مجلة اللسانيات الرياضية
Journal of Mathematical Linguistics
المجلـد 2 العـدد1/ديسمبر 2022
ص ص: 100-109

مقاربة اللسانيات الحاسوبية

Computational Linguistics Approach

عائشة بارة [1]*

[1] المركز الجامعي عبد الحفيظ بوالصوف، ميلة، (الجزائر)
baraaicha02@gmail.com

الاستلام: 2021/05/17 القبول:2021/07/15 النشر :2022/12/31

ملخص:

يهدف هذا المقال إلى شرح مجال علم اللغة الحاسوبي وتخصصاته، واه تقامه بالنمذجة الحسابية للغة الطبيعية، ودراسة المناهج الحسابية المناسبة للأسئلة اللغوية. بشكل عام، إذ يعتمد علم اللغة الحاسوبي على علم اللغة وعلوم الكمبيوتر والذكاء الاصطناعي والرياضيات والمنطق والفلسفة والعلوم المعرفية وعلم النفس المعرفي وعلم اللغة النفسي والأنثروبولوجيا وعلم الأعصاب، من بين أمور أخرى.

الكلمات المفاتيح: الحاسوبية، النمذجة، اللغة، الذكاء الاصطناعي، المناهج

Abstract:

This article aims to explain the field of computational linguistics and its specializations، its interest in computational modeling of natural language، and the study of computational approaches appropriate to linguistic questions. In general، computational linguistics draws on linguistics، computer science، artificial intelligence، mathematics، logic، philosophy، cognitive science،

* عائشة بارة

cognitive psychology، psycholinguistics، anthropology، and neuroscience، among others.

Keywords: computing، modeling، language، artificial intelligence، curriculum.

نشأة اللسانيات الحاسوبية:

قدم الحاسوب خدمات مرنة وسخية وجليلة للإنسان في مجالات عديدة، ويعد المجال اللغوي من ابرز تلك المجالات وهذه الاستفادة تزداد يوما بعد يوم. لذلك كان الحاسوب الركيزة الأساسية في هدا الفرع اللساني التطبيقي الحديث المتصل بالذكاء الاصطناعي باعتباره يركز على قواعد المعارف الأعمق والأشمل من قواعد البيانات، فاللغة العربية استفادت منه كثيرا في انجاز تطبيقات لغوية حققت بها قفزة نوعية،مما يفسر أنه لا حلّ لمعضلة اللغة دون اللجوء إلى أساليب الذكاء الاصطناعي وهندسة المعرفة، سعيا إلى عالميتها ووحدتها.

وترجع البداية الأولى للسانيات الحاسوبية عند الغرب إلى فترة ظهور الحاسوب 1948، إذ شكل أداة مسخرة لكل المعارف والمعالجات،لتكون اللغة العربية من اللغات الموجهة إلى المعالجة الآلية، حيث تم في هذه الفترة تحقيق ترجمة آلية باعتماد الحاسوب من لغة مصدر إلى لغة هدف .وقد مثلت اللغة الانجليزية المحطة الأولى للمعالجات الحاسوبية (مهديوي، 2016، صفحة 9).

أما عند العرب فلم يتأت إلا عام 1971، حين تم اعتماد الحاسوب قصد الدراسة الإحصائية للألفاظ، وهي عملية لا يمكن اعتبارها معالجة آلية، وإنما إحصاء لغوي باعتماد الحاسب.

اللسانيات الحاسوبية : له مصطلحات متعددة منها اللغات الحاسوبية،علم اللغة الحاسوبي، اللسانيات الآلية، اللسانيات الإعلامية، أما عبد الرحمان

الحاج صالح فيطلق عليها مصطلح الرتابية (الحاج صالح، 2007، صفحة 231).

