معالجة اللغة العربية آلياً قديماً وحديثاً: دراسة مسحية

محمد عدني حمد السيد
جامعة البطانة، السودان
mohdn111@gmail.com,mohdn111@sustech.edu

الخلاصة: تهدف هذه الدراسة إلى التقصي والبحث عن معالجة اللغة العربية وبدايتها وما آلت إليه الآن. وتكمن مشكلة البحث في عدم وجود دراسات توضح التطور التاريخي لمعالجة اللغة العربية، وما هي المؤثرات التي أسهمت في هذا التطور؟ وهل للتطور الذي حدث في البرمجيات واجهزة الحاسب كان له دور في ذلك؟ كذلك تبحث هذه الورقة عن الاتجاهات الحديثة لحوسبة اللغة العربية، وماهي مكامن الضعف فيها، وما هي الاتجاهات التي كان العمل فيها ضعيفا والتي لم تطرق من قبل؟ ولم؟ وقد استخدمت هذه الدراسة المنهج الوصفي الاستقرائي. وتوصلت إلى أن معالجة اللغة العربية تطورت من مرحلة الحرف العربي مروراً بالكلمات وصولًا لمعالجة الجملة وساعد في ذلك التطور الكبير في الأجهزة والبرمجيات. واحتوت هذه الورقة البحثية على عدة أمور هي: أجزاء معالجة البيانات، واللغة العربية ومميزاتها، وتطور الأجهزة والبرمجيات، ومعالجة اللغة العربية وتطبيقاتها، ثم تحدثت أخيراً عن الاتجاهات الحديثة لمعالجة اللغة العربية.

الكلمات الجوهرية: معالجة، معالجة اللغة الطبيعية، لغة عربية، تطبيقات، الذكاء الاصطناعي، التعلم العميق.

1. معالجة البيانات:

المعالجة هي العمليات التي تجرى على البيانات للحصول على المعلومة المطلوبة. وتأتي عملية المعالجة في المرحلة الوسطى بين إدخال البيانات وإخراج المعلومات. وتشمل عمليات كثيرة منها تخزين البيانات، واسترجاعها، وترتيب وتصنيف البيانات، وإجراء العمليات الرياضية والمنطقية. وقد بدأت معالجة اللغة العربية متأخرة نوعاً ما مقارنة باللغات الأجنبية عامة واللاتينية بصفة خاصة والتي بلغت مرحلة النضج وذلك لعدم الاهتمام بها من قبل العلماء وقلتهم في هذا المجال. وعدم الوعي بأهمية هذا المجال خاصة في المجالات الأخرى.

2. اللغة العربية:

اللغة العربية لغة سامية(Benajiba, Diab, and Rosso 2004) وهي لغة القرآن الكريم وترتيبها هو السادس في اللغات المستخدمة (Kanaan et al. 2009) ولها ثلاثة أنواع: التقليدية وهي لغة القرآن الكريم، واللغة العربية الحديثة (بدون تشكيل)، واللهجات العربية المحلية المستخدمة في كل منطقة (Alotaibi and Meftah 2013)، ولأن القرآن الكريم نزل باللغة العربية أصبحت اللغة واحدة من اسرع اللغات نمواً في العالم وأصبح تعلم اللغة العربية مهماً لجميع المسلمين لأن الإسلام حالياً هو الدين الأسرع انتشاراً في العالم لذا زادت الحاجة لتعلم اللغة العربية واستخدام التقنيات الحديثة كذلك لحوسبتها وذلك لاتجاه العالم نحو عصر الرقمنة. واللغة العربية لغة معقدة من ناحية القواعد اللغوية وذلك لأسباب عديدة منها: طول وتعقيد تركيب الجملة، إغفال الشكل في الكتابة، طبيعة الترتيب الحر للجملة العربية، وجود الضمائر(Shaalan 2005). ولحوسبة اللغة العربية أهمية قصوى في العديد من المجالات وذلك للانفجار المعلوماتي في الفترة الأخيرة، وحاجة الحكومات والافراد لتوفير المعلومة وسهولة استرجاعها ، التنافس الكبير بين الشركات لعرض منتجاتها لأكبر مجموعة من المستفيدين وبلغات متعددة، وأخيراً توفر وسائط التواصل الاجتماعي والتي يستطيع المستخدم أن يدلي برأيه بكل حرية وكذلك معرفة المتنافسين والحاجة لدراسة هذه الآراء والاستفادة منها.

3. مميزات اللغة العربية:

تتميز اللغة العربية بخصائص عديدة منها: نزل بها القرآن، اشتقاقية، ذات مفردات غزيرة، تعتمد على القوالب البنائية، وقوع الاشتراك والنحت والاتساع والترادف فيها(Anon n.d.).هذه الخصائص بعضها ساعد في المعالجة الآلية كالاشتقاق وبعضها ساعد في تعقيده كالاتساع والترادف.

4. تعريف معالجة اللغة الطبيعية ومقدمة تأريخية عنها:

4.1 معالجة اللغة الطبيعية Natural language processing (NLP):

معالجة اللغة الطبيعية Natural language processing (NLP)، أو اللسانيات الحاسوبية Computational Linguistics، وتقنيات اللغة الطبيعية الإنسانية Human Language Technologies جميعها أسماء لمسمى واحد، وهو علم يهتم بتطوير برامج ونظم لها القدرة على فهم وتوليد اللغة الطبيعية (محمد 2020). وهو علم يعتمد فيه على الحاسوب لدراسة اللغات الإنسانية وفهم طبيعتها، حيث توجه هذه الأنظمة لفهم لغة الإنسان والذكاء البشري. وتشمل ثلاث محاور أساسية هي تقنيات معالجة: النصوص، والكلام المنطوق، ومعالجة الصور(سعيد 2019).

