نحو محلّل صَرْفِيّ غير معجميٍّ للأفعال العربية

د. يحيى بن علي آل مريع عسيري
قسم اللغة العربية وآدابها–كلية الآداب والعلوم الإنسانية
جامعة الملك خالد– السعودية

المستخلص:

تقدم هذه الدراسة شرحًا مفصلا لطريقة بناء محلل صرفي غير معجمي للأفعال العربية. ويأتي هذا العمل في سياق التغلب على إحدى المشكلات التي تواجهها المحللات الصرفية العربية المتاحة بين يدي المستخدمين والباحثين، أعني مشكلة الاعتماد على المعجم التي جعلت هذه المحللات غيرَ قادرة على التعامل مع الكلمات التي لا تحتويها قواعدُ بياناتها المعجمية. وقد اعتمدتْ هذه الدراسة في تطوير هذا المحلل على أمرين: أحدهما الأوزان الفعلية لما تتسم به من انتظام يمكن الإفادة منه في بناء هذا المحلل الصرفي، والآخر هو تقنية الحالات المتناهية(Finite state technology) ، وتحديدًا خوارزميات المحولات الصرفية (morphological transducers) لما تتمتع به من سرعة وقدرة على التعامل مع الكلمة العربية وتحليلِها على عدة مستويات. وقد بينتْ نتائجُ تقييم هذا المحلل قدرتَه على تحليل الأفعال العربية صرفيًا سواء تلك الأفعال التي استطاعت المحللات الأخرى تحليلها لوجودها في معاجمها المصاحبة أو تلك الأفعال التي لم تستطع تلك المحللات تحليلاها لعدم وجودها في تلك القوائم المعجمية .

الكلمات المفتاحية: محلل صرفي، تقنية الحالات المتناهية، المحولات الصرفية، الأفعال العربية.

Towards a Non-Lexical Morphological Analyzer for Arabic Verbs

Dr. Yahya bin Ali Al Muriya Asiri
Department of Arabic Language and Literature
College of Arts and Humanities
King Khalid University - Saudi Arabia

Abstract:

This study presents a detailed description of a non-lexical morphological analyzer for Arabic verbs. The work aims to overcome one of the major problems facing Arabic morphological analyzers, which is their reliance on a lexical database. This reliance has made these tools unable to handle words that are not included in the lexical databases attached to these morphological analyzers. The development of the analyzer, presented in this study, relied on two things: 1) the verbal forms due to their regularity, and 2) finite state technology, specifically the morphological transducers and their algorithms that allow the analyzer to handle Arabic words at multiple levels efficiently and quickly. The evaluation of the system showed its ability to outperform two Arabic morphological analyzers; it was not only able to analyze words that other systems were able to analyze, but it was also able to handle contemporary verbs that were not included in the lexicons of these analyzers.

Keywords: morphological analyzers, Finite-state technology, morphological transducers, Arabic verbs

المقدمة:

يعد التحليل الصـــرفي مكونًا أســـاســـيًا من مكونات معالجة اللغات الطبيعية؛ إذ تعتمد عليه مكونات معالجة اللغات الطبيعية الأخرى كالمكون النحوي والمكون الدلالي. وترجع هذه الأهمية للتحليل الصـــرفي إلى كون اللغات الطبيعية تُشَـــفِّرُ كثيرًا من المعلومات على مســـتوى الكلمة عبر الوحدات المعجمية التي تحيل إلى معانٍ معجميةٍ تصـورية أو كياناتٍ خارجيةٍ وعبر ما يلتصـق بتلك الوحدات المعجمية من لواصقَ صرفيةٍ تحملُ كثيرًا من المعاني الوظيفية التي تختلف باختلاف اللغات. وتمييزُ هذه الوحدات/المورفيمات الصـــرفية المكونة للكلمة عن بعضـــها وتحديدُ المقولات التي تنتمي إليها وربطها بمدلولاتها يعد متطلبًا سـابقًا تعتمد عليه مكوناتُ تحليلِ اللغات الطبيعية على مستوى الجملة. وعلى هذا يمكن القول إنه إذا كانت معرفة دلالة الجملة، في جزءٍ كبيرٍ منها، تعتمدُ على التحليل النحوي، فإن التحليلَ النحوي نفسَه والتحليلَ الدلالي كذلك يستندان بشكلٍ أساسي على التحليل الصرفي. وإلى جانب هذه الأهمية للمحللات الصرفية في معالجة اللغات الطبيعية تبرز أهميةُ التحليل الصـــرفي في كونه أداةً لكثيرٍ من التطبيقات الأخرى كالترجمة الآلية وأنظمةِ اســـترجاعِ المعلومات وأنظمة البحثِ والتنقيبِ في النصوص والتصحيحِ الإملائي وغير ذلك من التطبيقات.

