معالجة اللغة الطبيعية في ضوء البرمجة اللغوية العصبية؛ التحديات والحلول

حلول اللغة العربية أنموذجا

Natural language processing in light of neurolinguistic programming; Challenges and Arabic language solutions.

سليم مزهود[1]

Salim Mezhoud[1]

[1] المركز الجامعي عبد الحفيظ بوالصوف. ميلة (الجزائر)، s.mezhoud@centre-univ-mila.dz

تاريخ الاستلام: 2022/05/10 تاريخ القبول: 2022/07/24 تاريخ النشر: 2023/01/01

ملخص:

تقدم اللغة تحديات خطيرة للباحثين ومطوري تطبيقات معالجة اللغة الطبيعية للنص والكلام باللغة العربية، على أساس أن المجموعات اللغوية يختلف بعضها عن بعض صوتيًا وصرفًا ونحويًا ودلالة، ويظهر ذلك في أثناء مناقشة السمات العامة للغة ما أو مجموعة من اللغات ذات انتماء واحد أو أصل واحد، ثم الخصائص المحددة للغة، والتحديات التي تطرحها هذه الخصائص على معالجة اللغة العربية الطبيعية.

يناقش المقال التحديات المختلفة التي تواجه معالجة اللغات الطبيعية في اللغة ويقدم بعض الحلول التي من شأنها توجيه الممارسين الحاليين والمستقبليين في مجال معالجة اللغة الطبيعية.

كلمات مفتاحية: لغة طبيعية، معالجة، برمجة لغوية عصبية.

Abstract:

The The language presents researchers and developers of natural language processing (NLP) applications for text and speech grounds that some languages is contrasts with others n lingos phonetically, morphologically, syntactically, and semantically, starting from a discussion the general features of the language, and then specific properties of the language, and the challenge these properties pose to ANLP.

This article discusses different challenges of NLP in Arabic and presents some solutions that would guide current and future practitioners in the field of Arabic natural language processing (ANLP)..

Keywords: Natural language; Processing; neurolinguistic programming.

المؤلف المرسل: د. سليم مزهود، **الإيميل:** s.mezhoud@centre-univ-mila.dz

1. مقدمة:

بالرغم من أنّ اللغة البشرية لغةٌ صعبة بسبب هيكلها اللغوي المعقّد، إلا أنها مثيرة للاهتمام بسبب تاريخها، والأهمية الإستراتيجية لمجتمعاتها والمنطقة الجغرافية التي تستوطن فيها، وتراثها الثقافي والأدبي.

ونظرًا للمكانة الّتي احتلّتها اللّغة منذ آلاف السنين، إلى وقتنا الحالي، بحسب الظروف الزمنية والمكانية التي احتوتها، على سبيل المثال، الظروف الدينية والسياسية والاقتصادية، والتاريخية والعلمية المتمثلة في مختلف الحضارات والاختراعات والاكتشافات العلمية، ووصل بها التطور إلى أن تبنّتها العديد من العلوم بالدراسة والتحليل والتقنين بغية الكشف عن أصلها وبنيتها وتكوينها، وتطويعها حاسوبيا، وأصبحت مطلبا جوهريا من منطلق لساني جعل منها موضوعا للدراسة العلمية.

ومن هنا نتساءل قدرة فطريّة كامنة في مخّ الكائن البشري، فإنّ اللّسانيات البيولوجيّة والعصبيّة قد عملتا على تقديم تفسيرات علميّة مرتبطة بالعمليّة العقليّة المصاحبة لها، وكذا كيفيّة إنتاجها في الدّماغ ومختلف المناطق المسؤولة عنها، وعلاقتها بها، والوظائف البيولوجيّة الّتي تؤدّيها لاكتساب وفهم وإنتاج اللّغة.

