالاسم، والأخرى الخاصة

بالفعل.

أما بالنسبة لطريقة التحليل في هذا البرنامج فإنه يعمل على إزالة علامات الشكل إذا كان النص مشكولاً شكلاً تاماً أو جزئياً، ثم يقوم بتجزئ النص إلى وحدات معجمية باعتماد مجموعة من المؤشرات من بينها: الفراغ، وعلامات الترقيم الموجودة في النص، ثم جرد كل الوحدات المعجمية، وبعدها ينتقل إلى عملية التحليل فيقدم تحليلاً

- السابقة
 - الجذع
 - اللاحقة⁽²⁾.
- سطحياً لكل من الاسم والفعل متمثلاً في تفكيكهما إلى ثلاثة أقسام تتمثل في:
لقد تمكن الفريق بعد إجراء مجموعة اختبارات من الحصول على نتائج جيدة يمكن أن تغني المستعمل لهذا البرنامج من الاعتماد الكلي على القواعد الصرفية التي يصعب أحياناً التعامل معها آلياً، والتي قد تؤدي إلى نتائج غير مرغوب فيها⁽³⁾.

(1) عبد الحميد الجمهاد وآخرون، (2009)، "محلل صرفي للأفعال العربية"، مرجع سابق.

(2) عبد الحميد الجمهاد وآخرون، (2007)، "التحليل الصرفي للأسماء العربية"

(3) عبد الحميد الجمهاد وآخرون، (2009)، "محلل صرفي للأفعال العربية"، بتصرف.

2.1. محلل الخليل:

برنامج الخليل هو نظام التحليل الصرفي لنصوص اللغة العربية بمعزل عن السياق، وقد صدر هذا المحلل في نسختين: كانت الأولى سنة 2011 في إطار شراكة بين جامعة محمد الأول، والمنظمة العربية للتربية والثقافة والعلوم، ومدينة الملك عبد العزيز للعلوم والتقنيات، ومن خلال مجموعة من الملاحظات التي أُبديت حول هذه النسخة عمل فريق المعالجة الآلية للغة العربية التابع لجامعة محمد الأول على الأخذ بتلك الملاحظات لتطوير النظام، فتم إصدار النسخة الثانية منه سنة 2016، وهي نسخة مفتوحة المصدر يمكن تجريبها مباشرة على الموقع أو يمكن تحميلها منه.⁽¹⁾

يقدم هذا النظام جميع الاحتمالات الممكنة للكلمة الواحدة مسطرة في جدول يضم مجموعة من الخصائص الصرفية لكل الاحتمالات المقدمة، والصورة الموالية توضح نموذجاً لتحليل كلمة "كتب" باعتماد هذا النظام (الصورة 1):

(1) <http://oujda-nlp-team.net/ar/programms-ar/alkhalil-morphology-2-ar/>

نتائج التحليل
Analysis Results

إحصائيات

الكلمات المدخلة المعالجة: | ktb | كُتب | الكلمات المدخلة غير المعالجة:

الكلمة المشكوك	السايق	الذاع	نوع الكلمة	وزن الكلمة	الفرع	وزن الفرع	الجزر	المدلة الإعرابية	التحق
كُتِبَ	#	كُتِبَ	اسم جامد	فَعْلٌ	كِتَاب	فَعَال	كتب	جمع مذكر مجرور في حالة الإضافة	#
كُتِبَ	#	كُتِبَ	اسم جامد	فَعْلٌ	كِتَاب	فَعَال	كتب	جمع مذكر مجرور نكرة	#
كُتِبَتْ	ك: حرف التشبيه	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر مدفوع في حالة الإضافة	#
كُتِبَتْ	#	كُتِبَتْ	مصدر أصلي	فَعْلٌ	كُتِبَ	فَعْل	كتب	مفرد مذكر مدفوع في حالة الإضافة	#
كُتِبَتْ	ك: حرف التشبيه	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر مدفوع نكرة	#
كُتِبَتْ	#	كُتِبَتْ	مصدر أصلي	فَعْلٌ	كُتِبَ	فَعْل	كتب	مفرد مذكر مدفوع نكرة	#
كُتِبَتْ	ك: حرف التشبيه	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر منصوب في حالة الإضافة	#
كُتِبَتْ	#	كُتِبَتْ	مصدر أصلي	فَعْلٌ	كُتِبَ	فَعْل	كتب	مفرد مذكر منصوب في حالة الإضافة	#
كُتِبَتْ	ك: حرف التشبيه	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر مجرور في حالة الإضافة	#
كُتِبَتْ	ك: حرف الجر	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر مجرور في حالة الإضافة	#
كُتِبَتْ	#	كُتِبَتْ	مصدر أصلي	فَعْلٌ	كُتِبَ	فَعْل	كتب	مفرد مذكر مجرور في حالة الإضافة	#
كُتِبَتْ	ك: حرف التشبيه	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر مجرور نكرة	#
كُتِبَتْ	ك: حرف الجر	نَبَتْ	مصدر أصلي	فَعْلٌ	نَبَتْ	فَعْل	تنب	مفرد مذكر مجرور نكرة	#
كُتِبَتْ	#	كُتِبَتْ	مصدر أصلي	فَعْلٌ	كُتِبَ	فَعْل	كتب	مفرد مذكر مجرور نكرة	#
كُتِبَتْ	ك: حرف التشبيه	نَبَتْ	فعل ماض مبني للمعلوم	فَعْلٌ	نَبَتْ	فَعْل	تنب	ثلاثي مجرد مسند إلى الغائب هو: لازم	#

© 2015 جميع الحقوق محفوظة

الصورة 1: نموذج لتحليل جملة بواسطة نظام الخليل

يبدأ النظام في خرجه باعتبار الكلمة اسما ويسرد كل الاحتمالات الممكنة لها، ثم يعقبه في كثير من الأحيان بالمصدر مع إعطاء مجموعة من الاحتمالات التي يمكن أن يأتي عليها المصدر، ويأتي الفعل في الأخير فيقدم الاحتمالات الممكنة له. وقد أوضحت مجموعة من الدراسات مدى تطور النسخة الثانية بالمقارنة مع النسخة الأولى، ومدى قدرتها على التعرف على الفئات النحوية،

2. المصرف الآلي:

لقد تم استثمار مجموعة من الموارد اللسانية في بناء العديد من النظم والتطبيقات من بينها المصرفات الآلية للأفعال العربية التي جاءت نتيجة الاهتمام

المتزايد من لدن الباحثين في مجال المعالجة الآلية للغات الطبيعية حيث ساهموا في بناء مجموعة من المصرفات، من بينها: نظام المصرف الآلى للأفعال العربية.

أنجز الدكتور الباحث عبد الحميد الجيهاد نظام المصرف الآلى للأفعال العربية ضمن مشاريع معهد الدراسات والأبحاث للتعريب بالمغرب، اعتمد في بنائه على مجموعة من الموارد اللسانية ضمت أزيد من 24175 فعل عربي، ومن 341 طبقة تصريفية، بالإضافة إلى القواعد الصرفية والصوتية العربيةتين.

تمكن الباحث عبد الحميد الجيهاد من صورنة الصرف العربي ثم برمجته ليقدم لنا نظاما مصرفا للأفعال العربية، عن طريق حصر كل أنواع الأفعال العربية بدءا من الفعل الثلاثي المجرد الصحيح بأوابه الصرفية، وأنواعه التي تضم السالم والمهموز والمضعف، وكذا المعتل الذي ينقسم إلى المثال الناقص والأجوف واللفيفين المقرون والمفروق، كما اعتمد في بناء النظام على أنواع الأفعال العربية من حيث عدد حروفها المتضمنة للثلاثي المجرد والرباعي المجرد، وجرى كل أوزان الأفعال الثلاثية والرباعية المزيدتين، ولم يقف هذا النظام في صورنته للقواعد عند الجانب الصرفي فحسب، بل أضاف أغلب القواعد الصوتية التي تؤثر في الفعل العربي وتغير بنيته بتغير تصريفه في الأزمنة العربية الثلاث خاصة الأفعال المعتلة والمضعفة والمهموزة، كما وظف الباحث قواعد كتابة الهمزة في الفعل باختلاف مؤضعها - معتبرا في ذلك الهمزة في بداية الفعل وفي وسط الفعل وكذا في آخره - لتجنب اللحن في تصريف الفعل.

ولا يقف هذا المصرف عند تصريف الفعل في الأزمنة الثلاث فقط، بل يقدم كذلك تصريفات الفعل مقترنا بضمائر المتكلم والمخاطب والغائب بأنواعها المفردة والثنية والجمع، فكان من اللازم ضبط القواعد الحاكمة في إسناد الضمائر للأفعال العربية ويتعلق الأمر بالقواعد الصوتية المتمثلة في: الإدغام، والإعلال، والإبدال.

أما بالنسبة للطريقة التي يصرف بها الفعل فإنه ينبغي وضع الجذع مشكولاً عوضاً عن الجذر الذي تعتمدة مجموعة من المصرفات الأخرى كمصرف " ACON ARABIC"، كما يقدم نظام تصريف الأفعال في اللغة العربية الأبواب الصرفية للفعل الذي تم وضعه في مكان الإدخال، ثم خصائص الفعل المراد تصريفه، بالإضافة إلى تصريفاته في جميع الأزمنة العربية المتمثلة في:

- الماضي المبني للمعلوم.
- المضارع المبني للمجهول.
- المضارع المبني للمجهول في حالة النصب.
- الماضي المبني للمجهول.
- المضارع المبني للمعلوم في حالة النصب.
- المضارع المبني للمجهول في حالة الجزم.
- المضارع المبني للمعلوم في حالة الجزم.
- الأمر⁽¹⁾

1) (Abdelhamid El Jihad, (2013) « نظام تصريف الأفعال في اللغة العربية »، Revue de l'Information Scientifique et Technique 20، no 2: 27-48.

ونمثل ذلك بالفعل "رَجَحَ": الواضحة في الصورة 2:

نظام تصريف الأفعال في اللغة العربية

أدخل الفعل المراد تصريفه: رَجَحَ البحث... رَجَحَ رَجَحَ رَجَحَ

الماضي للمعلوم الماضى المجهول المصارع للزمن المعلوم المصارع للمصوب المعلوم المصارع للزمن المجهول المصارع للمصوب المجهول

الأمر

مفرد	مثنى	مخاطبة	مخاطبة	مكلم	نائب	ناتبة
رَجَحَ	رَجَحَا	رَجَحِي	رَجَحِي	رَجَحْ	رَجَحْ	رَجَحْ
رَجَحْتَ	رَجَحْتُمَا	رَجَحِي	رَجَحِي	رَجَحِي	رَجَحِي	رَجَحِي
رَجَحَ	رَجَحَا	رَجَحِي	رَجَحِي	رَجَحْ	رَجَحْ	رَجَحْ

الجذر	ر ج ح
الصفة	تلاتي مجرد
نوع الفعل المجرد	سدا لم
الوزن	فعل - تَعَلُّم
الصدبة والنووم	لازم أو متحد

طباعة خروج

الصورة 2: مثال لتصريف الفعل "رَجَحَ" في نظام تصريف الأفعال في اللغة العربية

3. المحلل التركيبي:

يعتبر التحليل التركيبي من أهم تحديات المعالجة الآلية للغات الطبيعية ومرد ذلك إلى الخصوصية التركيبية لكل لغة خاصة اللغة العربية لما لها من تشعب الأبعاد وتعدد الجوانب نتيجة للخصوصية التركيبية التي تعرف بها اللغة العربية، بالإضافة إلى المنظومة النحوية التي تتعدد زواياها بتعدد النظريات النحوية القديمة والحديثة، ومما لا زال يعيق تقدم تطوير المحللات التركيبية الخاصة باللغة العربية إلى حد الآن هو الارتباط الحاصل بين المستوى التركيبي والمستويين الصرفي والمعجمي.

تتجلى وظيفة المحللات التركيبية في تحديد بنية الجملة من حيث هيكلية مكوناتها وتحديد وظائف عناصرها، وإيجاد قالب نحوي لها اعتماداً على ما تضمنته اللغة من قواعد نحوية أساسية من حيث تقسيم الجملة وتحديد مكوناتها وتقسيم كلماتها لإيجاد العلاقات النحوية فيما بينها، وقد أصبح المحلل التركيبي الآلي أداة أساسية لا غنى عنها في مجموعة من التطبيقات اللغوية الحالية من قبيل التطبيقات الخاصة بتحليل مضمون النصوص وفهمها آلياً، بالإضافة إلى التلخيص والفهرسة الآلية، ونظم الترجمة الآلية ونظم تحليل الآراء والمشاعر إلى غيرها من التطبيقات التي تهتم بمعالجة النصوص، وبالنظر إلى الجهود العربية فهي لا زالت تعاني من نقص كبير في توفير محللات تركيبية آلية نظراً لقلّة الأدوات اللازمة لبنائها. وسنعرض في ما يلي محلاً تركيبياً يُعتمد في تحليل النصوص العربية وهو مُحلل ستانفورد (Stanford Parser)

يعتبر محلل "ستانفورد بارسر" من أشهر المحللات التركيبية التي وضعت لتحليل اللغة الإنجليزية، وقد أثبت هذا البرنامج فعاليته بدرجات متفاوتة في تحليل لغات أخرى من بينها اللغة الإيطالية والبرتغالية والبلغارية والعربية. يعتمد محلل ستانفورد بارسر كغيره من المحللات النحوية على المقاربة الإحصائية في معالجة اللغة الطبيعية، فقد صمّمته مجموعة ستانفورد لمعالجة اللغة الطبيعية، واستخدمت فيه مجموعة من البيانات المكتوبة لعدة لغات كالإنجليزية، والألمانية، والعربية والصينية، وقد تم تطويره وصيانته منذ سنة 2002 من طرف Dan Klein و Christopher Manning⁽¹⁾ كما يعد من المحللات المفتوحة المصدر لتمكين الباحثين من إضافة

(1) محسن رشوان (2020)، "مدخل إلى حوسبة اللغة"، مجلة مباحث لغوية، "مقدمة في حوسبة اللغة العربية" ع: 55 تحرير: محسن رشوان والسعيد المعتز بالله، الطبعة: الأولى، مركز الملك

مقتضيات بنود رخصة جنو العمومية (GNU GPL). ويمكن الحصول عليه بزيارة موقع جامعة ستانفورد أو ⁽¹⁾ Stanford NLP Group ويتم تثبيته اعتماداً على (1.5 JAVA).

تستخدم في هذا المحلل مجموعة من البيانات كمدخلات؛ فبالنسبة للغة العربية يتم اعتماد بنك المشجرات العربي (ATB) في نسخه الثلاثة من أجل توسيم أجزاء الكلام اعتماداً على PCFG/ Probabilistic context-Free Grammar ، كما يتضمن مخرجات متعددة بناء على نوع التحليل الذي يرغب فيه المستخدم.

بلغت نسبة أداء محلل ستانفورد حوالي 83.5 بالمائة، إلا أنه رغم ذلك لا يتمكن من تمثيل العديد من أبنية الجمل في اللغة العربية، الأمر الذي يفتح مجال البحث عن وسائل وأساليب تقنية وعلمية مستقاة من واقع الدراسة اللسانية للغة العربية تهدف إلى الرفع من أدائه.

4. المدقق الإملائي:

المدقق الإملائي هو "آلية برمجية تتعامل مع الكلمة الموجودة في النص لغرض مراجعتها إملائياً، إذ يضع المدقق الإملائي سطراً أحمر على الكلمات الخاطئة ويقترح بدائل صحيحة للكلمة الخاطئة اعتماداً على ما هو مخزن في المعجم المعتمد"⁽²⁾، وقد

عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض - المملكة العربية السعودية، ص: 57، بتصرف.

(1) <https://nlp.stanford.edu/>

(2) جمعة عمرو (2016)، "المدقق الإملائي والنحوية-محركات البحث-آليات التعرف الآلي على الحروف - آليات الإحصاء اللغوية" ضمن "تقنيات اللغة العربية الحاسوبية معايير التقييم ورؤى

الفهر الآلي للغة العربية تحليل الأراء والمشاعر نموذجاً

كان إصدار أول مدقق إملائي سنة 1992 من طرف شركة كوليتك "Colect" استعمل مع نظام وورد "Word" لمعالجة النصوص بعدما اقتنته شركة مايكروسفت "Microsoft"،

وسرعان ما تضايف الإقبال على نظم التدقيق الإملائي نظراً للحاجة الماسة لها وتزايد استعمال الأجهزة الإلكترونية كالحواسيب والهواتف الذكية، إذ أصبحت نظم التدقيق الإملائي تُعتمد في العديد من مواقع البحث، ومواقع التواصل الاجتماعي وتطبيقاتها.

5. المدقق النحوي:

يختلف المدقق النحوي عن المدقق الإملائي في أن هذا الأخير يُعتبر آلية لكشف الخطأ على مستوى الكلمة فقط بينما يعتبر المدقق النحوي "آلية برمجية تتعامل مع الكلمة داخل التركيب الموجود في النص لغرض مراجعتها تركيبياً وإعرابياً، إذ تضع هذه الآلية سطراً أخضر تحت العبارة الخاطئة فيطلب من الكاتب أن يعيد صياغة العبارة، أو يقترح بدائل صحيحة لأجل الوصول إلى الصواب النحوي أو التركيبي"⁽¹⁾. وقد كان أول استعمال للمدقق النحوي في نظام وورد "Word" لمعالجة النصوص سنة 2000 والذي اشترته شركة الميكروسفت من شركة كولكت "COLECT".

6. المشكل الآلي:

التطوير"، دراسات: 4 ، الأولى (مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، المملكة العربية السعودية، ص:9.

(1) جمعة عمرو (2016)، "المدققات الإمائية والنحوية-محركات البحث-آليات التعرف الآلي على الحروف - آليات الإحصاء اللغوية"، مرجع سابق، ص:9.

يعتبر تشكيل النصوص في اللغة العربية أداة يُعتمد عليها في القراءة السليمة للكلمات، لأن "سبيل كل حرف أن يُوقى ما يستحقه من الحركة والسكون، والتشديد، وغير ذلك... كما جاء التشكيل لرد الخوف من الخطأ في الكلمة العربية، أو تصحيفها، أو تحريفها أو اللحن فيها" (1)، ونمثل لذلك بجملة: "كتب الدرس"، فهذه الجملة لها

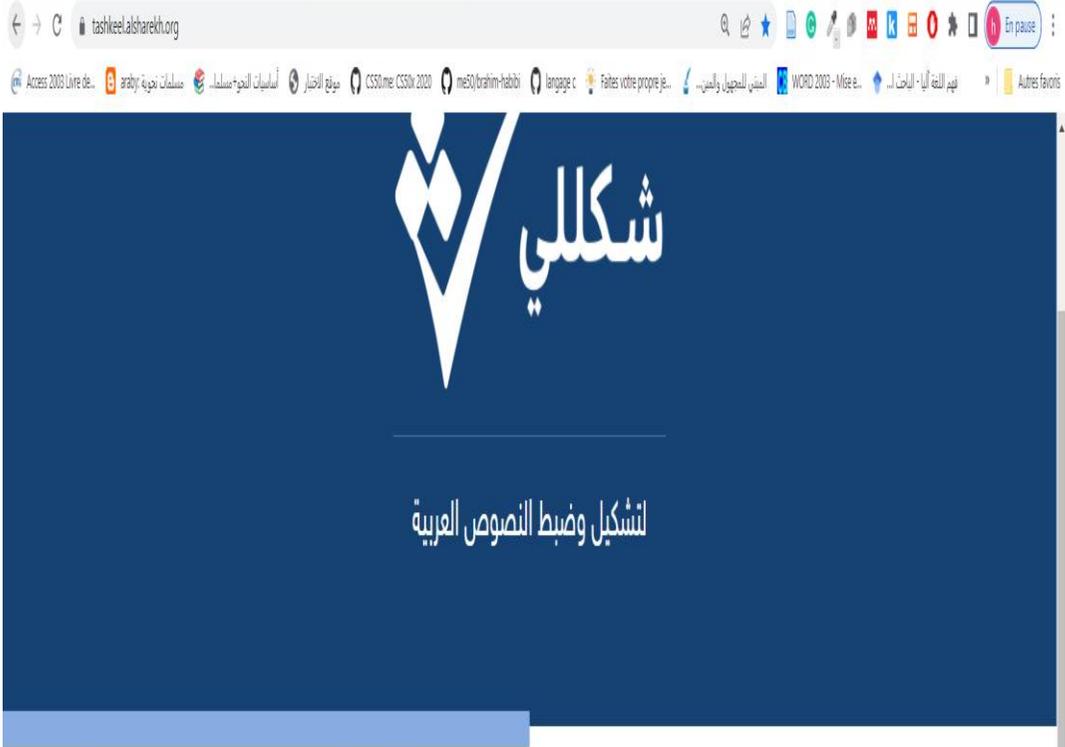
مجموعة من القراءات من بينها: "كَتَبَ الدَّرْسُ"، باعتبار الجملة جملة فعلية فاعلها ضمير مستتر و"الدرس" مفعول به، والقراءة الثانية هي: "كُتِبَ الدَّرْسُ"؛ باعتبارها جملة اسمية خبرها محذوف، وأما القراءة الثالثة فهي: "كُتِبَ الدَّرْسُ"، باعتبارها جملة فعلها مبني للمجهول، "الدرس" نائب الفاعل، إلى غيرها من القراءات.

إن تزايد النصوص العربية وإقبال غير العرب على تعلم العربية، كان دافعا قويا للتفكير في إيجاد آلية تساعد القارئ غير العربي على التمكن من قراءة الألفاظ العربية بدون لحن مما دفع فرق المعالجة الآلية للغة العربية إلى البحث في السنوات الأخيرة عن آليات لصورنة علامات الشكل العربي وبرمجتها من أجل حل مشكل هذه الظاهرة التي تتميز بها اللغة العربية، ففكروا في وضع نظم لتشكيل النصوص العربية آليا، وهو ما بادرت إليه شركة كوكل "google" سنة 2010، غير أن مبادرتها لم يكتب لها الاستمرار، مما دفع مجموعة من فرق البحث إلى بلورة العديد من نظم التشكيل الآلي من بينها: النظام الذي بلورته المنظمة العربية للتربية والثقافة والعلوم بإشراك مجموعة من الباحثين التابعين لجامعة محمد الأول بوجدة، من أجل توفير نظام للتشكيل الآلي حصل على المرتبة الأولى فيما يخص تطوير النظم المهتمة بالمعالجة الآلية للغة العربية.

(1) محمد سعيد ربيع الغامدي (2017)، " التشكيل في نظام الكتابة العربية: تاريخه ودوره الصوتي"، ضمن، " نظام الكتابة العربية: النشوء والتطور"، مباحث لغوية، ع: 33، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، المملكة العربية السعودية، الطبعة: الأولى، ص: 193.

