مجلة ألف: اللغة، الإعلام والمجتمع، مصنفة في فئة ب

بشير بويجرة سومية وسالمي عبد المجيد - جامعة الجزائر 2 Alger

اللسانيات الحاسوبية: بين التقنيات الحديثة للمعالجة الآلية للغة وممارسات اللسانيات التقليدية

La linguistique informatique : Entre la technologie moderne du traitement automatique des langues et les pratiques de la linguistique classique

Computational linguistics: Between modern technology of automatic language processing and the practices of classical linguistics

تاريخ النشر ASJP	تاريخ الإلكتروني	تاريخ الإرسال	
2024-06-05	2022-06-02	2023-06-15	ASJP Algerian Scientific Journal Platform

الناشر: Edile- Edition et diffusion de l'écrit scientifique

إيداع قانوني: 6109-2014

النسخة الورقية : 2023 06-05

https://www.asjp.cerist.dz/en/PresentationRevue/226

ترقيم الصفحات:519-505

ردمد: 2437-0274

النشر الإلكتروني: https://aleph.edinum.org

تاريخ النشر: 2024-06-02

ردمد-د: 2437 1076-

المرجعية على ورقة

بشير بويجرة سومية وسالمي عبد المجيد، « اللسانيات الحاسوبية: بين التقنيات الحديثة للمعالجة الآلية للغة وممارسات اللسانيات التقليدية» , 2024, 505-519. | Aleph, 11 (3-1)

المرجع الإلكتروني

بشير بويجرة سومية وسالمي عبد المجيد، « اللسانيات الحاسوبية: بين التقنيات الحديثة للمعالجة الآلية للغة وممارسات اللسانيات التقليدية»، : URL 2024 juin 02 mis en ligne le | [En ligne], Aleph https://aleph.edinum.org/9260

© حقوق الطبع والنشر: بشير بويجرة سومية وسالمي عبد المجيد- لا استخدام تجاري -لا تعديل دولي4.0
CC BY-NC-ND

اللسانيات الحاسوبية: بين التقنيات الحديثة للمعالجة الآلية للغة وممارسات اللسانيات التقليدية

La linguistique informatique : Entre la technologie moderne du traitement automatique des langues et les pratiques de la linguistique classique

Computational linguistics: Between modern technology of automatic language processing and the practices of classical linguistics

بشير بويجرة سومية وسالمي عبد المجيد
جامعة الجزائر 2 Alger

المقدمة

تعتبر اللسانيات الحاسوبية من العلوم البينية الحديثة، التي ظهرت منتصف الستينيات في الأكاديمية الوطنية للعلوم في الولايات المتحدة الأمريكية[1]، وهي نتاج تظافر جهود اللسانيين من جهة، والحاسوبيين من جهة أخرى. وإن اتّسمت بدايةَ ظهورها بنوع من التخفّي والخجل[2]، فإنها أصبحت اليوم من أكثر التخصصات تطورا بالنظر للأبحاث المتزايدة لاسيما في مجال معالجة اللّغات الطبيعية، والذكاء الاصطناعي، ومختلف التطبيقات الحاسوبية المستعملة في عديد المجالات العلمية والاجتماعية.

وتعرّف اللسانيات الحاسوبية على أنها علم فرعي ناتج عن علمي اللسان والحاسوب، يهدف إلى تصميم نماذج رياضية للأبنية اللسانية التي تمكن من معالجة اللغة بواسطة الحاسوب، ويعتبرها « Pierre M. Nugues »[3] بمثابة إضفاء الطابع الصوري على النظريات والنماذج اللسانية أو تنفيذها في الآلة (الحاسوب)، إذ يمكن اعتبارها أيضًا وسيلة لتطوير نظريات لسانية جديدة بمساعدة الحاسوب.

لذلك يتم تقديم الفترة الذهبية للتعاون بين اللسانيين والحاسوبيين، في العلاج الآلي للغة، على أنها الفترة التي قدم فيها اللغويون معرفة لسانية مهيكلة للحاسوبيين، فقاموا بدمجها في الأنظمة الحاسوبية[4]؛ على اعتبار أن المعرفة اللسانية التي تم إدخالها بهذه الطريقة زادت من أداء التطبيقات الحاسوبية، وفي المقابل زوَّد هذا التفاعل اللغويين بوسائل علمية أدت بصفة مباشرة إلى تحسين النماذج اللسانية المدروسة.

1. (الفيفي2017 : 5).
2. (Kay 2003 : 17).
3. (Nugues 2014 : 1)
4. (Habert 2004 : 12).

االإرسال2023- 06-15 النشر الإلكتروني 02-06-2024 تاريخ النشر 05-06-2024
https://aleph.edinum.org/9260

فالبحوث اللسانية الحديثة استفادت، من الناحية التقنية، من مساهمة التجهيزات والأدوات والموارد التي يوفرها الحاسوب بصفة عامة، والعلاج الآلي للغة بصفة خاصة؛ كما أنها استفادت أيضا، من ناحية اتساع مجال الدراسات اللغوية، وتنوّع مناهج البحوث اللسانية، الناتجيْن عن حوار المعارف المثمر، الذي حدث بين هذين التخصصين

يدرس هذا البحث الإشكالية المطروحة حول الاختلال بين دور كل من اللسانيين والحاسوبيين في اللسانيات الحاسوبية الحديثة، نتيجة الاستعمال الطاغي لمناهج العلاج الآلي للغة ووسائله وأدواته على حساب القواعد اللغوية الأصيلة، مما يجعلنا نتساءل عن التطورات الحاصلة في مجال اللسانيات الحديثة، خصوصا التقنية منها، التي يمكن أن تكون سببا مباشرا لهذا الاختلال؟ وهل أصبح هناك فعلا أدوار رئيسية وأخرى ثانوية لقطبي هذا العلم البيني الحديث؟

هذه التساؤلات تدفعنا إلى توجيه هذا المقال ليتطرق إلى دور أدوات العلاج الآلي للغة ووسائله في ظهور ما يطلق عليه اللسانيات الأداتية، وما أحدثته من ثورة في مصادر البحوث اللسانية ومناهجها ومجالاتها في السنوات الأخيرة، مما يساعدنا على تمييز منحى الدراسات اللسانية الحديثة وتحليله بين قطبي هذا العلم البيني الهام.