وتختلف أهميةُ التحليل الصـرفي باختلاف اللغات، فالتحليل الصـرفي للغات ذات الغنى الصـرفي أكثرُ أهميةً منه للغات الفقيرة صـــرفيًا. والتحليل الصـــرفي للغات التي تجمع بين الصـــرف الخطي/الإلصـاقي وغير الخطي يبدو كذلك أكثرَ أهمية منه للغات التي تعتمد على الصـرف الخطي ذي الطبيعة الإلصاقية فقط. والكلمة العربية تَتَّسِمُ بالغنى الصرفي والتعقيد في آن واحد؛ إذ تتصف بتنوع اللواصق الصرفية التي تشفر كثيرًا من المعاني، خصوصا الفعل، كما تتصف بالخاصية الصرفية الخطية وغير الخطية، وهو ما يعطي أبعادًا مختلفةً للإشـــكالات التي تواجه العاملين في حقل اللسانيات الحاسوبية ومعالجة العربية. ولا شك أن معالجة اللغة العربية، وتحديدًا على هذا المستوى الصـــرفي، قد حظيت بجهود كبيرة اختلفت باختلاف الأهداف وتنوعت بتنوع الطرق والخوارزميات المستخدمة، إلا أن الواقع يشـهدُ وجود تحدياتٍ كبيرةٍ تبرزُ أمام الجهود المبذولة في معالجتها. وتعود

بعض أسباب تلك التحديات إلى طبيعة اللغة ذاتها بينما يعود البعضُ الآخرُ منها إلى وجودِ بعض المشكلات المتعلقة بطبيعة المقاربات المتبعة في بناء المحللات الصرفية العربية.

ومن أبرز المشكلات المتعلقة بطبيعة المقاربات المتبعة في بناء المحللات الصرفية العربية مشكلةُ الاعتمادِ على وجود معجمٍ مصاحبٍ لهذا المحلل أو ذاك، وهو ما يحول دون قدرة المحلل الصرفي على تحليل الكلمات غير الموجودة في قاعدة البيانات المعجمية المصاحبة له. ولهذا تقدم هذه الورقة مقترحًا لبناء محلل صرفي عربي غير معجمي للأفعال العربية باستخدام تقنية الحالات المتناهية، وهو ما يسمح بتحليل أي فعل من دون معجم، وقد بني هذا المحلل باستخدام بتقنية فوما (foma) الذي طوره (Hulden 2009)، وهي تقنيةٌ مفتوحةُ المصدر تعتمد على خوارزميات XFST. وستقتصرُ هذه الورقة على تقديم محلل صرفي خاص بالأفعال لسببين: أولا، يعد الفعلُ مكونًا أساسيًا من مكونات الجملة؛ إذ يعتمدُ عليه مستخدمُ اللغةِ بشكل أساسي في تشفير الأحداث وأزمنتها والمشاركين فيها؛ ولذلك فإن تحليل الأفعال تحليلا صرفيًا تحتاجه كثيرٌ من التطبيقات الحاسوبية التي تهدف إلى معرفةِ طبيعةِ الأحداثِ التي تحتوي عليها النصوص وأزمنتِها والمشاركين فيها. وثانيًا، تتسمُ بنيةُ الفعل العربي بنظامٍ اشتقاقيٍّ توليدي مطردٍ إلى درجة يمكن معها التنبؤ بالأفعال المشتقة وبدلالتها بناء على معرفة الجذر والصيغة الصرفية، وهذا ما جعل كثيرًا من المعاجم العربية لا تدرج معظمَ هذه الأفعالِ ضمن مداخلها المعجمية مع أنها تشكل حيزًا كبيرًا من معجم الأفعال.

مشكلة الدراسة وهدفها:

من المشكلات التي ما زالت تواجه المحللات الصرفية العربية المتاحة للمستخدم والأكثر شيوعًا مشكلةُ الاعتماد على معجمٍ مصاحبٍ للمحلل الصرفي، بمعنى أن المحللَ الصرفيَّ يستندُ في تحليله على قاعدة بيانات معجمية[1] تحتوي على الوحدات المعجمية الاسمية والفعلية وغيرها؛ وبناء على ذلك تكونُ جودةُ هذا المحلل الصرفي أو ذاك معتمدةً بشكلٍ كاملٍ على غنى المعجم المصاحب لهذا المحلل الصرفي، فإذا لم تكن الوحدةُ المعجميةُ مدرجةً مسبقًا في المعجمِ المصاحبِ فإن هذا المحلل الصرفي لا يستطيع تحليل الكلمة المدخلة. ولعل أحد الأسباب التي دعت الباحثين إلى اتباع هذه

[1] . تُستخدم عبارات "قاعدة البيانات المعجمية" وعبارة "المعجم المصاحب" في هذه الدراسة بمعنى واحد، ويقصد بهما وجود معجم للوحدات المعجمية كمكون من مكونات المحلل الصرفي بحيث لا يمكن التحليل الصرفي من دونه.