كما نجد موقع شكلي (ينظر إلى الصورة(3)) وهو موقع يساعد على تشكيل النصوص العربية، ويتميز بدقة تصل بالنسبة للنصوص البسيطة إلى 98%، بينما تتراوح في النصوص التي تتميز باللغة الصعبة إلى ما يقارب 80% ونمثل له بالمثال الآتي:

"إِنَّ تَزَايِدَ النُّصُوصِ الْعَرَبِيَّةِ وَإِقْبَالَ غَيْرِ الْعَرَبِ عَلَى تَعَلُّمِ الْعَرَبِيَّةِ ، كَانَ دَافِعًا قَوِيًّا لِلتَّفَكُّيرِ فِي إِجَادِ الْيَّةِ تُسَاعِدُ الْقَارِئَ غَيْرَ الْعَرَبِيِّ عَلَى التَّمَكُّنِ مِنْ قِرَاءَةِ الْأَلْفَاظِ الْعَرَبِيَّةِ بِدُونِ لَحْنٍ مِمَّا دَفَعَ فِرْقَ الْمُعَالَجَةِ الْآلِيَّةِ لِلُّغَةِ الْعَرَبِيَّةِ إِلَى الْبَحْثِ فِي السَّنَوَاتِ الْأَخِيرَةِ عَنْ آيَاتِ لِحُونَةِ عِلَامَاتِ الشَّكْلِ الْعَرَبِيِّ وَبَرْمَجَتِهَا مِنْ أَجْلِ حَلِّ مُشْكَلِ هَذِهِ الظَّاهِرَةِ الَّتِي تَتَمَيَّزُ بِهَا اللُّغَةُ الْعَرَبِيَّةُ ، فَفَكَّرُوا فِي وَضْعِ نَظْمٍ لِتَشْكِيلِ النُّصُوصِ الْعَرَبِيَّةِ الْيَا".



نص جديد

إن تزايد النصوص العربية وإقبال غير العرب على تعلم العربية، كان دافعاً قوياً للتفكير في إيجاد آلية تُساعد القارئ غير العربي على التمكن من قراءة الألفاظ العربية بدون لحني مما دفع فريق المعالجة الآلية للغة العربية إلى البحث في السنوات الأخيرة عن آليات لصورة علامات الشكل العربي وبرمجتها من أجل حلٍ مُسكّلٍ هذه الظاهرة التي تتميز بها اللغة العربية، ففكروا في وضع نظامٍ لتشكيل النصوص العربية آلياً

الصورة 3: واجهة موقع شكلي

7. المعاجم الإلكترونية:

يُعرف المعجم الإلكتروني بكونه "تمثيلاً هندسياً لكفاية المتكلمين باللغة الطبيعية على خوارزميات التوليد والتحليل لاستخدامه في مختلف التطبيقات الهندسية للغة الطبيعية"⁽¹⁾ حيث يكون هذا المعجم مرتباً عموماً وفق ترتيب أفقي يضم كل سطر كلمة مصحوبة بخصائصها الصوتية والمعجمية والتركيبية والدلالية والصرفية المتمثلة في تناسب الجذر مع وزنه، وقد يخرج المعجم الإلكتروني عن هذا الترتيب إلى ما تقتضيه طبيعة البحث التي تشترط ترتيباً خاصاً، أو وضعه للمستعمل وفق طرق سهلة تمكنه من البحث في المعجم بطريقة سلسلة، وهذا ما جعل المعجم الإلكتروني يتميز عن المعجم الورقي، إذ يحتوي على إمكانيات هائلة لطرق البحث حسب الغرض المطلوب، وباللغة التي يختارها المستخدم، في وقت وجيز وبأقل تكلفة ممكنة.⁽²⁾

لقيت المعاجم الإلكترونية اهتماماً كبيراً من لدن مجموعة من الباحثين، نظراً لما يؤديه المعجم من دور في إرساء اللغة والحفاظ عليها من الزوال، حيث بدأ الأمر بوضع بنوك هدفت إلى جمع اللغة، وقد ذاع صيت هذه البنوك منذ التسعينيات من القرن الماضي شملت بنوكاً عربية وعربية نعرض من بينها:

● بنك المصطلحات الخاص بالمجموعة الأوروبية في لوكسمبورغ الذي

صدر عام 1975.

● بنك المصطلحات الكندي صدر عام 1997.

● بنك المصطلحات الأردني سنة 1985.

(1) محمد الحناش (2018)، "مصطلحات هندسة اللغة العربية بين التعريب والتأثيل"، مجلة اللسان العربي، ع: 79، المنظمة العربية للتربية والثقافة والعلوم ومكتب تنسيق التعريب، الرباط، ص: 214.

(2) حسن محمد السبومي، (2009)، "المعجم الحاسوبي للغة العربية وأبعاده التعليمية"، مؤتمر دولي، الرياض، بتصرف.

- البنك الآلي السعودي للمصطلحات سنة 1986. (1)
- وبعد ذلك تم الانتقال إلى ما يسمى بالمعاجم الإلكترونية، حيث تمكن العديد من الباحثين وكذا المؤسسات البحثية، والشركات من بناء العديد من المعاجم الإلكترونية نذكر منها:
- معاجم آلية فرنسية قدمها مختبر المعالجة الآلية والتوثيق اللساني (LADL) من بينها:
- المعجم العربي الإلكتروني الذي أنجزه الدكتور محمد الحناش. (2)
- موقع معاجم. (3)
- موقع الباحث العربي. (4)
- قاموس معجم المعاني. (5)
- معجم الوسيط الإلكتروني. (6)
- معجم اللغة العربية التفاعلي. (7)

(1) عبد الفتاح حمداني (2016)، "المعاجم العربية الآلية جرد وتقييم" ضمن كتاب، " المعجم العربي العصري وإشكالاته"، إشراف، أحمد بريسول وكنتزة بنعمر، بدون طبعة، معهد الدراسات والأبحاث للتعريب، جامعة محمد الخامس - السويسي - الرباط، ص: 251.

(2) عمر المهديوي (2009)، "المعجم الإلكتروني: قراءة في بعض التجارب الغربية والعربية"، ضمن الدراسات المعجمية، العدد السابع والثامن، منشورات الجمعية المغربية للدراسات المعجمية، الرباط - المغرب.

(3) www.maajim.com

(4) www.baheth.info

(5) www.almaany.com

(6) <http://sibawayh.emi.ac.ma/safar/applications>

(7) <https://sourceforge.net/projects/almuajim>

خلاصة:

استطاعت اللسانيات الحاسوبية العربية، بفضل جهود مجموعة من الطاقات البحثية والمراكز المهتمة بمعالجة اللغة العربية، صورنة المستويات اللسانية، وترجمتها إلى العديد من النظم والتطبيقات من أجل صياغة نماذج صورية محاكية لما هو موجود في الذهن البشري.

لقد استفادت المراكز البحثية المهتمة بالمعالجة الآلية للغة العربية من تطور تكنولوجيا المعلومات التي أبانت في العديد من المجالات قدرتها على محاكاة الذهن البشري، ويأتي المجال اللغوي في مقدمة الميادين باعتباره المجال الأكثر تأثراً بتكنولوجيا المعلومات ضمن العديد من اللغات، من بينها اللغة العربية التي حققت قفزة نوعية جعلتها تضع بصمة بارزة في مجال الصناعة اللغوية العالمية من خلال النظم التي قدمتها بالإضافة إلى نظم أخرى لم يسعفنا الحظ في الإشارة إليها، كبرامج القراءة الآلية...

خاتمة:

تشكل اللغة القيمة الجوهرية الأساسية لكل مجتمع، باعتبارها الأداة المترجمة لأفكار المجتمعات، والآلية التي تسهم في تحقيق عملية التواصل بين أفرادها، والوسيلة التي تمكن من الترابط والانسجام بين الناس، عبر صور كلامية، وقوالب لغوية ضابطة لذلك.

امتلكت اللغة العربية كغيرها من اللغات مجموعة من العناصر التي مكنتها من صياغة آليات وتوفير الوسائل التي اعتمدها العرب في تحقيق تواصلهم، وبناء انسجامهم، والحفاظ على الترابط في ما بينهم، وتتلخص كل تلك العناصر في أقسام الكلام العربي الثلاثة (الاسم، والفعل، والحرف)، والتي تمكن عبر ترابطها من تحقيق العمليات التواصلية بين الناس، وقد أشار العرب القدماء إلى أن إغفال أحد الأقسام لن يمكن من تحقيق الغرض المنشود، إن لم يكن ظاهراً فضمنياً، إذ أن المتكلم العربي أثناء كلامه بالاسم أو الفعل أو الحرف، فإنه لم يغفل باقي الأقسام.

الفهم الآلي للغة العربية تحليل الأراء والمشاعر نموذجاً

ونظراً لدقة مكونات اللغة العربية فقد تمكنت من مساهمة ركب باقي اللغات الطبيعية في ما تعلق بالتقنيات الحاسوبية؛ حيث أكدت جهود الكثير من المؤسسات البحثية، والعديد من التخصصات العلمية، كعلم الحاسوب تحقيق الهدف الذي ظهر في البداية صعباً لدى البعض، فتم تحقيق ترجمة للمستويات اللسانية العربية في إطار المعالجة الآلية للغة العربية، هذا التوجه المعرفي الذي أصبح يشكل أداة أساسية ووسيلة ضرورية من أجل مساهمة التقنية، واستيعاب النصوص في ظل الزخم الكبير الذي نعيشه في عصرنا، وقد توجت تلك الجهود بوضع مجموعة من النظم الحاسوبية التي ساعدت على استيعاب التفكير البشري، من ناحية الإنتاج اللغوي.

الفصل الثاني: الفهم الآلي للغات الطبيعية: تحليل الأراء والمشاعر العربية

المبحث الأول: تطبيقات الفهم الآلي للغات الطبيعية:

مدخل:

لقد أصبح إنتاج النصوص في زمن الثورة التكنولوجية الكبيرة أمراً غير محصور، حيث يتم إنتاج الملايين من النصوص يوميا، إلا أن تحليلها وفهم الخطاب أو السياقات التي وضعت من خلالها من طرف الإنسان يعتبر أمراً شبه مستحيل نظراً لغزارتها، فتم التفكير في استثمار الذكاء الاصطناعي لمحاولة تحليل تلك النصوص لغرض فهمها واستخلاص مضامينها، وتحديد معاني الكلمات التي تضمنتها، وعلى إثر ذلك تم انبثاق مجال معرفي جديد سمي بالفهم الآلي للغات الطبيعية.

يعتبر مجال فهم اللغات الطبيعية آلياً من المجالات الحديثة التي تهتم بحوسبة اللغة، كما أنه مجال متفرع عن الذكاء الاصطناعي الذي جاء لتوضيح كيفية اشتغال العقل البشري، وما دام الأمر يتعلق بالفهم الآلي للغات الطبيعية فإن الأمر يتعلق بمحاولة توضيح الكيفية التي يستطيع الإنسان من خلالها فهم الخطاب الذي يتلقاه من المتكلم، ومن ثم تصنيفه، أو تأويله، أو الإجابة عنه إذا كان الأمر متعلقاً بسؤال إلى غيرها من الوظائف التي تندرج ضمن فهم اللغة البشرية (الخطاب الإنساني) وعلى هذا الأساس فقد قامت مجموعة من المؤسسات والمعاهد البحثية، منذ الخمسينيات من القرن الماضي، ببلورة العديد من التطبيقات الحاسوبية التي تسعى إلى تبسيط فهم اللغة البشرية منطلقين من التطبيقات التي تمت بلورتها في إطار المعالجة الآلية للغات الطبيعية.

لقد مكنت النتائج الجيدة التي توصلت إليها نظم معالجة اللغة الطبيعية من المضي قدماً نحو تطوير نظم أخرى، بما فيها الفهم الآلي للغات الطبيعية مما حفز المعاهد والمختبرات البحثية على بلورة العديد من التطبيقات المتميزة بالذكاء الاصطناعي، والمساعدة على فهم اللغات الطبيعية، فعلى الرغم من أن المعالجة الآلية للغات الطبيعية تعود بوادها إلى الأربعينيات من القرن الماضي، حينما بدأ الباحثون

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

والمبرمجون الحاسوبيون في تجربة نمذجة اللغة وبرمجتها على مستوى الحاسوب لأسباب سياسية وحربية بالدرجة الأولى، وقد كانت إدخلات لغوية بسيطة تمثلت في بلورة نظام سيستران للترجمة الآلية وما جاء بعدها من تطبيقات أخرى، إلا أن فهم اللغة الطبيعية نجده قد بدأ يتطور في الستينيات من القرن الماضي لغرض تطوير وتمكين الحواسيب من فهم النصوص اللغوية الأكثر تعقيداً، لذلك فالفهم الآلي اللغات الطبيعية يعتبر فرعاً منبثقاً من مجال المعالجة الآلية للغات الطبيعية؛ إلا أن الفهم والمعالجة يختلفان من حيث الهدف؛ فإذا كان مجال المعالجة الآلية للغات الطبيعية يهتم بتحليل النص المدخل ألياً لتقديم معلومات صوتية، وتركيبية، ونحوية، ومعجمية، فإن الفهم الآلي للغات الطبيعية تهتم بالفهم الآلي للنص المدخل مع محاولة تقديم المعنى الحقيقي للنص من خلال تقديم ملخص له، وتحديد المجال الذي ينتمي إليه النص بناء عن الكلمات التي يتضمنها إلى غيرها من الوظائف الآلية، وهو ما حفز العديد من المؤسسات والشركات على إنشاء روبوتات وتطبيقات لها القدرة على التفاعل من الأشخاص، ومن بين الشركات التي تمكنت من تحقيق تقدم بارز في مجال فهم اللغات الطبيعية نجد أمازون وأبل وجوجل ومايكروسوفت وشركات أخرى ناشئة تسعى إلى تحقيق مشاريع خادمة للفهم الآلي للغات الطبيعية.

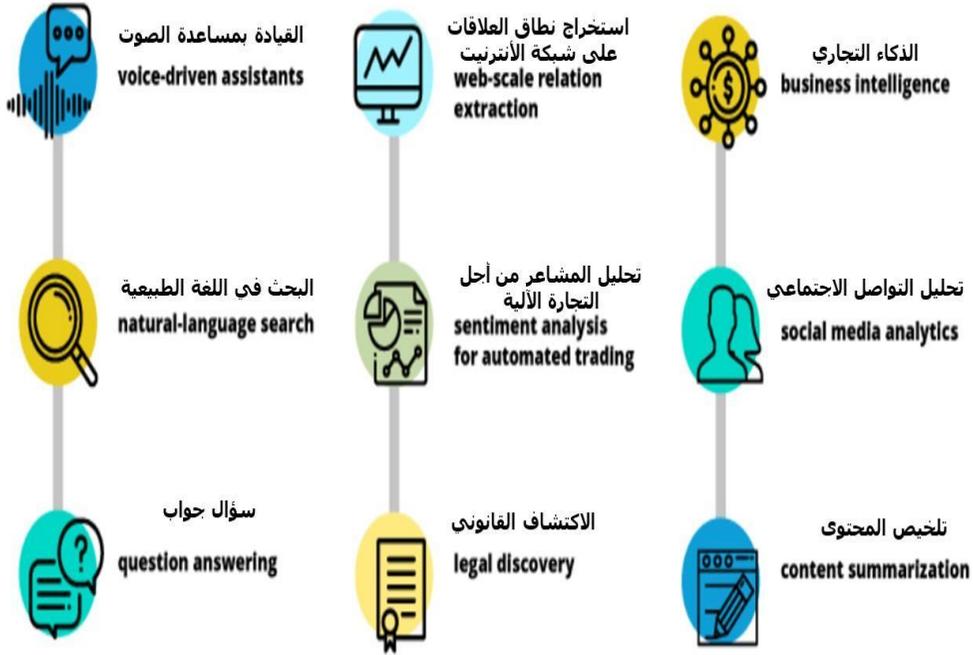
1. تطبيقات الفهم الآلي للغات الطبيعية

يستعين الفهم الآلي للغات الطبيعية، باعتباره أحد الموضوعات الأكثر تحدياً في مجال الذكاء الاصطناعي⁽¹⁾، بمجموعة من خوارزميات الذكاء الاصطناعي لتمكين النظم من الفهم الآلي للقدرات الإنسانية مثل: تحديد المواقف والمشاعر، وإنتاج النصوص أو

(1) Xiaolin Hu, « Meta Semantics: Towards better natural language understanding and reasoning » (arXiv, 20 avril 2023), <https://doi.org/10.48550/arXiv.2304.10663>.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

تلخيصها النصوص، إلى غيرها من الاهتمامات، والصورة (4) الموالية توضح بعض اهتمامات الفهم الآلي للغات الطبيعية⁽¹⁾.



الصورة 4: اهتمامات مجال الفهم الآلي للغات الطبيعية²

تبين لنا الصورة مدى تشعب مجالي حوسبة اللغات الطبيعية واللسانيات الحاسوبية، فبعدما استفاد العديد من الباحثين، والكثير من المؤسسات البحثية في الحديث عن الآليات والسبل لنمذجة المستويات اللسانية والتي استطاعوا من خلالها توفير العديد من النظم، أصبحت هذه المراكز توجه البحث نحو فهم اللغة الإنسانية التي تنبني

(1) <https://www.facebook.com/kdnuggets>, « NLP vs. NLU: From Understanding a Language to Its Processing », *KDnuggets* (blog), consulté le 3 juillet 2023, <https://www.kdnuggets.com/nlp-vs-nlu-from-understanding-a-language-to-its-processing.html>.

(2) <https://www.facebook.com/kdnuggets>.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

بالأساس على مبدأ التفاعل الذي راهن عليه مطورو شبكات الويب، وفي ما يلي سيتم عرض في بعض التخصصات المتعلقة بالفهم الآلي للغات الطبيعية في فصل خاص.

1.1. الذكاء التجاري (Business intelligence)

يعد الذكاء التجاري واحداً من الأنظمة التي تركز على البيانات المستجمعة من المعاملات التجارية لكل شركة، ونظراً لكثرة المعاملات الرقمية، فإن الشركات اتجهت نحو تجميع تلك البيانات الضخمة من أجل استثمارها في تطوير منتجاتها وتوسيع معاملاتها، لكن التعامل مع تلك البيانات يستدعي استثمار الذكاء الاصطناعي من أجل دقة تحليل وفهم تلك البيانات، ولهذا السبب تم دمج مجال التجارة والأعمال مع حقل الفهم الآلي للغات الطبيعية للتمكن من تحليل كل البيانات التي يتم جمعها، ومن ثم التمكن من تطوير المنتجات التي تنتجها الشركات، وكذا معرفة الفئات الاجتماعية التي يمكن أن يوجه لها المنتج الذي أنتجته الشركة، كما أنه بفضل الذكاء التجاري تمكنت الفرق البحثية من بلورة العديد من التطبيقات من بينها تطبيق "التعاون الجماعي" (Crowd-Sourcing)⁽¹⁾ والذي يستثمر في مجال التجارة من أجل السعي نحو إنجاز مهمة تجارية معينة، أو حل مشكلة متعلقة بمهمة تجارية معينة، عبر مساهمة عدد كبير من الأفراد عبر الإنترنت للحصول على حلول أو أفكار أو آراء ومعلومات يمكن استخدامها في إنجاز المهمة التجارية، ويستخدم هذا التطبيق في العديد من المجالات: كالصميم، والبرمجة، والبحث، والتسويق، وغيرها للاستفادة من التجارب الجماعية.

2.1. تحليل التواصل الاجتماعي (social media analytics)

في العقد الماضي أصبحت منصات وسائل التواصل الاجتماعي أكثر القنوات الرئيسية للتواصل الجماهيري، مما جعلها تكتسب شعبية كبيرة على الصعيد العالمي، فأصبح

(1) Musiri Kailasanathan Nallakaruppan et al., « Secured MCDM Model for Crowdsourcing Business Intelligence », Applied Sciences 13, no 3 (janvier 2023): 1511, <https://doi.org/10.3390/app13031511>.

الفهم الآلي للغة العربية تحليل الأسماء والمشتاعر نموذجاً

الأشخاص الذين ينخرطون في منصات التواصل الاجتماعي المزود الأول بالبيانات التي تستثمر في العديد من الدراسات، وما جعل تلك البيانات ذات قيمة هي الكفاءة التقنية والهندسة الحاسوبية لمنصات التواصل الاجتماعي الحالية، الشيء الذي جعل العديد من المؤسسات والمعاهد البحثية والحكومية تفكر في كيفية التعامل مع تلك البيانات، وهو ما شكل موضوعاً بحثياً منطلقاً من تدارس ما يتم تداوله في منصات التواصل الاجتماعي في العديد من الميادين: كالأعمال التجارية والسياحة والضيافة، وإدارة الكوارث وغيرها من المجالات، ومن بين ما يمكن التمثيل به على سبيل المثال لا الحصر، هو استثمار بيانات التواصل الاجتماعي من أجل تحليل الكوارث وذلك عن طريق تطوير وسائل التواصل الاجتماعي من أداة اتصال متعددة الاتجاهات إلى أداة للتعاون المشترك في استجابة الكوارث، بالإضافة إلى توفير أداة من شأنها أن تساعد السلطات والمجتمع في اتخاذ القرارات المستنيرة للتعامل مع التحديات في استجابة الكوارث⁽¹⁾

3.1. تلخيص المحتوى (content summarization)

يعتبر "لون" (Luhn) أول من اقترح فكرة تلخيص النصوص تلقائياً عام 1958، حيث دعا إلى البحث عن الكيفية التي يمكن من خلالها استخراج كيفية توزيع الكلمات داخل الجمل وتحديد الكلمات الرئيسية في الوثيقة، وقد شكلت تلك الفكرة نقطة انطلاق للبحث في مجال التلخيص الآلي للمحتوى النصي، حيث تم تطوير العديد من تقنيات التلخيص باستخدام مقاربات مختلفة⁽²⁾، ومن بين الأبحاث التي انصبحت نحو التلخيص الآلي نجد ما قام به كل من منيل بن عبد الله شفيق علو لمياء حديث بلغيث الذين

(1) Turgut Acikara et al., « Contribution of Social Media Analytics to Disaster Response Effectiveness: A Systematic Review of the Literature », Sustainability 15, no 11 (janvier 2023): 8860, <https://doi.org/10.3390/su15118860>.

(2) Yazan Alaya AL-Khassawneh et Essam Said Hanandeh, « Extractive Arabic Text Summarization-Graph-Based Approach », Electronics 12, no 2 (janvier 2023): 437, <https://doi.org/10.3390/electronics12020437>.

الفهم الآلي للغة العربية، تحليل الأسماء والمشاعر نموذجاً

اقترحوا منصة لتلخيص النصوص العربية تتضمن الوحدات التالية: ترميز الكلمات، تحليل الصيغة، تحليل بناء الجمل، استخراج الجمل ذات الصلة، ومراجعة ما تم استخراجها، وقد اعتمدوا على مجموعة متنوعة من النصوص (قصيرة ومتوسطة وطويلة) في تقييم هذه المنصة من حيث وقت التنفيذ، وتبين أن وقت تشغيل وحدات المنصة لنص محدد يتأثر بحجمه، حيث أنه كلما كان النص صغيراً، كانت سرعة التنفيذ قصيرة، وكلما كان حجم النص أطول إلا واستغرقت المنصة وقتاً أطول للقيام بعملية التلخيص⁽¹⁾.