1. ظهور ما يسمى باللسانيات الأداتية

ظهر مصطلح اللسانيات الأداتية «La Linguistique outillée» لأول مرة سنة 2004 في مقال للباحث الفرنسي «Benoît Habert»[5]، ليعبر عن اللسانيات الحاسوبية الحديثة التي تعتمد بصفة أساسية على أدوات، ووسائل، وتقنيات اعتمدها أو فرضها العلاج الآلي للغة.

1.1. تعريف اللسانيات الأداتية

1. التعريف :عرّف « Benoît Habert » اللسانيات الأداتية بأنها « اللسانيات المزودة بأدوات جديدة للملاحظة والحساب والمدعمة بصفة واسعة بالمناهج المعتمدة في مجال العلاج الآلي للغة ». يظهر من هذا التعريف توصيف اللسانيات الأداتية انطلاقا مما ورثته اللسانيات الحاسوبية -من خلال العلاج الآلي للغة- من إدماج لوسائل ومعدات متطورة، تعتمد في الأساس على مقاربات كمية، بالإضافة إلى معطيات تتميز بخاصتي التنوع والضخامة في آن واحد في هذا الإطار يعزى العديد من الباحثين، في مجال اللسانيات الحاسوبية، التطورات الكبيرة التي تحدث هذه السنوات الأخيرة، في اختصاص العلاج الآلي

5. (Habert, 2004 : 5).

للغة، إلى اعتماد هذا الأخير على المناهج الكمية والرياضية[6]، وكذا حجم الوسائل المعقدة والغامضة –أحيانا لغير المتخصصين في الحوسبة– لإنتاج تطبيقات تساعد على تبسيط تناول كميات ضخمة من المعلومات، وتسهيل معالجتها آليا. على هذا الأساس يرى Robert Martin[7] أن تطور علم الدلالة سيكون في المستقبل مرتبطا بأتمتتها «lié à son automatisation» في العلاج الآلي للغة، أين تلعب القواميس الحاسوبية -على الأقل في قسم من هذا التخصص- دورا مهما في التطور المنتظر مستقبلا. وهو تحول تقني مهم، يجعل من الإعلام الآلي –بالطبع كتقنية وليس كإطار نظري– ملازما أو لصيقا لكل البحوث العلمية الحديثة، أين سيكون موجِّها وموجَّها، مؤثِّرا ومتأثرا بنتائج هذه البحوث وإن كانت خارج مجال اختصاصه البحت.

2. مزايا اللسانيات الأداتية: تمنح اللسانيات الأداتية عديد من المزايا، سواء للسانيين أو الحاسوبيين، نذكر منها ثلاثا، نراها الأجدر بالتناول في هذا المقال:

- تعتبر السرعة في تنفيذ الاحتياجات المطلوبة في مجال المعالجة الآلية للغة أحد أهم الميزات الجذابة للسانيات الأداتية. فالقدرة على تقديم الحلول والاستجابة للمشكلات المطروحة من الأهمية بمكان في تطور كل علم حديث، إلا أن السرعة في ذلك غالبا ما تكون حاسمة ومفيدة بدون شك، في ظل التزايد الأسي للمعلومات وما يقابلها من احتياجات لمعالجتها.
- أيضا، تظهر ميزة المرونة في الولوج إلى المعلومات –مهما كان حجمها– ومعالجتها حسب احتياج المستعمل، كميزة رائدة للسانيات الأداتية، تجعل سمة القبول والرضا سائدة لدى المستغلين، خصوصا اللسانيين منهم.
- ميزة أخرى تتيحها اللسانيات الأداتية للباحثين اللسانيين خاصة، تتمثل في إمكانية رسم مخطط سريع لدراسة ظاهرة لغوية معينة؛ إذ بمجرد جمع البيانات تسمح أدوات معالجة المدونات بوضع نهج تقريبي على أساس أنماط معجمية أو نحوية، تمكن الباحث من ضبط كل محددات موضوع الدراسة. كتحديد الظواهر ذات الصلة، ومعرفة طبيعة ومكونات المجموعة، وحصر المشكلات والتباينات الموجودة، ... وغيرها من العناصر المهمة للشروع في أي دراسة لسانية.

6. (Tanguy 2014: 15-23).
7. (Martin 2001: 38).

2.1. الوسائل والتقنيات الحديثة للعلاج الآلي للغة

إن أهم ما واكب ظهور اللسانيات الأداتية توَفُّر مجموعة من المصادر الرقمية في مجال اللسانيات، يمكن الولوج إليها عبر الشابكة؛ كما تم وضع عديد المدونات اللغوية من بعض المؤسسات العلمية والبحثية وبمختلف اللغات.

