

## بناء موارد أساسية لتحليل الآراء والمشاعر الواردة في التعليقات باللغة العربية

ابراهيم حبيبي، عبد الحميد الجيهاد

معهد الدراسات والأبحاث للتعريب - جامعة محمد الخامس - المغرب

brahimhabibi43@gmail.com

تأريخ الاستلام: 27/01/2020

تأريخ القبول: 25/03/2020

**المستخلص** - نهدف من هذه الورقة إبراز خطوات بناء بعض الموارد اللسانية التي من شأنها أن توظف في بناء البرمجيات والتطبيقات الخاصة بتحليل الآراء والمشاعر الواردة في المواقع ثم تقييمها، من بين الموارد التي قمنا ببنائها معجم متخصص يضم كلمات تستعمل في إبداء الآراء والتعبير عن المشاعر، ثم لائحة حروف المعاني التي تؤثر في تغيير توجيه التعبيرات العربية، بالإضافة إلى المدونة المعتمدة في اختبار برمجيات تحليل الآراء والمشاعر. وللتمكن من بناء الموارد السابقة الذكر، فقد ركزنا في البداية على المنهج الوصفي لرصد متطلبات بناء نظم تحليل الآراء والمشاعر، كما اعتمدنا على المنهج اليدوي والمنهج الآلي في إعداد كل من المعجم، ولائحة الحروف، والمدونة، فتمكنا من تشكيل معجم قارب 21453 مدخل معجمي، وإعداد لائحة ضمت 9784 حرفاً، وجمع ما يقارب 100630 وسم منها 25% يدوياً.

**الكلمات المفتاحية:** تحليل الآراء والمشاعر، مدونة تحليل الآراء والمشاعر، نظرية الأبريسال، الحقول المعجمية، تأثير حروف المعاني،

**ABSTRACT** - This paper aims at highlighting the procedures of building some linguistic resources that would help to make and assess software and applications for opinion and sentiment analysis. Among the resources that we have built are a specialized dictionary that includes opinions and sentiments lexicon, a list of words and prepositions that orient the Arabic discourse, and the corpus adopted during the assessment of 'opinion mining' and 'sentiment analysis' software. And in order to build these linguistic resources, we basically used the descriptive approach so as to provide the tools required for opinion and sentiment analysis. We also adopted the manual and automated approach for preparing the specialized dictionary, the list of words and prepositions, and the corpus. So we were able to create a dictionary which is containing 21453 lexical entries, prepare a list which contains 9784 words and prepositions, and collect about 100630 comments which have 25% of manually tagging.

برمجيات الترجمة الآلية، برمجيات التصريف الآلي، برمجيات

الاشتقاق الآلي، أما فيما يخص برمجيات تحليل الآراء والمشاعر فلا زالت اللغة العربية تعاني من نقص في هذه البرمجيات، وهذا راجع إلى قلة الموارد اللسانية الموظفة في بنائها، عكس اللغات الهندوأوروبية التي حققت تقدماً كبيراً في توفير الموارد الضرورية لبناء برمجيات تحليل الآراء والمشاعر.

لقد أصبح تبادل الآراء والتعبير عن المشاعر اتجاه القضايا المعروضة في العديد من المواقع أمراً سهلاً، مما طرح إشكالات كيفية تتبع هذه الآراء وتحليلها ثم تقييمها لمعرفة رغبات المتتبع، وقد لوحظ أن تحليلها يدوياً يستغرق وقتاً طويلاً، لهذا الغرض تدخل الباحثون في مجال المعالجة الآلية للغات الطبيعية، قصد بناء برمجيات تقوم بتحليل هذه الآراء وتساعد على تقييمها بشكل آلي لمعرفة وجهات نظر المتتبعين حول مختلف ما يعرض في هذه

### 1. المقدمة:

تزامن التطور التكنولوجي الكبير والحاجة الماسة للحاسوب، على نطاق واسع وفي مختلف مجالات الحياة، مع تطور أدوات ( Web 2.0) وما ترتب عن ذلك من ثورة في مجال المعلومات والاتصالات المتمثلة أساساً في ظهور ما سمي بمواقع التواصل الاجتماعي، المنتديات، مواقع البيع والشراء وغيرها من المواقع، فكانت الحاجة ماسة إلى اعتماد مجال المعالجة الآلية للغات الطبيعية، من أجل تنظيم ما يتم تداوله في هذه المواقع ومعالجته في وقت وجيز وبأقل جهد وكلفة.

لقد استقادت اللغة العربية كباقي اللغات الطبيعية من التطور الذي حققته دراسات اللسانيات الحاسوبية في مختلف مستوياتها الصوتية والصرفية والنحوية والنصية والدلالية والمعجمية، توج ذلك بإنجاز مجموعة من البرمجيات مختلفة مثل المعاجم الإلكترونية،

الفصحى، في حين تم إهمال إنشاء المدونات التي تضم اللغة العربية الفصحى، لهذا فكرنا في بناء مدونة عربية متخصصة في مجال تحليل الآراء والمشاعر باعتماد بعض الوسائل المعلوماتية لتحميل التعليقات من مواقع مختلفة مع التركيز على اللغة العربية الفصحى وذلك راجع إلى مجموعة من الدوافع يمكن إجمالها فيما يلي:

✓ ندرة المدونات العربية المتخصصة في مجال تحليل الآراء والمشاعر.  
 ✓ أغلب المدونات المتوفرة لا تحتوي على الكثير من التعليقات.  
 ✓ التركيز على اللغة المحلية حسب الانتماء الجغرافي للباحثين.  
 ✓ غياب تنوع المصادر المعتمدة في بناء المدونات، إذ إن أغلبها يركز على موقع الفيسبوك وموقع تويتر.  
 ✓ لكل ذلك نهدف من خلال هذا العمل تقديم بعض الموارد الأساسية الخاصة بمجال تحليل الآراء والمشاعر والمتمثلة أساساً في:

✓ معجم عربي غني يضم كلمات عربية مصنفة وفق تصنيف يساعد على تقديم نتائج دقيقة في عملية تحليل الآراء والمشاعر.  
 ✓ لائحة حروف المعاني التي تساهم في تغيير توجيه التعليقات العربية.  
 ✓ مدونة عربية متخصصة بمجال تحليل الآراء والمشاعر تضم تعليقات عربية مأخوذة من مجالات مختلفة.  
 لتحقيق هذه الأهداف سنقوم بتقديم أهم الأعمال السابقة التي اهتمت بتوفير الموارد اللسانية الخاصة بمجال تحليل الآراء والمشاعر، ثم سنقدم تعريفاً بمجال تحليل الآراء والمشاعر، مع الإشارة إلى أهم العناصر المشكلة لعملية المعالجة الآلية للآراء والمشاعر، وإلى مختلف المستويات التي تشملها هذه العملية، ثم ننقل إلى الحديث عن أنواع معاجم تحليل الآراء والمشاعر، والإشارة إلى أهم المناهج المتبعة في بناء هذا النوع من المعاجم، لنفصل الحديث بعد ذلك في الخطوات التي اتبعناها في بناء المدونة العربية المتخصصة في مجال تحليل الآراء والمشاعر، مع الإشارة إلى المواقع المعتمدة في عملية تحميل التعليقات العربية، لنخصص الحديث عن عدد التعليقات المأخوذة من كل موقع، ثم نشير بعد ذلك إلى الخطوات الأساسية المتبعة في بناء معجم موجه لتحليل الآراء والمشاعر خاص باللغة العربية، كما سنتحدث بعد ذلك عن أهمية توظيف حروف المعاني في عملية تحليل الآراء والمشاعر، ونقدم الخطوات المتبعة في إعداد لائحة حروف المعاني العربية المسؤولة عن تغيير توجيه التعليقات العربية، لنختم عملنا باستنتاج خلاصات وكذا رهانات البحث المستقبلية.

المواقع، باعتماد بعض الموارد اللسانية الضرورية وقد شكل المعجم المتخصص أولى هذه الموارد المعتمدة في بناء مثل هذه البرمجيات، وتجاوزا للأبحاث العربية السابقة في هذا المجال والمرتكزة على ترجمة المعاجم المتوفرة بالعديد من اللغات أوفرها اللغة الإنجليزية، وعلى هذه الأساس ارتكز عملنا على مجموعة من الدوافع الوجيهة ترد إجمالاً إلى:

✓ قلة المعاجم الخاصة بمعالجة الآراء والمشاعر باللغة العربية ومحدوديتها مقارنة بالموثوث المعجمي.  
 ✓ الاعتماد على الترجمة الآلية لتوفير المعاجم المتعلقة بتصنيف الآراء والمشاعر المكتوبة باللغة العربية.  
 ✓ عدم اشتمال المعاجم المترجمة إلى اللغة العربية على أقسام كلام اللغة العربية الموظفة في التعبير عن الآراء والمشاعر.  
 ✓ أغلب المعاجم المتوفرة حالياً تعتمد على تصنيفات ضيقة: تصنيف ثنائي (موجب - سالب) أو ثلاثي (موجب - سالب - محايد) على الأكثر.

ونظراً لما تلعبه حروف المعاني العربية من دور في تغيير معاني الكلمات والجمل، واكتشاف الغياب التام لتوظيف هذا القسم الأساسي من أقسام الكلام في عملية تحليل الآراء والمشاعر من قبل العديد من الباحثين العرب، أو تركيز بعضهم على حروف النفي وإغفال المعاني العديدة التي تتميز بها اللغة العربية، فقد كان دافعاً للبحث في الموثوث العربي في العديد من المؤلفات التي اهتمت بحروف المعاني واستعمالاتها وكذا معانيها، حيث حاولنا جمع الحروف التي من شأنها تغيير توجيه تعليقات المتابعين في المواقع ضمن لائحة مفصلة مقترنة بالسوابق واللواحق الممكن ورودها مع هذه الحروف.

ونظراً للدور الفعال الذي لعبته المدونات في الدراسات اللغوية إثر توظيفها، فقد أخذت الدراسات اللسانية الحاسوبية على عاتقها الاهتمام بالمدونات اللغوية، فاهتمت في البداية بالمدونات العامة، لكن مع تزايد الطلب على مثل هذه المدونات والعمل بها في مجالات متخصصة، تم الاهتمام ببناء مدونات اصطلح عليها بالمدونات المختصة أو أحادية النوع<sup>[1]</sup>، ومن بين هذه المدونات نجد المدونة المختصة أو المتخصصة بمجال تحليل الآراء والمشاعر، إذ حققت اللغات الهندوأوروبية تقدماً كبيراً في هذا المجال، في حين أن اللغة العربية ما زالت في حاجة ماسة إلى إنشاء مثل هذه المدونات، فبعد الاطلاع على مجموعة من الأبحاث العربية التي اهتمت ببناء مدونات متخصصة في مجال تحليل الآراء والمشاعر وجدنا أن أغلبها ركزت على ضم تعليقات باللغة المحلية، أو الجمع بين اللغة المحلية واللغة العربية

## 2. الأعمال السابقة:

بإثراء المعجم بمجموعة من المرادفات باعتماد ثمانية مواقع هي: Reverso Dictionary, Bab.la, Atlas, Thesaurus, Ortolang, SensAgent, The Free Dictionary, and the Synonym website. فقام الفريق بحذف الكلمات المتماثلة، فتم الحصول على 141428 كلمة، وقد تعاقد فريق البحث في إطار شراكة مع أحد المتخصصين في مجال الترجمة لمراجعتهم الكلمات المترجمة بعد تصفيتهما، فانتجوا بذلك إلى 81757 كلمة وسم منها إلا 14129<sup>[3]</sup> مدخل مقسمة بين 11979 كلمة مفردة و2148 كلمة مركبة<sup>[4]</sup>، وتجدر الإشارة إلى أن هذا المعجم متوفر للتحميل على الرابط أسفله<sup>[5]</sup>.

### معجم MPQA منظور متعدد سؤال جواب:

تم إنشاء هذا المعجم من طرف كل من: تيريزا ويلسن، جنسي ويب وبوبول هوفمن (Wiebe, Janyce Theresa Wilson and Paul Hoffmann)<sup>[4]</sup>، وهي اختصار لمنظور متعدد سؤال جواب "Multi-Perspective Question Answering".