الطريقة تكمن في الخصائص الصرفية لبنية الكلمة العربية التي تتسم بتداخل الصرف الخطي/ الإلصاقي وغير الخطي في بنائها بالإضافة إلى ما يعتري بنية الكلمة الداخلية من تغيرات فونولوجية تطرأ عليها عند اتصال اللواصق الصرفية بها، وتحديدًا بنية الفعل العربي، وهذه أمور تشكل تحديات أمام أي محاولة لبناء محلل صرفي للكلمات العربية من دون وجود قاعدة بيانات معجمية. وعلى الرغم من أهمية المعجم لكثير من التطبيقات، إلا أن الاعتماد عليه بشكل كامل في التحليل الصرفي قد يشكل عقبة لعدة أسباب، منها:

1) أن العربية تتسم بالثراء المعجمي الناتج عن الخاصية التوليدية الاشتقاقية التي تسمح بتوليد مفردات متعددة من الجذر الواحد، ومعظم هذه المشتقات، وتحديدًا المشتقات الفعلية، لا تُدرج في المعاجم العربية القديمة أو الحديثة، كما أشرنا سابقًا، لكونها مفردات يمكن التنبؤ بدلالتها من صيغتها الصرفية ومن المعرفة السابقة بالجذر.
2) أن نظام الفعل العربي يتصف بقدرته على استيعاب الأفعال الأجنبية المقترَضة من لغات أخرى وتكييفها حسب هذا النظام، وهي مسألة معروفة عند المتخصصين نوقشت من بداية الدراسات العربية القديمة تحت ما يسمى (بالمعَرَّب)، وهذه المفردات الأجنبية المعَرّبة قد تكون أكثر شيوعًا واستخدامًا في اللغة المعاصرة من تلك المفردات العربية التي تحتوي عليها المعاجم العربية القديمة.
3) أن الاعتماد على قاعدة بيانات معجمية مستخلصة من المعاجم العربية أو من المدونات النصية تظل ناقصة مهما كان الجهد المبذول لاستيعاب أكبر قدر ممكن، وهنا تظل الحاجة إلى التحديث مستمرة، وهذا التحديث شيء لا نجده يصاحب المحللات المتاحة بين أيدينا؛ فكثيرٌ من الأفعال المعاصرة سواء العربية أو المعرّبة لم تستطع هذه المحللات التعامل معها، وفي الجدول [1] أمثلة توضح هذه المشكلة.

أفعال عربية أصلية	أفعال عربية مشتقة	أفعال مُعرّبة
بَتَكَ	استَجْرَلَ	هَكَّرَ
فَرَعَ	أَزْكَسَ	سَنَّبَ
مَشَعَ	كَيّفَ	أَثْمَتَ
نَمَصَ	انْقَلَعَ	كَنْسَلَ
ناشَ	احْتَلَعَ	شَيَّكَ

والجدول [1] يقدم أمثلة لبعض تلك الأفعال التي لم تستطع بعض المحللات الصرفية التي بين أيدينا تحليلها

ولهذا السبب تقدم هذه الدراسة محللا صرفيًا غيرَ معجمي للأفعال العربية يستطيع أن يتغلب على المشكلات السابقة، حيث يقوم بتحليل الفعل الذي يأتي على أوزان الفعل العربي، سواء أكان من الأفعال العربية القديمة أو المعاصرة أو من الأفعال المعّربة، دون الحاجة إلى الاستعانة بمعجم، وقبل أن نقدم وصفًا لفكرة هذا المحلل، سوف يتولى القسمُ القادمُ من هذه الورقة تقديمَ ملخصٍ للدراسات السابقة المعنية بالمحللات الصرفية العربية، وتحديدًا الأكثر شيوعًا، وبيان كيفية اعتمادها بشكل مباشر أو غير مباشر على قاعدة بيانات معجمية.