والسؤال المطروح؛ ما مفهوم اللسانيات العصبية، وكيف أسهمت في معالجة اللغة، وما هي حلول معالجة اللغة الطبيعية العربية؟

يحاول المقال الإجابة عن السؤال متتبعا المنهج الوصفي، هادفا إلى تحديد مفهوم اللسانيات العصبية وعلاقتها باللغة وحل مشكلاتها المرتبطة ببرمجتها في مختلف المجالات التي تكون فيها، وتحديد طرق حل مشكلات المعالجة اللغوية العربية، لأجل أن تحافظ اللغة بشكل عام والعربية بشكل خاص على مكانتها في ضوء التطور التكنولوجي.

2. مفهوم اللسانيات العصبية:

1.2 تعريف اللسانيات العصبية:

تنسب صياغة مصطلح "اللسانيات العصبية" Neurolinguistic إلى إديث كروويل تراجر Edith Crowell Trager وهنري هيكين Henri Hecaen وألكسندر لوريا Alexandr Luria، في أواخر الأربعينيات والخمسينيات، ومن المحتمل أن يكون عالِم الأعصاب: السوفيتي الكسندر رومانوفيتش لوريا Alexander Romanovich Luria (1902–1977م) أوّل من أشار إلى هذا المصطلح في كتابه الموسوم: "مشكلات في اللسانيات العصبية "Problems in Neurolinguistics" وهو أول كتاب يحتوي على علم اللغة العصبي في العنوان. ثم انتشرت اللسانيات العصبية في الولايات المتحدة الأمريكية في السبعينيات، حيث أسس هاري ويتاكر Harry Whitaker مجلة "الدماغ واللغة Brain and Language" في عام 1974م. (Peng, 1985, p. 8)

ويدرس علم اللغة العصبي الآليات العصبية في الدماغ البشري المتحكّمة في فهم اللغة وإنتاجها واكتسابها. كمجال متعدد التخصصات، وتستمد اللسانيات العصبيةُ الأساليبَ والنظريات من مجالات متنوعة ذات علاقة بها وأبرزها علم الأعصاب واللسانيات والفلسفة وعلم النفس العصبي والعلوم المعرفية واضطرابات التواصل. ويتخذ كثير من الباحثين في علم اللغة العصبي أعمالهم كنماذج في علم اللغة النفسي والنظري، ويركزون على التحقيق في كيفية تنفيذ الدماغ العمليات التي يقترحها علم اللغة النظري وعلم اللغة النفسي الضرورية في إنتاج اللغة وفهمها، ويدرس علماء اللسانيات العصبية الآليات الفيزيولوجية التي يعالج بها الدماغُ المعلوماتِ المتعلقةَ باللغة وفهمها وإنتاجها، ومؤثرات النطق والسمع، فيكشفون عن نظريات لغوية ونفس لغوية، باعتماد علم الحبسة وتصوير الدماغ والفيزيولوجيا الكهربية والنمذجة الحاسوبية (Nakai, Jeong, & Brown, 2017, p. 20)

وبالنسبة للإنسان الذي يعرف أكثر من لغتين، أو الذي يتحدث لهجته ويتحدث الفصحى، فإن اللسانيات العصبية تبحث عن كيفية عدم خلط الدماغ بينهما، وكيف يمكنه الانتقال من نظام لغوي إلى آخر بالرغم من الاختلاف بين اللغتين، وهل يختلف دماغ المتكلم لغتين فأكثر عن دماغ وحيد اللغة؟، ماذا لو فقد أحدنا الذاكرة، هل يمكن أن يتحدث اللغة نفسها حين يشفى ويتكلم مرّة أخرى، ماذا عن أنواع الكتابة من اليمين إلى اليسار كالعربية ومن اليسار إلى اليمين كالإنجليزية ومن الأعلى إلى الأسفل كاليابانية والصينية، بأي طريقة يختلف دماغ كل شخص عن الآخر؟، وماذا عن مشكلات القراءة والتلعثم ونقص السمع واضطراب الكلام وعيوب اللسان ومشكلاته المتنوعة؟

حقًّا؛ إن علم اللغة العصبي متشابك بعمق مع علم اللغة النفسي، في دراسة خطوات معالجة اللغة المطلوبة للتحدث وفهم الكلمات والجمل، وتعلم اللغات الأولى واللغات الثانية وغيرها الأخرى، وكذلك معالجة اضطرابات الكلام وعسر القراءة.