4.2 مقدمة تأريخية:

أدى التطور الكبير في التكنولوجيا في مجال الالكترونيات عموماً وخاصةً الحاسب الآلي إلى ظهور نهضة شاملة في جميع المجالات، وكان لذلك أيضاً أثرٌ كبير في مجال معالجة اللغات الطبيعية Natural language processing (NLP)، وكانت البدايات الأولى في الجانب النظري للتحليل النظري لقواعد النحو التوليدي اللغوي عام 1956 (Nadkarni, Ohno–Machado, and Chapman 2011) ومن ثم إعادة توجيه قواعد التوليد لتكون مبنية على الدلالات (Schank 1972). ظهرت الترجمة الآلية في البداية وبدأت بالترجمة من الروسية للإنجليزية بواسطة الحاسب الآلي Machine Translation (Jones 1994) في عام 1967 و 1968 كانت المحاولات الأولى لتصميم أنظمة لمخاطبة الآلة conversation (Schank 1972). تم تصميم لونار LUNAR عام 1971 (Allam and Haggag 2012) وهو نظام رد آلي لتمكين الجيولوجيين من الوصول بسهولة ومقارنة وتقييم بيانات التحليل الكيميائي على الصخور القمرية وتركيبة التربة. ظهر نظام SHRDLU الذي هو في الأساس محلل لغوي تفاعلي للحوار مع المستخدم باستخدام المصطلحات الإنجليزية لعرض الأشكال. ويعتبر لونار ونظام SHRDLU أنظمة تشغيلية بمستويات حقيقية للمعالجة اللغوية ، تلتها عملية التحليل المعجمي lexical Analyzer في فترة السبعينات(Nadkarni et al. 2011). وحتى هذه الفترة كانت التجارب بسيطة وذلك لأن علوم اللغة تحتاج لفهم عميق ولمعلومات دقيقة تكون خفية حتى على علماء اللغة في بعض الأحيان. في فترة الثمانينات ظهرت أدوات أفضل للقواعد اللغوية لتصميم قواميس لتكون على أفضل نحو. LIFER or LADDER تم تصميمه كواجهة لغوية طبيعية لقاعدة بيانات عن السفن البحرية الأمريكية. واستخدمت قواعد اللغة الدلالية لتحسين التعامل معه(Alexander, Rukshan, and Mahesan 2013). في فترة التسعينات ظهرت مجموعة من النماذج مثل ,basic theory of Hidden Markov Models (HMMs) الدراسات في مختلف المجالات وظهرت محاولات لمعالجة مشكلة الغموض اللغوي (Manaris 1998) كما ظهرت أنظمة اللغة المنطوقة spoken language systems كما ظهرت أنظمة فهم وتوليد اللغة الطبيعية؛ واجهات اللغة الطبيعية، إدارة الخطاب ، فهم القصة وتوليد النص ، والترجمة الآلية التفاعلية ومساعدي الكتابة الذكية (Bates 1995). مع ظهور الذكاء الاصطناعي وتطور تقنياته خاصة في العقدين الأخيرين ظهرت اتجاهات جديدة لتطبيقات وتقنيات الذكاء الاصطناعي منها على سبيل المثال : أنظمة الرد الآلي، تحويل النص لكتابة والعكس، تعلم الآلة، والتعرف على الكلام(Sajnani Harsh Kamani Hitesh Jeswani 2017). وأيضاً من المجالات التي أخذت حيزاً في معالجة اللغة الطبيعية التدقيق الإملائي (Gupta and Mathur 2012) وتصحيح القواعد اللغوية (Bhirud, Bhavsar, and Pawar 2017). كل هذه الدراسات اعتمدت على المصادر اللغوية كالذخائر اللغوية والموارد المعجمية الرقمية(Joseph et al. 2016).

5 تطور الأجهزة Hardware والبرمجيات وأثرها على معالجة اللغة العربية:

ظهر الكمبيوتر الرقمي الإلكتروني للبرامج المخزنة لأول مرة في الأربعينيات من القرن الماضي، وتطور سريعًا إلى معالجة المعلومات، بدءًا من الآلة الحاسبة وصولاً إلى آلة المعالجة والاتصالات التي أصبحت لا غنى عنها. حيث بدأت الأجهزة من آلة تملأ غرفة بأكملها، تكلف عدة ملايين من اليورو وقادرة على إجراء عمليات حسابية بسيطة فقط لأجهزة الكمبيوتر المحمولة، وأجهزة الكمبيوتر الشخصية، وأجهزة الكمبيوتر العملاقة القوية والمتعددة الاستخدامات التي نعرفها اليوم للأعمال و مهدت هذه التطورات لأربع مجموعات واسعة من التقنيات، غالبًا ما تكون متداخلة، هذه التقنيات هي: تقنيات للمساعدة في الحساب، وتقنيات الأتمتة والتحكم، وتقنيات معالجة المعلومات وإدارة المعلومات، وتقنيات الاتصال(Tatnall 2012). وأسهم ذلك التطور الكبير ليس فقط في الجوانب الاجتماعية والتعليمية بل في التجارية أيضاً، حيث أثرت تكنولوجيا المعلومات والاتصالات الجديدة على تطوير نماذج اقتصادية جديدة مثل التسليم في الوقت المناسب just–in–time (JIT) و TINA there is no alternative to globalization and liberalization لا يوجد بديل للعولمة والتحرير(Asaolu 2006). ساعد ذلك التطور الكبير في الأجهزة ونمو البرمجيات التي تستخدمها هذه الأجهزة بدءًا من البرمجيات الرقمية التي استخدمت لغة الآلة (0,1) ثم لغة التجميع التي استخدمت الرموز وصولًا للبرمجيات المعقدة جدًا التي استخدمت في جميع الأجهزة الحالية من الأجهزة البسيطة كالغسالات والثلاجات بواسطة المتحكمات controllers إلى الأجهزة البسيطة من حاسب شخصي، ودفتري، ومحمول، وصولًا لأجهزة الهاتف المحمول، وصولًا للنظم الكبيرة والمعقدة كالأنظمة الموزعة، وأجهزة الحاسبات الضخمة super computer. وأدى توفر هذه التقنيات إلى الحاجة لتطوير برمجيات لتشغيلها ومن ثم التطبيقات المناسبة للعمل على هذه الأجهزة مغلقة ومفتوحة المصدر والتي أعطت المطورين فرصة كبيرة للاستفادة مما هو متاح من البرمجيات مفتوحة المصدر وتقليل المجهودات، وحاجة المستخدم المناسبة لكل جهاز واستخداماته المختلفة. مما سبق نلاحظ توفر جميع البرمجيات المناسبة لكل مستخدم وأدى ذلك لتوفير برمجيات خاصة بحوسبة اللغة العربية ساعد في ذلك توفر أجهزة ذات كفاءة عالية تؤدي المطلوب منها. وسنتابع في الجزئية التالية معالجة اللغة العربية وبعض التطبيقات المهمة ودورها في معالجة اللغة العربية. وأدى أيضاً ظهور البرمجيات الخاصة بمعالجة اللغات الطبيعية كبرنامج sketch engine، والمكتبات التي وفرها المطورون في لغة بايثون لإجراء المعالجات المختلفة للغات عامة، والبرمجيات التي طورت في الفترة الأخيرة لمعالجة النصوص في اللغة العربية والتي تعتمد على علم الذكاء الاصطناعي والتي ساعدت بصورة كبيرة على معالجة اللغة العربية.

6 معالجة اللغة العربية:

معالجة اللغة الطبيعية تعنى بتوليد وفهم اللغة فهماً صحيحاً باستخدام علم الذكاء الاصطناعي، ولكن في بداية هذه التجربة عانت معالجة اللغة العربية، واعتبر الباحث دكتور نبيل علي أن مشكلة تخلف وقصور تكنولوجيا المعلومات والتي من ضمنها حوسبة اللغة العربية ترجع لأسباب عديدة أهمها ثلاثة أسباب رئيسة وهي: التبعية العلمية، والتكنولوجية، وقلة ميزانية البحوث. وأوضح الباحث أن بداية العمل في البرمجيات العربية كان عملية ترجمة لبعض البرامج المكتوبة باللغة الانجليزية (القشرة الظاهرية) ولكنها لم تصمم لتطبيقات عربية خالصة. كانت الشركة العالمية هي أول من أنتج برمجيات عربية في ذلك الوقت حيث عربت عدة نظم تشغيل وطورت لغات برمجية عربية وبرامج تعليمية، واختتمت جهودها بأهم الأشياء إطلاقاً وهي إنشاء وحدة متخصصة في اللسانيات الحاسوبية ومعالجة اللغة العربية (علي 1994). كانت البداية لهذه البرمجيات نظم القراءة الآلية للنص العربي. في جانب تعلم اللغة العربية ما زالت البرامج بدائية في الفترة الأولى (فترة الثمانينات والتسعينات) وفي غالبها اختيار من متعدد ولم تستخدم حتى هذه الفترة نظم الذكاء الاصطناعي كالصرف والنحو الآليين(علي 2001). في العقدين الأخيرين بدأ الاهتمام بحوسبة اللغة العربية ولكن بمجهودات فردية من قبل الباحثين العرب والغربيين في الجامعات الغربية وبعض المحاولات في الدول والمجامع العربية.

1.6 مقدمة عن تطبيقات معالجة اللغة العربية:

بدأت معالجة اللغة العربية متأخرة نوعاً ما حيث بدأت في المعالجات البسيطة المتمثلة في معالجة الحرف والكلمة باستخدام تقنيات بسيطة نوعاً ما، ثم بدأ الاهتمام بها في العقدين الأخيرين، ونجد أن غالب هذه المعالجات كانت عبارة عن مجهودات فردية في بعض الجامعات الغربية، لكن بدأ في العقد الأخير ظهور بعض الاهتمامات من قبل بعض الجامعات ومراكز البحوث المهتمة بأبحاث اللغة العربية كمنظمة الالسكو ومدينة الملك عبد العزيز للعلوم والتقنية. شملت هذه المعالجات مجالات كثيرة منها المكانز اللغوية، الترجمة الآلية، استرجاع المعلومات واستخلاص المعلومات، تحويل النص لصوت Text To Speech والعكس، وتدرجت هذه المعالجات من الحرفية البسيطة إلى مراحل استخدام التطبيقات الدلالية لتحسين جودة هذه المعالجات. كما توجد تطبيقات كثيرة متوفرة لكنها لا توضح طريقة عملها والمناهج التي تم استخدامها لعمل تلك التطبيقات. تحتاج اللغة العربية لكثير من المجهودات من الجميع أفرادا ومنظمات وحكومات للوصول لمراحل متقدمة في حوسبة اللغة العربية لمسايرة عصر المعلومات.