الدراسات السابقة: محللات صرفية معتمدة على المعجم

بدأ الاهتمام بالمحللات الصرفية العربية مع بداية معالجة اللغة العربية آليا. وقد اختلفت المقاربات باختلاف تقنيات التحليل الصرفي وذلك من النماذج القائمة على البرمجة المسبقة (Said وآخ., 2018) كالدراسات التي سعت إلى بناء محللات صرفية عربية باستخدام تقنية الحالات المتناهية (Attia وآخ., 2011; Beesley, 1996, 1998; Habash وآخ., 2005; Hulden, 2009) إلى النماذج المعتمدة على تعلم الآلة بما في ذلك خوارزميات الشبكات العصبية الاصطناعية (Ayed وآخ., 2012; Zalmout & Habash, 2017). وسنركز هنا على أشهر هذه المحللات الموجهة للعربية الفصحى، وأولها محلل باكولتر (Buckwalter, 2002)، وهو أكثرها استخدامًا في معالجة اللغة العربية والذي اعتمد عليه المحلل النحوي الشجري المعروف ببنك بنسلفانيا، وقامت عليه بعض المحللات الموجودة على الإنترنت كمحلل Araflex[2]. وكما هو معلوم فإن محلل باكولتر الصرفي يعتمد على قواعد بيانات مصاحبة، منها معجم للجذوع وآخر للسوابق وثالث للواحق. وأهم هذه المعاجم هو معجم الجذوع الذي يحتوي على أشكال الوحدات المعجمية عند اتصالها باللواصق، وليس الأصل المعجمي، ويحتوي على ما يقرب من 40,648 جذع، وقد اعتمد في تحديد هذه الجذوع على معاجم عربية قديمة، وخرجت نسخته المطورة باسم محلل ساما (Maamouri وآخ., 2010). وفي فترة لاحقة دُمج محلل ساما مع مصنف آلي سياقي عرف بـــ (MADA) (Habash وآخ., 2009) يقوم بإزالة

[2]- انظر الرابط: http://lexanalysis.com/araflex/araflex.html

اللبس الصرفي بالإضافة إلى قيامه ببعض المهام الأخرى مثل تقطيع الكلمة وتحديد الأصل المعجمي والتشكيل وتحديد نوع الكلمة، وظهرت النسخة المطورة منه في 2014 (Pasha وآخ., 2014)، وهي النسخة المستخدمة ضمن أدوات كامل (Obeid وآخ., 2020) الصادر عن معمل النمذجة الحاسوبية بالجامعة الأمريكي في أبوظبي[3].

وقد لُوحظتْ المشكلات المتعلقة بمحلل باكولتر بنسخه المختلفة كمشكلة الاعتماد على مفردات قديمة ومشكلة الاعتماد على المعاجم فقط دون الاعتماد على المدونات، ولمعالجة المشكلة الأولى قام (Attia وآخ., 2011) ببناء محلل ومولد صرفي للغة الفصحى المعاصرة يعتمد على تقنية الحالات المتناهية (Finite-state Technology) حرص فيه على اشتمال معجمه على مفردات من اللغة المعاصرة كما سعى إلى استبعاد كثيرٍ من المفردات الموجودة في المعاجم السابقة التي أظهرت المدونات عدم استخدامها في لغتنا المعاصرة. والمهم هنا هو أن هذا المحلل كغيره من المحللات الصرفية يعتمد بشكل أساسي على وجود معجم يحتوي على الوحدات المعجمية، إلا أنه اعتمد في معجمه هذا على ما يسمى بالأصل المعجمي أو الوحدة المعجمية (lemma) بدلا من الاعتماد على الجذر أو الجذع كما هو في المحلات السابقة؛ وذلك لكون هذه الطريقة أقلَ تعقيدًا من الطريقة المعتمدة على الجذر أو تلك المعتمدة على الجذع. ويقرب مجموع الأسماء والأفعال والصفات التي اشتمل عليها معجمه من 30587 وحدة معجمية بعد استبعاد الكلمات النادرة والمنقرضة. وقد بلغ دقة هذا المحلل كما ورد في تقريره 85.73%. ومن المحللات الصرفية الأخرى المعروفة في هذا المجال محلل الخليل (Boudchiche وآخ., 2017; Boudlal وآخ., 2010)، وهو محلل صرفي نحوي موجه لتحليل الكلمات العربية الفصحى، وقد صدرت نسخته الأولى عام 2010 والثانية عام 2017، وهو كغيره من المحللات يعتمدُ على وجود قواعد بيانات/معاجم بعضها يختص باللواصق السابقة وبعضها يختص بالجذوع، والبعض الآخر يشتمل على اللواحق. ويتلخص عمل هذا المحلل في مطابقة شكلِ الكلمةِ المدخلةِ مع قواعد البيانات الخاصة بالجذع واللواصق السابقة، من ثم تحليلها

[3] - للاطلاع على أدوات كامل(camel_tools) يرجى مراجعة الرابط: https://camel-tools.readthedocs.io/en/latest/

وفقًا لذلك. وفي نسخته الثانية، أُضيفتْ خاصية إدراج الوحدة المعجمية وصيغتها الصرفية في نواتج التحليل الصرفي. ومن الدراسات الحديثة التي سعت إلى تلافي بعض المشكلات الموجودة في المحللات السابقة (Alkhairy وآخ., 2020; Altantawy & Alsalman, 2020 وآخ., 2010, 2011; Ghembaza وآخ., 2018; Iazzi وآخ., 2021; Sawalha & Mahyoob, 2018 وآخ., 2013)، وكلها محللات تعتمد على قاعدة بيانات معجمية.