2.2 تاريخ اللسانيات العصبية:

ترجع جذور علم اللغة العصبي تاريخيًا إلى تطور علم الحبسة في القرن التاسع عشر، وهو دراسة حالات العجز اللغوي، أي فقدان القدرة على الكلام، الذي يحدث نتيجة تلف في جزء أو أجزاء من الدماغ (Colin & Sakai, 2005, p. 23)

يحاول علم الحبسة الكلامية ربط البنية بالوظيفة عن طريق تحليل تأثير إصابات الدماغ على معالجة اللغة (Wiśniewski, 2007)

كان بيير بول بروكا Pierre Paul Broca (1824–1880م) طبيبًا وعالم تشريح وأنثروبولوجيًا فرنسيًا. اشتهر بأبحاثه في منطقة بروكا التي سميت باسمه، وهي منطقة من الفص الجبهي، وهي معنية باللغة، وكشف عمله أنّ أدمغة المرضى الذين يعانون من فقدان القدرة على الكلام تحتوي على آفات في جزء معين من القشرة، في المنطقة الأمامية اليسرى.

وكان هذا أول دليل تشريحي لتوطين وظائف المخ، وقد أسهم عمل بروكا في تطوير الأنثروبولوجيا الفيزيائية، وعلم القياسات البشرية.

ومنطقة بروكا، هي منطقة في الفص الأمامي من نصف المخ، وعادة ما يكون الجانب الأيسر من الدماغ مع وظائف مرتبطة بإنتاج الكلام، أما منطقة Wernicke، التي تسمى أيضًا منطقة كلام Wernicke، هي أحد جزأين من القشرة الدماغية المرتبطين بالكلام، والآخر هو منطقة Broca. وهي تشارك في فهم اللغة المكتوبة والمنطوقة، على عكس منطقة بروكا التي تشارك في إنتاج اللغة.

وقد كان بحث بروكا هو أول من قدّم دليلًا تجريبيًا على أنّ مناطق الدماغ المختلفة تؤدي وظائف مختلفة وأن المناطق الأمامية من الدماغ هي التي تتحكم في اللغة غالبا (Dronkers.Et.all, 2007, p. 14). وقد تمّ وصف بحثه بأنه صناعة حقبة، ووصف بروكا بأنه صانع عهد جديد في مجالات علم اللغة العصبي والعلوم المعرفية Epoch-making (Broca, 2009, p. 21)

أما كارل ويرنيك Carl Wernicke (1848-1905م)، وهو طبيب أعصاب بولندي ألماني، اشتهر ويرنيك أو فيرنيك بأبحاثه في حقل التأثيرات المرضية لأشكال معينة من اعتلال الدماغ، ودراسة الحبسة الاستقبالية، وأسهمت أبحاث رفقة أبحاث بروكا في تحقيق إنجازات بارزة فيما يتعلق بتحديد مناطق الدماغ المسؤولة عن الوظائف اليومية، وبخاصة تلك المتعلقة بالنطق، على غرار منطقة ويرنيك التي سميت على اسمه. رأى ويرنيك أنّ مناطق مختلفة من الدّماغ هي متخصصة في مهام لغوية مختلفة مع منطقة Broca التي تتعامل مع الإنتاج الحركي للكلام، ومنطقة ويرنيك تتعامل مع فهم الكلام السمعي (Wiśniewski, Neurolinguistics , 2007, p. 35)

أسهم عمل بروكا وويرنيك في ضبط مجال الحبسة الكلامية، ورأوا أنه يمكن دراسة اللغة من خلال فحص الخصائص الفيزيائية للدماغ (Teter, 2000, p. 52).