2.6 تطبيقات معالجة اللغة العربية في العقدين الأخيرين:

تطورت التطبيقات المتعلقة بحوسبة اللغة العربية في العقدين الأخيرين وذلك نتيجة للاهتمام بعلم اللسانيات الحاسوبية والذي هو علم بيني وسيط بين علوم اللغة وعلوم الحاسب الآلي حيث ساعد ذلك في الحوسبة من ناحيتي المبنى والمعنى(عمر 2009). معظم الدراسات في هذا المجال كانت باللغة الانجليزية يغلب عليها الطابع الفني المختص بمعالجة الحرف العربي ولكن ليس مواضيع متقدمة(طاهر 2006). سنذكر في الجزء التالي بعض التطبيقات في الفترة الأخيرة والتطورات التي حدثت فيها مقارنة بسابقتها.

1.2.6 المكانز اللغوية:

تم تصميم مجموعة كبيرة من الذخائر النصية اللغوية الخاصة باللغة العربية لاستخدامها في البحوث والأغراض التعليمية (Atwell et al. 2004) ومفردها corpus وتجمع corpora. تعتبر الذخائر النصية (المكانز اللغوية) هي الأساس في بناء غالب التطبيقات اللغوية وذلك لإجراء المقارنات والحصول على المعلومات اللغوية المختلفة خاصة في المجالات النحوية والمعجمية(Arts et al. 2014)، ومعظم هذه المكانز غير متوفرة ولم تعلم tagged. ولكن سنركز على واحدة من هذه المكانز وسنتناولها بالشرح والتوضيح. ولتبيين أهميتها سندرس كل تطبيق من التطبيقات ومن ثم المكنز اللغوي المستخدم إن وجد.

وسنتناول في هذا الجزء المكنز اللغوي arTenTen وهو مكنز يتكون من 5.8 مليار كلمة وتم تعليمه tagged (تحديد بنية الجملة فعل، اسم، ...) وكلماته مشكلة وغير مشكلة بواسطة الأداة MADA وتم استخدام الأداة Sketch Engine لحساب الإحصاءات المختلفة وتم توليد الفهرس الأبجدي concordance والاستخدام للكلمات في الجمل التي وردت فيها، وتردد هذه الكلمات، وقائمة الكلمة وتكرار كل كلمة، والكلمات المرتبطة بها Collocate. وبالرغم من ذلك تم استخدام 115 مليون كلمة فقط في الأداة Sketch Engine لإجراء هذه العمليات مما قد يؤثر على النتائج والإحصاءات التي تعرضها الأداة، كما أنه لم يعلم كاملاً لمشكلة في الأداة MADA.

2.2.6 الترجمة الآلية:

هي نقل نص مترجم من لغة إلى أخرى باستخدام الحاسب، ولها طرق منها المقارنة المعجمية في معجم ثنائي اللغات، التحويلية بحيث تشمل التحليل والتحويل والتوليد، والترجمة الوسيطة باستخدام لغة أخرى، وأخيراً الترجمة الآلية الإحصائية التي تعتمد على المدونات. أوضح (Attia 2008) أن الترجمة الآلية تحتاج لأربع أنواع من المعارف هي: معرفة لغة المصدر، معرفة لغة الهدف (المعجم ، التركيب ، بناء الجملة ، وعلم الدلالة) معرفة العلاقة بين اللغتين للتحويل، وأخيراً معرفة الموضوع لمعرفة استخدام بعض المصطلحات ومعرفة السياق واعتبر (Abuelyaman et al. 2015) أن إزالة الغموض اللغوي للكلمة هو أحد المفاتيح الأساسية للترجمة الآلية.إن إحد الأدوات المهمة في الترجمة العربية هوِ Arabic Word

Net (AWN) والذي يمكن الرجوع إليه : (Al-Yahya and Al-Khalifa 2010). تعددت أدوات الترجمة المجانية كخدمات قوقل Google أو بالاشتراك والتسجيل كبرنامج ترجم tarjim الذي كان متاحاً مجاناً ثم أصبح بالاشتراك، ومن التي تترجم فقط الكلمات إلى النصوص والوثائق كاملة. من اشهر الأمثلة لهذه الأدوات مترجم قوقل الذي يستخدم لترجمة حوالي 90 لغة من ضمنها اللغة العربية ويستخدم الترجمة الآلية الإحصائية حيث يعتمد على ذخيرة لغوية تحتوي على 200 بليون كلمة (البواب 2015)، وبدأ كمترجم يترجم الكلمات فقط مما خلق لبسًا كبيراً في ترجمة الجملة ولكن مع استخدام الترجمة الآلية العصبية بدلًا من النماذج الإحصائية انتقلت قوقل إلى الترجمة الكاملة للجمل وكل من يستخدم هذا المترجم يلاحظ الفرق. يلاحظ أن كل الجهود المبذولة في الترجمة هي فردية أو تجارية. أما البحثية منها فإن طرقها تنوعت وشملت الآتي: المعتمدة على القواعد rule based، لغة وسيطة Interlingua، التحويل Transfer، المعتمدة على المثال Example based، المعتمدة على المعرفة Knowledge based، الطريقة الهجينة Hybrid(Hadla, Hailat, and Al-kabi 2014) ويلاحظ أن جميع هذه المجهودات عبارة عن مجهودات منفردة كل في جزئية بسيطة ولكن إن تكاملت هذه الجهود فقد تصبح هذه الترجمة أقوى وأنفع.