1.2.1. توفر المصادر الحاسوبية اللسانية

تعد المصادر الرقمية عنصرا هاما في تأسيس وتطوير أدوات المعالجة الآلية للغة، إذ تمثل الوسيلة الأحدث للتعرف على الإشكالات اللغوية، وتمنح الحلول الناجعة لمعالجتها آليا. ونسرد فيما يلي أهم هذه المصادر المتوفرة حسب مجالات العمل في العلاج الآلي للغة من خلال: موارد القواميس الحاسوبية، والشبكات الدلالية، ومصادر التعلم الآلي، وأخيرا نمذجة اللغة.

- مصادر القواميس الحاسوبية: نقصد بالقواميس الحاسوبية هنا المعجم المدوّن[8] « Dictionary » ، إذ يعتبر القاموس الحاسوبي مصدرًا أساسيًا لتطبيقات المعالجة الآلية للغات الطبيعية. فهو يوفر معلومات لغوية دقيقة مثل معاني الكلمات ومرادفاتها وترجماتها وكذا قواعد النحو والتركيب؛ التي يتم استخدامها في تطبيقات المعالجة والتحليل وإنتاج المحتوى بلغة طبيعية، بدقة واتساق. تستخدم التطبيقات الحاسوبية القواميس لأغراض متنوعة، كالترجمة الآلية التي يعمل فيها على توفير ترجمات دقيقة للكلمات والجمل من لغة إلى أخرى. كما أنه يسمح بفحص إملاء الكلمات في النصوص واقتراح تصحيحات في حالة وجود أخطاء ويعد القاموس الحاسوبي أداة أساسية في تحليل هياكل الجمل والتعرف على أجزاء الكلام والوظائف النحوية والصيغ الصرفية للكلمات، مما يدعم التطبيقات في تحليل البنية اللغوية للجمل وفهم معانيها. والنتائج المتحصل عليها حين تنفيذ مختلف التطبيقات ترتبط ارتباطا طرديا بنوعية القاموس من حيث جمعه ووضعه وقدرته على استحضار المواد اللغوية بطريقة سريعة ومرنة. أيضا، ما يميز القاموس الحاسوبي أن مادته أوسع من مادة القواميس التقليدية؛ حيث أنها تنهل من مصادر متنوعة، أبرزها المعاجم الورقية، المدونات النصية، المكتبات، والموسوعات والكتب الالكترونية، خدمة

.8: باعتبار أن هناك من يقسم العمل المعجمي الحاسوبي إلى المعجم الذهني «Mental Lexicon»، الذي يعنى باكتساب وتوليد واستعمال الثروة اللغوية؛ والمعجم اللغوي «Lexicon»، الذي يهتم بجمع الثروة اللغوية؛ وأخيرا المعجم المدون «Dictionary» الذي هو صناعة معجمية يحول الثروة اللغوية إلى مصدر معجمي حاسوبي. أنظر: د. محسن رشوان وآخرون، الموارد اللغوية الحاسوبية، مركز الملك عبد الله بن عبد العزيز، الرياض، الطبعة الأولى، 2019، ص 13 وما بعدها.

المعلومات على الخط، وكذلك مواقع الصحف والمجلات، وغيرها بشكل عام، يعد القاموس الحاسوبي أداة فعالة وحيوية في تعزيز كفاءة ودقة تطبيقات المعالجة الآلية من خلال توفير معلومات لغوية دقيقة.

- الشبكات الدلالية: تُعرف الشبكات الدلالية أيضًا باسم شبكات المعنى، وهي تهدف إلى تمثيل العلاقات الدلالية بين مختلف مفاهيم اللغة المتصلة ببعضها البعض، من خلال مخططات بيانية معرفية؛ بهدف تغذية الخوارزميات بالمعطيات والمفاهيم اللازمة، وبالتالي الوصول إلى الحلول الممكنة والمتاحة، وكذا تمكين محركات البحث من إيجاد البيانات المناسبة في ظل التزايد المستمر لحجم المعلومات المنشورة في شبكة الويب .تلعب الشبكات الدلالية دورًا مهمًا في تطبيقات المعالجة الآلية للغات في مجالات مختلفة، بما في ذلك فهم النصوص آليا، فهي تساهم في تحليل المعنى والمضمون الداخلي للنص وفهمه بشكل أعمق. كما أنها تُستخدم في استخلاص المعلومات من النصوص بأحجامها المختلفة. وتُساعد في تحسين دقة الترجمة الآلية وبرامج استرجاع المعلومات وتحليل المشاعر في النصوص ووسائل التواصل الاجتماعي. كما يمكن استخدام هذه المعلومات لاستخلاص الرأي وفهم العلاقات الاجتماعية بين الأفراد والمجموعات. ويتم استخدام الشبكات الدلالية في العديد من التطبيقات التجارية والطبية والتعليمية واللغوية وغيرها؛ مما يستدعي أن تستمد مُدخلاتها من مصادر متعددة في مختلف المجالات والتي تم تحويلها إلى الشكل المرقمن، وأصبحت متاحة من خلال:

1. مدونات نصية يتم إنشاؤها خصيصًا لمهام الشبكات الدلالية؛
2. قواعد بيانات لغوية يمكن استخدامها لاستخراج العلاقات التي تربط بين المفاهيم؛
3. موارد لغوية تحتوي على معلومات دلالية، مثل WordNet وFrameNet و ConceptNet وما إلى ذلك؛
4. يمكن استخدام منصات جمع البيانات بالتدرج للطلب من المشاركين توصيف العلاقات الدلالية؛
5. القواميس الحاسوبية...

باختصار، تعتبر الشبكات الدلالية مصدرا هاما في مجالات تطبيق اللسانيات الحاسوبية، فهي توفر إطارًا دقيقًا ومنهجيًا لفهم ومعالجة اللغة الطبيعية بشكل أكثر دقة وذكاء.