4.1. سؤال جواب (Question Answering):

تعتبر تطبيقات سؤال جواب من التطبيقات التي لقيت اهتماماً بالغاً من لدن فرق البحث، حيث أن معظم الأنظمة التي تم إنجازها، نجدها تنقسم إلى ثلاثة مكونات رئيسية متمثلة في: تحليل السؤال، استرجاع المعلومات، واستخراج الإجابة، وقد تم تطوير العديد من أنظمة سؤال جواب وأصبحت واسعة الانتشار والاستعمال خاصة بلغات مختلفة، نخص بالذكر اللغة الإنجليزية واللغة الفرنسية، وغيرها من اللغات، إلا أن اللغة العربية لا تزال تواجه مجموعة من التحديات في بلورة نظم متعلقة بالإجابة على الأسئلة العربية آلياً⁽²⁾، ومرد ذلك إلى ندرة المراد اللسانية التي تعتبر الحجر الأساس في بلورة تدريب واختبار النظم التي تهتم باللغة العربية، ومن بين التجارب العربية التي وجهت اهتمامها نحو بلورة نظم الإجابة على الأسئلة آلياً ما قام به كل من سارة محمد كامل وشيما حسن ولياء الرفاعي اللاتي بلورن نظاماً للإجابة على الأسئلة العربية وذلك عبر إنشاء مجموعة من البيانات العربية مرئية للإجابة على الأسئلة، حيث ضمت حوالي

(1) Manel Ben et al., «TOWARD A PLATFORM FOR ARABIC AUTOMATIC SUMMARIZATION », 3 juillet 2023.

2 AL-Khassawneh et Hanandeh, « Extractive Arabic Text Summarization-Graph-Based Approach ».

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

138 ألف سؤال مقرون بإجابته، ومتضمنة صوراً متعلقة بتلك الأسئلة، بالإضافة إلى الاستعانة بمجموعة من الخوارزميات⁽¹⁾.

5.1. البحث باللغة الطبيعية (Natural language search)

لقد أدى تطور مجال المعالجة الآلية للغات الطبيعية، إلى تطور المجالات المرتبطة بها، ونخص بالذكر مجال التوليد الآلي والفهم الآلي للغات الطبيعية، وأهم ما يمكن الحديث عنه في هذا الجانب هو تطوير أنظمة البحث في اللغات الطبيعية، حيث كان خرج البحث في الماضي يقتصر على الكلمة المدخلة في خانة البحث، بينما أصبح البحث حالياً في معظم محركات البحث يعتمد على ربط ما يتم البحث عنه بالعديد من الكلمات التي تتعلق به، وقد خاضت مجموعة من فرق البحث الغمار في تطوير نظم البحث في اللغات الطبيعية سعياً منها نحو تحديد فوائد وقيود البحث باللغة الطبيعية لتسهيل عملية البحث على المستخدم⁽²⁾.

1 Sarah M. kamel, Shimaa I. Hassan, et Lamiaa Elrefaei, « VAQA: Visual Arabic Question Answering », *Arabian Journal for Science and Engineering*, 24 mars 2023, <https://doi.org/10.1007/s13369-023-07687y>.

2 Jeroen Smienk, « Natural Language Search for Personal Information Retrieval : Benefits and Limitations to the User's Experience in Bookkeeping Software. » (info:eu-repo/semantics/masterThesis, University of Twente, 2023), <http://essay.utwente.nl/94310/>.

المبحث الثاني: تحليل وتقييم الآراء والمشاعر ضمن التجارب العربية

مدخل:

أصبح للأنترنيت دور مهم في الحياة اليومية، وازداد تعزيز هذا الدور مع تطوير تقنيات الويب (Web) التي اقتصرت في بدايتها على توفير أدوات تمكن من هندسة مواقع تقدم المعلومات، قصد الاطلاع عليها والاستفادة منها، وقد وجهت مجموعة من الانتقادات لمثل هذه المواقع أهمها: أنها لم تكن تتيح للمتتبع الحق في الإشارة إلى الأخطاء الواردة في المحتوى، أو التعليق عليها، أو حتى إبداء الآراء حول المواضيع المعروضة فيها، وهذه هي أهم العيوب التي سجلت عن هذا الجيل، فتم التفكير في تطوير تقنية الويب لسد الثغرات التي سُجلت عن الجيل الأول، فظهر الجيل الثاني الذي اعتمد فيه مجموعة من التقنيات، لتجاوز النقص الحاصل في الجيل الأول.

يعتبر الجيل الثاني من خلال الامتيازات التي اتصف بها، بأنه نسخة جديدة تمكنت من تحويل المواقع التي كانت تعمل على تقديم المعلومة فقط، إلى مواقع تشرك المتتبع في إنتاج المعلومة، وتحديد أهم المواضيع التي يرغبون في الاطلاع عليها، فيكون الجيل الثاني بذلك سببا في جعل هدف المواقع، متمثلا في خلق تفاعل بين القائمين على إدارة الموقع من جهة، وبين المتتبعين للموقع من جهة أخرى، ومع ظهور الهواتف الذكية، ازداد التشجيع في تطوير تقنيات الويب وتحويل مجموعة من المواقع إلى تطبيقات يسهل تثبيتها والاشتغال بها انطلاقا من الهواتف الذكية، وقد ساهمت كل هذه العوامل في تطوير العديد من النظم ضمن هذا الجيل أهمها:

- الويكي Wiki الكتابة التعاونية: وهو عبارة عن نظام يخول

إنشاء صفحات الويب عبر تحرير محتواها وربطه تشعبيا بمحركات البحث بطريقة سهلة، وعادة ما يتم توظيف هذه النظم في المواقع التي تركز في بنائها على التعاون، وكمثال على ذلك الموسوعة الحرة الأكثر شهرة والمعروفة

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

التي تختص بمجال البيع والشراء بتمكين زوارها من إبداء آرائهم حول المنتجات المعروضة للبيع، حتى أصبحت إمكانية إبداء الآراء سنة كل موقع يوفرها لزواره لغرض تطوير ما يقدمه.

لقد نتج عن ذلك النوع من المواقع إنتاج نصوص كثيرة ومتزايدة بتزايد زوارها والمنخرطين ضمنها، حتى أصبح الأمر لمهندسي تلك المواقع صعباً للاطلاع على كل ما يتم تداوله في المدونات، أو المجموعات المنشأة على مواقع التواصل الاجتماعي، أو حتى تلك الآراء الواردة في المواقع التجارية والإخبارية، وغيرها من المواقع، مما أدى في السنوات الأخيرة إلى ظهور حقل معرفي كبير، يهتم بتدبير محتوى الوثائق الرقمية التي يتم نشرها وتدارسها في مختلف المواقع، سمي هذا الحقل بحقل "استرجاع المعلومات" (**Information Retrieval**)، وقد حظي هذا الأخير بمكانة كبيرة في مجال نظم المعلومات نتيجة الإنتاج المتزايد للنصوص الرقمية، ومدى الحاجة لاسترجاعها بمرونة ومعالجتها بسهولة، ففي أوائل الستينيات تمت الاستعانة بمجال تصنيف نصوص اللغات الطبيعية ووسمها، باعتبارهما مجالين مساعدين على استرجاع المعلومات، وسيزداد الأمر اهتماماً في أوائل التسعينيات، حيث أصبح مجال تصنيف النصوص أحد أهم المجالات الفرعية لنظم استرجاع المعلومات، وذلك بفضل الاهتمام المتزايد بالنصوص الرقمية وتوفير الأدوات المساعدة على استرجاع المعلومات، بل وأصبح مجال تصنيف النصوص معتمداً في العديد من المجالات البحثية من بينها: فهرسة الوثائق باعتماد المعجم، تصفية الوثائق، إزالة غموض معنى الكلمات، نشر الفهرسة الهرمية موارد الويب، إلى غيرها من الاهتمامات التي تركز على تنظيم الوثائق.⁽¹⁾

(1) Grzegorz Dzikowski (2008), « Analyse des sentiments: système autonome d'exploration des opinions exprimées dans les critiques cinématographiques » (THÈSE DE DOCTORAT, PARIS, L'ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS), P: 7-8 . بتصرف.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

لقد شكل مجال تصنيف النصوص، نقطة التقاء بين كل من مجال التعليم الآلي واسترجاع المعلومات، وهو ما ساعد على ظهور بعض المجالات التي لها علاقة باسترجاع المعلومات، من بينها التنقيب في النصوص، واستخراج المعلومات، مما أدى إلى ظهور مجال معرفي جديد لا يقل أهمية عن مجال التنقيب في النصوص؛ ويتعلق الأمر بالتنقيب عن الآراء (Opinion Mining)، وهو عبارة عن مجموعة من المعالجات الحاسوبية التي تركز على استخراج المعرفة وفقاً لمعايير الجودة أو التشابه في النصوص التي ينتجها الإنسان، وقد انبثق عن مجال التنقيب عن الآراء مجموعة من الاهتمامات البحثية من بينها: تعليم التوجيه الدلالي للكلمات أو العبارات، والتحليل العاطفي للوثائق، وكذا تحليل الآراء والمواقف تجاه مواضيع أو منتجات معينة، إلى غيرها من الاهتمامات البحثية.

لقد ظهر أول استعمال لمفهوم التنقيب عن الآراء (Opinion Mining) سنة 2003 من طرف دافي (Dave) خلال مؤتمر الشبكة العنكبوتية المعروف بمؤتمر (www) أو (3 w)، والذي عرف إثره مجال التنقيب عن الآراء، باعتباره "معالجة مجموعة من نتائج البحث لحالة معينة عبر إنشاء قائمة من السمات (كالجودة، الخصائص، وما إلى ذلك)، وتجميع الآراء (سيئة، معتدلة، وذات نوعية جيدة)"⁽¹⁾، بل شمل العديد من التحليلات التي تهدف تقييم النصوص.

تزامنت شهرة مجال التنقيب في النصوص مع بروز مجال تحليل المشاعر، الذي استعمل لوصف التحليل التلقائي للنص ثم تقييمه من أجل البحث عن القيمة

(1) Kushal Dave, Steve Lawrence, et David M. Pennock (2003), « Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews », in Proceedings of the 12th International Conference on World Wide Web, WWW '03 (New York, NY, USA: ACM), 519–528, <https://doi.org/10.1145/775152.775226> بتصرف

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

التنبؤية للأحكام، ضمن أعمال كل من داس وشين (Das Et Chen) سنة 2001⁽¹⁾، وطنغ (Tang) 2001، وفي نفس السنة تم اعتماد هذا المجال لتحليل المشاعر في المجال الاقتصادي، ثم أشار كل من ترين (Turney) سنة 2002⁽²⁾ وكذا طنغ رفقة زملائه 2002⁽³⁾ بعض الأعمال التي وظفت مجال تحليل المشاعر، ومنذ تلك السنة تمت مناقشة مجموعة من الأوراق العلمية التي ركزت على تصنيف التعليقات وتحديد قطبيتها (موجب - سالب)، فأصبح بذلك كل من التنقيب عن الآراء وتحليل المشاعر، يشتركان في نفس مجالات البحث، وأصبح بذلك مجال تحليل المشاعر هو الأكثر شيوعاً، في حين قام بعض الباحثين في دمجهما، فنتج عن ذلك مجال تحليل الآراء والمشاعر.

ترتكز عملية تحليل الآراء والمشاعر على تحليل النصوص، لتقييم الآراء واستخلاص مختلف أنواع المشاعر، باعتماد مقاربات متنوعة، تمثلت في ثلاث مقاربات كبيرة: تتجلى الأولى في المقاربة الإحصائية التي تعتمد بالأساس على التعليم الآلي بشقيه الموجه وغير الموجه، أما الثانية فهي المقاربة اللسانية، في حين تتجسد الثالثة في المقاربة الهجينة التي تتأسس على دمج المقاربتين السابقتين.

وما دامت الاستعانة بحقل المعالجة الآلية للغات الطبيعية، فقد تم التركيز أثناء تحليل الآراء على الجانب النصي، دون الاستعانة بمناهج أخرى كالمنهج النفسي

(1) Sanjiv R. Das et Mike Y. Chen (2007), « Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web », Management Science 53, no 9: 1375-88 . بتصرف .

(2) Peter D. Turney (2001), « Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews » (Association for Computational Linguistics), 417, <https://doi.org/10.3115/1073083.1073153> . بتصرف .

(3) Bo Pang, Lillian Lee, et Shivakumar Vaithyanathan (2007), « Thumbs up?: Sentiment Classification Using Machine Learning Techniques », vol. 10 (Association for Computational Linguistics), 79-86, <https://doi.org/10.3115/1118693.1118704> . بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

والاجتماعي رغم اعتبارهما منهجين أساسيين في استخلاص الآراء والتعرف على دوافع التعبير عن المشاعر والعواطف والأحاسيس.

وقد خصصنا هذا المبحث، للتعريف بمجال تحليل الآراء والمشاعر، ثم الحديث بشكل مقتضب عن الاهتمامات المنطوية تحته، لنفصل بعد ذلك الحديث عن أهم مستويات التحليل المعتمدة في هذا المجال، ونشير كذلك إلى احتواء اللغة العربية لمجال تحليل الآراء والمشاعر، ثم نركز الحديث عن عقبات المعالجة القبلية لنصوص اللغة العربية ومدى تأثيرها، بالتمكن من تقييم الآراء واستنتاج المشاعر، وبعد ذلك سنشير إلى أهم التحديات التي تواجهها اللغة العربية في هذا المجال، ثم الحديث عن أهم المقاربات المعتمدة في مجال بناء نظم تحليل الآراء والمشاعر، لنختتم هذا المبحث بالإشارة إلى بعض الأعمال التي قدمت حول اللغة العربية.

1. التعريف بمجال تحليل الآراء والمشاعر:

يندرج مجال تحليل الآراء والمشاعر، ضمن حقل رئيسي وهو التنقيب عن الآراء (opinions mining)، هذا المجال الذي جاء من أجل دراسة النصوص، واستخلاص الآراء الواردة فيها، وتحديد التعابير الحاملة لتلك الآراء، وتحليل الجمل قصد الكشف عن قوة الرأي المعبر عنه، كما يتم كذلك الكشف عن بعض العناصر المكملة للرأي، من قبيل تحديد صاحب الرأي، والموضوع الذي وُضع من أجله الرأي، وقد يتعدى ذلك إلى تلخيص الرأي، وعموماً فإن مجال التنقيب عن الآراء ينقسم إلى قسمين أساسيين؛ الأول هو تحديد كيانات الرأي التي قد تتعلق إما بمنتج، أو خدمة، أو مكان، أو شخصية، أو حدث، أما القسم الآخر فيتجلى في المشاعر التي تتضمنها الآراء.

لقد قدمت تعاريف عدة لمجال تحليل الآراء والمشاعر، من بينها تعريف لي (Liu)، حيث قال بأن تحليل الآراء والمشاعر هو: "حقل دراسي تحلل وفقه آراء الأشخاص، ومشاعرهم، واثميناتهم، وانفعالاتهم، وعواطفهم، وتوجهات كياناتهم،

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

وسماتها المعبر عنها ضمن النص"⁽¹⁾، في حين قدمت أنا فلديفيا كرسيا (Ana Valdivia García) تعريفاً آخر تشير فيه إلى أن تحليل الآراء والمشاعر هو "عبارة عن مجموعة من التقنيات الحاسوبية لاستخراج، وتصنيف، وفهم، وتقييم الآراء المعبر عنها في مختلف مواقع الأخبار وضمن التعليقات الموجودة في مواقع التواصل الاجتماعي وغيرها من المحتويات"⁽²⁾، أما كرلوس شلستا (Karol Chlasta) فيدرج تحليل المشاعر أو تعدين الآراء ضمن سياق المعالجة الآلية للغات الطبيعية هو تطبيق منهجية تنقيب النصوص أو اللسانيات الحاسوبية لاستخراج وتحديد معلومات موضوعية، انطلاقاً من مصدر البيانات، ويتمثل الهدف الأساس من تحليل المشاعر في تصنيف القطبية التي يحيل عليها النص، انطلاقاً من المستند أو الجملة أو على مستوى الميزة/ الوجهة. وقد حدد مجموعة من أصناف تحليل المشاعر من بينها:

- تصنيف المستند حسب الآراء التي يتضمنها إلى موجب أو

سالب؛

- تحديد درجة المستند؛

- التركيز على العواطف لمعرفة طبيعتها انطلاقاً من الرأي...⁽³⁾

ومن بين التعاريف كذلك تعريف دعاء محي الدين محمد حسين، حيث قالت

بأنه مجال "يتم فيه توظيف تقنيات المعالجة الآلية للغات الطبيعية، ومعالجة

(1) Liu, B., & Zhang, L. (2012). «A survey of opinion mining and sentiment analysis». In Mining text data (pp. 415-463). Springer, Boston, MA بتصرف .

(2) Ana Valdivia García (2008), «Sentiment Analysis-Based Methods for Cultural Monuments» (Thesis Doctora, Espagne, Universidad de Granada, p:17-18. <http://digibug.ugr.es/handle/10481/54784> بتصرف .

(3) Karol Chlasta (2015), «Implementation of sentiment analysis model for Twitter data in Polish language », p:6, <https://doi.org/10.13140/RG.2.2.12715.69924> بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

النصوص، والتقنيات الحاسوبية للاستخراج، أو التصنيف الآلي للمشاعر⁽¹⁾، كما تعرف أسماء منتصر مجال "تحليل المشاعر أو التنقيب عن الآراء بأنه: مجال من مجالات البحث التي تركز على الآراء المعبر عنها في مواضيع مختلفة على صفحات الويب، ويتم تحليل هذه الآراء وفق تحديد واحدة من بين العناصر الآتية: إما التحليل الذاتي الذي يرمي إلى تحديد الآراء التي تتضمن تعابير ذاتية، أو تصنيف قطبيتها، أو تصنيف المشاعر، أو تحديد الانفعالات وقوة الآراء"⁽²⁾.

انطلاقاً من التعاريف التي سبق ذكرها، يمكن القول إن مجال تحليل الآراء والمشاعر - أو أحد المصطلحات التي تحيل عليه من قبيل الآراء، والذاتية، والموضوعية، والمشاعر، والعواطف، والأحاسيس- هو مجال يُعتمد في معرفة وجهات نظر الكاتب من خلال تدوينته، والتعرف على آرائه وتصنيفها إلى آراء إيجابية أو سلبية أو محايدة، واستخلاص ما إذا كانت كتابته تميل إلى الذاتية أو الموضوعية، كما يساعد هذا المجال على تحديد المشاعر الذاتية، وتصنيفها إلى حقول مختلفة باختلاف طبيعة الأحاسيس والانفعالات التي قد تمزج بين الفرح والحزن، بين الخوف والأمان، بين الحب والكرهية، إلى غيرها من الأحاسيس والعواطف والمشاعر التي يعترف بها الكاتب من خلال الألفاظ الموظفة في كتاباته، ويهتم هذا المجال بالأساس بما كتب على صفحات أو مواقع الويب التي تختلف باختلاف أهدافها من قبيل المنتديات، ومواقع التواصل الاجتماعي، ومواقع البيع والشراء، ومواقع الكتب إلى غيرها من المواقع.

(1) Doaa Mohey El-Din Mohamed Hussein (octobre 2018), « A Survey on Sentiment Analysis Challenges », Journal of King Saud University - Engineering Sciences 30, no 4: 330-38, <https://doi.org/10.1016/j.jksues.2016.04.002> . بتصرف .

(2) A. Mountassir, H. Benbrahim, et I. Berrada (2012), « A Cross-Study of Sentiment Classification on Arabic Corpora », in Research and Development in Intelligent Systems XXIX, éd. par Max Brammer et Miltos Petridis (Springer London, 259-72) . بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

إن الحديث عن تحليل الآراء والمشاعر، يقتضي الحديث عن مجموعة من العناصر التي يتضمنها الرأي المعبر عنه في المواقع؛ إلا أن معظم الأبحاث لا تركز إلا على عنصر واحد، متمثل في تحديد قطبية توجيه الرأي أو الشعور، وفي ما يلي إشارة إلى أهم العناصر التي يتضمنها كل رأي:

○ الموضوع (Object): هو الشيء الذي أُعطي فيه الرأي سواء كان منتجاً، أو خدمة، أو قضية، أو موضوعاً، أو شخصية، أو تنظيم، أو غيرها.

○ الجانب (Aspect): هو المسؤول عن الموضوع الذي أُعطي فيه الرأي، كالشركات التي تعرض منتجاتها في المواقع الإلكترونية.

○ موجّهات المشاعر (Sentiment Orientation): هو المؤشر الذي يحيل على أن الرأي إما موجبا، أو سالبا، أو محايدا.

○ صاحب الرأي (Opinion Holder): هو من قدم الرأي حول الموضوع سواء تعلق الأمر بشخص أو هيئة.

○ الوقت (Time): الوقت الذي نُشر فيه التعليق حول الموضوع.

(1)

2. اهتمامات مجال تحليل الآراء والمشاعر:

تم اكتشاف مجموعة من المهام، وبناء العديد من التطبيقات التي لها علاقة بمعالجة الآراء والمشاعر، وما زالت في تزايد مستمر بسبب كثرة المواضيع المتناولة على مختلف المواقع وتنوعها (السياسة، الرياضة، العلوم، السياحة، الاقتصاد، ...)، وهذه المعلومات من شأنها تحسين عملية صنع القرار لمجموعة من الشركات أو المنظمات، أو حتى الجهات الحكومية، كما أن النظم والتطبيقات التي تهتم بهذا المجال في تزايد يومي،

(1) N. Boudad, R. Faizi, R. Oulad Haj Thami, et R. Chiheb, (2017) « Sentiment analysis in Arabic: A review of the literature », Ain Shams Eng. J.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

نتيجة الالحاق المتزايد في معرفة آراء المدونين حول موضوع ما (منتج، خدمة، شخصية سياسية، منظمة، جمعية...)، والتي تشكل تحدياً كبيراً أمام الباحثين، إذ تجد بين الفينة والأخرى رواد المواقع يوظفون مجموعة من العبارات الجديدة، وكذا استعمال العامية، وهناك من يكتفي بالتعبير عن آرائه باستعمال الرموز التعبيرية (Emojing). ليشكل بذلك تحدياً كبيراً في التعرف على الآراء والمشاعر، والجدير بالذكر أن هناك مجموعة من التصنيفات المتعلقة بالآراء المدروسة من بينها:

تصنيف الذاتية (Subjectivity Classification): والمقصود بهذا التصنيف

تحديد الجمل التي تحوي الآراء الذاتية وتصنيفها إلى الموجب والسالب.

تصنيف القطبية (Polarity Classification): وهذا النوع من التصنيف،

يعتبر أكثر التصنيفات المعتمدة ضمن مجموعة من الأبحاث في مجال تحليل الآراء والمشاعر؛ إذ يتم تحديد قطبية النص الذي قد يكون عبارة عن مستند أو جملة (Sentence) أو فقرة (Phrase)، باعتماد أحد التوسيمات المعتمدة في التصنيف؛ إذ يمكن تصنيفها باعتماد تصنيف ثنائي (موجب - سالب)، أو تصنيف ثلاثي (موجب - سالب - محايد) أو باعتماد مقاييس النص (Text Scale) (موجب جداً - موجب - محايد - سالب - سالب جداً) أو درجة القوة (موجب قوي - موجب ضعيف - محايد - سالب ضعيف - سالب قوي)، كما قد يمكن اعتماد مقياس الأرقام ([1، 2، 3، 4، 5] - [0] - [-1، -2، -3، -4، -5])...