3.2.6 استرجاع المعلومات Information Retrieval واستخلاص المعلومات Information Extraction:

واحدة من أهم استخدامات الحاسب الآلي هي حفظ المعلومة وسهولة استرجاعها وذلك لدقته وسرعته. ومع الانتشار الكبير للمعلومات وتوفرها على شبكة الانترنت نجد أن اللغة العربية اسهامها قليل في محتوى الانترنت بالرغم من ثراء المكتبة العربية بجميع فنون وضروب المعرفة. أصبحت هناك مشكلة في استرجاع المعلومات العربية لأن معظم الأجهزة مصنوعة في الغرب ويستخدمون خصائص اللغة اللاتينية في اللغة العربية مع الاختلاف الكبير في خصائص هذه اللغات. فبعد أن كانت تستخدم نظم مطابقة الكلمات اتجهت حديثاً للبحث الدلالي حيث طور (Chen, Chu, and Chen 2010) نموذج دلالي لاسترجاع المعلومات وذلك لتحقيق الأهداف التالية: تحليل وتحديد السمات الدلالية للمحتوى، دليل استفسار المستخدم وتوسيع دلالاته الضمنية، إنشاء محتويات ذات دلالات تقريبية عن طريق المطابقة حيث حسنت المطابقة لكلمات الاستفسار الرئيسة وكذلك إجراء استخدام قائم على الدلالة وأمثلة لذلك محركات البحث كقوقل. في أنظمة استخلاص المعلومة كأنظمة الرد الآلي الذي يتم عبر مجموعة من المعالجات يجري عرض أفضل النتائج مختصرة مرتبة بدلاً عن عرض الوثائق فقط ومن ثم البحث داخلها بواسطة المستخدم ، كما تم تصميم منهجيات عديدة لتحسين عملها كاستخدام الأنماط (Adany, Ayed, and Atwell 2017) واستخدام الانطولوجيا لتحسن عملية البحث (Sheker et al. 2016)، واستخدام الـ tree bank في التمثيل النحوي والدلالي كدراسة (Dukes, Atwell, and Sharaf 2010).

4.2.6 المدقق الإملائي والنحوي Spelling and Grammar corrections:

المدقق الإملائي هو أداة تستخدم للتحقق من الأخطاء الإملائية والنحوية وتصحيح هذه الأخطاء في النص أو المستند(Kaur, Garg, and Tech 2014). وضح (Dickinson, Brew, and Meurers 2012) ان هناك نوعان للخطأ وهما: الكلمات غير الموجودة في اللغة Nonword والتي قد تنتج من الخطأ الطباعي أخطاء مطبعية (وقد تكون أخطاء إملائية) وهي أخطاء قد يكون الشخص يعرف فيها كيفية تهجئة الكلمة بشكل صحيح ، لكنه أخطأ في كتابتها، أخطاء الالتباس عندما تفقد المعلومة والمعرفة الضرورية، وأخطاء الكلمات الحقيقية: وتنتج من خطأ يحدث لكن ينتج منه كلمة موجودة في اللغة وهذه صعبة الاكتشاف لأنها تعتير من الأخطاء المنطقية والتي تعالج باستخدام المعالجات الدلالية واستخدام الذخائر اللغوية والطرق الإحصائية. أوضح العلماء أن هناك أربعة أنواع لأخطاء التعديل حيث يتم المقارنة بين الكلمة التي فيها الخطأ والتي نحاول أن نصل إليها بعد التصحيح (الهدف) وهي: الحذف، والإضافة، والتبديل، والنقل: حيث يتم تبديل حرفين متجاورين(Dickinson et al. 2012). وضعت استراتيجيات عديدة لمعالجة هذه الأخطاء مثل: استخدام القواميس، character n – gram وهي عبارة عن سلسلة متتالية من العناصر n من عينة معينة من النص أو الكلام. ولاكتشاف وتصحيح الأخطاء من خلال السياق يتم ذلك بالرجوع للقواعد النحوية والاستفادة في اكتشاف الكلمة المناسبة. وتركيب هذه القواعد النحوية من الصعوبة بمكان لإمكانية استخدام الترتيب: فعل، فاعل، مفعول (بترتيب مفتوح) مما يؤدي لمشاكل الغموض اللغوي. ويمكن استخدام الاشجار النحوية، واختبارات التناسق، وقواعد التراكيب التعبيرية واستخدام تطبيق WordNet لإزالة الخطأ اللغوي في الكلمة حسب السياق word sense disambiguation (WSD).

5.2.6 تحويل النص لصوت Text To Speech والعكس:

بالرغم من الاهتمام المتعاظم بالنظم الصوتية وتحويلها نص والعكس إلا أن هناك نقص كبير في اللغة العربية لهذه الأنظمة ، وتصمم في الغالب لمساعدة المكفوفين لحل بعض مشاكل التعامل مع الحاسب. وفي غالبها تعتمد على قواعد النطق والقواميس مع بعض الاستثناءات وذلك يؤدي في الغالب إلى فشل هذه الأنظمة. تم استخدام طرق هجينة لمعالجة هذه المشكلة وذلك بدمج بعض التقنيات مع مزج بعض التراكيب مع استخدام بعض نظريات اللغات الطبيعية(Zeki, Khalifa, and Naji 2010)، وتم استخدام الشبكات العصبية التي تستخدم نموذج التعلم غير خاضع للإشراف ومنهج تركيب الكلام المتسلسل لتوليف النص العربي باستخدام قاموس به 500 كلمة (Science and Publications 2009). أما في التحويل من نص لصوت تم تصميم نماذج مختلفة كالتي تستخدم ثلاثة نماذج واحد للصوت والثاني للغة والثالث للنطق كما في دراسة(Hamad and Hussain 2011)، كما تستخدم هذه التطبيقات في أنظمة التعرف على الصوت كما في دراسة(Ahmed and Ghabayen 2017).