وقد استفاد العمل المبكر في علم الحبسة الكلامية، من أعمال طبيب الأعصاب الألماني كوربينيان برودمان Korbinian Brodmann (1868–1918م) في أوائل القرن العشرين، الذي اشتهر برسم خرائط القشرة الدماغية، وتقسيم الدماغ إلى مناطق مرقمة بناءً على الهندسة المعمارية الخلوية لكل منطقة (بنية) ووظيفتها (McCaffrey, 2008, p. 13). وحدّد 52 منطقة تُعرف باسم مناطق برودمان، بناءً على خصائصها الهيكلية الخلوية (النسيجية)، ولا تزال مناطق برودمان مستخدمة على نطاق واسع في علم الأعصاب اليوم.

لقد توسع مجال اللسانيات العصبية بشكل كبير في السنوات الأخيرة، بدءا من توسع البحث وآلياته في مجال الحبسة الكلامية، بفضل ظهور تقنيات (أو تِقَنَات) جديدة لتصوير الدماغ؛ مثل التصوير المقطعي بالإصدار البوزيتروني Positron emission tomography؛ وهو تقنية تصوير وظيفية تستخدم المواد المشعة المعروفة باسم الكشف الإشعاعي بالراديو Radioactive، لتصوير وقياس التغيرات في عمليات التمثيل الغذائي، وفي الأنشطة الفيزيولوجية الأخرى بما في ذلك مجرى الدم، والكشف عن السرطان وما إلى ذلك. والتصوير بالرنين المغناطيسي الوظيفي Functional magnetic resonance imaging الذي يقيس نشاط الدماغ عن طريق اكتشاف التغيرات المتعلقة بمجرى الدم في المخ؛ فلَهُ علاقة بتنشيط الخلايا العصبية في الدماغ، واستخدام التقنيات الفيزيولوجية الكهربية الوقتية الحساسة؛ بالكشف التخطيطي الكهربائي للدماغ Electroence phalography هي طريقة مراقبة كهربية لتسجيل النشاط الكهربائي للدماغ، وعادة ما تكون غير جراحية، مع وضع الأقطاب الكهربائية على طول فروة الرأس ،على الرغم من استخدام الأقطاب الكهربائية الغازية في بعض الأحيان ، كما هو الحال في تخطيط كهربية القشرة ،

تسمى أحيانًا الكشف "الإلكتروإنسيفالوغرافي Electroencephalography" داخل الجمجمة، أي تخطيط أمواج الدماغ (Phillips & Saka, 2005, p. 33).

أضف إلى ذلك تقنية MEG أي التصوير المغناطيسي للدماغ Magnetoence phalography، والتي يمكنها تسليط الضوء على أنماط تنشيط الدماغ عندما يشارك الناس في مهام لغوية مختلفة.

والتصوير المغناطيسي للدماغ؛ هو تقنية تصوير عصبي وظيفية لرسم خرائط لنشاط الدماغ عن طريق تسجيل المجالات المغناطيسية التي تنتجها التيارات الكهربائية التي تحدث بشكل طبيعي في الدماغ ، باستخدام مقاييس مغناطيسية حساسة للغاية. وتشمل لأبحاث الأساسية في عمليات الدماغ الإدراكية والمعرفية، وتحديد المناطق المتأثرة بعلم الأمراض قبل الإزالة الجراحية ، وتحديد وظيفة أجزاء مختلفة من الدماغ، والارتجاع العصبي. ويمكن تطبيق هذا في بيئة سريرية للعثور على مواقع الشذوذ وكذلك في بيئة تجريبية لقياس نشاط الدماغ ببساطة.

وقد ظهرت تقنيات الفيزيولوجيا الكهربية على وجه الخصوص كطريقة قابلة للتطبيق لدراسة اللغة في عام 1980 وباكتشاف N400 تظهر أنّ استجابة الدماغ للمعاني والقضايا الدلالية حساسة جدا في فهم اللغة (Hall, 2005, p. 53)، وN400 هو أحد مكونات إشارات تخطيط أمواج الدماغ بالتكهرب المغلقة زمنياً والمعروفة باسم (ERP) الحادث المستجيب كهربائيا Event-Related Potential، باستجابة الدماغ المقاسة التي هي نتيجة مباشرة لحدث حسي أو معرفي أو حركي معين، أي استجابة كهربائية فيزيولوجية نمطية لمحفز، وتوفر دراسة الدماغ بهذه الطريقة وسيلة غير باضعة لتقييم أداء الدماغ (Kutas & Federmeier, 2011, p. 24).