الطريقة المقترحة: محلل صرفي غير معجمي

للتغلب على مشكلة اعتماد المحللات الصرفية العربية على المعجم، نشرح في هذا المبحث الطريقة التي اتبعناها في بناء محلل صرفي غير معجمي للأفعال. وقد بُني هذا المحلل باستخدام خوارزميات الحالات المتناهية (Finite-state technology) دون الحاجة إلى إضافة معجم مصاحب، وهو ما يسمح بتحليل أيِّ فعلٍ عربيٍّ منضبطٍ بأوزان العربية الفصحى سواء أكان عربيًا قديمًا أم عربيًا معاصرًا أم مُعرّبًا. وقبل الشروع في شرح هذا النظام، سنخصُّ القارئَ العربي غيرَ المطلع على تقنية الحالات المتناهية بمقدمةٍ موجزةٍ للتعريف بها وبطريقة عملها.

نبذة عن تقنية الحالات المتناهية (Finite-state Technology) :

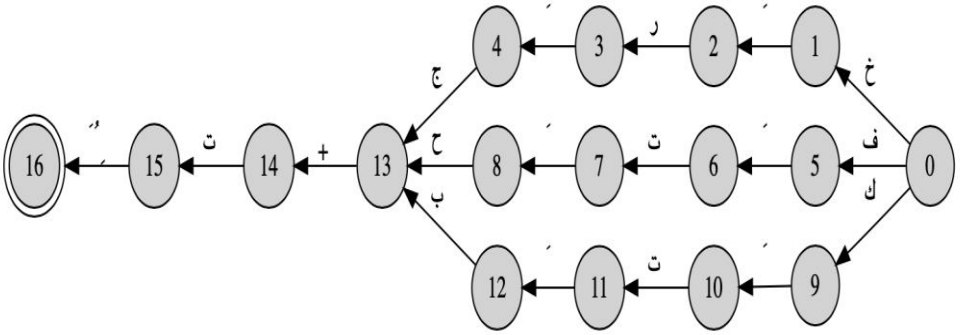
تعد خوارزميات الحالات المتناهية (Finite-state Technology) إحدى التقنيات الفعالة في معالجة اللغات صرفيًا لما تتسم به من القدرة على العمل في اتجاهين في آن واحد (Hulden, 2018). ونعني بالقدرة على العمل في اتجاهين القدرة على تحليل الكلمة إلى الوحدة المعجمية واللواصق الصرفية إلى جانب القدرة على عمل العكس وهو توليد الكلمة بمجردِ معرفةِ الوحدةِ المعجميةِ ومعرفةِ المعلومات الصرفية الملائمة للسياق المطلوب. وتستند خوارزمياتُ الحالات المتناهية على نظرية المجموعات الرياضية المستخدمة في وصف اللغات الصورية ووصف العلاقة بين المجموعات الرياضية وعلى ما يسمى بالتعابير الرياضية المنتظمة (Regular Expressions). ووفقًا لنظرية المجموعات، فإن اللغة، بمفهومها الصوري الرياضي، تُعرّفُ بأنها مجموعة من متواليات (string) تتشكلُ من عددٍ نهائي من الرموز الألفبائية (Alphabets)، وتعد هذه المجموعةُ مجموعةً جزئيةً من مجموع المتواليات الممكن توليدها من ذلك العدد النهائي من الرموز، ويقصد بمجموعة جزئية هنا

المتواليات السليمة/الصحيحة قواعديًا فقط من مجموع المتواليات السليمة وغير السليمة التي يمكن توليدها باستخدام تلك الألفبائية. ولو أخذنا تصريفات الأفعال العربية، كمثال للتوضيح، فإننا نستطيع القول بأنها مجموعة جزئية من مجموع الكلمات/المتواليات المولدة من الألفبائية العربية التي ستشتملُ بالتأكيد على كلماتٍ غير سليمة وكلمات سليمة. والذي يميز ما هو صحيح من غيره (أي ما يسمى لغة) هو القواعد والقيود الواصفة لنظام تلك اللغة، وهي عدد نهائي من القواعد أو القيود الواصفة المسؤولة عن توليد عدد لا نهائي من التعابير/الكلمات التي تعتبر صحيحة وفقا لتلك اللغة. وتُكتب هذه القواعد باستخدام القوانين الرياضية المستمدة من نظرية المجموعات كقانون الإلصاق(concatenation) وقانون الاتحاد (union) وقانون التقاطع(intersection) والفارق (differences)، وغيرها من القوانين والعمليات الرياضية المعروفة في نظرية المجموعات.