- مصادر التعلم الآلي: التعلم الآلي هو فرع من فروع الذكاء الاصطناعي، يهدف إلى تطوير نماذج وأنظمة قادرة على اكتساب المعرفة وتحسين الأداء، من خلال تفاعلها مع مختلف البيانات المتاحة. فهو يعتمد على خوارزميات ونماذج رياضية لتحليل البيانات والتنبؤ بالنتائج أو اتخاذ القرارات. ويُستخدَم التعلم الآلي في العديد من مجالات اللسانيات الحاسوبية، مثل التعرف على الصوت والصورة، وتحليل وتوليد النصوص، كإنشاء الملخصات والردود التلقائية وغيرها من أشكال النصوص التي تنتجها الآلة، بالإضافة إلى توصيات المنتجات، وتحليل المشاعر، والتنبؤ بالسلوكيات، وغيرها .ففي سياق معالجة اللغات الطبيعية، يتم استخدام التعلم الآلي في العديد من المهام الأخرى، بما في ذلك التعرف الآلي على الكلام وتحويله تلقائيًا إلى نصوص مكتوبة عن طريق تدريب نماذج تُسهل عملية تفريغ الخطابات وإنشاء الترجمات وتطبيقات أخرى لها علاقة بالكلام .ولتتمكن الآلة من التعلم الآلي، يجب توفر مجموعة كافية من البيانات التدريبية يتم اختيارها بعناية، لضمان تنوعها وتمثيلها الجيد للمشكلة المطروحة؛ وتحقيق نتائج دقيقة وموثوقة. فعلى سبيل المثال لا الحصر يمكن أن تشمل:

1. مقالات الأخبار،
2. منشورات وتغريدات وتعليقات وسائل التواصل الاجتماعي لتدريب نماذج على تحليل المشاعر مثلا وفهم الرأي العام والكشف عن الاتجاهات...
3. الروايات والقصص؛
4. الرسائل الإلكترونية ومحادثات الدردشة؛
5. التقارير الطبية والعلمية؛
6. بيانات المستخدمين التي يتم جمعها في المؤسسات والشركات من خلال استطلاعات الرأي أو النماذج الإلكترونية أو التعليقات أو أي تفاعلات أخرى.

وبناء على ما سبق، يعتبر التعلم الآلي مصدرًا أساسيًا في مجال معالجة اللغات الآلية

- نمذجة اللغة: تعنى نمذجة اللغة بإنشاء نماذج صورية لتمثيل مختلف المواد والقواعد والسمات الخاصة بلغة محددة. فالهدف الرئيس منها هو استخراج الخصائص اللغوية الصرفية والنحوية والدلالية وغيرها، حتى يتسنى تحليل النصوص وفهمها وتوليدها بطريقة آلية، ولا يمكن القيام بهذه المهام إلا بتوفر البيانات اللازمة لتمثيل اللغة تمثيلا شاملا وصارما ومن مصادر مختلفة ومتنوعة، نحو:

1. المدونات أو البنوك النصية التي تم إعدادها من خلال النصوص المستخرجة من مصادر متنوعة مثل الكتب والمقالات الصحفية ومواقع الويب والمنتديات ووسائل التواصل الاجتماعي وغيرها؛
2. المقالات العلمية؛
3. قواعد البيانات اللغوية التي تضم الجمل والكلمات والهياكل اللغوية في لغات مختلفة...

2.2.1. المدونات اللغوية الحاسوبية

تعرّف المدونات اللغوية بصفة عامة بأنها مجموعة من النصوص المكتوبة أو الشفوية والتي تمثل واقعا لغويا معينا، الهدف منها البحث في الظواهر اللغوية وتفسيرها، سواء للغة بعينها أو مجموعة من اللغات المشتركة في بعض خصائصها. على هذا ظهرت المدونات وتم الاعتماد عليها خصوصا في المجال المعجمي ودراسة دلالات الكلمات منذ القدم، حيث كانت أساسا في فهرسة وتبويب وحتى تفسير الكتب المقدسة قديما.

أما ما يسمى بالمدونات اللغوية الحاسوبية فهو تخصيص لهذا التعريف، فقط فيما يتعلق بالمدونات المحوسبة، أو المدخرة أو المكنزة في الحاسوب[9]. ومنذ المليون كلمة التي جمعتها أول مدونة لغوية حاسوبية سنة 1961 بجامعة براون الأمريكية[10]، تتجاوز مدونات محرك قوقل -فقط- حاليا مئات المليارات من الكلمات، جراء حوسبة مختلف المصادر المكتوبة والمنطوقة على الشبكة العنكبوتية[11]؛ فضلا عن مدونات المؤسسات البحثية والعلمية وبمختلف اللغات، ونذكر منها، على سبيل المثال لا الحصر، ما هو تحت تصرف اللسانيين العرب ما يلي:

- مشروع الذخيرة العربية[12] : يمثل هذا المشروع الذي أطلقه عالم اللسانيات الجزائري الأستاذ عبد الرحمن حاج صالح إنجازا أكاديميا مفتوحا يسير على هديه الكثير من المهتمين بأحوال اللغة العربية ومكانتها، كونه مرجعا ومصدرا لمختلف الدراسات والبحوث في جميع الميادين. تشرف عليه جامعة الدول العربية ومقرّه

9. لذا اتخذت لدى الباحثين اللسانيين العرب عدة مسميات منها: الذخيرة اللغوية (عبد الرحمان حاج صالح – الجزائر)، والمدونة النصية (مها الربيعة – السعودية)، والمكنز (عبد الغني أبو العزم – المغرب)، ولسانيات المتون (صالح بن فهد العتيبي – السعودية).