ويعد N400 جزءًا من استجابة الدماغ الطبيعية للكلمات والمحفزات الأخرى الواضحة، بما فيها الكلمات المرئية والسمعية ، وعلامات لغة الإشارة والصور والوجوه والأصوات البيئية والروائح (Hall, 2005, p. 54) ، ومنذ اكتشاف Electroence phalography و Magnetoence phalography يستخدم في نطاق واسع لإجراء أبحاث اللغة (Hagoort, Brown, & Osterhout, The neurocognition of syntactic processing., 1999, p. 26).

3.2 اللسانيات العصبية وعلم النفس اللغوي (قواعد السلوك):

ترتبط اللسانيات العصبية ارتباطًا وثيقًا بميدان علم النفس اللغوي الذي يسعى إلى توضيح الآليات المعرفية للغة وقواعد السلوك المعرفي، من خلال استخدام التقنيات (أو التِّقَنَات) التقليدية لعلم النفس التجريبي (Hagoort, 2003, p. 35)

ويهدف علم النفس اللغوي، أو علم اللغة النفسي إلى دراسة العلاقة المتبادلة بين العوامل اللغوية والجوانب النفسية، ويلتزم بشكل أساسي بالآليات التي يتم من خلالها معالجة اللغة وتمثيلها في العقل والدماغ؛ أي العوامل النفسية والعصبية الحيوية التي تمكن البشر من اكتساب اللغة واستخدامها وفهمها وإنتاجها.

أما علم النفس التجريبي فيهدف إلى العمل الذي يقوم به أولئك الذين يطبقون الأساليب التجريبية للدراسة النفسية والعمليات التي تكمن وراءها، ويستخدم علماء النفس التجريبيون مشاركين من البشر والحيوانات لدراسة العديد من الموضوعات النفسية المرتبطة بالأعصاب والبيولوجيا، بما في ذلك الإحساس والإدراك والذاكرة والتعلم والتحفيز والعاطفة، وعمليات النمو، وعلم النفس الاجتماعي، والركائز العصبية لكل هذه الموضوعات. وتتضمن أعمال اللسانيات العصبية الكثير من الاختبارات حول النظريات التي طرحها علم النفس اللغوي وتقييمها.

يقترح علم النفس اللغوي نماذج لشرح بنية اللغة وكيفية تنظيم المعلومات اللغوية وفق خوارزميات لشرح كيفية معالجة المعلومات اللغوية في العقل، في حين تعمل اللسانيات العصبية على تحليل نشاط الدماغ لاستنتاج أشكال الهياكل البيولوجية؛ كمجموعات الخلايا العصبية؛ حيث يحتوي الدماغ على ملايين الخلايا العصبية التي يتم تنظيمها في مناطق دماغية مختلفة، في مناطق فرعية مختلفة داخل المنطقة دماغية ، وكل منطقة صغيرة مقسمة إلى طبقات مختلفة، وداخل كل طبقة أنواع خلايا مختلفة. وتسمى: "مجموعات الخلايا العصبية"، وتهدف دراسة عمل هذه المجموعات وكيفية بناء هيكلها العصبي البيولجي، ليجيب عن التساؤلات عن نشاط جميع هذه الخلايا في هذه الطبقة من المنطقة الفرعية التي تكون من النوع الهرمي، في استجابة لمثير، مثلا حول إنسان أو حيوان يتلقى منبها بصريا أو سمعيا أو حسيا جسديا، ما رد فعل هذه المنطقة الفرعية، وكيف تستجيب منطقة الدماغ على سبيل المثال.

تقوم هنا المعالجة اللغوية النفسية بتنفيذ خوارزميات، مثلا تستخدم التجارب في معالجة الجملة لمعرفة استجابة الدماغ الفيزيولوجي، من خلال N400 التي تحدثنا فيها آنفا.