يضم هذا المعجم 8214<sup>[7]</sup> كلمة موزعة بين الأسماء والصفات وبعض الأفعال صنفت وفق تصنيف ثلاثي (موجب، سالب، محايد)، وقد تُرجم جزء منه إلى اللغة العربية، في حين قام فريق المعالجة الآلية للغات الطبيعية بجامعة صفاقس بتونس بترجمة الكلمات المتبقية، ثم إغناؤه بالاعتماد على (ArabicWordNet) و (SentiStrength)<sup>[8]</sup>، لكنهم أغفلوا توسيم الكلمات كما كانت موسومة في النسخة الإنجليزية، بالإضافة إلى ذلك فقد قام الفريق بتصنيف الكلمات إلى: "موجب قوي - موجب ضعيف - سالب قوي - سالب ضعيف".

### معجم (SentiWordNet):

يعد هذا المعجم من بين أشهر المعاجم استعمالاً في مجال تحليل الآراء والمشاعر، صنفت كلماته وفق تصنيف ثنائي (موجب - سالب)، وقد بني هذا المعجم وفق أربعة إصدارات<sup>[6]</sup> و<sup>[10]</sup>، وقد تُرجم إلى العديد من اللغات على رأسها اللغة العربية، حيث قام كل من سماح الحازم بلاك وماكنوت (Black, W. and McNaught, 2013) بإصدار النسخة العربية من هذا المعجم سميت ب (Arabic Senti WordNet)، عبر ثلاث مراحل هي:

✓ الخطوة الأولى: تحديث الإصدار الثاني Arabe WordNet 2.0 اعتماداً على الإصدار الثالث الخاص باللغة الإنجليزية.

(Anglais WordNet 3.0).

✓ الخطوة الثانية: إصدار النسخة الثالثة للغة العربية (Arabe WordNet 3.0).

يلعب كل من المعجم والمدونة وحروف المعاني موارد لسانية دوراً أساسياً في مساعدة الحاسوبي في بناء نظام تحليل الآراء والمشاعر، وعلى هذا الأساس بدأ الحاسوبيون أثناء بنائهم لبرمجيات تحليل الآراء والمشاعر بتوفير بعض من هذه الموارد اللسانية، وفيما يلي عرض لأهم الأعمال السابقة التي اهتمت بتوفير الموارد اللسانية الخاصة بمجال تحليل الآراء والمشاعر بدءاً بالمعجم باعتباره الركيزة الأساس في عملية التحليل، ثم الحديث عن الأعمال التي وظفت حروف المعاني في برمجياتها، وأخيراً الحديث عن أهم المدونات العربية التي وُضعت لهذا المجال. الأعمال السابقة الخاصة بالمعجم.

يعتبر المعجم من أهم الموارد اللسانية التي ينبغي توفيرها لبناء برمجيات تحليل الآراء والمشاعر، والباحث في هذا المجال سجد أن العديد من اللغات من بينها اللغات الهندوأوروبية وعلى رأسها اللغة الإنجليزية قد حققت تقدماً بارزاً في توفير عدد من المعاجم الخاصة بمجال تحليل الآراء والمشاعر واستثمارها في بناء العديد من البرمجيات وفق مقاربات مختلفة، مما دفع العديد من الباحثين إلى اعتماد الترجمة الآلية كأداة وظفوها في ترجمة معاجم اللغة الإنجليزية إلى العديد من اللغات من بينها اللغة العربية، وسنعرض فيما يلي أهم المعاجم التي ترجمت إلى اللغة العربية:

### معجم المجلس الوطني للأبحاث في كندا:

تم إنشاء هذا المعجم من طرف فريق بحث تابع للمجلس الوطني للبحوث بكندا وهو معجم يضم كلمات إنجليزية بلغ عددها 14182 كلمة تم تصنيفها يدوياً وفق ثمانية حقول معجمية مرتبطة بالمشاعر هي: "الغضب والترقب الإشمئزاز والخوف والفرح والحزن والدهشة والثقة"، مع تحديد الموجب منها والسالب.

تُرجم هذا المعجم إلى العديد من اللغات ناهزت أربعين لغة باعتماد قوئل للترجمة الآلية من بين هذه اللغات "اللغة العربية، والكتالانية، والصينية، والإسبانية، والفرنسية، والألمانية، واليونانية، والعبرية، والهندية، والإيرلندية، والإيطالية، واليابانية، والفارسية، والبرتغالية، والرومانية، والروسية، والصومالية، والسودانية، السويدية، والتايلاندية، والتركية، والأوكرانية،..."، ويمكن تحميل هذا المعجم على الرابط أسفله<sup>[2]</sup>.

### معجم "FEEL" مشاعر اللغة الفرنسية الموسع:

بني هذا المعجم اعتماداً على ترجمة معجم المجلس الوطني للبحوث بكندا وذلك بتوظيف ستة مواقع للترجمة وهي: "Google Translate, Bing Translate, Collins Translator, Reverso, Bab.la et Word Reference" ثم قام الفريق

الباحثين الاهتمام ببناء العديد من المدونات الخاصة بهذا المجال، إلا أن معظمها اشتمل على آراء مكتوبة عامة، في حين تم إغفال الآراء المكتوبة باللغة العربية الفصحى، وفيما يلي عرض لأهم الأعمال التي عملت على بناء المدونات العربية الخاصة بتحليل الآراء والمشاعر:

#### مدونات التعليقات العربية

تعتبر هذه المدونة من أشهر المدونات، التي وُظفت في العديد من الدراسات، وقد حاول رشدي صالح من خلالها تحميل مجموعة من التعليقات مأخوذة من منتديات ومواقع مختلفة تهتم بمجال الأفلام، وقد ضمت المدونة 250 تعليقا موجبا، و250 تعليقا سالبًا [17].

#### قاعدة بيانات لتغريدات المشاعر العربية

اشتملت هذه القاعدة ما يقارب 10000 تغريدة تم الحصول عليها خلال مرحلتين؛ المرحلة الأولى تم فيها تحميل 36000 تغريدة مأخوذة من "SocialBakers" وهي صفحة تهتم بنشاطات المنتمين لها، حيث ركز الباحثون على ثلاثين منخرطًا في هذه الصفحة في حدود شهر نونبر 2013، في حين ركز الباحثون في المرحلة الثانية في تحميل 48000 تغريدة من صفحة "EgyptTrends"، فكان مجموع التغريدات التي تم تحميلها هو 80000 تغريدة، وبالاعتماد على عملية التصفية وعملة المعالجة القبلية، أزيلت التغريدات غير الصالحة فتقلص العدد إلى 10000؛ إذ شملت عمليتي المعالجة والتصفية كل التغريدات المكتوبة باللغة الأجنبية، ثم صنفت هذه التغريدات وفق أربع فئات؛ "الموضوعية - الذاتية الموجبة - الذاتية السالبة - الذاتية المشتركة" [18].

#### مدونة عربية لتحليل الآراء

قامت الباحثة أسماء منتصر وآخرون [19]، [20]، ببناء هذه المدونة عبر تحميل مجموعة من التعليقات اشتملت مجالات مختلفة ضمن منتديات متعددة ركزوا فيها بالأساس على موقع الجزيرة (www.aljazeera.net) وعلى مدونة التعليقات العربية. لقد قام الباحثون بوضع المدونة في ثلاثة سجلات؛ ضم الأول تعليقات مدونة التعليقات العربية، أما السجل الثاني فاشتمل على تعليقات متعلقة بالمجال الرياضي بلغت 1492 تعليقا مستمدا من موقع الجزيرة، في حين ضم السجل الثالث تعليقات متعلقة بالمجال السياسي وصل عددها إلى 1028 تعليقا استمدت كذلك موقع الجزيرة، بعد ذلك عمل الفريق على تصنيف التعليقات اعتمادا على مقارنة التعليم الآلي إلى: تعليقات موجبة، تعليقات سالبة، تعليقات محايدة وتعليقات مكتوبة بالعامية، فألغوا التعليقات المحايدة والتعليقات المكتوبة بالعامية واكتفوا بالتعليقات الموجبة التي لم

✓ الخطوة الثالثة: إصدار النسخة العربية الخاصة بالكلمات المتعلقة بالمشاعر (Arabic Senti WordNet) اعتمادا على منهج المقابلة انطلاقا من النسخة الإنجليزية (Anglais SentiWordNet3.0). [11].

#### معجم مارتن وويت:

ترجم هذا المعجم إلى اللغة العربية من طرف بعض الباحثين العرب، من بينهم الباحثة غيثة يمينة تليلي التي قامت بترجمة 325 كلمة مأخوذة من كتاب الباحثين مارتن وويت (Martin and P. R. R. White)، لتطبيقها في معالجة الآراء المعبر عنها في مجال الاقتصاد [12]. كما قام الباحث أحمد عمار بنفس العمل مع إغنايه بالاعتماد على (Arabic WordNet) و(BabelNet) فحصل على 1400 كلمة [13].

#### الأعمال السابقة الخاصة بلائحة حروف المعاني:

لقد شهدت أبحاث اللغات الهندوأوروبية تقدما بارزا في توفير الموارد اللسانية اللازمة لتحليل الآراء والمشاعر، إذ أولت اهتمامها الكبير بتوفير لوائح الحروف المؤثرة في تغيير توجيه الكلام وركزت في ذلك على حروف النفي، في حين أغفلت باقي المعاني الأخرى مثل النهي، والاستفهام، وغيرها من المعاني، وفيما يلي أهم الأعمال العربية التي اهتمت بتوظيف لوائح تضم الحروف التي تغير توجيه التعبير العربية.

**محمد سيف وآخرون:** أنشأ محمد سيف رفقة زملائه لائحة تضم ستة حروف للنفي بالاعتماد على ترجمة لائحة الحروف الإنجليزية، ضمت هذه اللائحة خمسة حروف عربية وهي "ليس، لا، لم، لن، ما" والحرف السادس هو حرف خاص بالعامية المصرية وهو "مُش"، كما قام الفريق بإلصاق "ليس" بضائير المتصلة الخاصة بالمتكلم والمخاطب [14].

**رحاب الدويري ومحمد الشبول:** قاما بإنشاء لائحة تضم خمسة حروف عربية للنفي، وقسماها حسب استعمالها إلى مجموعتين؛ مجموعة تقترن بالأفعال فقط وهي "لا، لم، لن، ما"، والمجموعة الثانية اقتصرت على "ليس" لكونها تختص بالاسم والفعل [15].

**الخوالدة فاطمة:** قدمت الباحثة نفس حروف النفي المعتمدة من طرف محمد سيف والدويري، واختلفت عنهم في كيفية توظيفها، إذ وضعت أربع قواعد تضبط استخدامها في عملية تحليل الآراء والمشاعر [16].

#### الأعمال السابقة الخاصة بمدونة الآراء العربية

لقد شكلت المدونات أداة أساسية تم استثمارها في العديد من الدراسات اللغوية والحاسوبية، من بينها المدونات الموظفة في اختبار برمجيات تحليل الآراء والمشاعر، حيث أولى العديد من

يتجاوز عددها 819 تعليقا أما التعليقات السالبة التي بلغت 1263 تعليقا<sup>[18]</sup>. وكما سبقت الإشارة إلى أن المدونات المشتملة الفصحى، والجدول (1) يلخص مجموعة من المدونات: على آراء مكتوبة باللغة العربية الفصحى تظل قليلة بالمقارنة

الجدول (1): جدول بعض المدونات الخاصة بمجال تحليل الآراء والمشاعر

اسم المدونة	عدد التعليقات	المصدر	لغة التعليقات	التصنيفات المعتمدة	طريقة التصنيف	الإحالة
تغريدات (TGRD)	3017	Tweets	اللغة العربية الفصحى والعامية	الموضوعية - الذاتية الموجبة - الذاتية السالبة - الذاتية المشتركة	يدوية	[22]
تحرير (THR)	3008	Wikipédia TalkPages	اللغة العربية الفصحى	الموضوعية - الذاتية الموجبة - الذاتية السالبة - الذاتية المشتركة	يدوية	[23]
منتدى (MONT)	3097	Forums	اللغة العربية الفصحى والعامية	الموضوعية - الذاتية الموجبة - الذاتية السالبة - الذاتية المشتركة	يدوية	[23]
AWATIF	2855	WikipédiaTalk Pages/Forums	اللغة العربية الفصحى والعامية	الموضوعية - الذاتية الموجبة - الذاتية السالبة - الذاتية المشتركة	يدوية	[24]
تعليقات حول الفندق (HTL)	15572	TripAdvisor.com	اللغة العربية الفصحى والعامية	موجب - سالب - مشترك	آلي	[25]
تعليقات حول المطاعم (RES)	10970	Qaym.com	اللغة العربية الفصحى والعامية	موجب - سالب - مشترك	آلي	[25]
تعليقات الأفلام (MOV)	1524	Elcinemas.com	اللغة العربية الفصحى والعامية	موجب - سالب - مشترك	آلي	[25]
تعليقات المنتوجات (PROD)	4272	Souq.com	اللغة العربية الفصحى والعامية	موجب - سالب - مشترك	آلي	[25]
مدونة التغريدات العربية	8868	Tweets	اللغة العربية الفصحى والعامية	موجب - سالب - محايد - مشترك	يدوي	[26]
مدونة تغريدات سعودية	4700	Tweets	اللغة العربية الفصحى والعامية السعودية	موجب - سالب - محايد	يدوي	[27]
MIAK	4000	Tweets	اللغة العربية الفصحى والعامية المصرية	موجب - سالب - محايد	يدوي	[28]

## 2. منهج البحث:

الكلام العربي كلها عكس ما اشتملت عليه المعاجم السابقة الذكر والتي ركزت على بعض أقسام الكلام العربي، بالإضافة إلى حروف المعاني التي تلعب دورا كبيرا في تغيير توجيه الكلام العربي، وهو ما فصلناه في المحاور المتعلقة ببناء الموارد اللسانية الخاصة بمجال تحليل الآراء والمشاعر.