10. ما يعرف بـ «مدونة براون» (Corpus Brown)، التي أنجزها اللغويان التشيكي «هنري كوتشيرا» والأمريكي «نيلسون فرانسيس» بطلب من جامعة براون الأمريكية لتكون أول مدونة لغوية حاسوبية. أنظر: د. محسن رشوان وآخرون، مرجع سابق، ص 57.

11. (Tanguy 2014 : 16).

12. https://www.dhakhira.org/ le 10/05/2023

مدينة الجزائر العاصمة. فالذخيرة العربية هي عبارة عن بنك للنصوص، القديم منها والحديث وما لا يزال يُنشر باللغة العربية. والهدف من هذا المشروع الوصول إلى أهداف الأمة العربية في التقدم والنهضة والقفزة الحضارية بالإسهام في عالم المعرفة المعاصر بشتى أنواعه من خلال رفع مستوى المواطن العربي والنهوض باللغة العربية، وكذلك إعادة إحياء الفكر اللّغويّ التّراثي وتنميته ليتوافق مع متطلّبات حوسبة اللّغة العربية وتعليمها باستعمال الوسائل الحديثة التي تعد إحدى استحقاقات هذا العصر.

- المدونة القرآنية[13] : وهو مصدر حاسوبي لغوي يتضمن مدونة كلمات القرآن الكريم، ويعرض قواعد اللغة العربية وبناء الجملة والصرف لكل كلمة في المصحف الشريف. توفر المدونة القرآنية ثلاثة مستويات من التحليل: الوسم المورفولوجي، والأشجار التركيبية، والأنطولوجيات الدلالية.

2. بين المناهج العددية والقواعد اللغوية

بعد دراستنا لتأثير أدوات العلاج الآلي للغة في وسائل البحث في اللسانيات الحاسوبية، يبقى من المهم تناول التوتر الحاصل في مناهج وطرق البحث، التي أصبح الكثير من الباحثين يرون أنها –في السنوات الأخيرة- ذات صبغة عددية رياضية أكثر منها هيكلية لغوية.

1.2. دوافع سيطرة المناهج العددية

تتمثل دوافع سيطرة المناهج العددية في اللسانيات الحاسوبية الحديثة من خلال التفوق الظاهر للأساليب العددية في التحليل واستخلاص نتائج البحوث بالإضافة إلى ما أصبحت تفرضه البيانات الضخمة من معالجة آلية خاصة.

1.1.2. تفوق الأساليب العددية

على عكس الأساليب التقليدية في المعالجة الآلية للغة والتي كانت تعتمد على كمية معطيات محدودة، وبذلك معالجة هيكلية أفقية للمواد اللغوية أساسها القواعد النحوية والصرفية، ... وغيرها؛ فإن استغلال أحجام ضخمة من البيانات في اللسانيات الحاسوبية الحديثة رافقها، من دون شك، تغيرات جذرية في طريقة معالجة البيانات اللغوية. اعتمدت هذه الطريقة بالأساس على مناهج كمية وإحصائية.

على سبيل المثال، فإن إنشاء نظام تشكيل آلي للغة العربية يعتمد في الطريقة التقليدية على شرح العمليات التي يتعين تنفيذها في شكل قواعد يتم تطبيقها بالاستناد

13 https://corpus.quran.com/ le 15/05/2023.

إلى موارد ثابتة كالمعاجم والقواعد اللغوية .... وما إلى ذلك. أي أننا نقوم بالتطبيق الدقيق والحصري للقواعد النحوية والصرفية على خطوات المعالجة، ومقارنتها بالموارد المعتمدة من أجل الحصول على المعالجة المطلوبة. وهي كما نلاحظ خطوات تتطلب معرفة لغوية متقدمة، وبذلك تدخل خبير بشري في عملية المعالجة.

إن نظام التشكيل الآلي في اللسانيات الحاسوبية الحديثة يلجأ إلى تزويد النظام بنصوص ضخمة تم تشكيلها يدويا (مصادر)، والتي يعتمدها في تحديد مجموعة من الخصائص العامة للوحدات اللغوية، من خلال آلية الاحتمالات الرياضية والتكرارات، حيث يتأسس النظام على هذه الخصائص، التي تصبح القواعد الجديدة للتشكيل الآلي. طبعا دون إهمال القواعد اللغوية التي تعتمدها كثير من نظم التشكيل الآلي الحديثة كأسس ثانوية أو مكملة للمعالجة.

2.1.2. البيانات الضخمة Big data

هي مجموعة من البيانات التي تكون بحجم يفوق أو يصعب معالجته بالطرق التقليدية خلال فترة زمنية معينة؛ ويطلق عليها اسم البيانات الضخمة أو البيانات الكبرى حيث يتم الحصول عليها سواء أكانت فيديوهات، صورا، أصواتا، أو نصوصا بكميات كبيرة وتأتي بطريقه سريعة ومن أماكن متنوعة. وتستخدم هذه البيانات الضخمة في مجالات مختلفة؛ فهي تلعب دورا محوريا في تحسين الرؤى والتنمية والتخطيط الاستراتيجي وصناعة القرار وأتمتة العمليات. وكأغلب مجالات التكنولوجيا، مع انتشار وزيادة استخدام الناس والأجهزة للشابكة ومع تطور الاتصالات والحواسيب أصبحت أنظمة قواعد البيانات التقليدية -التي غالبا ما تكون مخزنة في مزود واحد وأحجامها لا تتعدى عددا من الجيجا بايت- لا تسعها من ناحية الإدارة والتحليل. مما أدى إلى ظهور مصطلح البيانات الضخمة، الذي يقصد به الكم اللامتناهي من البيانات التي أفرزتها الشابكة وسائر الأجهزة المتصلة. فبالرغم من حجمها وتنوعها أصبحت تخزن وترتب وتعالج بسرعات عالية جدا في عدد هائل من المزودات باستخدام أنظمة تتيح استثمارها والاستفادة منها وعلى رأسها Apache Hadoop.