اللسانيات الحاسوبية هي ذلك الإطار التقني الذي تنصهر داخله تجليات اللغة الطبيعية وتمظهراتها في تفاعل يحاول ربط اللغة بالحاسوب وأنظمته، أو هي الفرع الذي يبحث في مختلف طرق وكيفيات الاستفادة من قدرات الحاسوب في تحليل اللغة ومعالجتها وتعليمها وتعلمها. في حين نجد دافيد كريستال يعرف اللسانيات الحاسوبية على أنها فرع من الدراسات اللغوية الذي يوظف في تقنيات والمفاهيم الحسابية والآلية بهدف توضيح المشكلات اللغوية. ومنه فاللسانيات الحاسوبية هي محطة تدرس بنية اللغة الطبيعية في ارتباطها بجهاز الحاسوب. وقد عرفت اللسانيات الحاسوبية في أول مؤتمر دولي قد أقيم من اجلها سنة 1965 بأنها علم جديد تتقاطع فيه اللسانيات مع جهاز صوري تفرزه العلوم المنطقية الرياضية، ويخضع للقيود التي تفرضها الآلات المعدة للمعالجة الآلية للمعلومة، ويؤدي البحث في هدا المجال إلى إنشاء نموذج خوارزمي.

إن موضوع اللسانيات الحاسوبية يعتبر عملا لغويا آليا بامتياز، يعالج مثل كيان موحد ومنظم في قلب الحواسيب الالكترونية، كأنه إجراء دقيق لبعض المفاهيم الخوارزمية والرمزية داخلها في ارتباط تام يجمع بين مكونات الحواسيب ووسائطها في قالب يتناسب ومبادئ نظم الحاسوب.

اللسانيات الحاسوبية والمعالجة الآلية للغة: المعالجة الآلية للغة، تهتم المعالجة الآلية للغة بدراسة الجوانب الحاسوبية للغة والمشاكل اللسانية والحاسوبية التي تواجه هذه المعالجة سواء أكانت هذه اللغة منطوقة أو مكتوبة، وبناء نظام معالجة اللغة العربية مهمة معقدة وصعبة، لصعوبة إدماج المعارف الصوتية والصرفية والنحوية والدلالية في هذا النظام .

مجالات اللسانيات الحاسوبية ذات الصلة:

ظهرت اللغويات الحاسوبية كمجال للذكاء الاصطناعي يقوم به علماء الكمبيوتر الذين تخصصوا في تطبيق أجهزة الكمبيوتر على معالجة لغة طبيعية. مع تشكيل جمعية اللغويات الحاسوبية ، وإنشاء سلسلة مؤتمرات مستقلة، تم توحيد المجال خلال السبعينيات والثمانينيات.

وتعرف جمعية اللغويات الحاسوبية علم اللغة الحاسوبي على أنه: الدراسة العلمية للغة من منظور حسابي. يهتم علماء اللغة الحاسوبيون بتوفير نماذج حسابية لأنواع مختلفة من الظواهر اللغوية (Sproat, 2005, p. 12).

ويعتبر مصطلح "اللغويات الحاسوبية" في الوقت الحاضر (2021) مرادفًا تقريبًا لمعالجة اللغة الطبيعية (NLP) وتكنولوجيا اللغة (البشرية). تضع هذه المصطلحات تأكيدًا أقوى على جوانب التطبيقات العملية بدلاً من البحث النظري ومنذ العقد الأول من القرن الحادي والعشرين. في الممارسة العملية، قاموا باستبدال مصطلح "اللغويات الحاسوبية" في مجتمع البرمجة اللغوية العصبية، على الرغم من أنهم يشيرون بشكل خاص إلى المجال الفرعي لعلم اللغة الحسابي التطبيقي.

ويحتوي علم اللغة الحاسوبي على مكونات نظرية وتطبيقية. يركز علم اللغة الحسابي النظري على قضايا في علم اللغة النظري والعلوم المعرفية. يركز علم اللغة الحسابي التطبيقي على النتيجة العملية لنمذجة استخدام اللغة البشرية (USZKOREIT, 2000).