### 3. التعريف بمجال تحليل الآراء والمشاعر:

يعتبر تحليل الآراء المشاعر مجال بحث جديد يهدف إلى معالجة آراء، مشاعر، تقييمات، تسمينات، مواقف وعواطف الشخص حول منتج، أو خدمة، أو تنظيم، أو حدث، أو موضوع معين، ومن خلال هذا التعريف فإن عملية معالجة الآراء والمشاعر تحتوي خمسة عناصر:

نظرا لما تعاني منه اللغة العربية من نقص حاصل على مستوى توفير الموارد اللسانية الضرورية لبناء وتطوير برمجيات تحليل الآراء والمشاعر خاصة المعجم منها، وكذا الغياب الشبه التام لحروف المعاني العربية، علاوة على توفير مدونة تضم آراء مكتوبة باللغة العربية الفصحى يمكن الاعتماد عليها في اختبار برمجيات تحليل الآراء والمشاعر، فقد اعتمدنا المنهج الوصفي لرصد المشاكل التي تعاني منها اللغة العربية في هذا المجال، حيث انطلقنا من بناء مدونة عربية ضمت تعليقات مكتوبة باللغة العربية الفصحى، مكننا من استخلاص أقسام الكلام العربي الموظف في التعليقات والذي تبين أنه يمكن توظيف كل أقسام الكلام العربي، الشيء الذي يفرض توفير معجم يشتمل على أجزاء

يُدرج ضمن هذا النوع من المعاجم تلك المعاجم التي تضم كلمات اللغة العامية، وهذا ما ارتكزت عليه بعض الدراسات العربية [31].

**المعاجم الثنائية اللغة:** اهتمت المعاجم ثنائية اللغة التي لها علاقة بمجال تحليل الآراء والمشاعر ببناء معجم للكلمات أو التعبيرات التي تستعمل في التعبير عن المواقف والمشاعر والآراء، ويبدو ذلك جليا في الأعمال العربية، التي ركز العديد منها على بناء معاجم تشمل كلمات وتعبيرات من اللغة العربية الفصحى وأخرى عامية تهم منطقة أو دولة معينة، كما توجد بعض الأعمال التي ارتكزت على بناء معجم ثنائي يضم اللغة العربية والإنجليزية [32].

**المعاجم المتعددة اللغات:** نظرا لكون المتتبعين لمواقع التواصل الاجتماعي وغيرها من المواقع يبدون آراءهم في كثير من الأحيان باستعمال أكثر من لغة في رأي واحد، كانت الحاجة ماسة لإعطاء أهمية لإنشاء معجم متعدد اللغات، وترتكز عملية توفير هذا النوع من المعاجم أساسا على الترجمة الآلية لمعجم من المعاجم [33].

**المعاجم المتعددة المجالات:** جاءت فكرة وضع معجم متعدد المجالات انطلاقا من بناء مدونة متعددة المجالات، إذ في كثير من الأحيان تعتبر كلمة موجبة في مجال وسالبة في مجال آخر؛ ومثلا على ذلك فإذا أخذنا كلمة "الطول" التي يختلف توجيهها من مجال إلى مجال "مسافة السفر طويلة" التي يعتبر توجيهها في هذه الجملة سالبا، في حين إذا قلنا "مدة بطارية الهاتف طويلة" فتوجيه هذه الجملة يعتبر موجبا [22].

**المناهج المعتمدة في بناء معجم خاص بتحليل الآراء والمشاعر:** يعتبر بناء المعاجم المتخصصة لتحليل الآراء والمشاعر من بين أهم الموارد الضرورية والتي ينبغي توفرها في بناء برمجيات هذا المجال، وقد تعددت المناهج المعتمدة في بناء المعاجم وتصنيفها نلخصها فيما يلي:

#### المنهج اليدوي:

يعتبر من بين المناهج البسيطة والناجعة لبناء المعاجم، إذ يتأسس على المعرفة الإنسانية المتمثلة في جمع الكلمات الدالة على الآراء والمشاعر مع تحديد قطبية كل كلمة. إلا أن هذا المنهج رغم نجاعته يبقى مكلفا زمنيا، ولهذا الغرض تم اقتراح مناهج آلية وشبه آلية لبناء هذا النوع من المعاجم.

#### منهج المدونة:

تحتوي هذه المنهجية على مجموعة من الكلمات المنفصلة (Seeds) التي لها قطبية معروفة، ويتم حساب مماثلة الكلمات باستعمال التحليل الإحصائي لسياقات ترابط الكلمات فيما بينها لسانيا أو دلاليا، بالاعتماد على نماذج متجهات المسافات التي تقدم المحتوى الدلالي للكلمات على شكل متجهات، فيتم اعتبار

✓ الموضوع: هو الشيء الذي أُعطي فيه الرأي سواء كان منتوجا أو خدمة أو قضية أو موضوعا أو شخصية أو تنظيما أو غيرها.

✓ الجهة: والمقصود هنا الجهة المعنية بالتعليقات التي نشرت حولها.

✓ توجيه المشاعر: أي وجهة التعليق حول الموضوع إما موجبا، سالبا، أو محايدا.

✓ صاحب الرأي: والمقصود هنا الشخص الذي نشر التعليق عن الموضوع المعروض.

✓ الوقت: الوقت الذي نشر فيه الموضوع والتعليقات التي كتبت حوله [29]

كما تتبني عملية تحليل الآراء والمشاعر على واحد من المستويات الآتية:

◀ **التحليل على مستوى الوثيقة:** يتم التعامل مع الوثيقة في عملية تحليل الآراء والمشاعر حسب هذا المستوى على أساس أن الوثيقة ككل تمثل رأيا واحدا وتصنف الوثيقة حسب الفئة التي تعبر عنها إما إيجابية أو سلبية أو محايدة. وبما أن هذا المستوى يفترض في عملية التحليل على أن كل وثيقة تعبر عن رأي واحد؛ فإنه يصعب التعامل مع الوثيقة، خاصة إذا كانت تعبر عن أكثر من رأي.

◀ **التحليل على مستوى الجملة:** يهدف هذا المستوى من التحليل إلى تحديد ما إذا اشتملت الجملة على رأي أم لا، كما يهدف كذلك إلى محاولة التعرف على ما إذا كانت الجملة تعبر عن المشاعر، أو تقتصر على جملة الموضوعية، ويعتبر هذا المستوى بمثابة المحدد الأساسي لتوجيه الكلمة باعتماد السياق اعتمادا أساسيا، ويشمل التصنيف حسب هذا المستوى على جمل المقارنة وجمل المدح والذم إلى غيرها من التعبيرات.

◀ **التحليل على مستوى الكلمة:** يعتبر هذا المستوى أكثر تدقيقا في عملية التحليل؛ إذ الغرض منه الكشف عن العناصر الخمسة السابقة الذكر أو البعض منها، وذلك عبر مهمتين أساسيتين هما: استخراج الكلمة أو العبارة الدالة على الآراء والمشاعر، ثم تصنيفها حسب توجيهها إما إيجابية أو سلبية أو محايدة. [30]

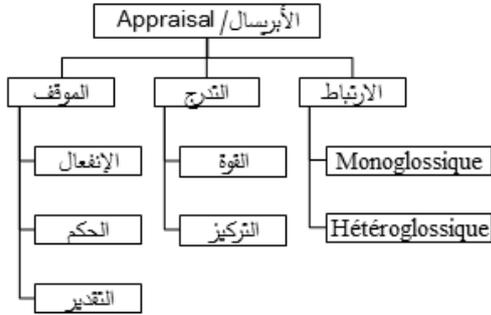
#### 4. معاجم تحليل الآراء والمشاعر:

##### أنواع معاجم تحليل الآراء والمشاعر:

أكدت الدراسات المنجزة أن هناك أربعة أنواع كبرى من المعاجم الخاصة بمجال تحليل الآراء والمشاعر وهي: معاجم أحادية اللغة، معاجم ثنائية اللغة، معاجم متعددة اللغات، ومعاجم متعددة المجالات:

**المعاجم الأحادية اللغة:** يضم المعجم الأحادي اللغة مجموعة من الكلمات أو العبارات الدالة على الآراء والمشاعر للغة معينة، كما

السابق الذكر بالتركيز على ثلاثة أنظمة رئيسية [32]. الشكل (1) وهي: نظام الموقف، ونظام الارتباط، ثم نظام التدرج، ونلخص فيما يلي هذه الأنظمة الثلاثة



الشكل (1): نظام الأبريسال [32]

- **نظام الموقف:** يتضمن مجمل التصنيفات التي إدراجها الباحثان في المعجم الدال على المشاعر، والتعبير عن العواطف، والأحكام، والتقدير، والسلوكيات والتصرفات، وكذلك تقييم الأشياء وغيرها، حيث جمعت في هذا النظام مجموعة من الكلمات الدالة على الآراء والمشاعر ثم صنفت وفق ثلاث حقول معجمية رئيسية يتفرع عن كل منها حقول فرعية عدة.

- **نظام التدرج:** يضم هذا النظام فعل التثمينات التي تمثل خاصية عامة للتدرج، ويسمح بتفاوت شدة الخصائص المتعلقة بدرجة الإيجابية والسلبية المرتبطة بالحكم، هذه الخاصية تتوافق مع الفئات الثلاثة للموقف المتمثلة في **العاطفة والحكم والتقدير**، كما يتوافق مع نظام الارتباط.

- **نظام الارتباط:** وهو يحيل على دلالات المتحدثين الذين يعترفون أو ينكرون اختلاف وجهات النظر حسب مواقفهم الخاصة وانطلاقاً من أماكن تواجدهم. وقد اعتمد أحد الباحثين هذه النظرية لتوضيح هذا الفرق على مفهومين أساسيين وضعهما ميخايل بختين هما «Monoglossic» و«Heteroglossic».

«Heteroglossic» يعمل على وصف تواجد الخطاب بين أسنة مختلفة ضمن سنن لساني واحد، وهذا الخطاب قد يشمل خطاب الشخصيات، خطاب الراوي، وخطاب الكاتب، في حين أن «Monoglossic» يهتم بتواجد الخطاب ضمن سنن واحد [32].

وبالنظر إلى نظام الموقف باعتباره محور دراستنا فقد قسمه كل من مارتن وويت تقسيماً أولياً إلى مجموعة من النظم الفرعية (أنظر الشكل (2))، لكن بعد قيامهما بتطبيقات على بعض النصوص تبين لهما أن بعض النظم الفرعية يمكن تقسيمها إلى نظم فرعية أخرى (أنظر الشكل (3))،

الكلمة موجبة إذا كانت قريبة من الكلمة الأصل، وتكون سالبة إذا كانت بعيدة عنها، وقد طبقت هذه المقاربة على مجموعة من المدونات الكبيرة الحجم.

#### المنهج الذي يعتمد على القاموس:

يعتمد في هذه المقاربة على القواميس أو الموارد اللسانية للمرادفات والأضداد مثل "وردنت" لتحديد التوجيه الدلالي لكلمات جديدة فانطلاقاً من كلمات منفصلة لها قطبيتها، يمكن الحصول على كلمات جديدة لها نفس قطبية الكلمات الأصلية باستعمال خوارزمية Bootstrapping الذي يُمكن من ربط الاتصال بالمرادفات والأضداد للمورد اللساني.

أما فيما يخص تصنيف كلمات المعجم فقد تعددت طرق تصنيفه والتي سنوضحها مختصرة في النقاط التالية:

✓ تصنيف المعجم وفق قطبية ثنائية: حيث يتم تصنيف كلمات المعجم إلى كلمات موجبة وكلمات سالبة.