ومع هذا التزايد المستمر لكمية البيانات الرقمية على غرار البيانات اللغوية والنصوص، تطلب الأمر وجود أدوات جديدة وأنظمة فعالة تتلاءم مع طبيعة اللغات، فتحتم على المهتمين استثمار تقنيات هذا المجال والعمل على صناعة مفاهيم دقيقة متوافقة مع متطلبات العصر وإنتاج برامج وآليات لها القدرة على النهوض باللغة، من خلال معالجتها في مختلف مستوياتها واستعمالاتها.

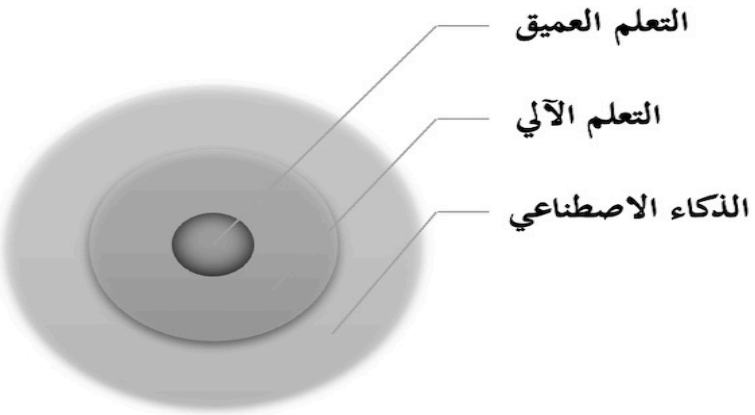
2.2. تراجع الاعتماد على القواعد اللغوية

يمكن القول أن سيطرة السمة العددية الرياضية في الأعمال اللسانية الحديثة قابلها أيضا، تراجع واضح لحضور القواعد اللغوية في هذه الأبحاث، إذ يعزو ذلك لأسباب عدة، نتناول منها ما يتعلق بموضوع مقالنا هذا أي المرتبط بنشاط العلاج الآلي للغة، من خلال ظاهرة التعلم العميق وأدوات ووسائل التحليل اللغوي.

1.2.2. ظاهرة التعلم العميق Deep Learning

هو فرع من فروع التعلم الآلي والذي يعتبر بدوره فرعا من فروع الذكاء الاصطناعي الذي يهدف إلى تمكين الآلة من محاكاة السلوك والأداء البشري. وتعلم الآلة أو التعلم الآلي غايته تحقيق الذكاء الاصطناعي من خلال خوارزميات مدربة على مجموعات كبيرة من البيانات الموسومة والمصنفة مسبقا بطريقة تقليدية، مستخدما في ذلك الشبكات العصبية الاصطناعية. ويلاحظ أن استخدام أساليب التعلم الآلي على النصوص في بعض الحالات يعد مشكلة فعلية نظرا لعمليات التصنيف واستخراج السمات من النصوص التي قد تكون مكلفة وتحتاج عمليات معالجة أولية، خاصة مع تزايد حجم البيانات المراد تحليلها؛ وهذا ما أدى إلى ظهور ما يسمى بالتعلم العميق الذي يستخدم أساليب تمكنه من استخراج مميزات النصوص ومكوناتها ووسمها بطريقة آلية ما يتيح تنبؤا أكثر دقة بالمشكلة المطروحة.

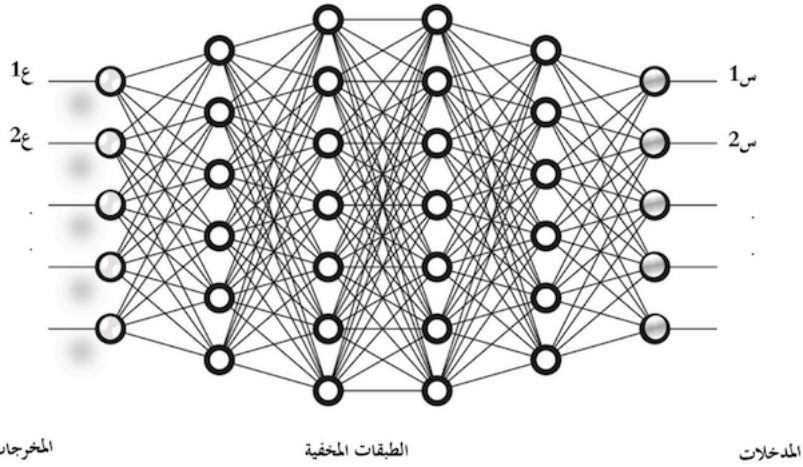
الشكل 1: التعلم العميق14



14. Simplifying The Difference: Machine Learning Vs Deep Learning. 2020. https://www.scs.org.sg/articles/machine-learning-vs-deep-learning. Le 10/05/2023.

يزداد أداء ودقة التعلم العميق كلما زاد حجم البيانات التي يتدرب عليها من خلال مجموعة من الطبقات المخفية على شكل شبكات عصبونية، فكلما زاد عدد هذه الطبقات ازداد عمق هذا النموذج[15].