وفي هذا الصدد تهدف اللسانيات العصبية إلى وضع تنبؤات جديدة حول بنية اللغة وتنظيمها بناءً على رؤى حول فيزيولوجيا الدماغ ، من خلال "تعميم معرفة الهياكل العصبية وبنية اللغة".

يتم إجراء أبحاث اللسانيات العصبية في مختلف المجالات الرئيسة لعلم اللغة والحقول الفرعية اللغوية الأساسية على النحو الآتي:

أولا؛ الصوتيات: هو حقل معرفي يُعنى بدراسة أصوات الكلام؛ وتتمحور أسئلة البحث اللساني العصبي فيه حول كيفية استخراج كيف يستخرج الدماغ أصوات الكلام من إشارة صوتية، وكيف يفصل الدماغ أصوات الكلام عن ضوضاء الخلفية.

ثانيا؛ الفونولوجيا ؛ دراسة كيفية تنظيم الأصوات في اللغة، كيف يتم تمثيل النظام الصوتي للغة معينة في الدماغ

ثالثا؛ علم الصرف؛ دراسة كيفية بناء الكلمات وتخزينها في المعجم العقلي؛ كيف يخزن الدماغ الكلمات التي يعرفها ويصل إليها

رابعا؛ علم الدلالة وبناء الجمل؛ دراسة كيفية ترميز المعنى في اللغة، وكيفية بناء الكلام متعدد الكلمات؛ كيف يجمع الدماغ الكلمات في مكونات الجملة؛ ويستخدم المعلومات الهيكلية والدلالية في فهم هذه الجمل.

3. أهم الحلول لمشكلات اللغة الطبيعية وفق البرمجة العصبية للغة:

1.3 حل مشكلة ازدواج اللغة:

إن أهم حل لازدواج اللغة في بلد ما، هو بناء موارد لأنواع مختلفة من لغة البلد. وقد قامت الدول الأقل نموا بالفعل ببناء موارد للعربية المصرية والشامية والعراقية، بالإضافة إلى ذلك، هناك مشروع مهم في جامعة كولومبيا لبناء بنك الشجرة من اللهجات العربية باستخدام الموارد المتاحة بالفعل للغة العربية الفصحى الحديثة. يستغل المشروع التخطيط المنهجي للغة العربية الفصحى الحديثة لبعض اللهجات على المستويات الصوتية والصرفية والمعجمية. يجب أن يكون باحثو ومطورو **ANLP** على دراية بآثار ازدواجية اللغة العربية في تطبيقاتهم نظرًا لأنه من الصعب بناء نظام يمكنه التعامل مع جميع أنواع اللغة العربية في وقت واحد. يجب أن يكون المطورون واضحين بشأن أي مجموعة متنوعة من اللغة العربية مناسبة لتطبيقاتهم الخاصة. على سبيل المثال ، سيحتاج تطبيق التعرف على الكلام للمحادثات الهاتفية العربية على الأرجح إلى موارد لهجة بينما يتطلب تطبيق آخر للمعالجة، إذاعات الأخبار العربية موارد عربية قياسية حديثة سواء في شكل معرفة لغوية أو في مجموعات لأغراض التدريب (Farghaly & Shaalan, 2009, p. 8).

2.3 حل مشكلة الترميز في اللغة العربية:

عادةً ما يتم تعريف الرمز المميز على أنه تسلسل مكون من حرف واحد أو أكثر مسبوقًا ومتبوعًا بمسافة، ويعمل هذا التعريف جيدًا مع اللغات التي لا تهتم كثير بتراص كلماتها مثل اللغة الإنجليزية، على عكس اللغة العربية التي يمثل الترميز فيها مشكلة بسبب التشكل الوافر والرصيد اللغوي الثري جدا والمعقد للغة العربية (ATTIA, 2016, p. 9) .