✓ تصنيف المعجم وفق قطبية ثلاثية: حيث يتم تصنيف كلمات المعجم إلى كلمات موجبة وكلمات سالبة وكلمات محايدة.

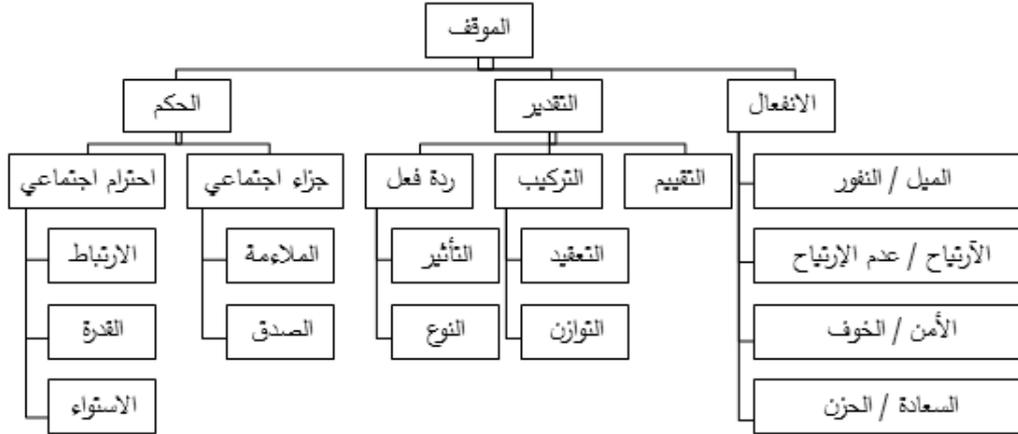
✓ تصنيف وفق الحقول المعجمية: ويتم في هذا النوع من التصنيفات اعتماد على مجموعة من حقول معجمية لها ارتباط بالآراء والمشاعر.

وقد اخترنا اعتماد التصنيف وفق الحقول المعجمية في سبرورة بناء معجمنا، وذلك راجع إلى الدقة التي يتميز بها عن التصنيفات الأخرى لأنه يساعد على تحديد وجهة نظر المتتبع تحديداً دقيقاً.

#### 5. نظرية الأبريسال

تعتبر نظرية الأبريسال من بين النظريات التي اهتمت بتحليل مشاعر الإنسان منذ القرن التاسع عشر ضمن حقل علم النفس وقد تم إحيائها من قبل كل من مارتن وويت باستثمارها لدراسة العلاقات الشخصية باعتبارها وظيفة من الوظائف الثلاثة التي جاء بها كل "هاليداي وماتيسين" (M. A. K. Halliday and C. M. I. M. Matthiessen) في إطار النحو الوظيفي [31]، وهي: وظيفة فكرية؛ وظيفية العلاقات الشخصية؛ والوظيفية النصية.

ومن بين هذه الوظائف الثلاثة اختار كل من الباحثين مارتن وويت التركيز على وظيفة العلاقات الشخصية في كتابهما المعنون ب: «The language of evaluation: Appraisal in English» كهدف لتبني نظرية الأبريسال، وتطبيقها على اللغة الإنجليزية فبعد خمس عشرة سنة من البحث استطاع الباحثان من إصدار الكتاب السابق الذكر سنة 2005 حيث وجها نظرية الأبريسال للوجهة الصحيحة من خلال المؤلف



الشكل (2): الحقول المعجمية لنظام الموقف والمعتمدة في الأبحاث الحالية.

وضع الكلمة ضمن نظامها فقط، بل هناك معايير أخرى تتمثل في: تحديد قطبيتها ودرجة قوتها وتوجيهها.

#### 6. بناء الموارد اللسانية:

سنقدم في هذا المحور الكبير أهم الموارد اللسانية التي قمنا ببنائها، حيث سنبدأ بالحديث عن الخطوات المتبعة في بناء المدونة العربية، ثم الانتقال بالحديث عن المراحل المتبعة في بناء المعجم العربي، وبعدها الإشارة إلى أهم الخطوات التي اتبعناها في بناء لائحة حروف المعاني التي تؤثر في توجيه الكلام العربي.

#### مدونات تحليل الآراء والمشاعر:

تقتضي عملية معالجة وتحليل الآراء والمشاعر التوفر على مدونة تضم مجموعة من التعليقات المعبر بواسطتها عن آراء حول قضية أو اتجاه معين، يتم تحميلها تلك التعليقات من عدة مواقع [36]، بهدف استثمارها في اختبار البرمجيات المسخرة لتحليل تلك الآراء، لهذا الغرض فقد تتعدد المناهج المتبعة في بناء المدونات المتعلقة بهذا المجال من بينها:

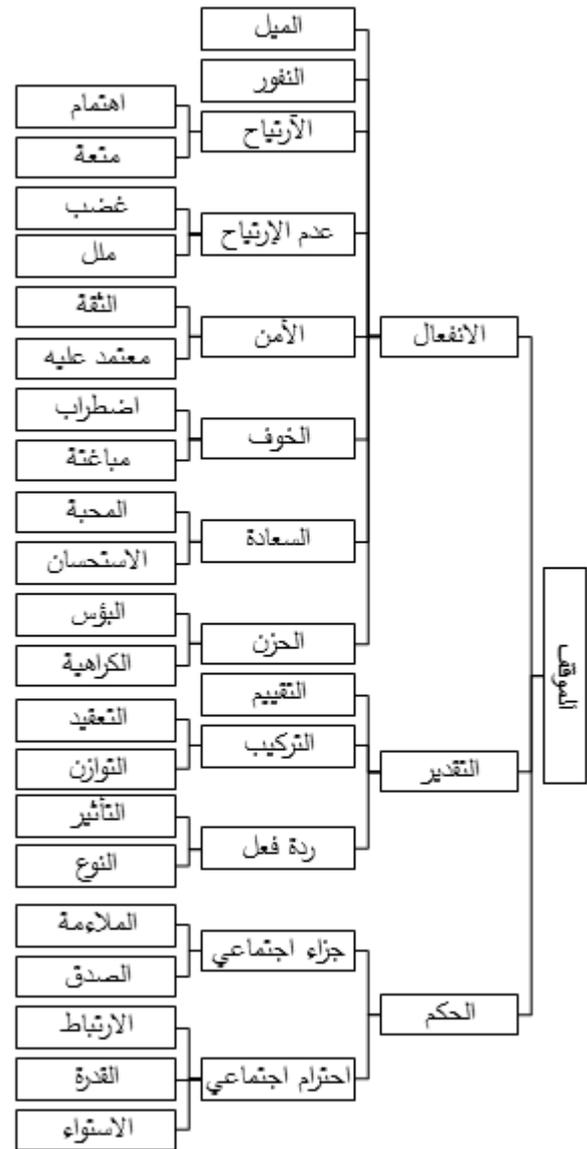
✓ بلورة برامج تعتمد على استعمال واجهة برمجة التطبيقات المعروفة في مجال المعلوماتيات بـ "API Application Programming Interface"، مع الأخذ بعين الاعتبار لهندسة المعمارية لكل موقع.

✓ استعمال البرامج النصية مثل برنامج "Perl".

✓ استعمال برنامج "ParseHub" وهو برنامج يمكن استعماله في عملية تحميل مجموعة من التعليقات من مواقع مختلفة.

#### بناء المدونة العربية:

ركزنا في بنائنا للمدونة على ثلاثة مواقع رئيسية؛ اثنين منها تهتم بمشاركة منخرطيه وتبادل آراءهم حول الكتب التي قرأوها، والتي عادت ما تكون هذه الكتب مصنفة في المواقع حسب عدة تخصصات كالأدب، والرواية، والشعر...، ويتعلق الأمر بكل من



الشكل (3): الحقول المعجمية لنظام الموقف

إلا أن معظم الأبحاث الحالية تعتمد في تصنيف كلماته على التصنيفات الموجودة في الشكل 2، وتصنيف الكلمات لا يقف عن

الجدول (3): عدد التعليقات المأخوذة من موقع (Hespress)

حسب المجالات

المجال	عدد التعليقات
صوت وصورة	3406
رياضة	232
مدارات	1690
حوادث	1411
سياسة	348
المجموع	7087

يبين الجدول (4) مجموع التعليقات التي تم تحميلها من المواقع المختارة، وكما سبق الإشارة إلى أن التعليقات التي تم تحميلها كتبت بأسلوب يطغى عليه في كثير من الأحيان الجانب الذاتي في تقييم الكتب، كما أنه في كثير من الأحيان تجد التعليقات الكبيرة تتضمن بعض الاستشهادات مأخوذة من الكتب المُقيّمة، أضف إلى ذلك الاستشهاد بالآيات القرآنية والأحاديث النبوية، والأمثال العربية، كما يبين الشكل (4) تفاوت التعليقات التي تم كسبها من المواقع الثلاثة السابقة الذكر.



الشكل (4): عدد التعليقات التي تم تحميلها في المواقع المعتمدة

الجدول (4): مجموع التعليقات

اسم الموقع	عدد التعليقات
(Hespress)	7087
(abjjad)	11616
(Goodreads)	81927
المجموع	100630

لقد شكلت التعليقات التي تم تحميلها من موقع (Goodreads) نسبة 82%، وترجع هذه النسبة إلى شهرة الموقع وحصوله على شعبية، أما موقع (abjjad) فقد شكلت عدد التعليقات المأخوذة منه ما يقارب نسبة 10%، في حين أن التعليقات المأخوذة من (Hespress) لم تتعدى 8% من مجموع التعليقات. بعد الانتهاء من عملية تحميل التعليقات من المواقع الإلكترونية السابقة الذكر قمنا بتخزينها ضمن قاعدة بيانات (ACCESS) تضم الحقول الآتية:

موقع (Abjjad)<sup>[38]</sup>، وموقع (Goodreads)<sup>[39]</sup>، كما اخترنا، موقع (Hespress)<sup>[40]</sup> باعتباره موقعا إخباريا يحتوي على مجموعة من المجالات ركزنا على البعض منها.

ولأن جمع هذه التعليقات يدويا سيتطلب وقتا طويلا، فقد اعتمدنا على أحد البرامج الحاسوبية التي تساعد على كشط (scraping) ما هو معروض في المواقع ونخص بالذكر برنامج (ParseHub)، وهو برنامج يقدم للمستعمل مجموعة من الامتيازات من بينها تحميل نصوص أو تعليقات، من مجموعة من المواقع الإلكترونية، بصيغ مختلفة؛ كإكسل (Excel) أو إكسل (XML)<sup>[41]</sup>.

#### طريقة الاشتغال:

انبنى منهجنا في كشط التعليقات باختيار أشهر الكُتاب الذين يحضون بعدد كبير من المتابعين، ثم حددنا الكتب التي حظيت بعدد كبير من التعليقات، فبالنسبة لموقع (abjjad) فقد اخترنا أشهر الكُتاب في تخصصات متنوعة شملت القصص والشعر والروايات، والكتب الدينية وغيرها، والمنهج نفسه اتبعناها في جمع التعليقات من موقع (Goodreads)؛ إذ اخترنا مجموعة من المؤلفين ثم تحميل التعليقات التي نشرت حول كتبهم، وقد بلغت نسب التعليقات التي تم تحميلها من هذا الموقع 82% من مجموع تعليقات المدونة، وذلك راجع لكونه موقع يضم عددا كبيرا من الكتب المتنوعة المكتوبة بلغات مختلفة من بينها اللغة العربية مما جعله يحظى بشعبية متمثلة في استقطاب مجموعة من القراء، ومن جهة أخرى يعتبر موقع (Goodreads) أقدم المواقع التي اهتمت بالكتب العربية والأجنبية، وكان هذان العاملان من أهم العوامل التي دفعتنا للتركيز على هذا الموقع وتحميل عدد كبير من التعليقات.

ويوضح الجدول (2) الآتي عدد التعليقات التي تم تحميلها من موقع (Goodreads) و (Abjjad):

الجدول (2): عدد التعليقات كل موقع.

الموقع	عدد الكتب	عدد التعليقات
(Goodreads)	698	81927
(Abjjad)	401	11729

أما موقع (Hespress) فهو موقع لجريدة مغربية متعددة المجالات، اكتفينا بالبعض منها في عملية تحميل التعليقات، وهذه المجالات هي: "صوت وصورة، رياضة، مدارات، حوادث، سياسة"، وفيما يلي الجدول (3) يفصل عدد التعليقات حسب المجالات المختارة.