الشكل 2: طبقات شبكة التعلم العميق16



2.2.2. تغير أدوات ووسائل التحليل اللغوي:

استمرارا للتطورات الحاصلة في مجال العلاج الآلي للغة، من خلال الأدوات والوسائل التي حفزت على إنشاء لسانيات المدونات الحاسوبية، فإن إمكانية معالجة البيانات الرقمية سمحت بالوصول إلى قواعد وترتيبات، خصوصا في مجال التحليل اللغوي، لم يكن من الممكن ملاحظتها عن طريق المعالجة بالعين المجردة أو دراسة اللغة من خلال اختلاف وتنوع مستعمليها (كحالة اللهجات في اللغة العربية).

في هذا الشأن أيضا، يرى (John Sinclair 1991)[17] أن تحليل نصوص ضخمة، مكتوبة ومنطوقة، كشفت من خلال المعالجة الآلية لها عن أنماط لغوية غير متوقعة تماما. مما جعله يصل إلى نتيجة أن اللغة تبدو مختلفة تماما عندما ننظر إلى الكثير منها دفعة واحدة.

15. (السلوم 2022).

16. Training Deep Neural Networks. https://colibris.link/A8bMB

17. (Sinclair 1991: 135).

3. مظاهر التباعد المتزايد بين اللغويين والحاسوبيين في تطبيقات اللسانيات الحاسوبية

كنتيجة لما تقدم، أصبح من الضروري، ليس تأكيد أو نفي التباعد الحاصل بين اللغويين والحاسوبيين في تطبيقات اللسانيات الحاسوبية، الذي أصبح بارزا، ولكن البحث في محصلة هذا التباعد ومنها إلى آفاق الدراسات اللسانية في ظل مجالات ومناهج البحث المستقبلية.

1.3. إشكاليات ومجالات بحث جديدة

تقاس أهمية استغلال البيانات والمعلومات المتزايدة وغير المتجانسة لدى المنظمات والمؤسسات باختلافها (علمية، اقتصادية، إدارية، عسكرية، .... إلخ) بقيمة المعرفة التي يمكن استخلاصها من المعالجة الآلية لكميات كبيرة من الوثائق والنصوص والتسجيلات الصوتية. لذا فقد طرحت هذه المنظمات والمؤسسات إشكاليات جديدة ودعت إلى مجالات بحث غير معهودة في ميدان المعالجة الآلية للغة، نذكر منها على سبيل المثال لا الحصر ما يلي:

- اعتماد مدونات المصطلحات المهنية؛
- تصنيف الوثائق وتحديد المعلومات المهمة فيها؛
- استغلال البيانات النصية التي تعود بالخبرة للمؤسسة (لتفادي الحوادث والأعطال، ...)، إذ تصبح المعالجة الآلية هنا بمثابة مورد أساسي لما يعرف بـ«تسيير المعرفة knowledge management» وهو مقياس أساسي في تطور المنظمات.
- تحسين وتثمين تصفح الوثائق المخزنة؛
- تحليل الوثائق والتسجيلات لمعرفة الانحرافات عن القواعد والإجراءات المهنية، وكذا مصادر الغموض ومشاكل عدم فهم اللوائح المكتوبة والمنطوقة؛
- تحسين قابلية قراءة اللوائح وتطبيقها، لاسيما النصوص ذات الطبيعة الإجرائية؛
- تصميم لغات موحد في حالة المواقف الخطرة[18].

في الحقيقة إن دور اللغويين في الكثير من هذه المجالات جد مفيد وحاسم، برأينا، إذا تم اللجوء إليهم. إلا أن الواقع يبين لجوء المنظمات والمؤسسات إلى الحاسوبيين لحل هذه الإشكالات وأخرى متجددة.

18عمل في هذا الشأن Bouffier (2009) على تسهيل الاطلاع على دليل الممارسة الجيدة في المجال الصحي. أنظر: -Bouffier Amanda, A textual approach for the analysis of health practice guide- lines, Revue Traitement Automatique de la Langue, vol. 50, no 1, 2009, p. 35-59.

2.3. منهجية عمل تجريبية غير اعتيادية

يركز الكثير من الباحثين[19] على التغير الجوهري الذي بدا في التعامل مع المدونات بين اللغويين والحاسوبيين. إذ يدرس اللغوي المدونات على اعتبارها مستودع للأمثلة المستخدمة، والتي من خلالها يمكن اختبار أو قياس مواد لغوية أخرى؛ في حين يستغل الحاسوبي المدونات كمصدر للبيانات، أين يمكن أن نبحث عن قواعد قابلة للتطبيق على نطاق أوسع. تظهر هذه القواعد على سبيل المثال من التكرارات او باستعمال أدوات إحصائية مهيئة على حسب أهداف الدراسة.

تظهر أهمية هذه القواعد، بصفة خاصة، من خلال اعتماد الحاسوبيين عليها في إجراء عمليات تعميم مشابهة تماما لما هو معمول به في العلوم التجريبية الأخرى. كما أن نسب صحة نتائجها تتوافق وحجم وكذا نوعية المدونات المعتمدة.

خاتمة

لقد كان للتطورات الكبيرة في وسائل وأدوات المعالجة الآلية للغة الأثر الواضح في علم اللسانيات الحاسوبية في السنوات الأخيرة، أين فرضت حتمية الاعتماد على الأساليب العددية وكذا البيانات الضخمة. وهو ما خلق فجوة بين قطبي هذا العلم البيني الحديث، إذ أحدثت مسافة بين هندسة اللغة من جهة ومعرفة المواد اللغوية من جهة أخرى.

إن أهمية اللغويين في اللسانيات الحاسوبية الحديثة لا يمكن تجاهلها خصوصا من خلال الأولوية في التصدي للكثير من الإشكاليات والمجالات الجديدة المطروحة من طرف المنظمات والكيانات لاستغلال بياناتها المكتوبة والمنطوقة.