6.2.6 التعرف الضوئي على الأحرف Optical character recognition OCR:

هو برنامج كمبيوتر مصمم لتحويل الصور مع النص إلى نص معالج آليًا. تتوفر أنظمة التعرف الضوئي على الحروف للعديد من اللغات بما في ذلك العربية، وقد تم تطوير وتحسين التعرف الضوئي على الحروف على مدار عقود، مما أدى في النهاية إلى عدد كبير من الأساليب ذات النتائج القوية، وفي بعض الحالات حوالي 99٪، بينما يمكن استخدام التعلم العميق في اللغة العربية التعرف الضوئي على الحروف. ينتج عنه دقة بنسبة 100٪ في وقت أقصر وموارد أقل لمعالجة الصورة(Alwaqfi and Mohamad 2020). ومن أفضل الاستخدامات لهذه التقنية الواضحة والمشهورة تطبيقات كتب قوقل[1] حيث يستخدم النص المكتوب في خانة البحث والبحث داخل الكتب والتي توجد في شكل رسم او صورة وتتم مقارنتها بالنص الأصلي ومن ثم إظهارها للمستخدم في الكتاب والكتب المختلفة، وأيضًا تم استخدام هذه التقنية في المعجم التاريخي للغة العربية[2] والذي يعتبر من أضخم الإنجازات الظاهرة في مجال حوسبة اللغة العربية والذي استخدم نفس التقنية، ولو أتيحت مادة المدونة النصية للباحثين في مجال حوسبة اللغة، فيمكن اعتبارها اكبر مصدر نصي متاح للباحثين في هذا المجال يمكّن الباحثين من دراسة الظواهر اللغوية المختلفة والبحث فيها.

7.2.6 تطبيقات مهمة في شتى ضروب اللغة العربية:

واحدة من هذه التطبيقات تطبيق ميازين[3] وهو تطبيق لوزن الشعر العربي ويستخدم بإدخال بيت الشعر لوزنه ولكنه يضع بعض الشروط لضمان صحة الوزن كوضع الشدة في ال الشمسية وغيرها من التعليمات ومن ثم عرض الوزن بتفعيلاته المختلفة. كما صممت منظمة الالسكو مجموعة من التطبيقات مثل: نظام الاشتقاق والتصريف في اللغة العربية، ونظام الخليل الصرفي، والمدقق الإملائي مفتوح المصدر، ومشكِّل آليٍّ مفتوح المصدر للغة العربية، لمعجم الحاسوبي التفاعلي للغة العربية. تم تصميم هذه البرامج بواسطة مجموعة من الجهات البحثية والجامعات العربية، كما لا يفوتنا أن نشير لمجهودات بعض الجامعات العالمية كجامعة ليدز في خدمة بحوث اللغة العربية.

7 الاتجاهات الحديثة لمعالجة اللغة العربية:

اتجهت التطبيقات الحديثة لحوسبة اللغة العربية بدلًا من الدراسات الإحصائية والتطبيقية إلى استخدام أحدث تقنيات مجال الذكاء الاصطناعي بعد التطور الكبير الذي حدث في برمجياته وهي تقنيات التعلم العميق deep learning الذي يسمح للآلة التعلم لوحدها وهو فئة من فئات التعلم الآلي التي تعمل بشكل

[1] https://books.google.com/?hl=ar ، تاريخ الوصول : 2021/11/28 11:33:49 ص
[2] https://www.almojam.org ، تاريخ الوصول : تاريخ الوصول : 2021/11/28 9:30:15 ص
[3] https://www.Mayazeen.com ، تاريخ الوصول : 2017/12/15 7:13:22 م

أفضل على البيانات غير المنظمة(Mathew, Amudha, and Sivakumari 2021) والتي أسهمت بشكل كبير في معظم التطبيقات الحديثة لمعالجة اللغة العربية مثل التعرف الضوئي على الحرف كما في دراسة (Alwaqfi and Mohamad 2020)، وتصنيف اللهجات العربية كما في دراسة (Lulu and Elnagar 2018)، وكذلك تصنيف النصوص وهو وضع كل مجموعة من النصوص أو الملفات المتشابهة في مجموعة خاصة بها كما في دراسة (Boukil et al. 2018)، وغيرها من الاستخدامات الكثيرة للتعلم العميق في مجال معالجة اللغة العربية.

8 النتائج:

أدى التطور الكبير في الأجهزة والبرمجيات خاصة برمجيات التطور الصناعي الخاصة بمعالجة اللغات الطبيعة وعلى الأخص برمجيات معالجة اللغة العربية إلى تطور برمجيات معالجة اللغة العربية، وظهور تطبيقات مختلفة في شتى ضروب اللغة العربية: شملت المعاجم، النحو والصرف، الأدب وغيره من المجالات المختلفة في اللغة العربية. ولكن لا زالت الدراسات في مجال المعاني والدلالات قليل مقارنة بالبنية والتركيب حيث أن معظم هذه الدراسات في الغالب اعتمدت على تركيب وبنية الكلمات. ومع التطور الكبير في البرمجيات ظهرت برمجيات تعلم الآلة كاتجاه حديث لمعالجة اللغة العربية. وساعد في ذلك الاهتمام الكبير من المختصين والباحثين في مجال اللغة العربية، واهتمام المنظمات بذلك على الأخص منظمة الالسكو (تصميم برمجيات خاصة بمعالجة اللغة العربية)، واتحاد المجامع اللغوية العربية (المعجم التاريخي للغة العربية)، وبعض الباحثين الأجانب المهتمين بحوسبة اللغة العربية مثل البروفيسور إريك أتول من جامعة ليدز وغيرهم.