الكتب النحوية والبلاغية [42] و [43]، مع إمكانية ورودها في التعبير عن الآراء والمشاعر، وسنورد فيما يلي أهم الخطوات التي تم اتباعها في عملية بناء معجمنا المتمثلة في ثلاث خطوات:

#### الخطوة الأولى: الدراسة القبليّة:

بعد الاطلاع على تعليقات المدونة التي تم بناؤها، تبين لنا أن الآراء والمشاعر يمكن التعبير عنها بالأسماء كما يمكن التعبير عنها كذلك بالأفعال بأنواعها، وكذا الأسماء العربية تتمثل في (المصادر، اسم المفعول والصفة المشبهة واسم الفاعل، وصيغة التفضيل) ويمكن الإشارة إلى أمثلة من هذه التعليقات: "هو حق وحي القلم، وهو عقيدة الرافعي في مختلف نواحي الحياة جمع فيه بين الأدب والتاريخ واللغة والدين فنراه يحسن الوصف ويدغدغ المشاعر تارة وتارة يدافع عن لغتنا العربية، وتارة يسوق من التاريخ ما له فائدة عامة، وهو يمسك في كتابه هذا على القراءة إذ أن لغته أبسط من رسائله الأدبية كما أنه متنوع القوة إذ أنه مجموعة مقالات على مر حياته الأدبية الحافلة فتختلف لغته بتقدمه في السن والعلم" [44].

"الأساتذة يتحملون جزءاً من المسؤولية سوف نرى مزيداً من الإعتداءات عليهم" [40]

وبعد الاطلاع على المدونة واستخلاص أنواع المفردات المستعملة في التعبير عن الآراء والمشاعر انتقلنا إلى البحث في المستوى الصرفي باعتباره المستوى اللساني الذي يهتم بتكوين ودراسة الكلمة العربية، فوجدنا أن الصرفيين العرب القدماء قد عرجوا على المباحث الدلالية التي تساعد في توجيه تشكيل الكلام العربي بما فيه الأبنية الثلاثية المجردة والمزيدة، ومن بين الكتب المعتمدة في الدراسة الدلالية للأفعال هناك:

✓ الرقم التعريفي لكل تعليق (ID).

✓ رقم الإحالة (Reference): أي الرقم الذي يحدد اسم الكتاب وصاحبه.

✓ التعليق.

✓ توجيه التعليق.

والصورة (1) الموالية توضح الحقول السابقة الذكر:

#### تصنيف المدونة:

تمكنا باعتماد برنامج (ParseHub) من جمع ما يزيد عن 100000 تعليق تتفاوت أجناس كُتابها بين الذكور والإناث، ومن مناطق مختلفة في العالم العربي، تتراوح أحجام هذه التعليقات بين:

✓ التعليقات الكبيرة والتي تتراوح عدد كلماتها ما بين 160 و300 كلمة.

✓ التعليقات المتوسطة والتي تتراوح عدد كلماتها ما بين 50 و159 كلمة.

✓ التعليقات الصغيرة والتي تتراوح عدد كلماتها بين 5 و49 كلمة.

بعد ذلك تأتي عملية تصنيف هذه التعليقات إلى تعليقات موجبة أو سالبة أو محايدة، والجدول (5) التالي يوضح مثالاً مأخوذاً من موقع أبجد يفسر تصنيفنا لهذه المدونة:

#### خطوات بناء المعجم:

لتجاوز ما تضمنته المعاجم الموجودة من مكامن النقص، خاصة أن التعبير عن الآراء والمشاعر لا يكتفي بتوظيف الصفة بل بتوظيف غيرها من أقسام الكلام، ارتأينا العمل على وضع مجموعة من المشتقات انطلاقاً من تعاريفها التي تمت الإشارة إليها في

Id	reference	commentaire	orientatic
1	table 0370	هذا الكتاب ووقوعي تحت سطوة سحره في تلك الأيام التي جرى فيها ذلك الخيال. ولكن لو أعدنا السؤال الآن هل سأختار "رجلي الفكرية"؟ ربما نعم...	Neutre
2	table 0370	عظمة أكثر من تفرقت أقصى أكبر من عقل أوسع من حياة هذا الكتاب ... و الكاتب مع كل مؤلف للمسيري أحاطب الواقع : " أنت منذ الآن غيوك "	Positive
3	table 0370	مصطلحاته ومعجمه الخاص وكيفية ابتكاره هذه المصطلحات. وهو ما كان يمكن أن يسبب ارتباك لمن يقرأ كتاب آخر للمسيري قبل أن يقرأ رحلته الفكرية	Neutre
4	table 0370	لا امتلك يوماً شجاعة أن أكتب انا مصطلحاً جديداً وأعطيته معني في أي شيء الا بعد ما اكتسبت هذه الشجاعة من المسيري ، وأصبح المسيري عندي م	Positive
5	table 0370	بأنه كلها من أجل أنكاره ناهيك عن الإحباطات وسوء الفهم أو عدمه الذي تكرر كثيرا سواء في موطنه الأصلي أو بلد الدراسة والتهديدات التي تعرض لها	Positive
6	table 0370	فعل في تنظيمات اللهم الا نظورات سياسية و فلسفية الرت الحركة الإسلامية و اضافت لها رصيد لا يستهان به و ما كان لها مثله الا من رجال كالمسيري	Neutre
7	table 0370	واحد من أمع ما قرأت لحد الآن وأبقى ألا ينتهي، رغم ضخامة حجمه	Positive
8	table 0370	بأننا وحسب، ولست مادة استعمالية تطحنها الحضارة الغربية، وتعيش على دماؤها. حقاً أنشد هذا الرجل بشدة، لم أشعر برجله، إلا أنني أتخسر عليه الآن	Positive
9	table 0370	إدع كتاب مثل هذا! منعة مستعمعة تمنني معاه ان صفحات الكتاب فجأة تزيد ! سأكتب عنه كامل واق في مدونتي ان شاء الله.. سأحدث التعليق هنا	Positive
10	table 0370	و في الجهد و قرأت المثني و كتبت مراجعة عنه واستفدت أيضا استفادة فنحن دارسي الأدب نستمد أصولنا العلمية من كتابات هذا الرجل و	Positive
11	table 0370	أن العقل كالمراة يعكس الواقع، وتبينت نموذجاً توليدياً في رؤيتي للواقع. وهكذا انتقلت من سداجة المادية واختزلتها إلى تركيبة الظاهرة الإنسانية .. حسناً.	Positive
12	table 0370	بإمه الله! وأعتذر أيضا مقدما عن المراجعة التي سأثرثر فيها فالرجل زالني أذهلي فلا أمك وسيلة أخرى - كفتاة- سوى الثثرة للتعبير عن الإعجاب	Positive
13	table 0370	وانتقاده الشديد للنظام الأمريكي المادي فليس كل ما يصدره لنا الغرب نعتبه مالا يجتذى به وغير ذلك من الأمثلة الكثيرة أعجبتني في عبد الوهاب لمسيري	Positive
14	table 0370	بنائية جديدة ثنائية الكتاب الإنسان الذي يختلف عن الطبيعة / للمادة و يتجاوزها في مقابل واجدية الكتاب الطبيعي / الكتاب للمادة رحك الله يا أساتذنا	Positive
15	table 0370	كانت رحلة متمعة و ثرية، صاحبتي طيلة شهر مضى ، يعتبر هذا الكتاب أهم كتاب قرأته هذا العام إلى الآن.	Positive

#### الصورة (1): نموذج من المدونة

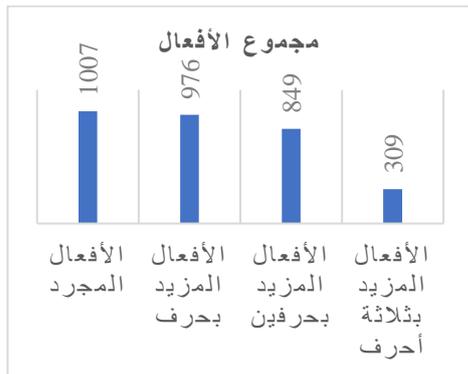
## الجدول (5): جدول إحصاء الحروف في اللاتحتين.

لا تحزن تأليف عائض القرني	
التعليق	التوجيه
سالب	"لولا اختلاف الأذواق لبارت السلع" بصراحة لم يشدني أسلوب الكتابة، بالنسبة لي كان مملا، جيد لكنه لم يترك أثراً في نفسي
موجب	اختصار التعليق على الكتاب إن كنت تريد أن لا تحزن فعليك بقراءته
موجب	كتاب جميل.. فيه من النصح والإفادة كثير جدا يمكن القول إن الاقتباسات فيه كثيرة جدا ودور المؤلف يمكن أن يكون عمل مزج فيه ما بين كل هذا لكن في مجمله مفيد يساعد في كثير من الأحيان على نسيان الحزن فعلاً أنصح به
سالب	قرأته منذ زمن بعيد.. مع توجيه من والدتي لأنها تراه الكتاب الوحيد الذي يجب أن يقرأ حقيقة حتى النجمة أشفق أنني ضحيت بها لأجل الكتاب.. اللغة لغة محاضرات أكثر منها لغة قراءه أولاً / التاريخ العربي والإسلامي عظيم لا نختلف ولكن أن نهدهد على الأكتاف دون أن نزيد الطريق.. هو كعادة الكتب الدينية كتاب تلقيني.. يلقي القصص ويرينا الإنسان العظيم الذي لا نلمسه في الواقع ثانياً   أظنه كتاب تجاري بحت.. وهذا واضح من عدد الطبع وتكرار القصص التي نسمعها وإنما بأسلوب آخر ثالثاً   الكتاب نثري بحت ولا جديد يعطيه للكاتب.. مقارنة بالسيد قطب والغزالي هو لا شيء رابعاً   الكتاب أفكاره متناثرة.. كان من الممكن أن يكون مرتب بشكل أكثر سهولة خامساً   نعود لموضوع حشو الآيات الكريمة والأحاديث النبوية الشريفة.. هناك كثير من الحشو بشكل مزعج سادساً   التقييم الهائل للكتاب أظنه عبثاً.. فهو لا يستحق كل هذا الزخم الانبهارين
موجب	كتاب جميل أنصح الجميع بقراءته
سالب	كلام جميل وعبارات متأقفة لكن قصة ضعيفة جدا

تضم 22856 فعل عربي، وكان ذلك بعد البحث عن معاني الأفعال المحددة باعتماد أحد مواقع معاجم اللغة العربية، ونخص بالذكر "معجم اللغة" وكان الدافع في اختيار هذا الموقع دون غيره من المواقع اعتماده على أربعة من المعاجم القديمة، وثلاثة من الحديثة، ونوردها على الشكل الآتي:

✓ المعجم القديمة: "لسان العرب لابن منظور، تاج العروس للزبيدي، والصحاح في اللغة للجوهري، والمختار الصحاح لمحمد بن أبي بكر بن عبد القادر الرازي.

✓ المعجم الحديثة: المعجم الوسيط الصادر عن مجمع اللغة العربية بالقاهرة، ومعجم اللغة العربية المعاصر لأحمد مختار عمر، ومعجم الرائد لجبران مسعود<sup>[47]</sup> التي يجاوز بين المعاجم العربية القديمة منها والحديثة مع تقديم هذه الأخيرة في عرض النتائج.



✓ "الكتاب" لسبويه أبي بشر عمرو بن عثمان بن قنبر.

✓ "دروس في التصريف" لمحمد محي الدين عبد الحميد.

✓ "أبنية الفعل دلالاتها وعلاقتها" لأبي أوس إبراهيم الشمسان.

هذه الكتب وغيرها بيّنت أن كل وزن من الأوزان الثلاثية له دلالات معينة، يمكن توضيحها على الشكل الآتي: فبنية فعل « بضم العين ولا يكون إلا لازماً، ولا يجيء إلا في أفعال الغرائز والطبائع، نحو: أدب، وأرب، وجنّب، ... »<sup>[45]</sup> وكذلك ما أشار إليه أبو أوس إبراهيم الشمسان حين قال « أما ما كان حسناً أو قبحاً فإنه مما يبني فعله على فعل يفعل وذلك قولك: قُبِحَ يَقبُحُ، وسُمَ يوسُمُ، جُمِلَ يَجُمَلُ »<sup>[46]</sup>، أما بنية فعل « بكسر العين ويجيء لازماً ومتعدياً، إلا أن لزمه أكثر من تعديه، ولذا غلب مجيء الأفعال على النعوت الملازمة وكبر الأعضاء ... ذرب لسانه ... عرج، وعوج ... »<sup>[37]</sup>، وبنية فعل « بفتح العين، من أخف الأبنية، لذا استعملوه للنعوت اللازمة، والأعراض، والأمراض، والألوان، ... نحو: بگى، ... ستر »<sup>[37]</sup>.