أيضا، يمكن تفادي هذا التباعد الحاصل بين اللغويين والحاسوبيين بالتركيز على تدريب اللغويين على المقاييس الحاسوبية ومتطلبات المعرفة التقنية المعقدة بشكل متزايد؛ وكذا تلقين الحاسوبيين المعارف اللغوية بصفة مكثفة، سعيا للتوفيق بين البيانات المعلوماتية والمعرفة اللغوية واسعة النطاق.

قائمة المراجع

الشريدة، صفاء، حيادرة، مصطفى. 2015. المعجم الحاسوبي أحادي اللغة حقيقته ومصادره وآفاق استخدامه، 2015. العدد1، المجلد 12. مجلة اتحاد الجامعات العربية للآداب

الفيفي، عبد الله بن يحي. 2017. مدخل إلى اللسانيات الحاسوبية. 2017. الطبعة الأولى. الرياض : مركز الملك عبد الله بن عبد العزيز الدولي

السلوم، كمال، أبو صالح، ناصر. 2022. استخدام التعلم العميق لتحليل المشاعر في اللغة العربية، 2022. المجلد 44، العدد 12. مجلة جامعة البعث للأبحاث العلمية، سلسلة العلوم الهندسية.

19. (Bonelli 2001).

رشوان، محسن، وآخرون. 2019. الموارد اللغوية الحاسوبية. 2019. الطبعة الأولى. الرياض : مركز الملك عبد الله بن عبد العزيز

Bonelli, Tognini. 2001. Corpus linguistics at work, Studies in Corpus Linguistics. 2001. vol. 6. Amsterdam: John Benjamins.

Bouffier, Amanda. 2009. A textual approach for the analysis of health practice guidelines. 2009. N° 1, vol. 50. Revue Traitement Automatique de la Langue,

Habert, Benoît. 2004. Outiller la linguistique : de l'emprunt de techniques aux rencontres de savoirs. 2004. N° 1 vol. IX, Revue française de linguistique appliquée.

Kay, Martin. 2003. Oxford Handbook of Computational Linguistics-Oxford University Press. 2003.

Martin, Robert. 2001. Sémantique et automate : L'apport du dictionnaire informatisé. 2001. Paris : Presses universitaires de France.

Nugues, Pierre M. 2014. Language Processing with Perl and Prolog. 2014. London: Springer Heidelberg.

Sinclair, John.1991. Corpus Concordance Collocation. 1991. Oxford: Oxford University Press.

Tanguy, Ludovic. Fabre, Cécile. 2014. Évolutions de la linguistique outillée : méfaits et bienfaits du TAL. 2014. Version 1. Revue l'Information Grammaticale.

https://www.dhakhira.org/ le 10/05/2023

Simplifying The Difference: Machine Learning Vs Deep Learning. 2020. https://www.scs.org.sg/articles/machine-learning-vs-deep-learning. Le 10/05/2023.

Training Deep Neural Networks. https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964. Le 10/05/2023.

https://corpus.quran.com/ le 15/05/2023.

مستخلص

لا تحتاج بحوث اللسانيات في كثير من الأحيان إلى أدوات معينة، إلا أن العلاج الآلي للغة بصفة خاصة والإعلام الآلي بصفة عامة ساهما في إدماج أدوات ووسائل تقنية في دراسات اللسانيات الحاسوبية، إلى أن أصبحت هذه الأدوات جزءا هاما، بل وحتمية لا رجعة فيها، في الأعمال اللسانية خصوصا في مجالات البحث والمناهج وحتى النتائج

دراسة التباعد الحاصل بين قطبي اللسانيات الحاسوبية والذي أحدثته ثورة الأدوات يسمح بتتبع أسبابه ومظاهره ما يمنح الرؤية لتصور سبل التعاون في آفاق الأعمال اللسانية وإعادة التنسيق بين اللسانيين والحاسوبيين

كلمات مفتاحيّة

لسانيات حاسوبية، لسانيات أداتية ، حاسوب، علاج آلي للغة، نمذجة

Résumé

Une grande partie de la recherche en linguistique ne nécessite aucun instrument spécifique. Toutefois, grâce aux progrès du Traitement Automatique de la Langue et de l'informatique, des outils et des méthodes techniques ont été intégrés aux études en linguistique informatique. Ces outils sont désormais devenus une composante essentielle, voire indispensable, des travaux linguistiques, influençant les domaines de recherche, les méthodes utilisées et même les résultats obtenus.

L'étude de l'écart entre les deux pôles de la linguistique informatique, engendrée par cette révolution des outils, permet de comprendre les causes et les manifestations de ce phénomène. Cela ouvre également la voie à une coopération plus étroite entre les linguistes et les informaticiens, en vue de réaligner les perspectives de recherche linguistique et de rétablir une coordination efficace entre les deux domaines.

Mots-clés

Linguistique Informatique, Linguistique Outillée, Ordinateur, Traitement Automatique de la Langue, Modélisation

Abstract

A significant portion of linguistic research does not require any specific instruments. However, thanks to advancements in Natural Language Processing and computer science, tools and technical methods have been integrated into computer linguistics studies. These tools have now become an essential, even indispensable, component of linguistic work, influencing research domains, the methods employed, and even the results obtained.

Studying the gap between the two poles of computational linguistics, brought about by this tool revolution, allows for an understanding of the causes and manifestations of this phenomenon. It also paves the way for closer collaboration between linguists and computer scientists, aiming to realign perspectives in linguistic research and reestablish effective coordination between the two fields.

Keywords

Computational Linguistics, Tooled Linguistic, Computer, Natural Language Processing, Modeling