تشمل اللسانيات الحسابية النظرية تطوير النظريات الرسمية للقواعد (الإعراب) والدلالات، وغالبًا ما ترتكز على المنطق الرسمي والنهج الرمزي (القائم على المعرفة). تشمل مجالات البحث التي يدرسها علم اللغة الحسابي النظري ما يلي:

التعقيد الحسابي للغة الطبيعية، المصمم إلى حد كبير على نظرية الأوتوماتا، مع تطبيق قواعد نحوية حساسة للسياق وآلات تورينج المقيدة خطيًا. وتشتمل دلالات الحساب على تحديد المنطق المناسب لتمثيل المعنى اللغوي، وبناءها تلقائيًا والاستدلال بها. و يهيمن التعلم الآلي على اللغويات الحاسوبية التطبيقية، التي تستخدم الأساليب الإحصائية تقليديًا، منذ منتصف عام 2010 من قبل الشبكات العصبية (Socher, 2011)

وقد لاقى هذا البرنامج التعليمي المبكر للتعلم العميق في 2012، اهتمامًا من قبل معظم المشاركين فيه، وتم رفض التعلم العصبي أساسًا بسبب افتقاره للتفسير الإحصائي. حتى عام 2015، ثم تطور التعلم العميق إلى إطار العمل الرئيس للغة البرمجة اللغوية العصبية. بالنسبة للمهام التي تتناولها اللغويات الحسابية التطبيقية، انظر مقالة معالجة اللغة الطبيعية. يتضمن ذلك المشكلات الكلاسيكية مثل تصميم POS-taggers (علامات جزء من الكلام)، ومحللات اللغات الطبيعية، أو مهام مثل الترجمة الآلية (MT)، والتقسيم الفرعي للغويات الحسابية الذي يتعامل مع جعل أجهزة الكمبيوتر تترجم بين اللغات . باعتبارها واحدة من أقدم وأصعب تطبيقات اللغويات الحاسوبية، تعتمد الترجمة الآلية على العديد من الحقول الفرعية والجوانب النظرية والتطبيقية. تقليديا، تم اعتبار الترجمة الآلية للغة فرعًا صعبًا من اللغويات الحاسوبية.

بصرف النظر عن الانقسام بين اللغويات النظرية والتطبيقية، توجد أقسام حسابية أخرى في مجالات رئيسية وفقًا لمعايير مختلفة، بما في ذلك:

وسيلة اللغة التي تتم معالجتها، سواء كانت منطوقة أو نصية: يتعامل التعرف على الكلام وتركيب الكلام مع كيفية فهم اللغة المنطوقة أو إنشائها باستخدام أجهزة الكمبيوتر.

المهمة التي يتم إجراؤها، على سبيل المثال، ما إذا كان تحليل اللغة (التعرف) أو توليف اللغة (الجيل): التحليل والتوليد هما قسمان فرعيان من اللغويات الحاسوبية يتعاملان على التوالي مع تفكيك اللغة ووضعها معًا.

تقليديا، تم وصف تطبيقات أجهزة الكمبيوتر لمعالجة مشاكل البحث في فروع علم اللغة الأخرى كمهام ضمن علم اللغة الحاسوبي. من بين الجوانب الأخرى، وهذا يشمل علم اللغة المدعوم بالحاسوب، والذي تم استخدامه منذ السبعينيات كوسيلة لإحداث تقدم تفصيلي في مجال تحليل الخطاب

المناهج التنموية:

اللغة هي مهارة معرفية تتطور طوال حياة الفرد. تم فحص هذه العملية التنموية باستخدام عدة تقنيات، ومن بينها النهج الحسابي. يوفر تطوير اللغة البشرية بعض القيود التي تجعل من الصعب تطبيق طريقة حسابية لفهمها. على سبيل المثال، أثناء اكتساب اللغة، يتعرض الأطفال إلى حد كبير للأدلة الإيجابية فقط. هذا يعني أنه خلال التطور اللغوي للفرد، يتم تقديم الدليل الوحيد على الشكل الصحيح، ولا يوجد دليل على ما هو غير صحيح. هذه معلومات غير كافية لإجراء اختبار فرضية بسيط للحصول على معلومات معقدة مثل اللغة (Elman، 1993، صفحة 75)، وبالتالي توفر حدودًا معينة لنهج حسابي لنمذجة تطوير اللغة واكتسابها في الفرد.