## الخطوة الثانية: مرحلة الجمع:

في هذه المرحلة عملنا على اختيار مجموعة من الأفعال التي يمكن اعتمادها في التعبير عن الآراء والمشاعر ضمن قاعدة بيانات أعدها معهد الدراسات والأبحاث للتعبير بالرباط والتي

✓ تحديد توجيه الكلمة ما إذا كانت موجبة أو سالبة أو محايدة.  
✓ تحديد درجة قوة الكلمة؛ هناك كلمات قوية، وأخرى متوسطة، وأخرى ضعيفة، ويتم هذا الوسم بالاعتماد على معاني هذه الكلمات في المعجم المعتمدة.

✓ تحديد قطبية الكلمة، ونشير هنا إلى أن كل كلمات المعجم لها نفس القطبية (unmarked) حيث يتم تغييرها أثناء عملية التحليل إلى (marked) إذا كان التعبير يشتمل على أحد الأدوات التي تغير معنى الكلمة من الموجب إلى السالب أو العكس وللمزيد من التوضيح نقترح الجداول (6 - 7 - 8) توضح كيفية تصنيفنا للمعجم:

#### لائحة حروف المعاني الخاصة بتحليل الآراء والمشاعر:

لقد غُيب اعتماد حروف المعاني في تحليل الآراء والمشاعر، إذ كانت في البداية تعتبر بمثابة كلمات إيقاف (StopWords) يتم حذفها أثناء المعالجة القبلية للنصوص، لكن سرعان ما تم تدارك قيمتها في تغيير توجيه الكلام العربي من قبل بعض الباحثين، حيث عملوا على إنشاء لائحة تضم بعض حروف المعاني العربية اقتصرنا فيها على حروف النفي في حين إلا أن اللغة العربية تشتمل على حروف معاني عدة غير النفي.

لذلك اعتبر إعداد لائحة من هذا النوع من أبرز التحديات التي تواجهها عملية معالجة الآراء والمشاعر العربية آليا، فإذا أخذنا عبارة "أنا أحبك"، فستكون نتيجتها موجبة على اعتبار أن فعل "أحب" مصنفة حسب توجيهه موجب، لكن إذا سبق فعل "أحبك" حرف من حروف المعاني التي تغير توجيهها وجهة معاكسة كأحد حروف النفي مثل: "أنا لا أحبك"، فإن النتيجة ستكون سالبة إذا أخذنا بعين الاعتبار حرف النفي في عملية التحليل، لكن سنظل النتيجة موجبة إذا تم تغيب حرف النفي "لا" [48]، كما يمكن أن تكون لحروف المعاني دورا في تأكيد الكلام العربي ومن ثم الزيادة من قوة العبارة أو الجملة، وكمثال على الجملة السابقة قولنا:

"أني أحبك" ففيها تأكيد للحب، عكس دخول حرف التحقيق مثلا والتي تفيد الاحتمال في قولنا: "قد أحبك" إذ أنه إذا لم يعتد بحرف التحقيق "قد" سيجعل العبارة موجبة، في حين إذا اعتمد حرف "قد" في عملية التحليل سيقدم النتيجة دقيقة والتي تحيل على الحياد، فيما يلي شرح لكيفية إعداد اللائحة.

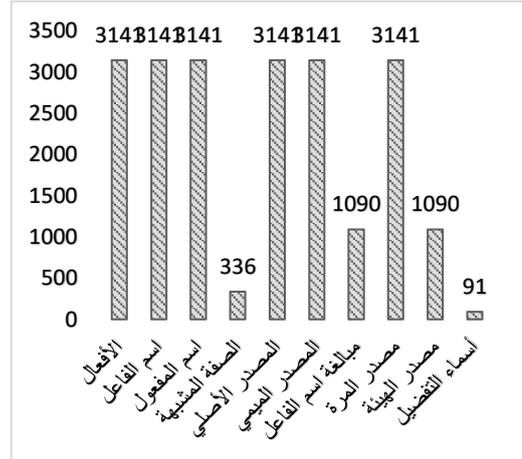
#### إعداد لائحة حروف المعاني الخاصة بتحليل الآراء والمشاعر:

حاولنا من خلال إعدادنا لهذه اللائحة حصر حروف المعاني مفردة ومركبة، إذ ركزنا على الحروف التي تؤثر في تغيير توجيه التعبير من السليبي إلى الإيجابي أو المحايد أو عكس

#### الشكل (5): عدد الأفعال حسب صنفها

هذه العملية مكنتنا من الحصول على 1007 فعلا مجردا و976 فعلا مزيدا بحرف واحد، و849 فعلا مزيدا بحرفين، و309 فعلا مزيدا بثلاثة أحرف، ونلخص مجموع هذه الأفعال في الشكل: (الشكل (5))

كما تم الحصول على مشتقات هذه الأفعال باعتماد برنامج الاشتقاق الذي أنجز بمعهد الدراسات والأبحاث للتعبير، والشكل (6) يوضح الأرقام بشكل مفصل.



الشكل (6): عدد الكلمات حسب كل فئة نحوية

#### الخطوة الثالثة: مرحلة التصنيف

عرفت مجموعة من الدراسات المعجمية تطورا ملحوظا نتج عنه إصدار مجموعة من المعجم المتخصصة التي ساعدت على توجيه الكاتب الوجهة السليمة في اختيار الكلمات الملائمة تبعا للسياق الذي يعبر بها عن آرائه أو مشاعره، الشيء الذي مكنتنا من الاعتماد على بعضها خاصة معجم "تجعة الرائد وشرعة الوارد في المترادف والمتوارد" لإبراهيم اليازجي اللبناني و"المكنز الكبير" لأحمد مختار عمر، (الشكل (7)) بحيث اعتمدنا على سبعة حقول معجمية وهي: "النباس الأمر واتضاحه، اتفاق الرأي، سهولة الأمر/ صعوبة الأمر، الشجاعة والجبن، قوة البنية/ ضعف البنية، حسن المنظر وقبحه، والخلق"، بينما تم الاقتصار فقط على حقلين اثنين هما: "الإيمان/ الكفر، الإرهاب/ التهذئة".

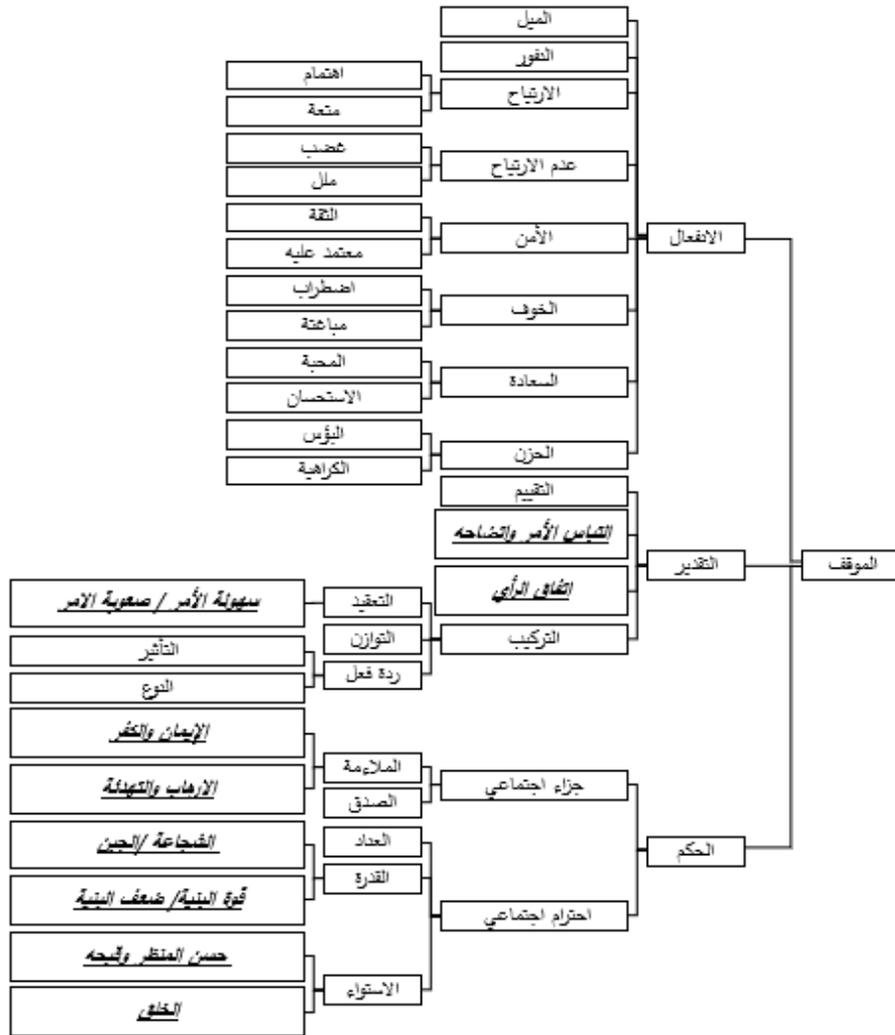
وفيما يلي التصنيف النهائي لكلمات المعجم وفق الحقول المعجمية المذكورة سلفا، مع وسم كل كلمة بالسلمات الموالية:

✓ تحديد جذع الكلمة.

✓ تحديد الخاصية الصرفية لكل كلمة حسب المشتقات التي تم اختيارها وهي: اسم الفاعل، اسم المفعول، الصفة المشبهة، مبالغة اسم الفاعل، المصدر الأصلي، المصدر الميمي، مصدر المرة، ومصدر الهيئة.

✓ "علوم البلاغة: البيان والمعاني والبديع" لأحمد مصطفى المرادي.  
 ✓ "معاني الحروف"، لأبي الحسن علي بن عيسى الرماني النحوي.  
 ارتكز إنشاؤنا على وضع لائحتين؛ لائحة بسيطة تضم حروف المعاني، ولائحة مركبة اشتملت على حروف اللائحة البسيطة مقرونة باللواصق الممكنة، وفيما يلي تفصيل لما اشتملت عليه اللائحتان:

ذلك مثل حروف النفي، والنهي، والاستفهام، والاستثناء، التحسر، ... وغيرها من المعاني التي اتقت عليها مجموعة من الكتب العربية المهمة بحروف المعاني، من بينها:  
 ✓ "الأزهرية في علم الحروف"، لعلي بن محمد النحوي الهروي، عبد المعين الملوحي.  
 ✓ "الجنى الداني في حروف المعاني"، للحسن بن قاسم المرادي.  
 ✓ "رصف المباني في شرح حروف المعاني"، أحمد بن عبد النور المالقي.



الشكل (7): الحقول المعجمية لنظام الموقف بعد إضافة بعض الحقول

• الجدول 6 يوضح الحكم حقل الإيمان والكفر.

الجدول (6): مثال لحق الإيمان والكفر.