تلخيص الآراء (Opinion Summarization): وهو محاولة تلخيص

مجموعة من التعليقات.

تصوير الآراء (Opinion Visualization) وهي عملية تنبني بالأساس على

عكس المحتوى النصي إلى رسوم بيانية.

كشف التهكمات (Sarcasm Detection): وتهدف هذه المهمة إلى التعرف

على السخریات التي يتضمنها النص؛ إذ عادة ما يتم الكشف عن المفارقات بين البشر

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

عن طريق الصوت أو إشارات الوجه، إلا أن هذه المهمة تجعل من هذا النوع من التصنيفات أشد صعوبة.

استخراج كيان الرأي وصاحبه، ووقت التعليق (Entity, Opinion Holder And Time Extraction): والهدف من هذا التصنيف، هو تفصيل المعلومات المستخرجة من الآراء، والمتمثلة في معرفة من كتب الرأي، وأين كتب الرأي، وعلى ماذا يتحدث.

التعرف على قرار الجماهير وتحديد غموض دلالة الكلمة (Coreference Resolution And Word Sense Disambiguation): ويتعلق الأمر في هذا النوع من التصنيف بالمفهوم أو الفكرة التي تتعدد دلالاتها، وهذا الأمر يهم بعض اللغات التي تتضمن نفس الكلمة لكن لها دلالات متعددة، لذلك فمن المهم اعتماد التصنيف الآلي للكلمات المتعددة الدلالات في النص إذا كان من المفروض الحصول على خلاصة دقيقة (1).

كشف هدف الرأي (Opinion Target Detection): يبنى كشف هدف الرأي على تمييز الأمر الذي يتعلق به الرأي، باعتماد منهجيتين؛ أولهما منهجية من الأعلى إلى الأسفل (Top-Down) وتوظف عندما يكون الهدف محددا مسبقا عن طريق سمة أو خاصية من خصائص الهدف المحددة، ضمن ما يسمى بالأنطولوجيات أو أنظمة التصنيف، وكمثال على ذلك الفندق الذي يضم مجموعة من الخصائص التي تكون عرضة للتقييم كالغرف، وموقع تواجد الفندق، أو أن يتعلق الأمر بالكاميرات وما يمكن تقييمه ضمنها كالسعر، والحجم، وعمر البطارية، إلى غيرها، أما المنهجية الثانية فتتمثل وفق مفهوم من الأسفل إلى الأعلى (Botton-Up)، وهو عكس المنهجية السابقة

(1) Ana Valdivia García (2008), «Sentiment Analysis-Based Methods for Cultural Monuments» (Thesis Doctora, Espagne, Universidad de Granada, p:17-18. <http://digibug.ugr.es/handle/10481/54784>.p: 18-19 بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

إذ أن الأهداف المحتملة لا تكون معروفة مسبقاً، بل يتم استخراجها من النص بشكل آلي، باعتماد مرحلة ضمن أحد مراحل المعالجة الآلية للغات، خاصة التحليل اللغوي منها، والذي لا يزال يشكل صعوبة في هذا النوع من التحديد نظراً لضعف نظم التحليل اللغوي الحالية.⁽¹⁾

3. مستويات التحليل في مجال تحليل الآراء والمشاعر:

اعتمدت مجموعة من الأبحاث العديد من مستويات التحليل، لاستخلاص الآراء والمشاعر وتقييمها، وسنحاول التعريف ببعضها في ما يلي:

1.3. التحليل على مستوى المستند (Document level):

يهتم هذا المستوى بتصنيف المستند حسب ما تضمنه من آراء، إلى مستند موجب أو سالب أو محايد، ولذلك فقد واجه هذا المستوى من التحليل، انتقاداً ضمن مجموعة من الدراسات، نظراً لعدم دقته في النتائج المقدمة في التحليل، إذ يصعب تطبيقه على المستندات التي تتضمن أكثر من رأي، أو تضم آراء متعارضة، أو تعبر عن مشاعر مختلفة، كما يصعب تطبيق هذا المستوى من التحليل عن الوثائق التي تتحدث عن مقارنة شيء بشيء آخر⁽²⁾، ورغم الانتقادات التي وجهت لهذا المستوى إلا أنه لم يمنع بعض الباحثين من العمل وفقّه، واتخاذهم مستوى تحليل لأعمالهم مثل الأعمال التي قام بها أحمد عباسي رفقة زملائه، حيث أعدوا نظاماً لتحليل الآراء والمشاعر المتعددة اللغات، فأجروا تجاربهم على اللغة العربية واللغة الإنجليزية، معتمدين في ذلك على مقارنة التعليم الآلي، فاستطاعوا تحقيق نتائج جيدة فاقت في اللغة العربية

(1) Diana Maynard et Kalina Bontcheva (2016), «Natural Language Processing for the Semantic Web», éd. par Isabelle Augenstein (Morgan & Claypool Publishers, p:76 بتصرف).

(2) Alawya Adnan Alawami (2017), «Aspect Extraction For Sentiment Analysis In Arabic Dialect», (Thesis Doctora, University of Pittsburgh, p: 11 بتصرف).

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

(91%) على اللغة الإنجليزية (90%)⁽¹⁾، ومن هذه الأبحاث التي اعتمدت هذا المستوى، نجد بحثاً لطارق المسلم ونزلية عمر، ومحمد البارد في عملهم الذي ركز على تصنيف تعليقات في مجالي الخدمات والإنتاج في الولايات المتحدة، وقد ضمت المدونة المعتمدة ما يقارب 3450 تعليقا، خصص منها 3000 تعليقا للتدريب، وخصص منها 450 تعليقا للاختبار، اعتمادا على مقارنة التعليم الآلي، وقد تمكنوا من الحصول على نتائج جيدة بلغت نسبة 97%⁽²⁾.

2.3. التحليل على مستوى الجملة (Sentence level):

يعتبر التحليل على مستوى الجملة أقرب تحليل ضمن مجال تحليل الآراء والمشاعر⁽³⁾، إذ يعمل هذا المستوى على تصنيف الجمل - حسب الرأي الذي عبرت عنه - إلى إيجابي، أو سلبي، أو محايد⁽⁴⁾، كما أشار كل من ليو (Liu) وذنغ ويو (Yu)، إلى أن هذا المستوى يظل مرتبطا ارتباطا وثيقا بالتحليل الذاتي (Subjectivity Analysis)، فيتم من خلاله التعرف على العواطف، والقناعات، والتمنيات، كما يهتم هذا المستوى كذلك بالتصنيف الموضوعي، الذي تُعبر فيه الجملة عن معلومات واقعية وهنا ينبغي الإشارة إلى أن الجمل الذاتية، لا تعادل الجمل المعبرة عن المشاعر، كما أن الجمل

(1) Ahmed Abbasi, Hsinchun Chen, et Arab Salem (2008), « Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums », ACM Trans. Inf. Syst. 26: 12-12, <https://doi.org/10.1145/1361684.1361685> . بتصرف .

(2) Nazlia Omar et al. (1 janvier 2013), « Ensemble of Classification Algorithms for Subjectivity and Sentiment Analysis of Arabic Customers Reviews », International Journal of Advancements in Computing Technology 5, 77-85 . بتصرف .

(3) Emma Haddi (2015), «Sentiment Analysis: Text Pre-Processing, Reader Views And Cross Domains» (Thesis Doctora, BRUNEL UNIVERSITY, p: 18 . بتصرف .

(4) Bing Liu (23 mai 2012), « Sentiment Analysis and Opinion Mining », Synthesis Lectures on Human Language Technologies 5, no 1: 1-167, <https://doi.org/10.2200/S00416ED1V01Y201204HLT016> . بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

الموضوعية من شأنها أن تحيل على رأي ولو بشكل ضمني، مثل "اشترينا السيارة الشهر الماضي، وقد سقطت لها ماسحة الزجاج الأمامي" فهذا المثال يحيل على رأي يقضي بالحكم؛ إما برداءة نوع السيارة وإن كان ضمنياً، أو برداءة ماسحة الزجاج وهو الغالب، ويعد هذا المستوى من التحليل، بمثابة المحدد الأساسي لتوجيه الكلمة، باعتماد السياق اعتماداً أساسياً، ويشمل التصنيف حسب هذا المستوى جمل المقارنة وأسلوب المدح والذم إلى غيرها من الأساليب. ورغم ذلك فقد أشار الباحثون أن مستوى الجملة ما زال محتاجاً إلى المزيد من الدراسات.

لقد اعتمدت مجموعة من الأبحاث هذا المستوى من التحليل في بداية تحليلها للتعليقات أو المستندات الطوال؛ إذ يتم تقسيم التعليق أو المستند إلى جمل أو عبارات اعتماداً - في غالب الأمر - على علامات الترقيم، أو على بعض حروف المعاني، ولتوضيح ذلك نقدم المثال الآتي: "الفندق جميل، الاستقبال رائع لكن الأفرشة رديئة" ففي هذا المثال يتم تقسيمه إلى ثلاث جمل:

- الفندق جميل؛
- الاستقبال رائع؛
- الأفرشة رديئة؛

فيتم اعتماد التحليل التجميعي (Collective Analysis) الذي يقضي إلى تجميع درجات المشاعر الإيجابية والسلبية لكل جملة، ثم تقديم الدرجة النهائية، أو اعتماد التحليل الجماعي ((Analysis Aggregated⁽¹⁾، ومن أهم الباحثين الذين ركزوا على هذا المستوى من التحليل نذكر محمد عبد المجيد وبعض زملائه في مجموعة متنوعة من الأعمال، أهمها ما طبقه رفقة منى دياب على مدونة متعددة الأجناس (AWATIF)

(1) Diana Maynard et Kalina Bontcheva (2016), «Natural Language Processing for the Semantic Web», éd. par Isabelle Augenstein (Morgan & Claypool Publishers, p: 78 بتصرف).

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

(1)، بالإضافة إلى العمل الذي اعتمد فيه رفقة منى دياب، على هذا المستوى لاختبار أداء معجم متعدد العاميات أسموه ب. (2) (SANA)

3.3. التحليل على مستوى الكيان والجانب (aspect level):

يعتمد هذا المستوى من تحليل الآراء والمشاعر على تحديد الكيان (Entity) والجانب (Aspect) المرتبط به، فإذا كنا بصدد التعليق على خدمة أو منتج على سبيل المثال، فإنه يمكن التعبير عن شعورنا تجاهه بالإشارة إلى عدة جوانب مرتبطة به، اعتماداً على الجمل أو الوحدات النصية الصغرى، حتى يتم إنهاء التعبير عن المنتج؛ فالمقهي على سبيل المثال إذا أردنا أن نعبر عن شعورنا تجاهها، فإن شعورنا قد يتعلق بمكان تواجد المقهى أو الخدمة، أو الاستمتاع بشرب المشروب داخلها، أو نظافة المكان، أو الحمام، إلى غيرها من الجوانب المتعلقة بكيان المقهى، ولهذا الغرض، فإن عملية التحليل وفق الكيان والجانب تكون أكثر تفصيلاً؛ إذ يتم الاعتماد على مجموعة من المعلومات المقدمة في التعليق، فيتم إثر ذلك اكتشاف الجوانب الجديدة الواردة في التعليقات، بدلا من تكرار الجوانب السابق تقييمها في تعليقات سابقة.

(1) Muhammad Abdul-Mageed (2019). «AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis.»، consulté le 11 septembre 2019. https://www.academia.edu/12105231/AWATIF_A_Multi_Genre_Corpus_for_Modern_Standard_Arabic_Subjectivity_and_Sentiment_Analysis . بتصرف .

(2) Muhammad Abdul-Mageed et Mona Diab (2014), «SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis», in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014) (LREC 2014, Reykjavik, Iceland: European Languages Resources Association (ELRA), 2014), http://www.lrec-conf.org/proceedings/lrec2014/pdf/919_Paper.pdf . بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

ينبغي هذا المستوى من التحليل على مقاربتين رئيسيتين: استخراج الجانب (Aspect Extraction) وهو تحديد الكيانات والجوانب التي تم تقييمها في الرأي، أما المقاربة الثانية، فتتمثل في تصنيف جانب المشاعر (Aspect Sentiment Classification) ويرتكز على مهمة تصنيف قطبية الجوانب، فحسب هاتين المقاربتين تم الكشف عن كل من الموضوع، والجانب، وقطبية المشاعر الواردة في التعليق أو التعليقات⁽¹⁾، ويعد محمد الارنواطي وسمير عبد الرحمان وعلي فهيمي، من بين الأوائل الذين اعتمدوا هذا المستوى من التحليل في بحثهم الذي ارتكز على استخراج أصحاب الرأي في الأخبار العربية التي تم تحميلها من موقع أسفله⁽²⁾، وبالاعتماد على معجم (MPQA) بعد ترجمته من اللغة الإنجليزية إلى اللغة العربية⁽³⁾.

4.3. مستوى المستخدم:

بالإضافة إلى المستويات السابقة من مستويات التحليل، ركزت بعض الدراسات على اعتماد مستوى المستخدم في عملية تحليل الآراء والمشاعر، وذلك بالنظر إلى شبكات المواقع التي يزورها متصفحوا الأنترنت أو تلك التي يحيل عليها ضمن مواقع التواصل الاجتماعي⁽⁴⁾.

(1) B. Liu. (2012) « Sentiment Analysis and Opinion Mining », Synth. Lect. Hum. Lang. Technol., vol. 5, no 1, p. 1-167 .بتصرف .

(2) <http://www.alzoa.com>

(3) Mohamed Elarnaoty, Samir AbdelRahman, et Aly Fahmy (6 avril 2012), « A Machine Learning Approach for Opinion Holder Extraction in Arabic Language », IJAIA International Journal of Artificial Intelligence & Applications, <https://doi.org/10.5121/ijaia.2012.3205> .بتصرف .

(4) Amitava Das, Sivaji Bandyopadhyay, et Björn Gambäck (2012), « Sentiment analysis: what is the end user's requirement? », in WIMS, <https://doi.org/10.1145/2254129.2254173> .بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

أصبح من السهل في عصرنا الحالي، الحصول على معلومات مرتبطة بالعلاقة التي يكوها المستخدم، لأن المحتوى الذي ينشئه المستخدم يظهر غالباً في سياق مواقع التواصل الاجتماعي، مثل Twitter الذي لا يحتفظ بالتغريدات التي يدونها المستخدم ويشاركها مع أصدقائه فحسب، بل كذلك يقوم بالاحتفاظ بالمتابعين لتلك التغريدات والمواقع التي يحيل عليها المستخدم، فمن خلالها يمكن معرفة طبيعته الذاتية وميولاته الموضوعية.⁽¹⁾

ومع ذلك فإن عدداً من الدراسات تستخدم نماذج مشتركة، تجمع بين مستويين أو ثلاثة مستويات مختلفة من مستويات التحليل السابقة الذكر، في عملية تحليلها للآراء والمشاعر؛ كدمجها بين مستوى المستخدم، ثم الانتقال إلى مستوى المستند، وتختتم بمستوى الجملة، أو الكيان.⁽²⁾

4. اللغة العربية وتحليل الآراء والمشاعر:

أشرنا سابقاً إلى أن تحليل الآراء والمشاعر، هو عملية تهدف إلى كشف الآراء الواردة في النصوص، وتصنيف المشاعر المعبر عنها في تلك النصوص، باعتماد مجموعة من الآلية والأدوات، المتمثلة أساساً في أدوات المعالجة الآلية للغات الطبيعية، وتقنيات التعليم الآلي، من أجل استخلاص نظرة عامة حول الآراء والمشاعر، والمواقف المعبر عنها من طرف الإنسان تجاه بعض الموضوعات في مجالات مختلفة، بالإضافة إلى الأحكام الصادرة، ثم تصنيفها حسب ما عبرت عنه إلى تعليقات موجبة أو سالبة أو محايدة⁽³⁾، كما أنه مجال يمكن من تحديد الكيانات التي تخضع للتقييم

(1) Chenhao Tan et al. (2011), « User-Level Sentiment Analysis Incorporating Social Networks » (ACM Press, 1397, <https://doi.org/10.1145/2020408.2020614>) . بتصرف.

(2) Ryan T. McDonald et al. (2007), « Structured Models for Fine-to-Coarse Sentiment Analysis », in ACL . بتصرف.

(3) R. T. Khasawneh et al. (2013), « Sentiment analysis of arabic social media content: a comparative study », in 8th International Conference for Internet Technology and

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

بواسطة التعليقات، وعادة ما يرتبط ذلك بالمجال الاقتصادي، المتعلق بالشركات التجارية التي تستفيد بشكل كبير من هذا النوع من النظم، بالإضافة إلى الجانب الخدماتي كالفنادق وغيرها⁽¹⁾،⁽²⁾.

أصبحت اللغات الهندوأوروبية متقدمة في مجال تحليل الآراء والمشاعر، نظراً لما حققته مجموعة من الأبحاث من نتائج ماهرة، فأضحى لزاماً على الباحثين المهتمين باللغة العربية، أن يوفرُوا الموارد اللسانية، وتبسيط الأدوات الحاسوبية، ومناهج التعليم الآلي الأساسية قصد تعبيد الطريق، وتسهيل سبل السير في دروب هذا المجال الشاسع، للتمكن من الوصول إلى نتائج جيدة بالنسبة للغة العربية وذلك نظراً لما أصبح يشكل مجال تحليل الآراء والمشاعر من أهمية في عدة مجالات، كما سبقت الإشارة لذلك، خاصة أن طبيعة التركيبات العربية تختلف عن نظيرتها في اللغات الهندوأوروبية⁽³⁾، كالتذكير، والتأنيث، والمثنى، والجمع، والاشتقاق، وتعدد المعاني، والوظائف الصرفية للكلمة الواحدة، بالإضافة إلى القواعد الصرفية والنحوية، زيادة

Secured Transactions (ICITST-2013),101-6,

<https://doi.org/10.1109/ICITST.2013.6750171> بتصرف.

(1) Bing Liu (23 mai 2012), « Sentiment Analysis and Opinion Mining », Synthesis Lectures on Human Language Technologies 5, no 1: 1-167, <https://doi.org/10.2200/S00416ED1V01Y201204HLT016> بتصرف.

(2) N. A. Abdulla et al.(2013), « Arabic sentiment analysis: Lexicon-based and corpus-based », in, IEEE 2013 Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 1-6, <https://doi.org/10.1109/AEECT.2013.6716448> بتصرف.

(3) H. Al-Rubaiee, R. Qiu, et D. Li (2016), « Identifying Mubasher software products through sentiment analysis of Arabic tweets », in 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), <https://doi.org/10.1109/ICCSII.2016.7462396> بتصرف.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

على ذلك بعض الاختصارات الموظفة في اللغة العربية إلى غيرها من أوجه الاختلافات بين اللغة العربية، ونظيرتها الهندوأوربية.⁽¹⁾

تعتبر البحوث المنجزة في اللغة العربية، ضمن مجال تحليل الآراء والمشاعر، محدودة مقارنة مع البحوث المنجزة في اللغات الهندوأوربية، ويرجع ذلك إلى جدة المجال في الأوساط العربية، ورغم ذلك فإن مجموعة من المحاولات تمكنت من إنارة الطريق، والتعريف بالعقبات والتحديات، والقضايا التي ما زالت تعترض المهتمين بهذا المجال.

بعد الاطلاع على مجموعة من الأبحاث العربية، استنتجنا أنها تركز على العامية لمختلف المناطق العربية، وقد يُرد ذلك إلى الهدف الربحي أكثر مما هو متعلق بخدمة اللغة العربية، وما يعزز ذلك هو كون العديد من النصوص موجهة إلى الجانب الاقتصادي بالدرجة الأولى، ووجهات أخرى ارتبطت بالجانب السياسي والفني والرياضي، باعتبارها مواضيع الساعة، مما جعل دائرة الكلمات المعتمدة، قليلة بالمقارنة مع ما تزخر به اللغة العربية، ونضيف إلى ذلك العزوف الشبه التام عن المقاربة اللسانية، رغم ما تمثله من مكانة في مجال المعالجة الآلية للغة العربية بصفة عامة، وتحليل الآراء والمشاعر على وجه الخصوص، ويعزز هذا التوجه بالإشارات التي وردت في العديد من الأبحاث، والتي تؤكد على ضرورة حضور الجانب اللساني، للتمكن من تقديم نتائج تشرف اللغة العربية في عملية تحليل الآراء والمشاعر.⁽²⁾

(1) K. Ahmad et Y. Almas (2005), « Visualising sentiments in financial texts? », in Ninth International Conference on Information Visualisation (IV'05), 363-68, <https://doi.org/10.1109/IV.2005.143> بتصرف.

(2) N. A. Abdulla et al.(2013), « Arabic sentiment analysis: Lexicon-based and corpus-based », in, IEEE 2013 Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 1-6, <https://doi.org/10.1109/AEECT.2013.6716448> بتصرف.

5. تحليل الآراء والمشاعر والمعالجة القبلية للغة العربية:

نظراً للتحديات التي سبقت الإشارة إليها، تعتبر المعالجة القبلية للغة العربية أمراً صعباً، في تحليل النص بصفة عامة، وتحليل الآراء والمشاعر على وجه الخصوص، وذلك راجع إلى استعمال نفس الخوارزميات التي تستعمل في المعالجة القبلية للغات الهندوأوروبية والتي لا تتوافق مع النص العربي، كحذف مجموعة من العناصر اللغوية التي من شأنها أن تكون عنصراً أساسياً في تحديد توجيه التعليقات العربية، وهنا يتجلى الدور الأساسي في الاستعانة باللغوي؛ ففي كثير من الأحيان قد يبدو للحاسوبي أن إزالة بعض العناصر اللغوية أثناء المعالجة القبلية أمر هين، خاصة في مجال تحليل الآراء والمشاعر، الذي يقتضي النظر في الجملة على أنها عنصر واحد؛ ومن الأمثلة على ذلك حذف كلمات الإيقاف (Stop Words)، التي تضم مجموعة من الحروف التي تؤثر في معاني الكلام العربي، أو ما يعرف في الأدبيات اللغوية العربية بحروف المعاني، وهذا أمر صعب نظراً للدور الكبير الذي تلعبه في توجيه معنى الجمل الداخلة عليها، وكما هو معروف أن اللغة العربية تتميز بغنى حروف المعاني، التي اهتم بها اللغويون العرب القدماء أشد الاهتمام، وأحصوها فاختلفوا في إحصائها حسب الوظائف التي وضعت لها، استناداً إلى ما توصلوا إليه من كلام العرب وأشعارهم، وكذا ما تضمنه القرآن الكريم من حروف متعددة المواضع والمعاني حسب سياقات الآيات التي وردت فيها.

تتضمن اللغة العربية مجموعة من القواعد النحوية من بينها التوابع، وضمن هذه التوابع توجد قاعدة التوكيد، التي توظف إذا أراد المتكلم/الكاتب تأكيد ما أشار إليه، ومعلوم في أدبيات قاعدة التوكيد أنه ينقسم إلى قسمين:

- توكيد معنوي: ويتجلى بإتباع ما ينبغي توكيده بإحدى الكلمات الآتية:

"نفس، عين، كل، جميع، كلا، كلتا";

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

- توكيد لفظي: ويتم بتكرار اللفظ المراد توكيده، وقد يكون التكرار لاسم، أو فعل، أو حرف، أو جملة.