9 الخاتمة:

ونختم هذا البحث المتواضع بأن نقول إن بدايات معالجة اللغة العربية بدأت حرفياً ومن ثم انتقلت لمرحلة المعالجة التركيبية (الكلمة والجملة) ثم الآن هي في مرحلة استخدام العلاقات الدلالية واللسانيات الحاسوبية للوصول لمرحلة النضج ومرحلة الحوسبة التي هي أعلى من المعالجة ولكن ذلك يحتاج لتكاتف الجهود جميعها والدعم غير المحدود ليتم ذلك الأمر وبأسرع وقت حتى نصل لمراحل متقدمة في هذا المجال لخدمة القرآن والسنة ولمنافسة اللغات الأخرى ومواكبتها بعد توفر الباحثين والمصادر الرقمية المختلفة.

10 المراجع:

Abuelyaman, Eltayeb, Limia Rahmatallah, Wafaa Mukhtar, and Mona Elagabani. 2015. “Machine Translation of Arabic Language: Challenges and Keys.” *Proceedings - International Conference on Intelligent Systems, Modelling and Simulation, ISMS* 2015-Septe:111–16.

Adany, Mohamed, Hamdels Ayed, and Eric Atwell. 2017. “Quran Question Answering System Using Arabic

Number Patterns (Singular, Dual, Plural)." *International Journal on Islamic Applications in Computer Science And Technology* 5(2):1–12.

Ahmed, Basem H. A., and Ayman S. Ghabayen. 2017. "Arabic Automatic Speech Recognition Enhancement." *Proceedings - 2017 Palestinian International Conference on Information and Communication Technology, PICICT 2017* (January 2018):98–102.

Al-Yahya, M., and H. Al-Khalifa. 2010. "An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns in the Holy Quran." *The Arabian Journal for Science and Engineering* 35(2):21–35.

Alexander, Rukshan, Prashanthi Rukshan, and Sinnathamby Mahesan. 2013. "Natural Language Web Interface for Database (NLWIDB)." (July):6–7.

Allam, AMN, and MH Haggag. 2012. "The Question Answering Systems: A Survey." *International Journal of Research and Reviews in Information Sciences, IJRRIS* 2(3).

Alotaibi, Yousef, and Ali Meftah. 2013. "Review of Distinctive Phonetic Features and the Arabic Share in Related Modern Research." *Turkish Journal of Electrical Engineering and Computer Sciences* 21(5):1426–39.

Alwaqfi, Yazan M., and Mumtazimah Mohamad. 2020. "A Review of Arabic Optical Character Recognition Techniques & Performance." *International Journal of Engineering Trends and Technology* (1):44–51.

Anon. n.d. "التحديات."

Arts, Tressy, Yonatan Belinkov, Nizar Habash, Adam Kilgarriff, and Vit Suchomel. 2014. "ArTenTen: Arabic Corpus and Word Sketches." *Journal of King Saud University - Computer and Information Sciences* 26(4):357–71.

Asaolu, Olumuyiwa Sunday. 2006. "On the Emergence of New Computer Technologies." *Educational Technology and Society* 9(1):335–43.

Attia, Ma. 2008. "Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation." *Linguistics*.

Atwell, Eric, Latifa Al-sulaiti, Saleh Al-osaimi, and Bayan Abu Shawar. 2004. "JEP-TALN 2004 - Session on Arabic Language Processing A Review of Arabic Corpus Analysis Tools Un Examen d ' Outils Pour l ' Analyse de Corpus Arabes." *JEP-TALN 2004, Arabic Language Processing. Fez* (April):19–22.

Bates, M. 1995. "Models of Natural Language Understanding." *Proceedings of the National Academy of Sciences* 92(22):9977–82.

Benajiba, Yassine, Mona Diab, and Paolo Rosso. 2004. "ARABIC NAMED ENTITY RECOGNITION : AN SVM-BASED APPROACH."

Bhirud, Nivedita S., R. .. Bhavsar, and B. .. Pawar. 2017. "Grammar Checkers for Natural Languages : A Review." *International Journal on Natural Language Computing* 6(4):1–13.

Boukil, Samir, Mohamed Biniz, Fatiha El Adnani, Loubna Cherrat, and Abd Elmajid El Moutaouakkil. 2018. "Arabic Text Classification Using Deep Learning Technics." *International Journal of Grid and Distributed Computing* 11(9):103–14.

Chen, Ming Yen, Hui Chuan Chu, and Yuh Min Chen. 2010. "Developing a Semantic-Enable Information

Retrieval Mechanism." *Expert Systems with Applications* 37(1):322–40.

Dickinson, Markus, Chris Brew, and Detmar Meurers. 2012. *Language and Computers*. Wiley-Blackwell.

Dukes, Kais, Eric Atwell, and Abdul-baquee M. Sharaf. 2010. "Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank." *In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt* 1822–27.

Gupta, Neha, and Pratistha Mathur. 2012. "Spell Checking Techniques in NLP: A Survey." *International Journal of Advanced Research in Computer Science and Software Engineering* 2(12):2277–128.

Hadla, Laith S., Taghreed M. Hailat, and Mohammed N. Al-kabi. 2014. "Evaluating Arabic to English Machine Translation." *International Journal of Advanced Computer Science and Applications (IJACSA)* 5(11):68–73.