Roots	Words	Grammatical Category	Orientation	Force	Polarity
أثم	مُؤْتَمٌ	اسم المفعول	Negative	Median	Unmarked
أثم	أْتَمَ	فعل	Negative	Median	Unmarked
أثم	إِيْتَامٌ	المصدر الأصلي	Negative	High	Unmarked
أثم	مُؤْتَمٌ	اسم الفاعل	Negative	Median	Unmarked
أثم	مُؤْتَمٌ	المصدر الميمي	Negative	Median	Unmarked
أثم	إِيْتَامَةٌ	المصدر المرة	Negative	Median	Unmarked

## • التقدير: الجدول 7 يوضح اتفاق الرأي

الجدول 7: مثال لحق اتفاق الرأي

Roots	Words	Grammatical Category	Orientation	Force	Polarity
ترك	مَنْزُوكٌ	اسم المفعول	Negative	Median	Unmarked
ترك	تَرَكَ	فعل	Negative	Median	Unmarked
ترك	تَرَكَكَ	مبالغة اسم الفاعل	Negative	High	Unmarked
ترك	تَرْكٌ	المصدر الأصلي	Negative	Low	Unmarked
ترك	تَارِكٌ	اسم فاعل	Negative	Median	Unmarked
ترك	تَرْكَةٌ	المصدر المرة	Negative	Low	Unmarked
ترك	تَرْكَةٌ	مصدر الهيئة	Negative	Low	Unmarked
ترك	مَنْزُوكٌ	المصدر الميمي	Negative	High	Unmarked

## • الانفعال: الجدول 8 يوضح حقل المتعة

• الجدول (8): مثال لحق المتعة

Roots	Words	Grammatical Category	Orientation	Force	Polarity
ألق	مُتَأَلِّقٌ	اسم المفعول	Negative	Median	Unmarked
ألق	تَأَلَّقَ	فعل	Negative	Median	Unmarked
ألق	تَأَلَّقٌ	المصدر الأصلي	Negative	Low	Unmarked
ألق	مُتَأَلِّقٌ	اسم فاعل	Negative	Median	Unmarked
ألق	تَأَلَّقَةٌ	المصدر المرة	Negative	Low	Unmarked
ألق	مُتَأَلِّقٌ	المصدر الميمي	Negative	High	Unmarked

## اللائحة المركبة:

تعتبر اللغة العربية لغة غنية صرفيا إذ يتمثل غناها في خاصية الإصاق؛ حيث يمكن لكل من الاسم والفعل والحرف إصاقها بمجموعة من اللواصق، لهذا السبب تطلب الأمر وضع لائحة مركبة حاولنا من خلالها إصاق جميع حروف اللائحة البسيطة بالسوابق واللواحق الممكنة، وكمثال على ذلك نأخذ عبارة "فلولاكم"، الذي يتضمن سابقة العطف "ف" ولاحقة ضمير جمع المذكر المتكلم "كم"، واسم الاستثناء "لولا"، ولهذا الغرض، ومن أجل الدقة في المعلومات فقد استعنا ببرنامج ساعد على إصاق حروف اللائحة البسيطة بكل السوابق واللواحق العربية، فبلغ عدد الحروف مقرونة باللواصق أزيد من 350244، فقمنا بعد ذلك بالتحقق من المستعمل منها بالاستعانة بمجموعة من المراجع أهمها: القرآن الكريم، وأشعار العرب، كما استعنا ببعض لوائح كلمات الإيقاف المعتمدة في عملية التحليل الآلي للنصوص العربية. وفيما يلي، جدول (الجدول 10) يوضح عدد الحروف في اللائحتين معا:

الجدول (10): جدول إحصاء الحروف في اللائحتين.

عدد الحروف	
44	اللائحة البسيطة
4848	اللائحة المركبة

## اللائحة البسيطة:

ركزنا في هذه اللائحة على جمع مجموعة من الحروف حسب معانيها من المؤلفات العربية السابقة الذكر، ومن بين هذه المعاني:

- ✓ حروف الاستثناء؛
- ✓ حروف الاستقهام؛
- ✓ حروف الشرط؛
- ✓ حروف النفي؛
- ✓ حروف التحسر والتمني؛
- ✓ حروف التأكيد؛
- ✓ حروف الترجي؛

وبعد الانتهاء من جمع الحروف قمنا بشكلها لنخلص باللائحة إلى سطرين؛ يضم السطر الأول الحروف مشكولة، والسطر الثاني يضم الحروف بدون شكل، والجدول 9 يوضح بعض الأمثلة على ذلك.

الجدول 9: بعض الحروف المتضمنة في اللائحة الأولى.

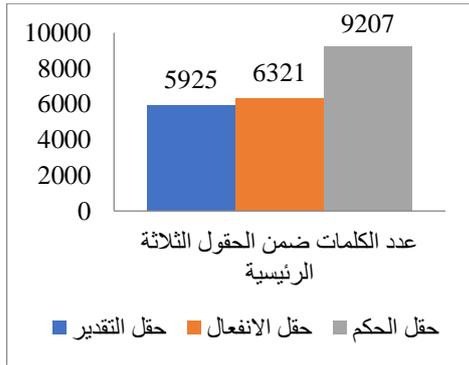
مع الشكل	بدون شكل
لُؤْلَا	لولا
لُم	لم
لُن	لن
لا	لا
مَا	ما
هَلْ	هل

ما يقارب 21453 كلمة، موزعة بين الحقول المعجمية الثلاثية الرئيسية: حقل الإنفعال، وحقل التقدير، وحقل الحكم (الجدول 13)

الجدول 13: عدد الكلمات وفق الحقول المعجمية الثلاثة الرئيسية

الحقل المعجمي	حقل الإنفعال	حقل التقدير	وحقل الحكم	المجموع
عدد الكلمات	5925	6321	9207	21453

ويبين الرسم البياني نسبة تفاوت عدد الكلمات حسب الحقول الثلاثة الرئيسية (الشكل 8):

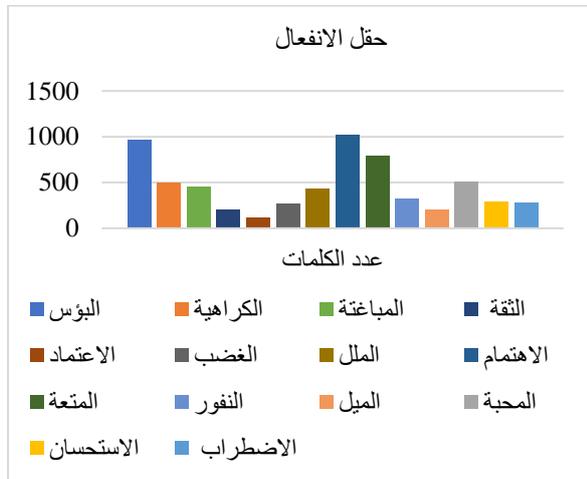


الشكل 8: عدد الكلمات ضمن الحقول الثلاثة الرئيسية

وفيما يلي رسوم بيانية توضح نسب تفاوت الحقول المعجمية الفرعية التي تنفرع عن الحقول الثلاثة الرئيسية ثم الإشارة في الأشكال الآتية، حقل الإنفعال (الشكل 9)، وحقل التقدير، (الشكل 10)، وحقل الحكم (الشكل 11).

ومن جهة أخرى، حاولنا تدارك النقص الحاصل في استعمال حروف المعاني التي من شأنها أن تؤثر في الكلام العربي، بل وتغيير توجيهه من تعبير موجب إلى تعبير سالب، أو العكس، أو ربما جعله محايداً؛ فالكره كمثل دال على توجيهه سالب،

#### ○ حقل الإنفعال:



الشكل 9: مجموع حقل الإنفعال

#### 7. النتائج والأبحاث المستقبلية:

##### النتائج:

عملنا في هذا البحث على تقديم الخطوات المتبعة في بناء الموارد اللسانية الخاصة بمجال تحليل الآراء والمشاعر؛ فكانت البداية الحديث عن المدونة وكما سبقت الإشارة إلى أننا وسمنا من هذه المدونة ما يقارب 25% أي حوالي 24800 تعليقا وسمت وسمنا يدويا توزعت على المواقع الثلاثة. شملت هذه التعليقات تنوعا بين التعليقات الموجبة، والتعليقات السالبة، والتعليقات المحايدة، والجدول (11) يوضح عدد التعليقات التي تم وسمها حسب المواقع الثلاثة:

الجدول (11): عدد التعليقات الموسومة يدويا حسب كل موقع

اسم الموقع	عدد التعليقات الموسومة يدويا
goodreads	15000
abjjad	8000
hespress	1800
المجموع	24800

أما من حيث نسب كل من التعليقات الموجبة والتعليقات السالبة، والتعليقات المحايدة، فنلخصها في الجدول (12):

الجدول (12): نسب التعليقات الموسومة يدويا.

نوع التعليق	عدد التعليقات	النسبة المئوية
تعليقات موجبة	14888	60%
تعليقات سالبة	7440	30%
تعليقات محايدة	2480	10%
المجموع	24800	100%

وبعد ذلك انتقلنا للحديث عن خطوات بناء ووسم معجم عربي اعتمادا على نظرية الأبريسال؛ إذ حاولنا من خلاله تدارك النقص الحاصل في المعاجم المتوفرة التي تُرجمت إلى اللغة العربية باعتماد الترجمة الآلية، كمنهجية لبناء معاجم موجهة لتحليل الآراء والمشاعر باللغة العربية مع إغفال خصوصية كل لغة، أضف أن أغلب المعاجم اقتصرت على الصفات والصفات القليلة منها من أضاف بعض الأفعال،

من هذا المنطلق ارتكزت فكرتنا على بناء معجم عربي موجه لتحليل الآراء والمشاعر باعتمادنا ثلاث خطوات أساسية؛ تمثلت الأولى في الاطلاع على المباحث الصرفية وما لها من دور في بناء المعجم العربي، أما الثانية فتجلت بجرد ما يقارب 3141 من الأفعال الثلاثية المجردة والمزيدة التي يمكن استعمالها في التعبير عن الأحاسيس والآراء والمشاعر من قاعدة البيانات التي أعدها معهد الدراسات والأبحاث للتعريب بالرباط ضمت أكثر من 22000 فعل، وانتقلنا بعد ذلك لإثراء المعجم اعتمادا على عملية الاشتقاق التي وظفنا فيها برنامج اشتقاقي مكنا من الحصول على

الجدول 14: عدد الحروف المفردة والمركبة التي تشتمل عليها

اللائحة		
عدد الحروف غير مشكولة	عدد الحروف مشكولة	الحروف
44	44	الحروف المفردة
1487	1487	الحروف المركبة باعتماد المصادر العربية
937	937	الحروف المركبة باعتماد اللوائح المعتمدة في المعالجة الآلية للغة العربية

أما فيما يخص

#### الأبحاث المستقبلية

في إطار الرقي باللغة العربية والمضي بها قدما في مجال تحليل الآراء والمشاعر، ومن خلال هذه الموارد التي قمنا ببنائها، فإننا نأمل من خلال الأبحاث المستقبلية إثراءها؛ فأما المعجم فسناحاول إثراء بالجدوع الرباعية ثم تصنيفها وفق الحقول المعجمية المعتمدة.

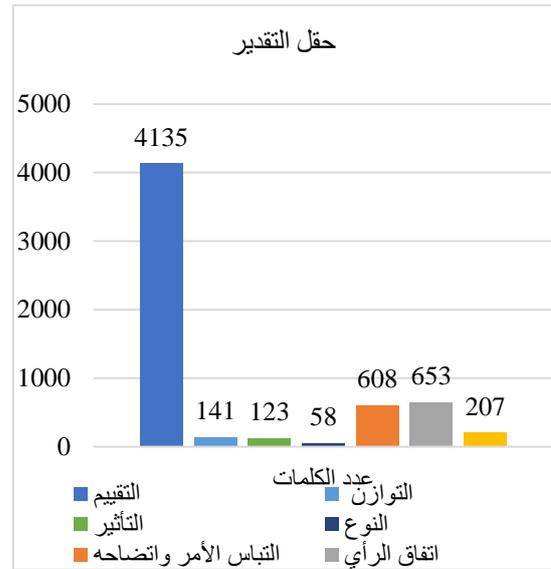
وفي إطار وضع تصور لساني واضح نهدف من خلاله توفير الموارد قصد استثمارها في بناء البرامج الخاصة بتحليل الآراء والمشاعر، فإننا سنحاول تحديد قواعد عمل حروف المعاني العربية دلاليا من خلال استقراء ما جاء في الكتب اللغوية والبلاغية قصد تسهيل عمل برمجتها واعتمادها في البرامج الخاصة بتحليل الآراء والمشاعر.

وبما أن المدونات التي تحظى بمصادقية هي التي تكون موسومة وسما كاملا، فإننا سنعمل على استثمار مقاربة تعليم الآلة لوسم ما تبقى من التعليقات، ثم مراجعتها يدويا ووضعها مفتوحة المصدر ليتمكن الباحثون من استثمارها في أبحاثهم المستقبلية.