وتعمل تقنية الحالات المتناهية (Finite State Technology) وفقًا لخوارزمية تقوم بقراءة أي مُدخل (input) ومطابقته شكليًا مع القواعد التي تحكم تلك اللغة والمعَرَّفة مسبقًا، وهذه القراءة تبدأ من بداية الكلمة/التعبير إلى نهايته على شكل حالات متوالية(تمثل هذه الحالات الدوائر الموضحة في الشكل [1])، أي بقراءة هذا التعبير المدخل حرفًا حرفًا ومطابقته مع الشكل الذي يمثل القاعدة، فإذا أمكن الوصول إلى الحالة النهائية (الدائرة المضعفة في الشكل) التي تمثل النقطة النهائية المعرفة مسبقًا، فإن هذا التعبير يُعد مطابقًا لقواعد وقوانين تلك اللغة، وإن لم يكن كذلك بحيث لم تتمكن هذه الآلة من الوصول إلى النقطة/الحالة النهائية فإن هذا التعبير سيكون مرفوضًا وفقًا لهذه الخوارزميات، ومن ثم يعتبر غيرَ منسجمٍ مع قواعد تلك اللغة، والشكل الموضح في [1] يبين كيف تقرأ هذه الخوارزمية الكلمات العربية [خَرَجتُ، خَرَجتَ، خَرَجتِ، فَتَحْتُ، فَتَحْتَ، فَتَحْتِ، كَتَبْتُ، كَتَبْتَ، كَتَبْتِ].

شكل 1

يوضح هذا الشكل تمثيل تصريفات بعض الأفعال العربية باستخدام تقنية الحالات المتناهية.

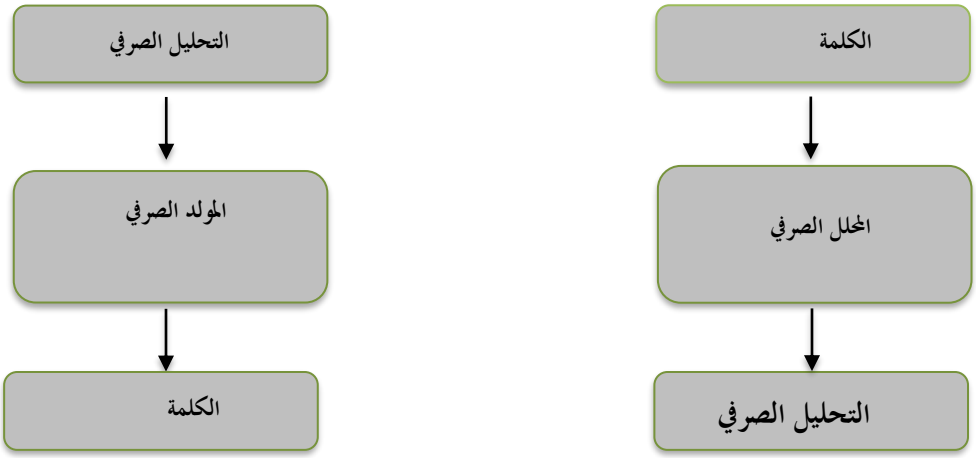


وإلى جانب استخدام تقنية الحالات المتناهية في تحديد/اختبار التعابير أو الكلمات وقياس مدى قاعديتها (أي انتمائها إلى لغة ما)، تستخدم هذه التقنية ما يسمى بالمحولات (Transducers)، وهي تمثيل مجرد على شكل مستويين كل مستوى ينتمي لمجال معين يمثل لغةً أو مجموعةً جزئية من لغة، وهذه الخاصية تعد أساسًا مهماً لما يُعرف بالمحولات الصرفية (morphological transducers) المستخدمة فيما يسمى، في معالجة اللغات الطبيعية، بالتحليل والتوليد الصرفيين. وأساس هذه المحولات الصرفية يقوم على فكرة الربط أو إقامة علاقة رياضية بين مجالين، أحدهما يمثل مجموعة الكلمات الصحيحة التي تنتمي إلى لغة معينة (العربية مثلا) والمجال الآخر يمثل مجموعة التحليلات/الواسمات الصرفية الصحيحة المطابقة للمورفيمات العربية. ولو أخذنا تصريف الفعلَ الماضي مثالا على ذلك، لأمكن التمييز بين مجالين أحدهما يمثل مجموعة الأفعال الماضية الصحيحة بتصريفاتها المختلفة (وهي مجموعة جزئية من مجموعة تصريفات الأفعال العربية)، أما المجال الثاني فهو مجال التحليل الصرفي المشتمل على تتابعٍ لترميز المعلوماتِ الصرفية الخاصة الفعل الماضي والتي تشتمل على الجذر والصيغة والزمن أو الجهة والشخص والجنس والعدد وغير ذلك من المعلومات التي تشفرها اللواصق الاشتقاقية أو التصريفية الفعلية. ودور الخوارزميات في هذه المحللات هو الربط بين المورفيمات المكونة للكلمة الفعلية (كلمة كتبوها مثلا) وما يناسبُ أو يقابلُ كلَّ مورفيم من تحليل صرفي (وَاسِمَة) صحيح موجود في مجال التحليلات الصرفية الصحيحة والممكنة الخاصة بالفعل