على سبيل المثال قد تحتوي كلمة عربية واحدة على ما يصل إلى أربعة رموز مختلفة، وبالتالي، يتطلب الترميز معرفة القيود المفروضة على تسلسل الأسماء ومعانيها وأضدادها في الكلمات العربية، ويجب التمييز بين الكلمات التي تشكل وحدات نحوية، واللواحق التي تشكل التصريفات النحوية، والزمن والعدد والصفات، ولا بد من إيجاد طريقة نجمع فيها بين التحليل الصرفي والنحوي والترميز في عملية واحدة.

وعلينا البحث عن كيفية نعوض بها غياب الكتابة بالأحرف الكبيرة للدلالة على أسماء الأشخاص والأماكن والعناوين الكبيرة، إما بزيادة في شفرة الحرف الأول، على نحو زيادة في شفرة حرف السين والشين، أو تكبير الحرف الأول على سبيل المثال. أو زيادة علامة ما للحرف للدلالة على أنه حرف كبير.

3.3 حل مشكلة الأوصاف النحوية للغة الفصحى المعاصرة:

على الرغم من أن الأوصاف النحوية للغة العربية الفصحى الحديثة لم تكن مكتوبة من وجهة نظر حسابية، إلا أنها مفيدة جدا في معالجة اللغة العربية المعاصرة، إذ تعتبر المجموعات العربية المشروحة التي تم تطويرها في أقل البلدان نمواً ذات قيمة كبيرة لتطبيقات معالجة اللغة الطبيعية العربية ANLP، حيث تم تحليل النصوص العربية القياسية الحديثة وفق قواعد اللغة العربية التقليدية، والنظريات اللغوية الحديثة (RYDING, 2005, p. 74).

ومن طرق حل مشكلة الأوصاف النحوية عمل بنك الشجرة النحوي العربي في براغ، وفق منهج وظيفي وصفي لتحليل اللغة العربية الفصحى الحديثة، وبنك الشجرة العربي في جامعة كولوبيا.

وقد كانت هناك محاولات عديدة لحل مشكلة الأوصاف النحوية والتحليل الصرفي العربي، باستعمال تطبيق تقنية الحالة المحدودة لاستعادة الجذور العربية، حيث يقوم التطبيق بالتحليل والتوليد، ويمكن إدخال الكلمات العربية باستخدام النص العربي.

4. خاتمة

هناك ميزات للغة العربية تشكل تحديًا بطبيعتها للباحثين والمطورين في معالجة اللغة الطبيعية العربية، تتضمن هذه الميزات الطبيعة غير الموصولة للصرف العربي، وغياب التمثيل الهجائي للحروف المتحركة العربية القصيرة من النصوص العربية المعاصرة، والحاجة إلى قواعد نحوية صريحة للغة العربية الفصحى تحدد المكونات اللغوية في غياب علامات الحالة، ويجب أن تقدم القواعد الجديدة وصفا للجمل غير الموضوعية، والعلاقات اللغوية الجذابة وتحليل الخطاب، وأساسيات بنية الكلمات والجمل، والعلاقات بين عناصر الجملة.

وعلى الرغم من أن اللغة العربية هي لغة صوتية بمعنى أن هناك تخطيط واحد لواحد بين الحروف في اللغة والأصوات المرتبطة بها، إلا أنها لا تخصص في الكلمة العربية حروفًا لتمثيل حروف العلة القصيرة، ولا يوجد مفهوم للكتابة بالأحرف الكبيرة مما يحتم فهم السياق لفهم اللغة، وهذا يزيد من صعوبة تعلم اللغة العربية للناطقين بغيرها، ويطرح تحديات أمام تحليل الكلمات العربية من الناحية المورفولوجية، فإن بنية الكلمات غنية جدا، ومضغوطة بحيث يمكن أن تمثل كلمة واحدة معنى جملة كاملة. وأما من الناحية النحوية، فإنّ الجملة العربية طويلة ومعقدة، مما يسهم في فشل نظام الترجمة الآلية في بعض

الحالات في تحديد السوابق الصحيحة، كما أن الكلمة الواحدة، تنقسم في سياق ما إلى سابقة والجذر، وفي سياق آخر تصير تلك السابقة من الجذر، كما أن بعض السوابق لا يمكن فصلها، فإذا عرفنا الإله فيمكن تقسيمه الكلمة إلى (ال) التعريف+ إله، لكن لا يمكن تجزئة كلمة الله بتلك الطريقة.