#### 9. الخاتمة:

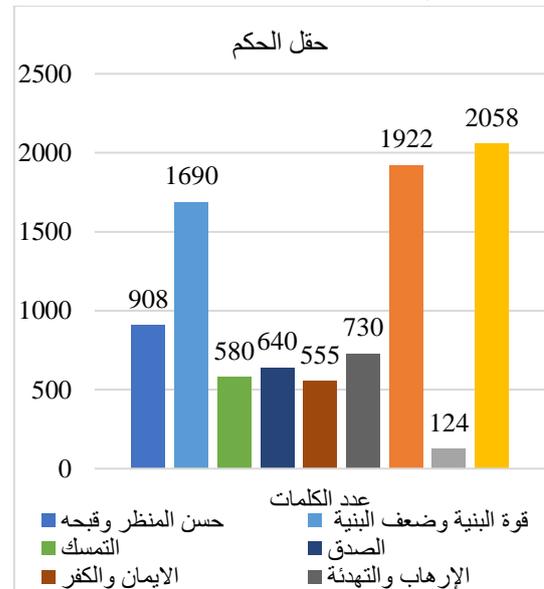
قدما في هذه الورقة تعريفا بمجال تحليل الآراء والمشاعر إذ يعتبر من بين المجالات البحثية الحديثة ضمن مجال المعالجة الآلية للغات الطبيعية، فمن خلال الاطلاع على مجموعة من الأبحاث تبين بأن البحث في مجال تحليل الآراء والمشاعر يتطلب في البداية توفير مجموعة من الموارد اللسانية التي من شأنها إنجاح عملية تحليل الآراء والمشاعر، تكمن أهم هذه الموارد في توفير معجم يضم كلمات وتعبيرات توظف في التعبير عن الآراء والمشاعر، ولوائح تشمل حروف المعاني المؤثرة في وسم التعليقات، بالإضافة إلى مدونة موسومة تشتمل على تعليقات يتم تحميلها من عدة مواقع لغرض توظيفها في اختبار البرنامج، فلاحظنا أن اللغات الهندوأوربية، خاصة منها اللغة الإنجليزية، قد حققت تقدما كبيرا في توفير هذه الموارد، بينما لازالت اللغة العربية تقنقر لمثل هذه الموارد، لهذا الغرض حاولنا المساهمة في توفير

○ حقل التقدير:



الشكل (10): مجموع حقل التقدير

○ حقل الحكم:



الشكل (11): مجموع حقل الحكم

إذا دخل عليه حرف النفي "لا" مثلا يصبح موجبا، أما إذا دخل عليه حرف تحقيق "قد"، فيصبح محايدا، لهذا الغرض قمنا بجمع مجموعة من الحروف وإصاقها باللواصق العربية، ثم التأكد من استعمالها باعتماد مجموعة من المصادر والمراجع العربية، ثم اعتماد بعض اللوائح المعتمدة في عملية المعالجة الآلية للغة العربية، ونظرا لغياب الشكل في مجموعة من الكتابات العربية، وبروزه في البعض، لذلك قمنا بوضع لائحتين؛ الأولى تشتمل على الحروف مشكولة، والثانية تضم الحروف بدون شكل، وقد بلغ مجموع الحروف 9784 حرفا، والجدول 14 يوضح عدد الحروف التي تم جمعها والتأكد من استعمالها حسب المصادر المعتمدة:

[10] A. Esuli, "Automatic generation of lexical resources for opinion mining: models, algorithms and applications," in *ACM SIGIR Forum*, 2008, vol. 42, pp. 105–106.

[11] S. Alhazmi, "Arabic SentiWordNet in Relation to SentiWordNet 3.0," p. 11, 2013.

[12] A. karima and G. Y. Tlili, "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: [http://www.citala.org/citala2014/papers/paper\\_50.pdf](http://www.citala.org/citala2014/papers/paper_50.pdf).

[13] A. Amar and A. Lachkar, "Vers la Construction d'un Lexique d'Appraisal en Langue Arabe pour l'Analyse des Sentiments dans les Réseaux Sociaux," *Ressour. Langagières L'arabe Pour TAL Constr. Stand. Gest. Exploit.*, p. 6, 2015.

[14] S. Kiritchenko, S. Mohammad, and M. Salameh, "SemEval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases," 2016, pp. 42–51.

[15] R. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment Analysis in Arabic tweets," in *2014 5th International Conference on Information and Communication Systems, ICICS 2014*, Apr. 2014, pp. 1–6.

[16] F. T. AL-Khawaldeh, "A Study of the Effect of Resolving Negation and Sentiment Analysis in Recognizing Text Entailment for Arabic," p. 5.

[17] M. Rushdi-Saleh, M. T. Martín-Valdivia, L. A. Ureña-López, and J. M. Perea-Ortega, "OCA: Opinion corpus for Arabic," *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, no. 10, pp. 2045–2054, Oct. 2011.

[18] M. Nabil, M. Aly, and A. Atiya, "Astd: Arabic sentiment tweets dataset," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2515–2519.

[19] A. Mountassir, H. Benbrahim, and I. Berrada, "A cross-study of Sentiment Classification on Arabic corpora," in *Research and Development in Intelligent Systems XXIX*, 2012, pp. 259–272.

[20] A. Mountassir, H. Benbrahim, and I. Berrada, "An empirical study to address the problem of unbalanced data sets in sentiment classification," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, 2012, pp. 3298–3303.

[21] Asmaa MOUNTASSIR, "Arabic Sentiment Classification: Addressing the problems of Unbalanced Data sets and Huge Dimensionality," Université Mohamed V, Rabat – Maroc, ENSIAS, 2015.

[22] M. Abdul-Mageed, M. T. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, pp. 20–37, 2014.

[23] M. Abdul-Mageed, M. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, Jan. 2014.

[24] M. Abdul-Mageed and M. T. Diab, "AWATIF: A Multi-Genre Corpus for Modern Standard Arabic

البعض من هذه الموارد اللسانية الخاصة باللغة العربية، فكانت بدايتنا في هذه الورقة تقديم الخطوات المتبعة في بناء معجم خاص بتحليل الآراء والمشاعر بالانطلاق من الموروث المعجمي، واعتماد مجموعة من البرمجيات الحاسوبية في تطويره، كما انطلقنا في تصنيف دخلات المعجم من نظرية الأبريسال، ثم إضافة بعض الحقول المعجمية المستمدة من المعاجم العربية المؤلفة وفق الموضوعات، كما أشرنا بعد ذلك إلى أهمية لائحة حروف المعاني في عملية تحليل الآراء والمشاعر، الشيء الذي دفعنا إلى إعداد لائحة حاولنا من خلالها جمع مجموعة من الحروف التي من شأنها أن تؤثر في توجيه الكلام العربي. كما قمنا ببناء مدونة عربية تشتمل على أزيد من 100000 تعليقات مكتوبة باللغة العربية الفصحى تم تحميلها من مواقع مختلفة، وقمنا بوسم ما يقارب 25% من التعليقات يدويا. وتجدر الإشارة في الأخير إلى أن أحد الباحثين في فريق هندسة اللغة العربية ونمذجة التعليم بالمدسة المحمدية للمهندسين يقوم ببلورة نظام خاص بتحليل الآراء والمشاعر باستعمال الموارد التي قمنا بإنشائها، كما تجدر الإشارة إلى أننا نعمل على إثراء المعجم بإضافة بعض الأفعال الرباعية واشتقاقاتها.

#### 9. المصادر والمراجع:

[1] A. Partington, "Corpora and discourse, a most congruous beast," *Corpora Discourse*, pp. 11–20, Jan. 2004.

[2] "NRC Emotion Lexicon." <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm> (accessed Mar. 29, 2020).

[3] A. Abdaoui *et al.*, "ADVANCE: Sentiment, Opinion and Emotion Analysis in French Tweets," in *DEFT: Défi Fouille de Texte*, Caen, France, Jun. 2015, Accessed: Dec. 16, 2018. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01222629>.

[4] A. Abdaoui, J. Azé, S. Bringay, and P. Poncelet, "FEEL: a French Expanded Emotion Lexicon," *Lang. Resour. Eval.*, vol. 51, no. 3, pp. 833–855, Sep. 2017.

[5] <http://www.lirmm.fr/~abdaoui/FEEL.csv> (accessed Mar. 29, 2020).

[6] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis," *Comput. Linguist.*, vol. 35, pp. 399–433, Sep. 2009.

[7] "MPQA Resources." <http://mpqa.cs.pitt.edu/> (accessed Mar. 29, 2020).

[8] A. Bayoudhi, H. Koubaa, H. Ghorbel, and L. Hadrich Belguith, "Vers un lexique arabe pour l'analyse des opinions et des sentiments," in *Proc of the 5th International Conference on Arabic Language Processing CITALA*, 2014, vol. 14.

[9] A. Esuli and F. Sebastiani, "SentiWordNet: a high-coverage lexical resource for opinion mining," *Evaluation*, vol. 17, pp. 1–26, 2007.

- Arabic Microblogs,” in *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Atlanta, Georgia, 2013, pp. 55–64, Accessed: Dec. 18, 2018. [Online]. Available: <http://aclweb.org/anthology/W13-1608>.
- [37] R. Mihalcea, C. Banea, and J. Wiebe, “Learning Multilingual Subjective Language via Cross-Lingual Projections,” in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic, 2007, pp. 976–983, Accessed: Dec. 18, 2018. [Online]. Available: <http://aclweb.org/anthology/P07-1123>.
- [38] “أبجد.” <https://www.abjjad.com/> (accessed Mar. 29, 2020).
- [39] “Goodreads | Meet your next favorite book.” <https://www.goodreads.com/> (accessed Mar. 29, 2020).
- [40] “Hespress - هسبريس جريدة إلكترونية مغربية.” <https://www.hespress.com/> (accessed Mar. 29, 2020).
- [41] “Using ParseHub – ParseHub Help Center.” <https://help.parsehub.com/hc/en-us/categories/202638628-Using-ParseHub> (accessed Mar. 29, 2020).
- [42] لطيفة ابراهيم النجار, دور النبوة الصرفية في وصف الظاهرة النحوية وتقعيدها, الأولى. عمان - الأردن: دار البشير, 1994.
- [43] عدنان عبد الكريم جمعة, and ميثاق علي عبد الزهرة الصيمري, “أبنية المشتقات في نهج البلاغة دراسة دلالية,” جزء من متطلبات نيل درجة ماجستير في اللغة العربية وآدابها, كلية الآداب, البصرة, 2002.
- [44] “الوحي القلم.” [https://www.goodreads.com/work/best\\_book/2960155](https://www.goodreads.com/work/best_book/2960155) (accessed Mar. 29, 2020).
- [45] محمد محي الدين عبد الحميد, دروس التصريف. بيروت - لبنان: المكتبة العصرية للطباعة والنشر, 1995.
- [46] إبراهيم الشمسان أبو أوس, أبنية الفعل دلالاتها وعلاقتها, الأولى. جدة: دار المدني, 1407.
- [47] S. SARL, قاموس معاجم قاموس عربي عربي, معجم عربي, “عربي” <https://www.maajim.com> (accessed Mar. 29, 2020).
- [48] B. Heerschoop, P. van Iterson, A. Hogenboom, F. Frasinca, and U. Kaymak, “Analyzing Sentiment in a Large Set of Web Data While Accounting for Negation,” in *Advances in Intelligent Web Mastering – 3*, vol. 86, Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2011, pp. 195–205.
- Subjectivity and Sentiment Analysis.,” in *LREC*, 2012, pp. 3907–3914.
- [25] H. ElSahar and S. R. El-Beltagy, “Building Large Arabic Multi-domain Resources for Sentiment Analysis,” in *CICLing*, 2015.
- [26] E. Refaee and V. Rieser, “An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis,” in *LREC*, 2014.
- [27] A. Assiri, A. Emam, and H. Al-Dossari, “Saudi twitter corpus for sentiment analysis,” *Int. J. Comput. Electr. Autom. Control Inf. Eng. World Acad. Sci. Eng. Technol.*, vol. 10, pp. 242–245, 2016.
- [28] H. S. Ibrahim, S. M. Abdou, and M. Gheith, “MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis,” in *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, Jul. 2015, pp. 353–358.
- [29] B. Liu, “Sentiment analysis and opinion mining,” *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [30] N. Boudad, R. Faizi, R. Oulad Haj Thami, and R. Chiheb, “Sentiment analysis in Arabic: A review of the literature,” *Ain Shams Eng. J.*, Jul. 2017.
- [31] F. H. H. Mahyoub, M. A. Siddiqui, and M. Y. Dahab, “Building an Arabic Sentiment Lexicon Using Semi-supervised Learning,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 4, pp. 417–424, Dec. 2014.
- [32] S. R. El-Beltagy and A. Ali, “Open issues in the sentiment analysis of Arabic social media: A case study,” in *2013 9th International Conference on Innovations in Information Technology (IIT)*, Mar. 2013, pp. 215–220.
- [33] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, “Multilingual sentiment analysis: from formal to informal and scarce resource languages,” *Artif. Intell. Rev.*, vol. 48, no. 4, pp. 499–527, Dec. 2017.
- [34] M. A. K. Halliday and C. M. I. M. Matthiessen, *An introduction to functional grammar*, 3rd ed. London : New York: Arnold ; Distributed in the United States of America by Oxford University Press, 2004.
- [35] J. R. Martin and P. R. R. White, *The language of evaluation: appraisal in English*. New York: Palgrave Macmillan, 2005.
- [36] A. Mourad and K. Darwish, “Subjectivity and Sentiment Analysis of Modern Standard Arabic and