ومعلوم بأن في المعالجة القبلية يتم حذف الكلمات المتكررة، إلا أن هذا الحذف قد يمس بمعنى التعليق المراد تحليله، وهذا الأمر قد يغيب عن الحاسوبي أثناء هندسته للنظم المعتمدة في المعالجة القبلية، خاصة في مجال تحليل الآراء والمشاعر. إن المطلع على التعليقات التي يتم نشرها في مواقع التواصل الاجتماعي وتطبيقاته، يتبين له أن هناك العديد من المتابعين يفضلون التعليق بالاكْتفاء بوضع رموز تعبر عن آرائهم ومشاعرهم تجاه القضية المعروضة، فكثير من الدراسات تستبعد هذه الرموز في عملية تحليلها للآراء.⁽¹⁾

6. تحديات استخدام تحليل الآراء والمشاعر في اللغة العربية:

لقد ازدادت الحاجة إلى الانخراط في مواقع التواصل الاجتماعي في الآونة الأخيرة، بفعل ما أصبح يقدمه من خدمات للإنسان، حتى أصبحت هذه المواقع والتطبيقات مادة خام للعديد من الحقول المعرفية أهمها، حقل المعالجة الآلية للغات الطبيعية، التي ولدت مجالاً معرفياً خصباً سعي بتحليل الآراء والمشاعر، امتاز بقدرته على تحليل كل النصوص المتداولة في العديد من المواقع، وبناء العديد من التطبيقات التي تمكن من الحصول على آراء المستخدمين، ضمن مجالات متنوعة نسردها بينها، الأعمال الإلكترونية، والتعليم الإلكتروني، والسياسة، والتجارة، والرياضة، ونقد الكتب، ... وذلك يفيد تلك المجالات في تحسين وتطوير كل ما تقدمه بناء على الآراء الموجهة لها.

(1) Hamed Saad AL-Rubaiee (2018) , «New Techniques And Framework For Sentiment Analysis And Tuning Of Crm Structure In The Context Of Arabic Language » (Thesis Doctora، s. d.), p: 31 بتصرف .

الفهرس الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

لقد أبانت مجموعة من الأبحاث جدارتها انطلاقاً من النتائج المتميزة، في حين أن اللغة العربية لا زالت تعاني من نقص الأبحاث وضعف النتائج، وذلك راجع إلى بعض التحديات التي لا يزال الباحث العربي يواجهها أثناء بحثه في هذا المجال، من بينها الوفرة المعجمية التي جعلت اللغة العربية من بين أغنى اللغات معجمياً، وكذا التعقيد اللغوي والتعدد التركيبي للعبارات. ومن بين العوامل التي تعيق المضي قدماً لبناء النظم من هذا النوع، عامل تغير شكل كتابة الحروف العربية بتغير موضعها في الكلمة وتميز اللغة العربية بخاصية الاشتقاق.

ومن بين التحديات كذلك التي تواجه عملية تحليل الآراء والمشاعر، قلة الموارد اللسانية، وتعدد مقاربات بنائها، والتي في غالبية الأمر تخل من المعنى اللغوي للكلام العربي، نقصد هنا مقارنة الترجمة الآلية التي اعتمدها العديد من الأبحاث في توفير بعض المعاجم للغة العربية، فيؤدي البعض من هذه المعاجم إلى ضعف نتائج عملية التحليل، أضف إلى ذلك إشكالية التعرف على نوع الآراء المقدمة في تلك المواقع وطبيعة العبارات التي يعبر بها الإنسان عن انفعالاته، خاصة عندما يستعين ببعض الألفاظ التي تولد شيئاً من الغموض في التعرف على قطبيتها، والتي عادة ما تكون في مجال انتقادات الكتب، أو المنتديات أو المجموعات التي أنشئت لتدارس ما يخص الأجناس الأدبية.⁽¹⁾

كما يشكل التفرع التخصصي، عائقاً كبيراً في المضي قدماً إلى بناء وتطوير هذا النوع من النظم، وغيرها مما يبني في مجال المعالجة الآلية للغة العربية، وإن أصبحنا نلمس في السنوات الأخيرة انخراط العديد من التخصصات العلمية، حيث

(1) R. Duwairi, M. Al-Refai, et N. Khasawneh (2007), « Stemming Versus Light Stemming as Feature Selection Techniques for Arabic Text Categorization », in 2007 Innovations in Information Technologies (IIT), 446-50, <https://doi.org/10.1109/IIT.2007.4430403> بتصرف.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

أصبحت مجموعة من مراكز البحث تركز على المتخصصين في اللغة باعتبارها المفتاح الأساسي في مثل هذه الدراسات، رغم خبرة بعض الحاسوبيين في المجال اللغوي نتيجة البحث المتزايد، ورغم كل ذلك فما زالت مراكز البحث تفتقر إلى ذوي التخصص اللغوي.

ومن جهة أخرى، قد أضى الاستعمال اليومي للعامة في المواقع الاجتماعية على حساب اللغة العربية الفصحى من بين أهم المعوقات التي تعرقل التقدم في بناء نظم تهتم بالمكتوب باللغة العربية، إذ جعل العامل الاقتصادي محفزاً في السير نحو العامة على حساب اللغة العربية، في توفير بعض النظم التي تحلل الآراء وتقييمها، كما يعيق الاستعمال المفرط للعامة، صعوبة بناء نظم موحدة نتيجة تنوع وتعدد عاميات الدول العربية.

علاوة على التحدي الكبير والمتمثل في توظيف الحروف اللاتينية في كتابة اللغة العربية، مما يدفع مجموعة من الباحثين إلى حذف كل ما كتب بغير الحروف العربية أثناء المعالجة القبلية، وإن كان بعض الباحثين سعوا إلى إحداث بعض النظم التي تحول التعليقات العربية المكتوبة بالحروف اللاتينية إلى حروف عربية، ومن بين هذه الأبحاث ما قام به الدويري رفقة زملائه، بتحليل الآراء والمشاعر المعبر بها حول المجال الاقتصادي انطلاقاً من موقع تويتر.⁽¹⁾

7. المقاربات المعتمدة في معالجة الآراء والمشاعر:

لقد أدى تعدد المجالات المرتبطة بمجال تحليل الآراء والمشاعر وكثرتها، إلى تنوع المقاربات المعتمدة في عملية التحليل، فبعد الاطلاع على مجموعة من الأبحاث، استنتجنا بأن هذه المقاربات قد لُخصت في ثلاث مقاربات أساسية، وهي المقاربة

(1) R. M. Duwairi et al. (2014), « Sentiment Analysis in Arabic tweets », in 2014 5th International Conference on Information and Communication Systems (ICICS). 2014. 1-6, <https://doi.org/10.1109/IACS.2014.6841964> بتصرف.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

اللسانية، ومقاربة تعليم الآلة والمقاربة الهجينة، رغم اختلاف المسميات التي سماها الباحثون في أبحاثهم، ومن بينهم ألكسندر باك (Alexander Pak)، الذي قسم المقاربات إلى مقاربتين؛ فجعل الأولى مقاربة تعتمد على المعجم، متمثلة في معجم المشاعر، ومورد وردنت (Wordnet)، بالإضافة إلى المعاجم التي يتم بناؤها آلياً. أما المقاربة الثانية، فهي المقاربة الإحصائية، وتحدث من خلالها على المصنفات واللوغاريتمات المعتمدة⁽¹⁾، أما حميد سعد الربيع فقد قسم هذه المقاربات إلى ثلاث مقاربات هي: مقاربة تعليم الآلة التي تعتمد على المدونة في عملية التدريب، وبعدها يتم الاستعانة باللوغاريتمات المعتمدة في التصنيف، أما المقاربة الثانية فتتمثل في مقاربة المعجم أو المقاربة اللسانية، وأما المقاربة الثالثة فتتجسد في المقاربة الهجينة التي تدمج بين المقاربتين السابقتين⁽²⁾، وفي ما يلي تعريف لأكثر المقاربات استعمالاً في مجال تحليل الآراء والمشاعر:

1.7. مقاربة تعليم الآلة: (Machine Learning Approach)

تعتبر مقاربة تعليم الآلة أحد أهم المقاربات في مجال الذكاء الاصطناعي، إذ توظف التقنيات الإحصائية لتمكين النظم الحاسوبية من التعلم، انطلاقاً من مجموعة من قواعد البيانات، وتعتبر مقاربة تعليم الآلة أكثر المقاربات استعمالاً في تصنيف الآراء والمشاعر، ويتفرع عن هذه المقاربة مجموعة من المقاربات الفرعية، نعرض أهمها:

- التعليم الموجه (Supervised Learning).

(1) Alexander Pak (2012), « Automatic, Adaptive, and Applicative Sentiment Analysis» (Thesis Doctora, PARIS, Université Paris Sud, p : 37-45, HAL (<https://tel.archives-ouvertes.fr/tel-00717329>)). بتصرف.

(2) Hamed Saad AL-Rubaiee (2018) , « New Techniques And Framework For Sentiment Analysis And Tuning Of Crm Structure In The Context Of Arabic Language » (Thesis Doctora, s. d.), p: 31. بتصرف.

- التعليم نصف الموجه (Semi-Supervised Learning).

- التعليم غير الموجه (Unsupervised Learning).

1. التعليم الموجه:

تعتمد مقارنة التعليم الموجه على معطيات موسومة يدوياً، ليتم بعد ذلك وسم معطيات أخرى آلياً، وكمثال على ذلك، إذا كان عندنا تصنيف مجالات في مكتبة معينة، يتم على إثر هذا التصنيف، تصنيف مستندات جديدة حسب التخصصات الموجودة في التصنيفات المسبقة، وقد تم اعتماد هذه المقاربة من طرف العديد من الباحثين من بينهم بينغ وآخرون⁽¹⁾ ووإنسون وفليت⁽²⁾.

2. التعليم نصف الموجه:

تعمل مقارنة التعليم نصف الموجه على الجمع بين التعليم الموجه وغير الموجه، أي الاعتماد على المعطيات الموسومة وغير الموسومة، لتحسين نوعية التصنيف ويمكن

(1) B. Pang, L. Lee, et S. Vaithyanathan, (2002) «Thumbs Up? Sentiment Classification Using Machine Learning Techniques», in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, Stroudsburg, PA, USA, p. 79–86.

(2) E. Eensoo et M. Valette, (2012) «Sur l'application de méthodes textométriques à la construction de critères de classification en analyse des sentiments », in TALN, vol. 2, p. 367–374.

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

الإحالة على بعض الأبحاث التي اعتمدت هذه المقاربة مثل بحث عبد المجيد وآخرون⁽¹⁾ وارتغوسا وآخرون⁽²⁾.

ج. التعليم غير الموجه:

تتجلى مقارنة التعليم غير الموجه في عدم اعتمادها على معطيات موجودة مسبقاً، عكس المقاربتين السابقتين

2.7. المقاربة اللسانية: (Linguistics Approach)

تعددت الأسماء لهذه المقاربة حسب ما يوظف من المستويات اللسانية، ومن بين هذه المسميات نجد "المقاربة الرمزية" أو "المقاربة المعجمية" أو "المقاربة اللسانية"، إلى غيرها من المسميات، كما تدخل نظرية الأبريسال ضمن المقاربات اللسانية كذلك، حيث تعتمد على معجم مصنف وفق حقول مفصلة، تساعد على تقديم نتائج دقيقة.⁽³⁾ وقد اعتمد العديد من الباحثين هذه المقاربة في معالجتهم للآراء والمشاعر الواردة في تعاليق مأخوذة من مواقع تفاعلية متعددة المجالات، من بينهم فيراري⁽⁴⁾ وكروول وآخرون⁽⁵⁾ وكردين وآخرون⁽¹⁾ وبلوم⁽²⁾ ووايثلاو وآخرون.⁽³⁾

(1) M. Abdul-Mageed et M. T. Diab, (2012) « AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. », in LREC, p. 3907–3914 .

(2) A. Ortigosa, J. M. Martín, et R. M. Carro, (2014) « Sentiment analysis in Facebook and its application to e-learning », Comput. Hum. Behav, vol. 31, p. 527-541 .

(3) J. R. Martin et P. R. R. White, (2005) "The language of evaluation: appraisal in English". New York : Palgrave Macmillan .

(4) S. Ferrari, Y. Mathet, T. Charnois, et D. Legallois, (2008) « Analyse d'opinion : discours évaluatif et classification de documents », Actes L'atelier FODOP, vol. 8, p. 23–36.

(5) J. Read et J. Carroll, (2012) « Annotating expressions of Appraisal in English », Lang. Resour. Eval., vol. 46, no 3, p. 421-447.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

تستثمر المقاربة اللسانية عن طريق توظيف مجموعة من المستويات اللسانية، أهمها المستوى الصرفي والمستوى المعجمي، والمستوى التركيبي. وفي هذا الصدد يمكن الإحالة إلى بعض الأعمال التي اعتمدت هذه المقاربة لتحليل الآراء والمشاعر مثل موريل⁽⁴⁾ والبحيري وآخرون⁽⁵⁾ وبرون وآخرون⁽⁶⁾.

3.7. المقاربة الهجينة: (Hybrid Approach)

وهي المقاربة التي تجمع بين المقاربة اللسانية ومقاربة تعليم الآلة؛ حيث تعمل على الجمع؛ إما بين اللسانيات باعتبارها مكملة لتعليم الآلة، أو تعليم الآلة باعتباره مكملًا لللسانيات أو التوفيق بينهما؛ فالطريقة الأولى تتضمن استعمال الأدوات اللسانية، لإعداد المدونة قبل تصنيف النصوص بمساعدة أدوات تعليم الآلة. تمكن هذه المقاربة

(1) P. Gardin, (2009) « Application de la théorie de l'Appraisal à l'analyse d'opinion », MajecSTIC Majecstic2009 Univ-Avignon Fr.Actes Majec STICRJCP.

(2) K. Bloom, (2011) « Sentiment analysis based on appraisal theory and functional local grammars », Illinois Institute of Technology.

(3) C. Whitelaw, N. Garg, et S. Argamon, (2005) « Using appraisal groups for sentiment analysis », in Proceedings of the 14th ACM international conference on Information and knowledge management, p. 625–631.

(4) S. Maurel, P. Curtoni, et L. Dini, (2007) « Classification d'opinions par méthodes symbolique, statistique et hybride », Actes Trois. DÉfi Fouille Textes, p. 121.

(5) S. Alhumoud, T. Albuhaire, et W. Alohaideb, (2015) « Hybrid Sentiment Analyser for Arabic Tweets using R », p. 417-424.

(6) Caroline Brun, Diana Nicoleta Popa, et Claude Roux, « Un Système Hybride Pour l'analyse de Sentiments Associés Aux Aspects », 2015, 228-34, <https://www.aclweb.org/anthology/2015.jeptalnrecital-court.34>.

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

من الاحتفاظ بقوة التعليم الآلي، وتوجيهه في نفس الوقت أساس التدريبات في الترتيب اليدوي المعتمد في المقاربة اللسانية.⁽¹⁾

8. بعض الأعمال السابقة:

بعد الاطلاع على مجموعة من الأبحاث المتعلقة بنظم تحليل الآراء والمشاعر، سواء الخاصة باللغة العربية الفصحى أو العامية العربية أو اللغات غير اللغة العربية، لاحظنا أنه يمكن تقسيم هذه الأبحاث إلى ثلاث مجموعات رئيسية؛ الأولى وضعت وفق مقاربة تعليم الآلة، والثانية وضعت وفق المقاربة اللسانية، أما الثالثة فوضعت وفق المقاربة الهجينة، وسنكتفي في هذا المقام بالإشارة إلى بعض الأبحاث العربية؛ سواء تعلق الأمر باللغة العربية الفصحى أو العامية.

1.8. أعمال وفق مقاربة تعليم الآلة:

قدم مجموعة من الباحثين، على رأسهم عبد العزيز محمد علبة بحثاً حول تحليل الآراء المتعلقة بمجال الخدمات الصحية، التي تم جمعها من Twitter، وقد ارتبطت تلك التعليقات بتحسين الجانب الصحي، ودواعي إغلاق المستشفيات وفتحها، وآراء أخرى حول الصحة بصفة عامة، وقد تم جمع ما يقارب 126959 تعليق، وبعد عملية التصفية تم تقليص العدد إلى 2026 تعليق، تم وسمها يدوياً من طرف ثلاثة باحثين، وبعدها تم اعتماد الوسم المشترك بين باحثين أو ثلاثة، فكان مجموع التعليقات بعد الوسم 628 تعليقا موجبا، 1398 تعليقا سالبا وبعد توظيف مصنفات تعليم الآلة والشبكات العصبية العميقة، تم الحصول على نتيجة قاربت 91%.⁽²⁾

(1) S. Maurel, P. Curtoni, et L. Dini (2007), « Classification d'opinions par méthodes symbolique, statistique et hybride », Actes Trois. Défi Fouille Textes, p. 121, 2007.

(2) A. M. Alayba et al. (2017), « Arabic language sentiment analysis on health services », in 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR) (2017 1st International Workshop on Arabic Script Analysis and

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

وقد أشرفت بسمة العسوطي رفقة زملائها على بحث اعتمدوا فيه مسافة ليفينشتين (Levenshtein Distance)، لمعالجة المشاعر المكتوبة بالعامية المغربية بلغت 902 تعليقا إيجابيا و588 تعليقا سلبيا، واستنادا لتمثيل حقيبة الكلمات ومسافة ليفينشتين بلغت دقة النظام هي 62٪⁽¹⁾.

ومن بين المجالات التي شملها هذا النوع من الأبحاث نجد مجال الصحة، إذ قام فريق من الباحثين ترأسهم أسماء بكوش، بتطوير أداة تساعد على تحديد توجيه تغريدات مكتوبة، وأخرى على شكل رموز مأخوذة من موقع التويتر، ذات صلة بالصحة لثلاث لغات وهي: اللغة العربية، والفرنسية، والإنجليزية، شملت تلك التغريدات، وقد اعتمدوا على التعليم العميق كمقاربة في إنجاز تلك الأداة.⁽²⁾

كما ركز كل من صدام العزاني والسيد محمد الألفي، على تصنيف المشاعر انطلاقا من الرموز (Emojis)، باستخدام الشبكات العصبية العميقة المتكررة، كمقاربة تساعد في بناء نظم تمكن من الكشف عن قطبية المشاعر للمدونات، وقد ارتكز عمل الباحثين على مدونة صغيرة ضمت 2091 تعليقا، مكون من خمس مدونات سابقة،

Recognition (ASAR), Nancy, France: IEEE, 2017), 114-18, <https://doi.org/10.1109/ASAR.2017.8067771> . بتصرف .

(1) B. Essatouti et al. (2018), « Arabic Sentiment Analysis Using a Levenshtein Distance Based Representation Approach », in 2018 IEEE 5th International Congress on Information Science and Technology (CiSt) (2018 IEEE 5th International Congress on Information Science and Technology (CiSt), Marrakech, Morocco: IEEE, 2018), 270-73, <https://doi.org/10.1109/CIST.2018.8596379> . بتصرف .

(2) A. Baccouche, B. Garcia-Zapirain, et A. Elmaghraby (2018), « Annotation Technique for Health-Related Tweets Sentiment Analysis », in 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, USA: IEEE, 2018), 382-87, <https://doi.org/10.1109/ISSPIT.2018.8642685> . بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

بالإضافة إلى 843 تعليق يشتمل على رموز أخذت من موقع اليوتيوب (YouTube) ووسمها يدويا، لذلك تم اعتماد معجم لرموز تترجم مجموعة من المشاعر والعواطف الإنسانية كالفرح والحزن... مصنفة حسب قطبيتها، وقد تم الحصول على نتائج جيدة باستخدام مجموعة من قواعد التعليم المُؤَلَّف⁽¹⁾.

وقد اشتغلت ثلة من الباحثين ترأسهم دنيا جمال، على إثبات نجاعة خوارزمية (N-Gram) في تصنيف الآراء والمشاعر، ومن أجل إثبات ذلك فقد قاموا بتطبيق بعض الخوارزميات، من بينها الخوارزمية السابقة في تصنيف مجموعة من التغريدات مختلفة العاميات المأخوذ من موقع تويتر صنفت آليا، حيث حوت ما يقارب 438931 تعليقا موجبا وسالبا، تم اختيارها بشكل متوازي لتدريب النظام 75774 تعليقا موجبا، و75774 تعليقا سالبا. وبعد القيام بالمعالجة القبلية للتعليقات وتطبيق الخوارزميات المحددة من طرف الباحثين، أكد مدى نجاعة كل من: (Unigram, Bigram, Trigram) حيث بلغت دقة أدائها 99.96%⁽²⁾.

وانطلاقا من خوارزمية (N-Gram)، عمل كل من بلقاسم ابراهيمي ومحمد طوهري وعبد الكامل طاري على تحسين نظام لتحليل الآراء، بتوظيف (Bigram) و (Skip-Ngram)، باعتبارهما خوارزميتين يقدمان نتائج جيدة ودقيقة، وقد ارتكز عمل الفريق على

(1) S. Al-Azani et E. El-Alfy (2018), « Emojis-Based Sentiment Classification of Arabic Microblogs Using Deep Recurrent Neural Networks », in 2018 International Conference on Computing Sciences and Engineering (ICCSE) (2018 International Conference on Computing Sciences and Engineering (ICCSE), Kuwait City, Kuwait: IEEE, 2018), 1-6, <https://doi.org/10.1109/ICCSE1.2018.8374211> بتصرف .

(2) Donia Gamal et al. (1 janvier 2019), « Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features », Procedia Computer Science, Proceedings of the 9th International Conference of Information and Communication Technology [ICICT-2019] Nanning, Guangxi, China January 11-13, 2019, 154: 332-40, <https://doi.org/10.1016/j.procs.2019.06.048> بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

مدونة التعليقات العربية (OCA)، فبعدها تم تلخيصها باعتماد طريقتين؛ ركزت الأولى منها على أقل التعليقات وضوحاً والتي اختزلت في أربع كلمات أو أقل منها، مثل "فيلم جيد" أو ببساطة "جيدة"، في حين ارتكزت الطريقة الثانية على استخراج التعليقات الصريحة المتعلقة بالكيان السينمائي، مركزين على الجمل التي ترد فيها مصطلحات تحيل على الأفلام مثل "الفيلم" و "السلسلة"، لقد حاول الباحثون تحليل الآراء الواردة في التعليقات المرتبطة بالأفلام مع التركيز على بعض الجوانب في العملية التقييمية وهي جمالية الأفلام، والهدف منها، والمتعة المحققة من خلالها، كما اعتمدوا في تصنيف تلك التعليقات على المعجم العربي للمشاعر (Nilulex) الذي تم بناؤه من طرف جامعة النيل، وبعد تطبيق الخوارزميات المقترحة تم الحصول على نتائج جيدة بلغت من الدقة 96٪.⁽¹⁾

بحث آخر ركز فيه مروان العمري رفقة زملائه على تحديد مؤشرات المشاعر الدالة على الترددي اللوجستيكي في لبنان، وقد انطلقوا من مجموعة من التعليقات التي تم أخذها من موقع (Google Mapp) لتحديد مجموعة من الأماكن المتعلقة بالخدمات، وهي " المقاطعات الخمس الرئيسية في لبنان: بيروت وجبل لبنان والنبطية وشمال لبنان وجنوب لبنان"، حيث بلغ مجموع التعليقات التي أخذت بفضل هذا الموقع 3500، تعليقا، كما تم الاعتماد على موقع (Zomato) والذي لا يتيح إلا أفضل خمس مراجعات لكل مزود خدمة، لذلك قام الفريق بجمع 416 تعليقا يدويا خلال فترة ممتدة بين 23 أكتوبر 2018 و22 نوفمبر 2018، ليكون مجموع تعليقات المدونة هو 3916 تعليقا حول مجموعة من الخدمات، همت الخدمات العامة، مثل الفنادق والمطاعم والمتاجر وغيرها من المجالات، وبعد القيام بمجموعة من المراحل المتعلقة بالمعالجة القبلية لتلك التعليقات، والتي بلغ عددها سبعة مراحل، وتطبيق

(1) Belgacem Brahimi, Mohamed Touahria, et Abdelkamel Tari (8 août 2019), « Improving sentiment analysis in Arabic: A combined approach », Journal of King Saud University - Computer and Information Sciences, <https://doi.org/10.1016/j.jksuci.2019.07.011> .بتصرف .