وعلاوة على ذلك ، يمكن تبديل مكونات الجملة العربية (ترتيب الكلمات الحرة) دون التأثير على البنية أو المعنى، مما يضيف مزيدًا من الغموض النحوي والدلالي، ويتطلب تحليلًا أكثر تعقيدًا.

انطلاقا من خصائص اللغة العربية الثرية والشائكة المتعددة، فإننا نقترح لمعالجة اللغة الطبيعية العربية ما يأتي:

– زيادة حجم الشفرة الأولى من الحروف ذات الشفرات، نحو السين والشين، أما بقية الحروف فيتم تحجيمها أكثر حين تكون الأولى وبقية الحروف بعدها تكون صغيرة، في الكلمات الدالة على أسماء الأعلام والحيوان والأماكن والأزمان ونحو ذلك.

– العمل على تقنية التشكيل الآلي، وتطويره في الترجمة الآلية، للتمكن من الوصول إلى ترجمة أكثر قربا من المعنى المقصود.

– إنجاز إحصائيات دقيقة وشاملة لأفضل اقتراحات التصويب.

– العمل على إحداث المصحح النحوي في الترجمة الآلية، على نحو ما هو معمول به في اللغة الإنجليزية مثلا، برنامج Grammarly الذي يصحح الأخطاء النحوية في اللغة الإنجليزية، وفي العربية نقوم بإعداد برنامج لتصحيح الأخطاء النحوية، التي تصحح العلامة الإعرابية (الضمة والفتحة والكسرة)، مع الأخذ بعين الاعتبار الحالات الإعرابية التي تكون علامتها حروفا على نحو (الواو والألف والياء) في الرفع والنصب والجر، علاوة على الجزم بحذف حرف العلة أو السكون.

العمل ضمن فريق بحث، لاكتشاف الأخطاء اللغوية في كثير من برامج التصحيح اللغوي وتصحيحها، بحيث إن هناك بعض البرامج تصحح كلمة صحيحة لتقدم أخرى خاطئة.

5. قائمة المراجع:

ATTIA, M. (2016, 02 17). *Arabic tokenization system. In Proceedings of the Association of Computational Linguistics* . Consulté le 05 04, 2022, sur sourceforge: https://sourceforge.net/projects/sarf

Broca, P. P. (2009). Who Named It?

Colin, P., & Sakai, K. L. (2005). Language and the brain . *Yearbook of Science and Technology. McGraw-Hill Publishers* , 166-169.

dfgf. (fgf). *gdf.* fdg: dfg.

Dronkers.Et.all. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain. , 130* (5), 1432-1441.

Farghaly, A., & Shaalan, K. (2009). Arabic Natural Language Processing. *ACM Transactions on Asian Language Information Processing , 8* (4), 8-20.

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing".*NeuroImage* , 19.

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). *The neurocognition of syntactic processing.* Oxford: Brown & Hagoort. The Neurocognition of Language.

Hall, C. J. (2005). *An Introduction to Language and Linguistics. . p. 274.* New York: Continuum International Publishing Group.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology* (62), 621-647.

McCaffrey, P. (2008). *CMSD 620 Neuroanatomy of Speech, Swallowing and Language.* California: California State University, Chico.

Nakai, Y., Jeong, J., & Brown. (2017). Three- and four-dimensional mapping of speech and language in patients with epilepsy. 140 (5): 1351-1370. . *Brai , 140* (5), 1351-1370.

Peng, F. (1985). (What is neurolinguistics? *Journal of Neurolinguistics* , 7-30.

Phillips, C., & Saka, K. L. (2005). *Language and the brain.* New York: Yearbook of Science and Technology. McGraw-Hill Publishers.

RYDING, K. C. (2005). *A reference grammar of modern standard Arabic. C.* Cambridge UK: ambridge University Press, Cambridge, UK.

Teter, T. (2000). Pierre-Paul Broca. *Muskingum College* .