Hamad, Mazin, and Mustafa Hussain. 2011. "Arabic Text-to-Speech Synthesizer." *Proceedings - 2011 IEEE Student Conference on Research and Development, SCOReD 2011* (May):409–14.

Jones, Karen Sparck. 1994. "Natural Language Processing : A Historical Review." *Natural Language Processing : A Historical Review* 9(10):53–59.

Joseph, Sethunya R., Hlomani Hlomani, Keletso Letsholo, Freeson Kaniwa, and Kutlwano Sedimo. 2016. "Natural Language Processing: A Review." *International Journal of Research in Engineering and Applied Sciences* 6(3):1–8.

Kanaan, Ghassan, Awni Hammouri, Riyad Al-Shalabi, and Majdi Swalha. 2009. "A New Question Answering System for the Arabic Language." *American Journal of Applied Sciences* 6(4):797–805.

Kaur, Jaspreet, Kamaldeep Garg, and M. Tech. 2014. "Hybrid Approach for Spell Checker and Grammar Checker for Punjabi." *International Journal of Advanced Research in Computer Science and Software Engineering* 4(6):2277–128.

Lulu, Leena, and Ashraf Elnagar. 2018. "Automatic Arabic Dialect Classification Using Deep Learning Models." *Procedia Computer Science* 142:262–69.

Manaris, Bill. 1998. "Natural Language Processing: A Human-Computer Interaction Perspective." *Advances in Computers* 47(C):1–66.

Mathew, Amitha, P. Amudha, and S. Sivakumari. 2021. "Deep Learning Techniques: An Overview." *Advances in Intelligent Systems and Computing* 1141(January):599–608.

Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. 2011. "Natural Language Processing: An Introduction." *Journal of the American Medical Informatics Association* 18(5):544–51.

Sajnani Harsh Kamani Hitesh Jeswani, Anmol B. 2017. "A Study on Natural Language Processing for Human Computer Interaction." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 6(11):2278.

Schank, Roger C. 1972. "Conceptual Dependency: A Theory of Natural Language Understanding." *Cognitive Psychology* 3(4):552–631.

Science, Computer, and Science Publications. 2009. "An Arabic Text-To-Speech System Based on Artificial Neural Networks Ghadeer Al-Said and Moussa Abdallah Department of Electronics Engineering , Princess

Sumaya University for Technology ,." 5(3):207–13.

Shaalan, Khaled F. 2005. "Arabic GramCheck: A Grammar Checker for Arabic." *Software – Practice and Experience* 35(7):643–65.

Sheker, Mustefa, Saidah Saad, Rehab Abood, and Mohanaad Shakir. 2016. "Domain-Specific Ontology-Based Approach for Arabic Question Answering." *Journal of Theoretical and Applied Information Technology* 83(1).

Tatnall, Arthur. 2012. "History of Computer Hardware and Software Development." *Encyclopedia of Life Support Systems (EOLSS)* 1(1):1–18.

Zeki, Mustafa, Othman O. Khalifa, and A. W. Naji. 2010. "Development of an Arabic Text-to-Speech System." *International Conference on Computer and Communication Engineering, ICCCE'10* (May):11–13.

.البواب, مروان. 2015. "الترجمة الآلية." *محاضرة ألقيت بمجمع اللغةالعربية بدمشق* 1–27

م. st ed. edited by سعيد, محسن رشوان، أحمد راغب، محمد عطية، سامح الأنصاري، المعتز بالله. 2019. *مقدمة في حوسبة اللغة العربية*. 1 .ا. ع. ا. ا. ل. ا. العربية. الرياض

.طاهر, بربارة سهيلة زوجة بن. 2006. "الترجمة بساعدة الحاسوب من الانجليزية إلى العربية." جامعة الجزائر

.علي, د. نبيل. 1994. *العرب و عصر المعلومات*

.علي, د. نبيل. 2001. *الثقافة العربية وعصر المعلومات*. Vol. 104.

".عمر, ديدوح. 2009. "فعالية اللسانيات الحاسوبية العربية

المجموعة العربية .st edمحمد, أسماء السيد محمد عبد الصمد، كريمة محمود. 2020. *تطبيقات الذكاء الاصطناعي ومستقبل تكنولوجيا التعليم*. 1 .للتعريب والنشر

11. جدول الالفاظ

English	عربي
Natural language processing	معالجة اللغة الطبيعية
lexical Analyzer	محلل معجمي
corpus	ذخيرة نصية
Optical character recognition OCR	التعرف الضوئي على الأحرف

12. الملخص باللغة الانجليزية

Automatic processing of the Arabic language, ancient and modern: a survey study

Mohamed Adany Hamdelsayed
Albutana University
mohdn111@gmail.com,mohdn111@sustech.edu

Abstract

This study aims to investigate and research the treatment of the Arabic language, its beginning, and what it is now. The problem of the research lies in the absence of studies showing the historical development of Arabic language processing, and what are the influences that contributed to this development? Did the development that occurred in software and computers have a role in this? This paper also searches for modern trends in computing the Arabic language, and what are the weaknesses in, and what are the directions in which work was weak and which were not addressed before? and did not? This study used the descriptive and inductive method. This study concluded that the processing of the Arabic language developed from the stage of the Arabic letter, through the words, to the processing of the sentence, which helped in the great development in hardware and software. This research paper contained several parts of data processing, the Arabic language and its features, the development of hardware and software, processing the Arabic language and its applications, and then finally talked about modern trends in processing the Arabic language.

Keywords: processing; natural language processing; Arabic language; applications; artificial intelligence; deep learning.