الماضي. وإذا كنا قد ذكرنا سابقًا أن مما يميز تقنية الحالات المتناهية، خصوصًا في مجال الصرف، قدرتَها على العمل في اتجاهين (التحليل والتوليد)، فتوضيح ذلك هو أن المدخل (input) قد يكون كلمة فعلية معينة، وهنا سيكون عمل هذه الخوارزميات هو تحديد المعلومات الصرفية (outputs) التي تُشَفِّرُها هذه الكلمة (أي القيام بعملية التحليل الصرفي)، وقد يكون المدخل (input) تحليلا صرفيًا محددًا، وهنا سيكون عمل هذه الخوارزميات هو ربط هذا التحليل بالكلمة المناسبة له، وتزويدنا بهذه الكلمة (أي القيام بعملية التوليد)، كما هو موضح في الشكل [2].

شكل 2

طريقة عمل كل من المحلل والمولد الصرفيين



1.1. خطوات بناء المحلل الصرفي

لبناء هذا المحلل الصرفي بطريقة غير معجمية سوف نستعين بأمرين: أحدهما الأوزان الفعلية والآخر هو تقنية الحالات المتناهية وتحديدًا المحولات وما يلازمها من عمليات رياضية. وسيشرحُ هذا المبحث طريقة بناء ذلك المحلل في عدة خطوات. الخطوة الأولى حُدِّدت فيها المجموعات التي سيتم الربط بينها باستخدام تلك المحولات، وفي الخطوة الثانية بُنيت المحولات الصرفية اللازمة، وعددها خمس محولات، للربط بين الشكل والمعلومة الصرفية التي تناسبه. وفي الخطوة الثالثة دُمجت تلك المحولات مع بعضها لتكون محولا واحدًا يتكون من مستويين فقط: أحدهما يمثل الكلمة المدخلة والآخر يمثل المعلومات الصرفية (الجذر، والصيغة، والزمن/الجهة، والشخص والجنس والعدد، ونحو

ذلك).وقد استُخدم في ذلك الدمج عملية التوليف التراكمي (incremental composition) وعملية الإلصاق الخطي (concatenation)، وبهاتين العمليتين تمكنا في النهائية من الحصول على محول واحد يقرأ الكلمة المدخلة في أحد المستويات، بعد أخذ القواعد الفونولوجية الإملائية في الحسبان، ثم يزودنا بالتحليل الصرفي المناسب لتلك الكلمة في مستوى التحليل الصرفي، وسيتولى هذه المبحث تفصيل تلك الخطوات فيما يأتي:

1.1.1.الخطوة الأولى: تحديد المجموعات

أول خطوة في هذه الخوارزمية هي تعريف المجموعات، وأولها مجموعة الصوامت العربية وأشباه العلل التي تُشكل الجذورَ العربية للوحدات المعجمية، وذلك بوضعها في مجموعة رُمِزَ لها بالرمز (C)، وكذلك تعريف وتحديد الصوائت القصيرة كمجموعة أخرى مستقلة ورُمزَ لها بالرمز (V). أما الواسمات التي ترمُزُ إلى المعلومات الصرفية الجهة والشخص والعدد ونحو ذلك فقد عُرِّفت في مجموعة مستقلة باسم (Tags)، وقد استُخدِم في ذلك الترميزُ المتعارفُ عليه في التحليل الصرفي. وقد أُضيفَت الصيغةُ الصرفيةُ للفعل إلى مجموعة الواسمات لكون صيغة الفعل العربي ذات دلالة صرفية، وهو ما يعني ضرورة إدراجها ضمن المعلومات الصرفية التي سيزودنا بها المحلل الصرفي.

1.1.2.الخطوة الثانية: بناء المحولات الصرفية.