بذلت محاولات لنمذجة العملية التنموية لاكتساب اللغة لدى الأطفال من زاوية حسابية، مما أدى إلى قواعد نحوية إحصائية ونماذج اتصال. تم اقتراح العمل في هذا المجال أيضًا كطريقة لشرح تطور اللغة عبر التاريخ. باستخدام النماذج، فقد ثبت أنه يمكن تعلم اللغات من خلال مجموعة من المدخلات البسيطة المقدمة بشكل تدريجي حيث يطور الطفل ذاكرة أفضل ومدى انتباه أطول. تم طرح هذا في وقت واحد كسبب لفترة نمو الأطفال البشرية الطويلة.

تم استخلاص كلا الاستنتاجين بسبب قوة الشبكة العصبية الاصطناعية التي أنشأها المشروع.

تم أيضًا نمذجة قدرة الأطفال على تطوير اللغة باستخدام الروبوتات من أجل اختبار النظريات اللغوية. تم تمكينه للتعلم مثل الأطفال، تم إنشاء نموذج بناءً على نموذج تحمل التكاليف حيث تم إنشاء التعيينات بين الإجراءات والتصورات والتأثيرات وربطها بالكلمات المنطوقة. بشكل حاسم، كانت هذه الروبوتات قادرة على الحصول على تعيينات فعالة من كلمة إلى معنى دون الحاجة إلى بنية نحوية، وتبسيط عملية التعلم إلى حد كبير وإلقاء الضوء على المعلومات التي تعزز الفهم الحالي للتطور اللغوي. من المهم ملاحظة أنه لا يمكن اختبار هذه المعلومات تجريبيًا إلا باستخدام نهج حسابي (Salvi, 2012, p. 78).

نظرًا لأن فهمنا للتطور اللغوي للفرد خلال العمر يتم تحسينه باستمرار باستخدام الشبكات العصبية وأنظمة التعلم الروبوتية، فمن المهم أيضًا أن نأخذ في الاعتبار أن اللغات نفسها تتغير وتتطور بمرور الوقت. لقد كشفت الأساليب الحسابية لفهم هذه الظاهرة عن معلومات مثيرة للاهتمام للغاية. باستخدام معادلة السعر وديناميكيات Pólya urn، أنشأ الباحثون نظامًا لا يتنبأ بالتطور اللغوي المستقبلي فحسب، بل يعطي أيضًا نظرة ثاقبة للتاريخ التطوري للغات العصر الحديث. تم تحقيق جهد النمذجة هذا، من خلال علم اللغة الحسابي، ما كان يمكن أن يكون مستحيلًا.

من الواضح أن فهم التطور اللغوي لدى البشر وكذلك عبر الزمن التطوري قد تحسن بشكل خيالي بسبب التقدم في علم اللغة الحاسوبي. إن القدرة على نمذجة وتعديل الأنظمة حسب الرغبة توفر للعلم طريقة أخلاقية لاختبار الفرضيات التي من شأنها أن تكون مستعصية على الحل.

المناهج الهيكلية:

لإنشاء نماذج حسابية أفضل للغة، فإن فهم بنية اللغة أمر بالغ الأهمية. تحقيقا لهذه الغاية، تمت دراسة اللغة الإنجليزية بدقة باستخدام الأساليب الحسابية لفهم أفضل لكيفية عمل اللغة على المستوى الهيكلي. من أهم أجزاء القدرة على دراسة التركيب اللغوي هو توافر مجموعات أو عينات لغوية كبيرة. يمنح هذا اللغويين الحسابيين البيانات الأولية اللازمة لتشغيل نماذجهم واكتساب فهم أفضل للهياكل الأساسية الموجودة في الكم الهائل من البيانات الموجودة في أي لغة واحدة. واحدة من أكثر المؤسسات اللغوية الإنجليزية استشهادا هي بن تريبانك. مستمدة من مصادر مختلفة على نطاق واسع، مثل أدلة كمبيوتر IBM والمحادثات الهاتفية المكتوبة، تحتوي هذه المجموعة على أكثر من 4.5 مليون كلمة من اللغة الإنجليزية الأمريكية. تم شرح هذه المجموعة بشكل أساسي باستخدام علامات جزء من الكلام والأقواس النحوية وقد أسفرت عن ملاحظات تجريبية كبيرة تتعلق ببنية اللغة. كما تم تطوير المناهج النظرية لتركيب اللغات. تسمح هذه الأعمال لعلم اللغة الحسابي أن يكون لها إطار عمل من خلاله لوضع فرضيات من شأنها تعزيز فهم اللغة بعدد لا يحصى من الطرق. اقترحت إحدى الأطروحات النظرية الأصلية حول استيعاب القواعد وهيكل اللغة نوعين من النماذج. في هذه النماذج، تزداد قوة القواعد أو الأنماط المكتسبة مع تكرار مواجهتها (Braine, 1971, p. 6).

خلق العمل أيضًا سؤالًا يجيب عنه اللغويون الحسابيون: كيف يتعلم الرضيع قواعد محددة وغير طبيعية (صيغة تشومسكي العادية) دون أن يتعلم نسخة مفرطة التعميم ويتعثر؟ . الجهود النظرية مثل هذه تحدد اتجاه البحث

للذهاب في وقت مبكر من عمر مجال الدراسة، وهي ضرورية لنمو هذا المجال.

تسمح المعلومات الهيكلية حول اللغات باكتشاف وتنفيذ التعرف على التشابه بين أزواج من نطق النص. على سبيل المثال، ثبت مؤخرًا أنه استنادًا إلى المعلومات الهيكلية الموجودة في أنماط الخطاب البشري، يمكن استخدام مخططات التكرار المفاهيمية لنمذجة وتصور الاتجاهات في البيانات وإنشاء مقاييس موثوقة للتشابه بين الكلمات النصية الطبيعية. هذه التقنية هي أداة قوية لمزيد من التحقيق في بنية الخطاب البشري. بدون النهج الحسابي لهذا السؤال، فإن المعلومات المعقدة للغاية الموجودة في بيانات الخطاب كانت ستظل بعيدة عن متناول العلماء.

ومن دون نهج حسابي لهيكل البيانات اللغوية، فإن الكثير من المعلومات المتاحة الآن ستظل مخفية تحت اتساع البيانات داخل أي لغة واحدة. تسمح اللغويات الحاسوبية للعلماء بتحليل كميات هائلة من البيانات بشكل موثوق وفعال، مما يخلق إمكانية لاكتشافات لا مثيل لها في معظم الأساليب الأخرى.

المصادر والمراجع:

Braine, M. (1971). *On two types of models of the internalization of grammars.* New York: Academic Press.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition , 48* (1), 77-99.

Member Portal | The Association for Computational Linguistics Member Portal". www.aclweb.org. Retrieved 2020-08-17.

Salvi, G. (2012). Language bootstrapping: learning word meanings from the perception-action association. *IEEE Transactions on Systems* , 660-671.

Socher, R. (2011). *Parsing Natural Scenes and Natural Language with Recursive Neural Networks.* USA: ICML.

Sproat, R. S. (2005). *What is Computational Linguistics?". The Association for Computational Linguistics.* Cambridge: Prentice Hall.

USZKOREIT, H. (2000). *WHAT IS COMPUTATIONALLINGUISTICS.* Germany: Department of Computational Linguistics and Phonetics of Saarland University.

الحاج صالح ,ع .ا .(2007) .*بحوث ودراسات في اللسانيات العربية* . الجزائر :مؤسسة الكتاب.

مهديوي ,إ .(2016) .*اللسانيات الحاسوبية، رقمنة اللغة العربية ورهان مجتمع المعرفة* .دار العلم.