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

الخوارزميات المقترحة المتمثلة في (TF*IDF) و (Bag-Of-Words (Bow))، أظهرت النتائج ثلاث نقاط رئيسية؛ تمثلت الأولى في كون المصنف يمكن الوثوق بنتائجه في تحديد التعليقات الموجبة، أما النقطة الثانية فأشارت إلى انحياز النظام عند التنبؤ بالتعليقات ذات المشاعر السلبية، في حين شملت النقطة الأخيرة كون النظام يمكنه تحديد الترددي اللوجستيكي، انطلاقاً من النسبة المئوية المنخفضة للتعليقات السلبية (1).

ومن بين الأبحاث كذلك ما قامت به رؤى إسماعيل رفقة مجموعة من الباحثين، سعياً لتحليل تغريدات تويتر المكتوبة باللهجة العربية السودانية بلغت 4712 تغريدة، مأخوذة من تويتر ما بين سنة 2012 و 2017، توزعت تلك التعليقات بعد وسمها إلى 716 تعليق موجب، و3358 تعليق سالب، و638 موضوعي، وبعد القيام بالمعالجة القبلية وتوظيف بعض حروف المعاني المزدوجة بين اللغة العربية الفصحى والعامية، والتي بلغت 626 حرفاً، وبعد القيام بمجموعة من التدريبات التي ارتكزت على أربعة مصنفات (Naïve Bayes, SVM, KNN And Multinomial Logistics Regression) خلصوا إلى أن أحسن المصنفات دقة هو (KNN)⁽²⁾.

2.8. أعمال وفق المقاربة اللسانية:

(1) M. Al Omari et al. (2009), « Sentiment Classifier: Logistic Regression for Arabic Services, Reviews in Lebanon », in 2019 International Conference on Computer and Information Sciences (ICCIS) (2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, Saudi Arabia: IEEE, 2019), 1-5, <https://doi.org/10.1109/ICCISci.2019.8716394> بتصرف.

(2) R. Ismail et al. (2018), « Sentiment Analysis for Arabic Dialect Using Supervised Learning », in 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 2018, 1-6, <https://doi.org/10.1109/ICCCEEE.2018.8515862> بتصرف.

الفهر الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

يتطلب العمل وفق المقاربة اللسانية، التوفر على مجموعة من الموارد اللغوية والقواعد التركيبية، من أجل بلورة نظام خاص بتحليل الآراء والمشاعر، ومن بين الأبحاث العربية التي أنجزت وفق هذه المقاربة، نجد ما قدمته غيثة يمينة تليبي رفقة زملائها في دراستهم، على بناء معجم لم تتعدى مصطلحاته 325 أخذت من كتاب ووايث ومارتن ثم قاموا بترجمتها، وبالاعتماد على «Arabic Wordnet» تم تحديد خصائص المصطلحات المترجمة والمتمثلة في: القوة (Force) – التركيز (Focus) – القطب (Polarité) – الموقف (Attitude)⁽¹⁾. ومن بين الأبحاث كذلك ما قدمه أحمد عمار وعبد المنعم لشقر، اللذين اعتمدا في بناء المعجم الموظف في عملية التحليل خطوتين؛ تتمثل الأولى في ترجمة المصطلحات الواردة في كتاب مارتن ووايث ثم الاعتماد على Arabic Wordnet Et BabelNet لإغنائها، وذلك بالبحث عن مجموع مرادفات الكلمات التي تمت ترجمتها، ثم تصنيفها وفق المعايير المحددة في النظرية، فحصل على 553 مصطلح، أما الخطوة الثانية فتمثل في العمل على ترجمة المرادفات التي تم الحصول عليها من مصدرى Arabic Wordnet Et BabelNet فيبلغ عدد الكلمات المحصل عليها حوالي 1400 مصطلح.⁽²⁾

3.8. أعمال وفق المقاربة الهجينة:

(1) Guiassa Yamina Tlili (2014), « L'application de la théorie de l'Appraisal pour l'identification de l'opinion dans le texte arabe. », Int. Conf. Arab. Lang. Process., 2014, [Online]. Available: consulté le 4 avril 2016, http://www.citala.org/papers/paper_50.pdf .بتصرف

(2) Ahmed Amar et Abdelmonaime Lachkar (26 novembre 2015), « Vers la Construction d'un Lexique d'Appraisal en Langue Arabe pour l'Analyse des Sentiments dans les Réseaux Sociaux » (Institut d'Etudes et de Recherches pour l'Arabisation. Rabat) .بتصرف

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

تتأسس المقاربة الهجينة، على تكامل مقارنة تعليم الآلة والمقاربة اللسانية، وتعتبر المشاريع التي أنجزت وفق هذه المقاربة قليلة، نقدم من بينها ما قام به محمد الحربي بمعية بعض الباحثين، الذين ركزوا من خلاله على تحليل الآراء والمشاعر الواردة في تعليقات مدونة ضمت 5400 تغريدة تم تحميلها من موقع التويتر، تنوعت تلك التغريدات بين اللغة العربية الفصحى والعامية السعودية، وقد تم تصنيف هذه التعليقات عن طريق تحديد قطبيتها، وتتضمن (موجب، سالب، محايد، متفق، موضوعي، غير مهم، ليس مؤكداً)، بالإضافة إلى استخلاص العواطف الستة الأساسية التي جاء بها إكمن (Ekman)، والمتمثلة في (الغضب، الخوف، الاشمئزاز، الحزن، الفرح، المفاجأة، ليس بعاطفة، غير متأكد)، وقد بلغت دقة النظام بعد اعتماد خوارزميات التعليم الآلي نسبة 65%⁽¹⁾.

بحث آخر لتصنيف الآراء والمشاعر، ضمن مجالات مختلفة، اعتمد فيه فريق بحث مكون من محمد الصمادي، محمود الأيوب، ياسر الجرارة، وعمر القواسمة على مدونة (16Semeval ABSA)، تضم مجموعة من التعليقات حول مجموعة من المجالات من بينها: المطاعم وأجهزة الكمبيوتر المحمولة، والهواتف المحمولة، والكاميرات الرقمية والفنادق، والمتاحف، والاتصالات السلكية واللاسلكية، كما غطت هذه المدونة ثمان لغات وهي: العربية والإنجليزية والصينية والهولندية والفرنسية والروسية والإسبانية والتركية، وقد ركزوا في عملية التحليل استخراج الميزات الصرفية باستخدام محلل مداميرا (MADAMIRA)، بالإضافة إلى الميزات التركيبية والدلالية، المتمثلة في تحديد أصناف لكل الكلمات مثل: الجنس، والموقع والمنظم ثم تحديد أسماء الكيانات،

(1) A. Al-Thubaity et al. (2018), « A Saudi Dialect Twitter Corpus for Sentiment and Emotion Analysis », in 2018 21st Saudi Computer Society National Computer Conference (NCC) (2018 21st Saudi Computer Society National Computer Conference (NCC), Riyadh, Saudi Arabia: IEEE, 2018), 1-6, <https://doi.org/10.1109/NCG.2018.8592998> .بتصرف

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

وباعتماد مجموعة من خوارزميات تعليم الآلية، تمكن فريق البحث من الحصول على نتائج جيدة.⁽¹⁾

خلاصة:

لقد أصبح عصر الرقمنة التي يشهدها العالم، مجالاً يضع المؤسسات البحثية والجهات الحكومية أمام مجموعة من التحديات، عقب تطور شبكة الويب 3.0، أو ما اصطلح عليه بالشبكة الدلالية، ونجمل هذه التحديات بالأساس في طوفان المعلومات الذي يغمر المواقع الإلكترونية، ويأتيها من كل حدب وصوب، بالإضافة إلى كثرة الإشهارات الاقتصادية، والثقافية، والفنية المحفزة، التي أصبحت تغطي كل المواقع الإلكترونية، مما ساهم في اقتحام مجال معرفي حاسوبي جديد، عرف "بالبيانات الضخمة" "big data" التي تضم الملايين من النصوص، مما دفع تلك المؤسسات إلى التفكير في هيكلتها وتصنيفها، لغرض تطوير مجموعة من المجالات المتمثلة في المجال الثقافي، والفني، والاقتصادي، والسياسي والاجتماعي، وغيرها من المجالات، مما مكن مجموعة من المؤسسات البحثية من تطوير مجموعة من نظم تحليل الآراء والمشاعر، التي تعمل على تحليل وتقييم الآراء والمشاعر، انطلاقاً من التعليقات التي يتم نشرها في المواقع الإلكترونية، سواء تعلق الأمر بمواقع التواصل الاجتماعي، أو المواقع الثقافية، أو مواقع البيع والشراء.

(1) Mohammad Al-Smadi et al. (1 mars 2019), « Enhancing Aspect-Based Sentiment Analysis of Arabic Hotels, reviews using morphological, syntactic and semantic features », Information Processing & Management, Advance Arabic Natural Language Processing (ANLP) and its Applications, 56, no 2: 308-19, <https://doi.org/10.1016/j.ipm.2018.01.006> . بتصرف .

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

لقد أصبح مجال تحليل الآراء والمشاعر، من أحدث المجالات التي ينبغي السعي إلى تطوير آلياته بعدما أبانت مجموعة من الأبحاث التي تم ذكرها في هذا المبحث، مدى فعاليتها في تحليل التعليقات المكتوبة بعاميات مناطق متعددة، في حين أن النظم التي تحلل التعليقات المكتوبة باللغة العربية الفصحى تكاد تكون نادرة لقلة الموارد اللسانية التي تمكن من إنشاء نظم من هذا النوع، بالإضافة إلى إغفال العمل بالمقاربة اللسانية، كما أن توفير الموارد اللسانية باعتماد المقاربة اليدوية، تعد نادرة الاستعمال لما تتطلبه من وقت، رغم كفاءتها ودقتها العالية التي أبانت عنها في بعض الدراسات في هذا المجال.

المبحث الرابع: خطوات تحليل الآراء والمشاعر من خلال نظام (IERA)

لتحليل الآراء والمشاعر:

مدخل :

يعتبر نظام (IERA) من بين التجارب العربية التي انطلقت من الموروث العربي ومن النظريات اللسانية لبلورة نظام تحليل الآراء والمشاعر، وقد تجلت المساهمة في عمليتين كبيرتين، يتجلى الأولى في العمل اللساني الذي ارتكز على إعداد معجم مصنف حسب الحقول المعجمية المطابقة لنظرية الأبريسال، وحقول معجمية جديدة مستمدة

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

من معاجم عربية منتقاة لهذا الغرض¹، كما تم التركيز على إعداد لائحة حروف المعاني العربية لما لها من تأثير في تغيير توجيه الكلام العربي من الموجب إلى السالب، أو المحايد أو عكس ذلك²، بالإضافة إلى إنشاء مدونة عربية تتضمن تعليقات باللغة العربية الفصحى حول الكتب والأخبار الفنية، والثقافية وكذا الحوادث من أجل استعمالها في عملية اختبار نظم تحليل الآراء والمشاعر، بينما تمثل العمل الثاني في المجهود الحاسوبي، ففي إطار شراكة بحثية عقدها معهد الدراسات والأبحاث للتعريب مع فريق هندسة اللغة العربية ونمذجة التعليم بالمدرسة المحمدية للمهندسين لبلورة نظام تحليل الآراء والمشاعر استثمر فيه الموارد التي أنجزت في إطار العمل اللساني.

سنقدم في هذا المبحث وصفا للنظام الذي تمت بلورته من طرف معهد الدراسات والأبحاث للتعريب بالمغرب، حيث تم استثمار مجموعة من الموارد اللسانية المتمثلة في:

- معجم خاص بمجال تحليل الآراء والمشاعر اشتمل على 31373 دخلة معجمية؛
- لائحة جمع فيها مجموعة من الحروف والكلمات الوظيفية التي تؤثر في توجيه الكلام العربي؛
- مدونة عربية اشتمل على أزيد من 127116 تعليق، كُتِبَ معظمها باللغة العربية الفصحى تم تحميلها من موقع كودريدرز وموقع أبجد وموقع هيسبريس، وقد تمحور معظم تلك التعليقات حول الكُتُبِ والمؤلفين بمختلف موضوعاتهم، بالإضافة إلى بعض المجالات الأخرى، المتمثلة في الرياضة والفن والحوادث، كما قمنا بوسم ما يقارب 25 %

Vol. 80-81, pp. 335-365. 2019-2020. (1) B. Habibi, A. El Jihad " مجلة اللسان العربي "

(2) B. Habibi, A. El Jihad, "A complex Arabic negators-words list design" The Third National Doctoral Symposium On Arabic Language Engineering (JDILA'2018) FSBM, Hassan II University of Casablanca, 1st November 2018

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

أي ما يقارب 27441 تعليق بعد حذف التعليقات التي كتبت باللغات الأجنبية والمكتوبة بالعامية، وقد تم استثمارها في اختبار أداء النظام. لقد ارتكزت عملية تحليل الآراء والمشاعر ضمن هذا النظام على خمس خطوات وهي:

- عزل الاقتباسات.
- مرحلة المعالجة القبلية.
- مرحلة التجديع.
- مرحلة تحديد توجيه التعليق.
- تصنيف التعليقات.

1. عزل الاقتباسات:

تم العثور في المدونة التي اعتمدت في تدريب النظام على مجموعة من التعليقات التي يستند فيها مُبَدِي الرأي، إلى وضع اقتباس أو اقتباسات توضح حكمه في الرأي المعبر عنه، وقد يكون مضمون تلك الاقتباسات مختلفاً عن توجيهه ضمن التعليق؛ حيث قد يدل رأيه من خلال العبارات الموظفة على رأي موجب، وأما الاقتباس فيتضمن توجيهها سالبا، لهذا السبب تم عزل تلك الاقتباسات؛ خاصة أن صاحب الرأي يستعمل بعد الألفاظ، وعلامات التنصيص للإحالة على الاقتباسات.

تتنوع الاقتباسات بين القرآن الكريم، والأحاديث النبوية، وكذا الأبيات الشعرية، وأقوال العلماء والمفكرين، والكتاب؛ وتنبني عملية التعرف على الاقتباسات على بعض المؤشرات من قبيل:

- أقتبس "
- من بين الاقتباسات "
- ما أعجبي "
- ما لفت انتباهي قول الكاتب "
- مقتطفات أعجبتني "

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

إلى غيرها من العبارات التي وضعناها في لائحة يعتمدها النظام في إزالة التعليقات، كما ضمنا اللائحة العبارات التي تنتهي وفقها الاقتباسات من قبيل:

- "

- كانت تلك في الاقتباسات التي أثرت في

- كانت جملة الاقتباسات التي أعجبتني

ونمثل لذلك بالتعليق الذي تم تحليله بواسطة النظام ويتضمن اقتباساً

(الجدول 2):

الجدول 2: مثال تم تحليله بواسطة النظام

التعليق	مصدر التعليق	فئة التعليقات
إن تحفة الرافي تستحق القراءة لمرات عديدة، جميبيبييلة وشبيهه كما توقعت، انصح الجميع بقراءتها. مقتطفات أعجبتني من أوراق الورد: <u>أريدها لاتعرفني ولا أعرفها .. لا من شيء إلا انها تعرفني و أعرفها .. تتكلم ساكتة و أرد عليها بسكوتي .. صمت ضائع كالعبث لكن له في القلبين عمل كلام طويل .. " ان ساعة كتابتي إليك هي ساعة من الحياة معك وان كنا</u>	Table 0468	تعليق يتضمن الاقتباس

6. حذف كلمات الإيقاف؛

الجدول 3: التعليق بعد القيام بعملية المعالجة القبليّة

Token	#	Token	#	Token	#	Token	#	Token	#
مقطفا ت	13	انصح	10	جميلة	7	القراءة	4	إن	1
أعجبتني	14	الجميع	11	شبيهه	8	لمرات	5	تحفة	2
		بقراءتها	12	توقعت	9	عديدة	6	تستحق	3

3. مرحلة التجذيع (Stemming)

في هذه المرحلة تم اعتماد مجذع الخليل، وهو مجذع تم إنجازه من طرف فريق بحث المعالجة الآلية للغة العربية بجامعة محمد الأول بوجدة، حيث كان الهدف من هذه العملية هو الحصول على جذع كلمات التعليق، فمن خلال التعليق السابق فإن عملية التجذيع تمكننا من الحصول على جذع الكلمات التي تتركب منها التعليق؛ على الشكل الآتي (الجدول 4)

الجدول 4: جذوع الكلمات التي تتركب منها التعليق

Type	Stem	Word	#	Type	Stem	Word	#
ROOT	شهي	شبهية	8	—	إن	إن	1
ROOT	وقع	توقعت	9	ROOT	تحف	تحفة	2
ROOT	نصح	انصح	10	ROOT	سحق	تستحق	3
ROOT	جمع	الجميع	11	ROOT	قرأ	القراءة	4
ROOT	قرأ	بقراءتها	12	ROOT	مرر	لمرات	5
ROOT	قطف	مقطفات	13	ROOT	عدد	عديدة	6

7	جميلة	جمل	ROOT	14	أعجبتني	عجب	ROOT
---	-------	-----	------	----	---------	-----	------

4. مرحلة تحديد توجيه الكلمات المشكلة للتعليق:

تعد هذه المرحلة من أهم مراحل النظام؛ فبعد القيام بعزل الاقتباسات، والقيام بالمعالجة القبليّة للتعليقات، والحصول على جذوع كل كلمة رُكِبَ منها التعليق، يبدأ النظام بالبحث عن الجذوع من أجل بلوغ المداخل المعجمية، التي تتركب منها التعليق وتحديد معلوماتها التي ضُمنت في المعجم؛ فإذا كانت موجودة يتم تحديد معلوماتها المتعلقة بها؛ والمتمثلة في: الحقل المعجمي، التوجيه، والقوة، وذلك من أجل التمكن من تحديد توجيه التعليق توجيهها دقيقاً، وتجدر الإشارة إلى أنه خلال هذه المرحلة، يعمل النظام على تحديد الكلمات التي لم يعثر عليها في المعجم، من أجل النظر فيما إذا كانت كلمة أساسية يجب إضافتها للمعجم أم لا، ونمثل لهذه العملية بالجدول الموالي (5):

الجدول 5: تحديد الحقول المعجمية لكل مدخل معجمي

#	Word	Stem	Found/ Not Found	Lexical Field	Orientation	Force
1	تحفة	تحف	Found	النوع	Positive	High
2	تستحق	سحق	Found	الاهتمام	Positive	Median
3	القراءة	قرأ	Found	التقييم	Neutre	Median

—	—	—	Not Found	مرر	لمرات	4
—	—	—	Not Found	عدد	عديدة	5
High	Positive	التقييم	Found	جمل	جميلة	6
Median	Positive	التقييم	Found	شهبي	شهيه	7
Median	Neutre	الشك واليقين	Found	وقع	توقعت	8
Median	Positive	الاهتمام	Found	نصح	انصح	9
—	—	—	Not Found	جمع	الجميع	10
Median	Neutre	التقييم	Found	قرأ	بقراءتها	11
Median	Neutre	التقييم	Not Found	قطف	مقتطفا ت	12
Median	Positive	المتعة	Found	عجب	أعجبتي	13

بعد تحديد المعلومات المتعلقة بكل مدخل معجمي، يتم النظر فيما إذا اشتمل التعليق على حروف المعاني أو الكلمات الوظيفية التي وضعناها في لائحة حروف المعاني، التي تؤثر في الكلام العربي أم لا؛ وإذا نظرنا إلى هذا التعليق، نجده يتضمن حرف "إن" المصنف ضمن الحروف التي تزيد من قوة التعليق، بالإضافة إلى الكلمات الوظيفية "مرات - عديدة" التي ساهمت كذلك في زيادة القوة في هذا التعليق (الجدول6).

الجدول 6: تحديد المعلومات المتعلقة بالحروف التي تؤثر في توجيه الكلام العربي

الوظيفة	الكلمة
الزيادة من قوة التعليق	إن
الزيادة من قوة التعليق	مرات
الزيادة من قوة التعليق	عديدة

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

وفي الأخير يتم جمع توجيه جمل التعليق ككل، فإن تعددت فيها التوجيهات، في هذه الحالة يتم تحديد التوجيه حسب أغلبية توجيه الكلمات الواردة في التعليق؛ فإذا كانت أغلبية الكلمات ذات توجيه موجب سيكون توجيه التعليق موجبا، فإذا كانت أغلبية الكلمات ذات توجيه سالب سيكون توجيه التعليق سالبا، فإذا كانت أغلبية الكلمات ذات توجيه محايد سيكون توجيه التعليق محايدا.

5. تصنيف التعليقات:

من خلال التعليق السابق، فإن النظام، قد صنف التعليق ضمن صنف الموجب القوي؛ فعلى الرغم من أن قوة المداخل المعجمية أغلبها متوسطة، إلا أن حرف التوكيد "إن" وكذلك "مرات - عديدة" زادت من قوة التعليق، ومن خلال ذلك يتضح الدور الذي تؤديه حروف المعاني والكلمات الوظيفية في التعابير العربية.

الخاتمة:

لقد حظيت اللغة العربية باهتمام كبير من لدن اللغويين قديماً ومحدثين، حيث تمكنوا من الحفاظ عليها كمورث حضاري يجسد الهوية والثقافة العربيتين، فبعدما تمكن القدماء من التعقيد لها، وحفظ موروثها المعجمي وتعقيد العديد من العلوم، فقد استطاع المحدثون من سبر أغوار العلوم العربية بمنطق العلمية التي استمدوها من العلوم الحديثة أساسها اللسانيات، باعتبارها العلم الذي يدرس اللغة دراسة علمية، كما استطاعت اللسانيات الانفتاح على العديد من العلوم المعرفية؛ من بينها الفلسفة وعلم النفس وعلم الحاسوب وغيرها من العلوم.

لقد ظهور الحاسوب وانبثاق علم قائم به عاملاً مهماً في التطورات الحاسوبية المتسارعة التي أصبحنا نشهدها في السنوات الأخيرة مما نتج عنه ظهور حقول معرفية جديدة كرس فيها الباحثون جهودهم في خدمة اللغات الطبيعية؛ فنتج عن ذلك انبثاق كل من المعالجة الآلية والفهم الآلي، وكذا التوليد الآلي للغات الطبيعية، بما فيها اللغة العربية، وقد تزامن كل ذلك مع تطوير تقنيات الويب الرامية إلى خلق علاقة تفاعلية بين القائمين على المواقع وروادها، وهذا ما ساهم في انبثاق دينامية معلوماتية، اهتمت بالعديد من المجالات التي تتطلب فهماً سريعاً ودقيقاً للغة، وقد كان مجال تحليل الآراء والمشاعر أحد أهم تلك المجالات لما له من دور في تطوير محتوى المواقع الإلكترونية، ذات المحتوى الثقافي، والأخباري، والسياسي، والتجاري، إلى غيرها من أنواع المحتويات، كما يساعد منتج المحتوى على تطوير ما ينتجه ويعرضه من خلال تدارك ما أشار إليه الرواد ضمن تعليقاتهم.

وعلى هذا الأساس، حاولنا في هذا الكتاب الإشارة إلى أهم ما تتميز به اللغة العربية عن باقي اللغات، كما أشرنا إلى الجهود البحثية العربية التي استطاعت جعل اللغة العربية لاحقاً باللغات الطبيعية الأخرى على مستوى نمذجة المستويات اللسانية، كما وضحت أن المستويات اللسانية العربية يصعب عزل مستوى عن آخر،

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

وهذا ما ساهم في تطوير الدراسات الحاسوبية المتعلقة بنمذجة المستويات اللسانية على مستوى الحاسوب، فتم التمكن من بلورة العديد من النظم الحاسوبية التي تساعد على معالجة اللغة العربية آلياً، كما أشرنا إلى بعض من تلك النظم؛ كنظم التحليل الصرفي، وتلك التي تهتم بالمعاجم الالكترونية، وأخرى وظيفتها تولي عملية التشكيل الآلي للنصوص وغيرها من الأنظمة المعتمدة في تحليل أو فهم اللغة العربية.

كما تحدثنا بعد ذلك عن الفهم الآلي للغات الطبيعية باعتباره حقلاً معرفياً جاء للمساعدة على بلورة نظم تساعد على بلورة نظم آلية تساعد على خلق التفاعل بين الإنسان والآلة، في إطار الفهم الآلي للنصوص العربية، وقد توفقنا في العديد من النظم التي تندرج ضمن هذا الحقل المعرفي؛ حيث قدمنا البعض منها في الفصل الثاني من هذا الكتاب، كما أفردنا الحديث عن حقل تحليل الآراء والمشاعر العربية، وملامسة بعض قضاياها، المتمثلة أساساً في الإشارة إلى الموارد اللسانية، والمقاربات المتنوعة التي يحتاجها كل مهتم بهذا المجال، كما وقوفنا عند بعض التجارب العربية، كما تمت الإشارة إلى بعض العوائق التي تعترض سبيل البحث في هذا المجال أهمها قلة الموارد اللسانية الضرورية التي تساعد في تحقيق نتائج جيدة على مستوى تدريب واختبار النظم الحاسوبية، وختمنا هذا الكتاب بالإشارة إلى أهم الخطوات التي تقطعها نظم تحليل الآراء والمشاعر من خلال عرض تجربة نظام (IERA) التي تمت بلورته من طرف معهد الدراسات والأبحاث للتعريب بشراكة مع المدرسة المحمدية للمهندسين بالمغرب، وقد كانت تجربة رائدة، ميزها الانطلاق من الموروث العربي والاستغناء عن المورد الغربية التي تهافت عليها مجموعة من الباحثين، كما سبق الإشارة إلى ذلك.

خلاصة القول، تؤدي الموارد اللسانية دوراً كبيراً في تحسين دقة النظم الآلية المتعلقة بمعالجة اللغات الطبيعية، بما فيها اللغة العربية، سواء تعلق بالنظم بالمعالجة الآلية أو الفهم الآلي، أو التوليد الآلي، وهي نظم معالجة أصبحت متاحة

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

بفضل تطوير مقاربات الذكاء الاصطناعي، ورغم ما تم التوصل إليه في الآونة الأخيرة من تزايد النظم المعلوماتية القيمة خاصة مع ثورة الروبوتات التي أبانت عن وجود تجارب رائدة في توليد اللغات الطبيعية إلا أن اللغة العربية لا زالت في حاجة ماسة لإغناء مكتبتها بالموارد اللسانية الضرورية لبقائها وتقوية حظوظ تنافسيتها بين باقي لغات العالم، ومهمة جسيمة كهذه، لن تجد من يتولاها بالأمانة المطلوبة والمسؤولية الكبيرة غير مجموعات وفرق المختصين في مجال اللسانيات؛ لما يحوزونه من دراية واسعة باللغة العربية وبمختلف علومها، ولعل من أبرز هذه المهام التي يمكن اقتراحها لإنجاز بحوث مستقبلية ما يلي:

- الاهتمام الدقيق بتوفير الموارد اللسانية وتوسيمها دلاليا وتركيبيا؛
- استثمار الأدوات الحاسوبية والمنصات العربية التي تتبنى الأعمال اللسانية حفظاً للخصوصية؛
- العمل المتواصل بين الفرق اللسانية والحاسوبية من أجل حل كل الإشكالات التي تعترض تقدم بلورة النظم الحاسوبية الخادمة للغة العربية؛
- الاهتمام بالجانب الدلالي، باعتبار المستوى الأهم الذي تركز على معظم الأبحاث الحالية؛

المصادر والمراجع

2. المصادر باللغة العربية

- أبو العزم، عبد الغني، السعدية أيت الطالب، وفوزية بن جلون. « المعجم الحاسوبي العربي التصور، والمنهجية ». في الدراسات المعجمية. العدد السابع والثامن. الرباط - المغرب: منشورات الجمعية المغربية للدراسات المعجمية، 2009.
- أحمد روبي، محمد. البنك الشجري النحوي. الأولى. المملكة العربية السعودية - الرياض: مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، 1438.
- الجهاد، عبد الحميد، عبد الفتاح حمداني، توفيق اليزيدي، و سعيد الحسني. « محلل صرفي للأفعال العربية »، 1-3. الرباط - المغرب: المدرسة المحمدية للمهندسين ومعهد الدراسات والأبحاث للتعريب، 2009.
- الحناش، محمد. « مصطلحات هندسة اللغة العربية بين التعريب والتأثيل ». في دورية اللسان العربي، 195-230. الرباط: المنظمة العربية للتربية والثقافة والعلوم ومكتب تنسيق التعريب، 2018.
- الخولي، فاتن. « المعجم بين اللغة والحاسوب بناء المعجم الإسلامي من منظور لساني حاسوبي ». في المعجمية العربية قضايا وآفاق، تحرير حافظ إسماعيلي، محمد الملاح، محمد إسماعيلي، و أمين منتصر، الطبعة الأولى، 98-177. سلسلة المعرفة اللسانية، الجزء الثاني. عمان - الاردن: دار كنوز المعرفة، 2014.
- الزجاجي، عبد الرحمن بن إسحاق/أبي القاسم. الإيضاح في علل النحو Dar Al Kotob Al Ilmiyah دار الكتب العلمية، 2012.
- السبومي، حسن محمد. في المعجم الحاسوبي للغة العربية وأبعاده التعليمية. الرياض، 2009.
- الغامدي، محمد سعيد ربيع. نظام الكتابة العربية: النشوء والتطور. الأولى. مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض: دار وجوه للنشر والتوزيع، s. d.

الفهرس الآلي للغة العربية، تحليل الأراء والمشاعر نموذجاً

المهديوي، عمر. « المعجم الإلكتروني: قراءة في بعض التجارب الغربية والعربية ». في الدراسات المعجمية. العدد السابع والثامن. الرباط - المغرب: منشورات الجمعية المغربية للدراسات المعجمية، 2009.

النحوي، الشيخ كمال الدين الانباري. اسرار العربية. دار الارقم بن ابي الارقم - بيروت / لبنان، 2016.

اليوبي، بلقاسم. « اللسانيات الحاسوبية: مفومها و تطورها و مجال تطبيقاتها ». في مجلة مكناسة. عدد 12. مكناس: كلية الآداب والعلوم الانسانية، 1999.

بن يشو، جيلالي. « حوسبة المعجم العربي الواقع والآفاق ». شبكة الصورة العربي، 19 août 2011. <https://www.voiceofarabic.net/ar/articles/2662>

consulter le 03/07/2019 a 23:43.

بوزويغ، عبد الكريم، و عبد الفتاح حمداني. « المعاجم الرقمية العربية: من المصطلح إلى الدلالة ». المنظمة العربية للتربية والثقافة والعلوم - مكتب تنسيق العرب، مجلة اللسان العربي، 76 ع (2016): 79-93

جمعة، عمرو. تقنيات اللغة العربية الحاسوبية معايير التقييم ورؤى التطوير: دراسة لغوية حاسوبية. الأولى. مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض: دار وجوه للنشر والتوزيع، 2016.

حمداني، عبد الفتاح. « المعاجم العربية الآلية جرد وتقييم ». في المعجم العربي العصري وإشكالاته، تحرير أحمد بريسول و كنزة بنعمر، بدون طبعة. جامعة محمد الخامس - السويسي - الرباط: معهد الدراسات والأبحاث للتعريب، 2016.

دكيكي، حبد الواحد. « التخطيط المعجمي للجملة في العربية، السمات والتوزيع في التعابير المسكوكة نموذجاً ». في البحث المعجمي ورهانات تأهيل اللغة العربية قضايا في النظرية والتخطيط والهندسة، تحرير عبد الواحد دكيكي، محمد السهول، محمد وحيد، و نور الدين بوعمالي، طنعة جديدة ومنقحة. 380-407. وقائع الندوة العلمية الدولية الأولى للسانيات الحاسوبية والمعجم. الكلية المتعددة التخصصات بالرشيدية - جامعة مولاي إسماعيل المغرب:

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

فريق البحث هندسة اللغات الطبيعية وتكنولوجيا الحوسبة GELINFO،
2015.

روبي محمد عبد الرحمان، أحمد. « التحليل النحوي ». في مدخل إلى اللسانيات
الحاسوبية، تحرير عبد الله بن يحيى الفيقي، الأولى، 75-100. مباحث لغوية
30. المملكة العربية السعودية - الرياض: مركز الملك عبدالله بن عبدالعزيز
الدولي لخدمة اللغة العربية، 1438.

شندول، محمد. « القاموس الإلكتروني: قاموس Le petit Larousse
Multimédia » الإلكتروني نموذجاً. في الدراسات المعجمية. العدد السابع
والثامن. الرباط - المغرب: منشورات الجمعية المغربية للدراسات المعجمية،
2009.

علي أحمد الرفاعي، إشراق. « التحليل الدلالي ». في مدخل إلى اللسانيات الحاسوبية،
édité par أحمد روبي محمد عبد الرحمان، الطبعة الأولى، 107-25. مباحث
لغوية 30. المملكة العربية السعودية - الرياض: مركز الملك عبدالله بن
عبدالعزیز الدولي لخدمة اللغة العربية، 2017.

علي، نبيل. العرب وعصر المعلومات. عالم المعرفة 184. الكويت: المجلس الوطني
للثقافة والفنون والأدب مطابع السياسة الكويتية، 1994.

أسامة الخولي، اللغة العربية والحاسوب دراسة بحثية. دار الفكر العربي، 1988.
علي، نبيل، و نادية حجازي. الفجوة الرقمية. بدون طبعة. عالم المعرفة 318. مطابع
السياسة - الكويت: المجلس الوطني للثقافة والفنون والأدب، 2005.

فتيحة محمد الدبابسة، و هاني صبري البطاط. « نهاد الموسى وجهوده اللغوية ».
المجستير، جامعة الخليل، J.-C211.apr.

فريق المعالجة الآلية للغة العربية. « التحليل الصرفي للأسماء العربية ». تحرير عبد
الفتاح حمداني، 51-63. الرباط - المغرب: معهد الدراسات والأبحاث للتعريب،
2007.

الفهم الآلي للغة العربية تحليل الأسماء والمشاعر نموذجاً

قدراوي، أمين. « نحو بناء معجم إلكتروني للمعالجة الآلية للغة العربية مفردات ألفاظ القرآن الكريم أنموذجاً مقارنة لسانية حاسوبية ». بحث لنيل درجة الماجستير، جامعة أبي بكر بلقايد - تلمسان، 2009.

محمد، الهيلالي. « اللغة العربية والحاسوب ». بحث لنيل شهادة الإجازة، جامعة مولاي إسماعيل، 1998.

مهديوي، عمر، و سلوى السيد حمادة. « المعالجة الدلالية الآلية للغة العربية: نحو بناء قاعدة بيانات معجمية للعلاقات الدلالية بين الكلمات ». مجلة فكر ونقد، 2008.

3. المصادر باللغات الأجنبية

- Abbasi, Ahmed, Hsinchun Chen, et Arab Salem. « Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums ». *ACM Trans. Inf. Syst.* 26 (2008): 12-12. <https://doi.org/10.1145/1361684.1361685>.
- Abdulla, N. A., N. A. Ahmed, M. A. Shehab, et M. Al-Ayyoub. « Arabic sentiment analysis: Lexicon-based and corpus-based ». In *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 1-6, 2013. <https://doi.org/10.1109/AEECT.2013.6716448>.
- Abdul-Mageed, Muhammad. « AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. » Consulté le 11 septembre 2019. https://www.academia.edu/12105231/AWATIF_A_Multi-Genre_Corpus_for_Modern_Standard_Arabic_Subjectivity_and_Sentiment_Analysis.
- Abdul-Mageed, Muhammad, et Mona Diab. « SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis ». In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. Reykjavik, Iceland: European Languages Resources Association (ELRA), 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/919_Paper.pdf.
- Acikara, Turgut, Bo Xia, Tan Yigitcanlar, et Carol Hon. « Contribution of Social Media Analytics to Disaster Response Effectiveness: A Systematic Review of the Literature ». *Sustainability* 15, n° 11 (janvier 2023): 8860. <https://doi.org/10.3390/su15118860>.

- Ahmad, K., et Y. Almas. « Visualising sentiments in financial texts? » In *Ninth International Conference on Information Visualisation (IV'05)*, 363-68, 2005. <https://doi.org/10.1109/IV.2005.143>.
- Ahmed Amar et Abdelmonaime Lachkar. « Vers la Construction d'un Lexique d'Appraisal en Langue Arabe pour l'Analyse des Sentiments dans les Réseaux Sociaux ». Institut d'Etudes et de Recherches pour l'Arabisation, Rabat, 26 novembre 2015.
- Alawami, Alawya Adnan. « Aspect Extraction For Sentiment Analysis In Arabic Dialect ». Thesis Doctora, University of Pittsburgh, 2017.
- Alayba, A. M., V. Palade, M. England, et R. Iqbal. « Arabic language sentiment analysis on health services ». In *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, 114-18. Nancy, France: IEEE, 2017. <https://doi.org/10.1109/ASAR.2017.8067771>.
- Al-Azani, S., et E. El-Alfy. « Emojis-Based Sentiment Classification of Arabic Microblogs Using Deep Recurrent Neural Networks ». In *2018 International Conference on Computing Sciences and Engineering (ICCSE)*, 1-6. Kuwait City, Kuwait: IEEE, 2018. <https://doi.org/10.1109/ICCSE1.2018.8374211>.
- AL-Khassawneh, Yazan Alaya, et Essam Said Hanandeh. « Extractive Arabic Text Summarization-Graph-Based Approach ». *Electronics* 12, n° 2 (janvier 2023): 437. <https://doi.org/10.3390/electronics12020437>.
- Al-Rubaiee, H., R. Qiu, et D. Li. « Identifying Mubasher software products through sentiment analysis of Arabic tweets ». In *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*, 1-6, 2016. <https://doi.org/10.1109/ICCSII.2016.7462396>.

- AL-Rubaiee, Hamed Saad. « New Techniques And Framework For Sentiment Analysis And Tuning Of Crm Structure In The Context Of Arabic Language », s. d.
- Al-Smadi, Mohammad, Mahmoud Al-Ayyoub, Yaser Jararweh, et Omar Qawasmeh. « Enhancing Aspect-Based Sentiment Analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features ». *Information Processing & Management, Advance Arabic Natural Language Processing (ANLP) and its Applications*, 56, n° 2 (1 mars 2019): 308-19. <https://doi.org/10.1016/j.ipm.2018.01.006>.
- Al-Thubaity, A., M. Alharbi, S. Alqahtani, et A. Aljandal. « A Saudi Dialect Twitter Corpus for Sentiment and Emotion Analysis ». In *2018 21st Saudi Computer Society National Computer Conference (NCC)*, 1-6. Riyadh, Saudi Arabia: IEEE, 2018. <https://doi.org/10.1109/NCG.2018.8592998>.
- Baccouche, A., B. Garcia-Zapirain, et A. Elmaghraby. « Annotation Technique for Health-Related Tweets Sentiment Analysis ». In *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 382-87. Louisville, KY, USA, USA: IEEE, 2018. <https://doi.org/10.1109/ISSPIT.2018.8642685>.
- Barr, Avron, et Edward A. Feigenbaum, éd. « Chapter IV - Understanding Natural Language ». In *The Handbook of Artificial Intelligence*, 223-321. Butterworth-Heinemann, 1981. <https://doi.org/10.1016/B978-0-86576-089-9.50009-0>.
- Ben, Manel, Abdallah Chafik, Aloulou Lamia, et Lamia Belguith. « TOWARD A PLATFORM FOR ARABIC AUTOMATIC SUMMARIZATION », 3 juillet 2023.

- Brahimi, Belgacem, Mohamed Touahria, et Abdelkamel Tari. « Improving sentiment analysis in Arabic: A combined approach ». *Journal of King Saud University - Computer and Information Sciences*, 8 août 2019. <https://doi.org/10.1016/j.jksuci.2019.07.011>.
- Brun, Caroline, Diana Nicoleta Popa, et Claude Roux. « Un Système Hybride Pour l'analyse de Sentiments Associés Aux Aspects », 228-34, 2015. <https://www.aclweb.org/anthology/2015.jeptalnrecital-court.34>.
- Chlasta, Karol. « Implementation of sentiment analysis model for Twitter data in Polish language », 2015. <https://doi.org/10.13140/RG.2.2.12715.69924>.
- Das, Amitava, Sivaji Bandyopadhyay, et Björn Gambäck. « Sentiment analysis: what is the end user's requirement? » In *WIMS*, 2012. <https://doi.org/10.1145/2254129.2254173>.
- Das, Sanjiv R., et Mike Y. Chen. « Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web ». *Management Science* 53, n° 9 (2007): 1375-88.
- Dave, Kushal, Steve Lawrence, et David M. Pennock. « Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews ». In *Proceedings of the 12th International Conference on World Wide Web*, 519-28. WWW '03. New York, NY, USA: ACM, 2003. <https://doi.org/10.1145/775152.775226>.
- Duwairi, R., M. Al-Refai, et N. Khasawneh. « Stemming Versus Light Stemming as Feature Selection Techniques for Arabic Text Categorization ». In *2007 Innovations in Information Technologies (IIT)*, 446-50, 2007. <https://doi.org/10.1109/IIT.2007.4430403>.

- Duwairi, R. M., R. Marji, N. Sha'ban, et S. Rushaidat. « Sentiment Analysis in Arabic tweets ». In *2014 5th International Conference on Information and Communication Systems (ICICS)*, 1-6, 2014. <https://doi.org/10.1109/IACS.2014.6841964>.
- Dziczkowski, Grzegorz. « Analyse des sentiments: système autonome d'exploration des opinions exprimées dans les critiques cinématographiques ». THÈSE DE DOCTORAT, L'ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS, 2008.
- El Jihad, Abdelhamid. « نظام تصريف الأفعال في اللغة العربية ». *Revue de l'Information Scientifique et Technique* 20, n° 2 (2013): 27-48.
- Elarnaoty, Mohamed, Samir AbdelRahman, et Aly Fahmy. « A Machine Learning Approach for Opinion Holder Extraction in Arabic Language ». *IJAIA International Journal of Artificial Intelligence & Applications*, 6 avril 2012. <https://doi.org/10.5121/ijaia.2012.3205>.
- Essatouti, B., H. Khamar, S. E. Fkihi, R. Faizi, et R. O. H. Thami. « Arabic Sentiment Analysis Using a Levenshtein Distance Based Representation Approach ». In *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 270-73. Marrakech, Morocco: IEEE, 2018. <https://doi.org/10.1109/CIST.2018.8596379>.
- Fellbaum, Christiane. « La représentation des verbes dans le réseau sémantique WordNet ». *Langages* 33, n° 136 (1999): 27-40. <https://doi.org/10.3406/lgge.1999.2210>.
- Gamal, Donia, Marco Alfonse, El-Sayed M. El-Horbaty, et Abdel-Badeeh M. Salem. « Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features ». *Procedia Computer Science*, Proceedings of the 9th International Conference of

- Information and Communication Technology [ICICT-2019] Nanning, Guangxi, China January 11-13, 2019, 154 (1 janvier 2019): 332-40. <https://doi.org/10.1016/j.procs.2019.06.048>.
- Haddi, Emma. « SENTIMENT ANALYSIS: TEXT PRE-PROCESSING, READER VIEWS AND CROSS DOMAINS ». Thesis Doctora, BRUNEL UNIVERSITY, 2015.
- Hajic, Jan, Otakar Smrz, Petr Zemánek, Jan Šnidauf, et Emanuel Beška. « Prague Arabic dependency treebank: Development in data and tools ». In *Proc. of the NEMLAR Intern. Conf. on Arabic Language Resources and Tools*, 110-17, 2004.
- Han, Chung-hye, Na-Rae Han, Eon-Suk Ko, Martha Palmer, et Heejong Yi. « Penn Korean Treebank : Development and Evaluation », 69-78, 2001. <http://hdl.handle.net/2065/12256>.
- <https://www.facebook.com/kdnuggets>. « NLP vs. NLU: From Understanding a Language to Its Processing ». *KDnuggets* (blog). Consulté le 3 juillet 2023. <https://www.kdnuggets.com/nlp-vs-nlu-from-understanding-a-language-to-its-processing.html>.
- Hu, Xiaolin. « Meta Semantics: Towards better natural language understanding and reasoning ». arXiv, 20 avril 2023. <https://doi.org/10.48550/arXiv.2304.10663>.
- Hussein, Doaa Mohey El-Din Mohamed. « A Survey on Sentiment Analysis Challenges ». *Journal of King Saud University - Engineering Sciences* 30, n° 4 (octobre 2018): 330-38. <https://doi.org/10.1016/j.jksues.2016.04.002>.
- Ismail, R., M. Omer, M. Tabir, N. Mahadi, et I. Amin. « Sentiment Analysis for Arabic Dialect Using Supervised Learning ». In *2018 International*

- Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, 1-6, 2018. <https://doi.org/10.1109/ICCCEEE.2018.8515862>.
- kamel, Sarah M., Shimaa I. Hassan, et Lamiaa Elrefaei. « VAQA: Visual Arabic Question Answering ». *Arabian Journal for Science and Engineering*, 24 mars 2023. <https://doi.org/10.1007/s13369-023-07687-y>.
- Khasawneh, R. T., H. A. Wahsheh, M. N. Al-Kabi, et I. M. Alsmadi. « Sentiment analysis of arabic social media content: a comparative study ». In *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*, 101-6, 2013. <https://doi.org/10.1109/ICITST.2013.6750171>.
- Khurana, Diksha, Aditya Koli, Kiran Khatter, et Sukhdev Singh. « Natural Language Processing: State of The Art, Current Trends and Challenges ». *arXiv e-prints* 1708 (1 août 2017): arXiv:1708.05148.
- Larkey, Leah S., Lisa Ballesteros, et Margaret E. Connell. « Light Stemming for Arabic Information Retrieval ». In *Arabic Computational Morphology: Knowledge-Based and Empirical Methods*, édité par Abdelhadi Souidi, Antal van den Bosch, et Günter Neumann, 221-43. Text, Speech and Language Technology. Dordrecht: Springer Netherlands, 2007. https://doi.org/10.1007/978-1-4020-6046-5_12.
- Liu, Bing. « Sentiment Analysis and Opinion Mining ». *Synthesis Lectures on Human Language Technologies* 5, n° 1 (23 mai 2012): 1-167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>.
- Maamouri, Mohamed, Ann Bies, Tim Buckwalter, et Wigdan Mekki. « The penn arabic treebank: Building a large-scale annotated arabic

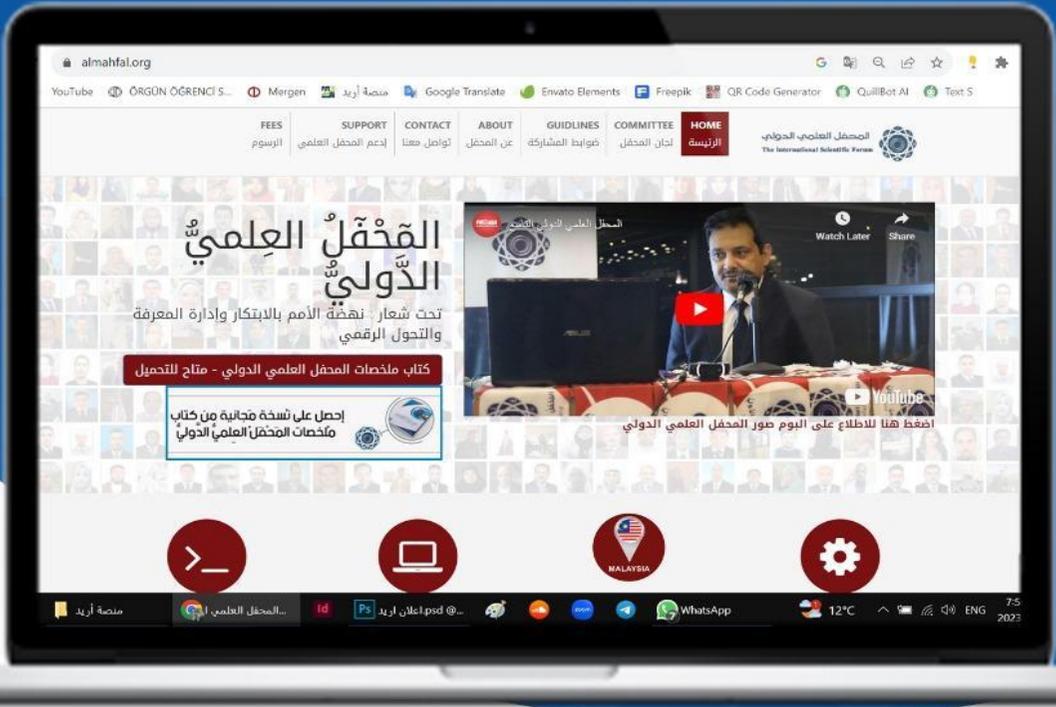
- corpus ». *NEMLAR Conference on Arabic Language Resources and Tools*, 1 janvier 2004.
- Manning, Christopher D, et Hinrich Schiitze. *Foundations of Statistical Natural Language Processing*. Second printing. Cambridge États-Unis: The MIT Press Cambridge, Massachusetts London, England, 2000.
- Maynard, Diana, et Kalina Bontcheva. *Natural Language Processing for the Semantic Web*. Édité par Isabelle Augenstein. Morgan & Claypool Publishers, 2016.
- McDonald, Ryan T., Kerry Hannan, Tyler Neylon, Mike Wells, et Jeffrey C. Reynar. « Structured Models for Fine-to-Coarse Sentiment Analysis ». In *ACL*, 2007.
- Mountassir, A., H. Benbrahim, et I. Berrada. « A Cross-Study of Sentiment Classification on Arabic Corpora ». In *Research and Development in Intelligent Systems XXIX*, édité par Max Bramer et Miltos Petridis, 259-72. Springer London, 2012.
- Nallakaruppan, Musiri Kailasanathan, Ishita Johri, SivaRamaKrishnan Somayaji, Surbhi Bhatia, Areej A. Malibari, et Aliaa M. Alabdali. « Secured MCDM Model for Crowdsorce Business Intelligence ». *Applied Sciences* 13, n° 3 (janvier 2023): 1511. <https://doi.org/10.3390/app13031511>.
- Nebia, BENZATER. « Analyse Morphologique du Texte Arabe pour Son Indexation Sémantique. » Magistère, Université des Sciences et de la Technologie - Mohamed Boudiaf, 2014.

- . « Analyse Morphologique du Texte Arabe pour Son Indexation Sémantique. » Magistère en Informatique, Université des Sciences et de la Technologie - Mohamed Boudiaf, 2014.
- Omar, Nazlia, Mohammed Albared, Adel Al-Shabi, et Tareq Al-Moslmi. « Ensemble of Classification Algorithms for Subjectivity and Sentiment Analysis of Arabic Customers' Reviews ». *International Journal of Advancements in Computing Technology* 5 (1 janvier 2013): 77-85.
- Omari, M. Al, M. Al-Hajj, N. Hammami, et A. Sabra. « Sentiment Classifier: Logistic Regression for Arabic Services' Reviews in Lebanon ». In *2019 International Conference on Computer and Information Sciences (ICCIS)*, 1-5. Sakaka, Saudi Arabia, Saudi Arabia: IEEE, 2019. <https://doi.org/10.1109/ICCISci.2019.8716394>.
- Pak, Alexander. « Automatic, Adaptive, and Applicative Sentiment Analysis ». Thesis Doctora, Université Paris Sud, 2012. HAL (<https://tel.archives-ouvertes.fr/tel-00717329>).
- Pang, Bo, Lillian Lee, et Shivakumar Vaithyanathan. « Thumbs up?: Sentiment Classification Using Machine Learning Techniques », 10:79-86. Association for Computational Linguistics, 2002. <https://doi.org/10.3115/1118693.1118704>.
- Reiter, Ehud, et Robert Dale. « Building Applied Natural Language Generation Systems ». *Natural Language Engineering* 3, n° 1 (mars 1997): 57-87. <https://doi.org/10.1017/S1351324997001502>.
- Semaan, Paul. « Natural Language Generation: An Overview ». *Journal of Computer Science* 1, n° 3 (2012): 8.

- Smienk, Jeroen. « Natural Language Search for Personal Information Retrieval : Benefits and Limitations to the User's Experience in Bookkeeping Software. » Info:eu-repo/semantics/masterThesis, University of Twente, 2023. <http://essay.utwente.nl/94310/>.
- Tan, Chenhao, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, et Ping Li. « User-Level Sentiment Analysis Incorporating Social Networks », 1397. ACM Press, 2011. <https://doi.org/10.1145/2020408.2020614>.
- Taylor, Ann, Mitchell Marcus, et Beatrice Santorini. « The Penn Treebank: An Overview ». In *Treebanks*, édité par Anne Abeillé, 20:5-22. Dordrecht: Springer Netherlands, 2003. http://www.springerlink.com/index/10.1007/978-94-010-0201-1_1.
- Tlili, Guiassa Yamina. « L'application de la théorie de l'Appraisal pour l'identification de l'opinion dans le texte arabe. » Consulté le 4 avril 2016. http://www.citala.org/papers/paper_50.pdf.
- Turney, Peter D. « Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews », 417. Association for Computational Linguistics, 2001. <https://doi.org/10.3115/1073083.1073153>.
- Valdivia García, Ana. « Sentiment Analysis-Based Methods for Cultural Monuments ». Thesis Doctora, Universidad de Granada, 2018. <http://digibug.ugr.es/handle/10481/54784>.
- Xue, Naiwen, Fei Xia, Fu-Dong Chiou, et Marta Palmer. « The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus ». *Natural Language Engineering* 11, n° 2 (juin 2005): 207-38. <https://doi.org/10.1017/S135132490400364X>.



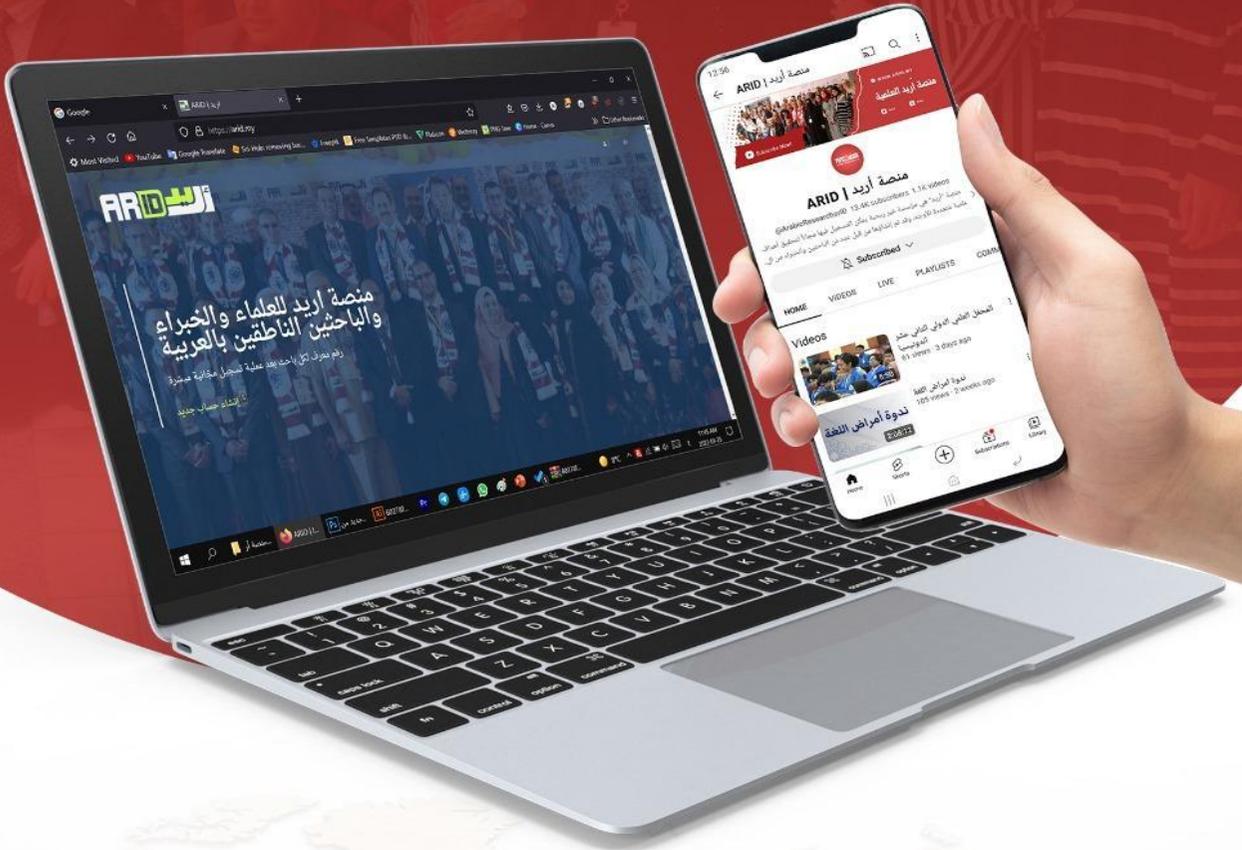
المحفل العلمي الدولي



يُعد "المحفل العلمي الدولي" وعاءً علميًّا خاصًّا بأعضاء منصة أريد، يستوعب ما يمكنهم القيام به من أنشطة وفعاليات علمية، وثقافية، وفكرية، وتواصلية، تُحقِّق الفائدة للمنصة وأعضائها في ضوء ما أقره أعضاء اللجنة التحضيرية لهذا المحفل من قيم، ورسالة، ورؤية.

منصة أريد العلمية

ARID Scientific Platform



منصة أريد هي منظمة غير ربحية تتيح التسجيل المجاني لتحقيق أهداف علمية متعددة الأوجه. تم تأسيسها من قبل باحثين وخبراء مهتمين بتطوير البحث العلمي.

مقرات المنصة: بريطانيا - ماليزيا - العراق - تركيا (13467806) ARID SCIENTIFIC LTD
All rights reserved - Address: 128 City Road, London, EC1V 2NX

● WWW.ARID.MY للتسجيل مجاناً في المنصة



Arid



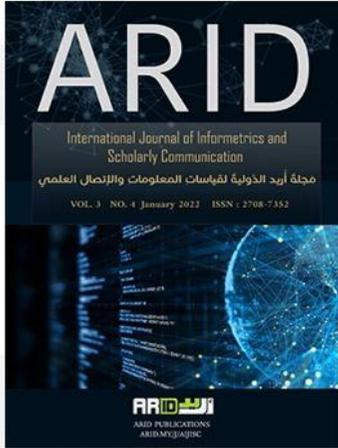
Arid

مجلات أريد الدولية

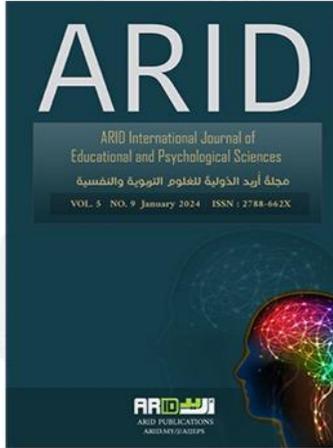
مجلات علمية تهدف الى الرقي بالمستوى العلمي
للناطقين باللغة العربية

WWW.ARID.MY

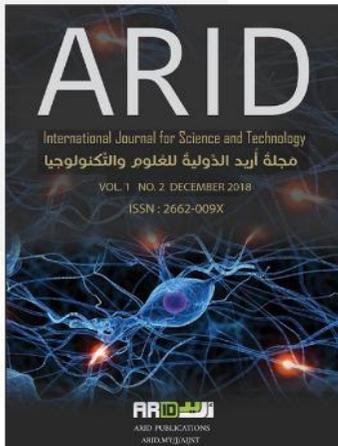
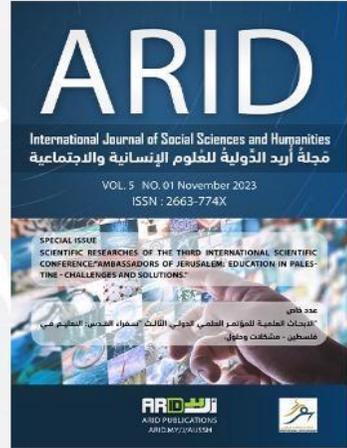
مجلة أريد الدولية لقياسات
المعلومات والاتصال العلمي



مجلة أريد الدولية للعلوم
التربوية والنفسية



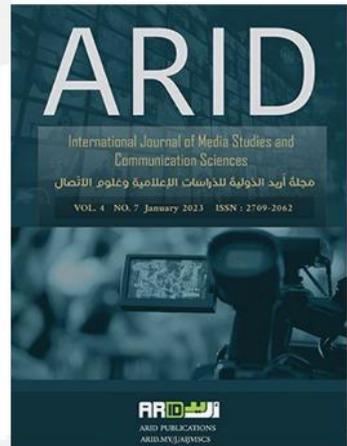
مجلة أريد الدولية للعلوم
الإنسانية والاجتماعية



مجلة أريد الدولية للدراسات
الإعلامية وعلوم الإتصال

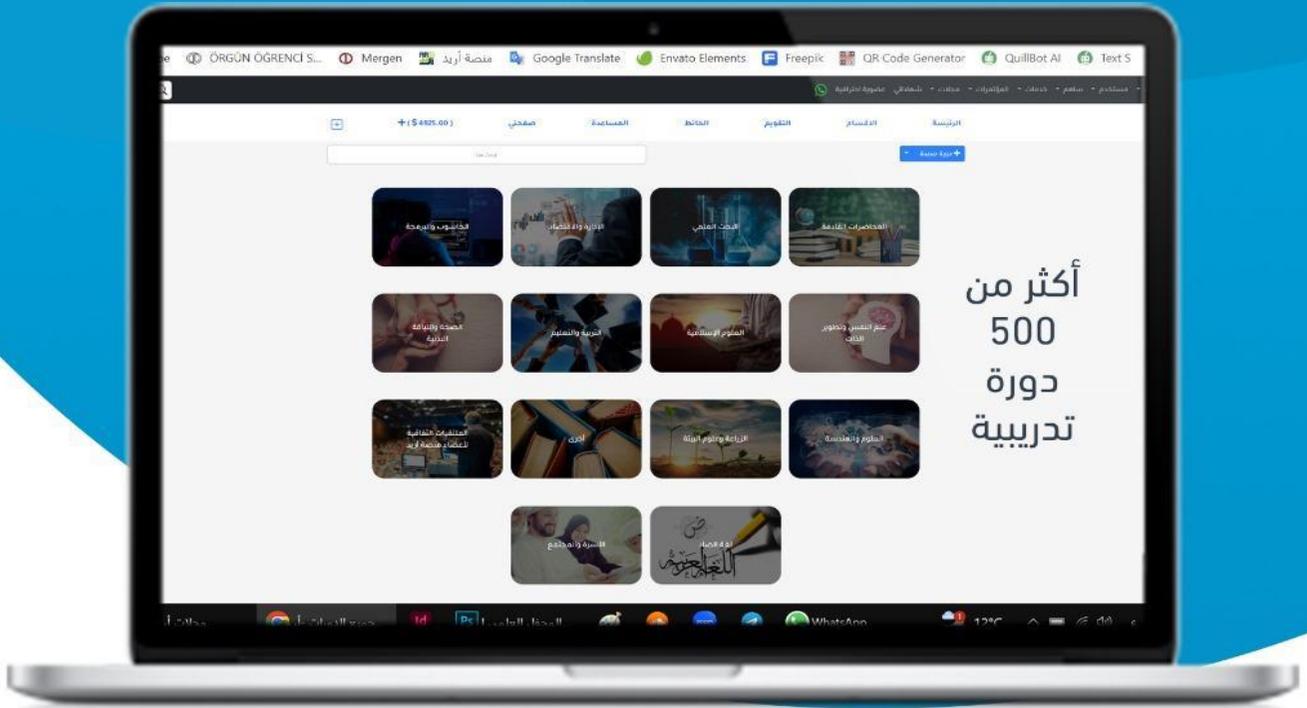


مجلة أريد الدولية للعلوم
والتكنولوجيا



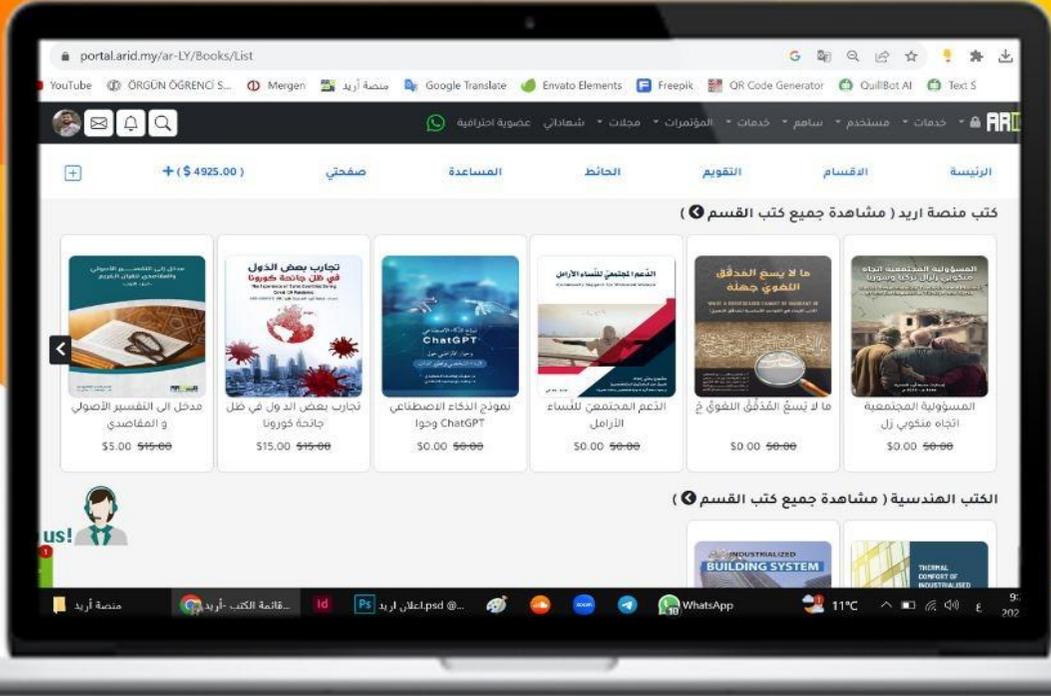
نظام عليم للتعليم الالكتروني في منصة أربد العلمية

أكثر من 500 دورة تدريبية ومحاضرة علمية



المكتبة الرقمية في منصة أريد

مكان واحد لجميع إصدارات أريد العلمية



الموقع الرسمي للمكتبة الرقمية

portal.arid.my/ar-LY/Books/List



أريد
ARID
Arabic Researcher ID

WWW.ARID.MY

أوسمة منصة أريد العالمية

سجل في منصة أريد العلمية واحصل على الأوسمة العلمية مع الشواهد وكتب التأييد

الفهم الآلي للغة العربية تحليل الآراء والمشاعر نموذجاً

تقدم في هذا الكتاب بعض ما توصلت إليه الأبحاث والتجارب الخاصة باللغة العربية في إطار حلق اللسانيات الحاسوبية، منطلقين من تحديد مميزات وخصائص اللغة العربية، ثم الإشارة إلى المستويات اللسانية العربية التي يسرت سبل نمذجة اللغة العربية في الحاسوب، الشيء الذي حول بلورة العديد من التطبيقات الخاصة بالمعالجة الآلية للغة العربية، ثم التأصيل إلى الفهم الآلي للغات الطبيعية والإشارة إلى بعض النظم المرتبطة بفهم اللغة العربية آلياً، مركزين على مجال تحليل الآراء والمشاعر المكتوبة بها، حيث عرضنا الخطوات التي تقطعها نظم تحليل الآراء والمشاعر، انطلاقاً من أحد التجارب البحثية العربية التي اعتمدت على الموارد اللسانية العربية في بلورة نظامها .