بعد تحديد/تعريف المجموعات السابقة، تأتي الخطوة الثانية، وهي بناء المحولات الصرفية، وعددها خمسة محولات. وأول هذه المحولات محولُ الجذر والصيغة الصرفية الذي سيقوم بتمثيل معلومات الجذر والصيغة في مستويين: المستوى العلوي (مستوى المعلومات الصرفية)، ويشمل الجذرَ والصيغةَ والمستوى السفلي وهو المستوى الذي يمثل الجذر مع رمز الحركات القصيرة (V)، ولو أخذنا الجذر [ك/ت/ب]، على سبيل المثال، والصيغة (CVCVC) التي عُرّفت في هذا المحول بـ (C 0:V C 0:V C)، فإن تمثيلها في هذا المحول سيكون على النحو الآتي:

0 ?*": CVCVC"?*) V b:b :k:k 0:V t:t 0(

ويجدر أن نلاحظ هنا أن ما قبل العلامة (:) من جهة اليسار يمثل المستوى العلوي الذي يظهر فيه التحليل الصرفي، (وهو هنا الجذر ورمز الصيغة فقط)، ويمثل ما بعدها المستوى السفلي الذي

يمثل شكل الكلمة مع الحركات، وهو في هذا المثال (kVtVb). وقد استُخدم (*?) الذي يسبق رمزَ الصيغةِ، وهو رمز رياضي يعني وجود عدد 0 أو أكثر من أي حرف/رمز؛ وقد استخدم ليمثلَ أيَّ لاصقةٍ صرفيةٍ لاحقةٍ لجذع الكلمة يمكن أن نضيفها كما استُخدم كذلك بعد الصيغة ليتيح لنا إضافة أيَّ واسمة أخرى بعد واسمة الجذر والصيغة. وللتغلب على صعوبة تعامل تقنية الحالات المتناهية مع تضعيف العين في الأفعال المضعَّفة، قمنا بأمرين: اعتبرنا العين المضعفة متوالية من الصامت وحركة الشدة التي يرمز لها بالرمز (~) المتعارف عليه في ترميز باكولتر، وذلك على النحو (C~) بحيث يتم إحلالها محل عين الفعل حينما تواجه الآلة بعض الأفعال المضعفة. ولكي يحصل التمييز بين مستوى الجذر ومستوى الصورة السطحية للفعل المضعف، عرفنا التضعيف الذي تمثله الشدة في المستوى السفلى فقط الذي يمثل الصورة السطحية للفعل المضعف، وقد مكننا ذلك من مقابلة الصوت الثاني من الصامت المضعف بالرمز ε الذي يرمز له الصفر في المستوى العلوي من التمثيل، بحيث يكون تمثيل صيغة فَعَّلَ ($C_1VC_2C_2VC_3$) على النحو الآتي:

C 0: V Geminate 0: V C

أما المحولُ الثاني فقد صُمم ليقوم بالتعرف على نمط الحركات القصيرة في البنية الداخلية للفعل وربطه بالتحليل الصرفي المناسب لهذه الحركات، وفي هذه المرحلة تُضاف للمستوى العلوي المعلوماتُ الصرفيةُ التي تشفرها الحركات (حركة الفاء وحركة العين) وتحديدا البناء للفاعل والبناء لغير الفاعل، خصوصا في حالة الفعل الماضي. وقد مُثِّل كل نمط من أنماط الحركات القصيرة (حركة فاء الفعل وعينه) في هذ المحول على النحو الآتي:

:0 ?*"Act":0 "Perf"?* ?* [V:a|V:0] ?*[V:a|V:i|V:u|V:o]

ونلاحظ هنا أن الواسمة ("Perf") التي ترمز لجهة الماضي والواسمة (Act) التي ترمز للبناء للفاعل قد أُضيفتا بشكل خطي في المستوى التحليلي العلوي. وعند بناء هذا المحول، كان من الضروري الفصل بين الجهتين (جهة المضي/ التمام وجهة المضارع/ أو اللاتمام)؛ لأن معالجة الفعل الماضي تختلفُ عن معالجة الفعل المضارع لما قد يحدث من تغير في الحركات الداخلية للفعل عند إضافة

اللواصق التصريفية الخاصة بالفعل المضارع ولكون صيغة البناء للفاعل/لغير الفاعل تُدرك من حركة اللاصقة/السابقة التي تسمى حرف المضارعة. هذا بالإضافة إلى أن الكلمة الناتجة عن عملية دمج هذا المحول مع المحول السابق تختلف في الماضي عنها في المضارع؛ ففي حالة الماضي تكون النتيجةُ النهائيةُ كلمةً حقيقيةً بينما في حال المضارع ينتجُ عن ذلك كلمة غير حقيقة تمثل مستوى متوسطًا في التحليل يكتملُ بإضافة اللواصق التصريفية الخاصة بصيغة المضارع.

وقد خصص المحول الثالث لإدراك اللواصق التصريفية وربطها بالواسمات التي تمثل وظائف هذه اللواصق. ولو أخذنا الفعل (كَتَبَتْ katabat) مثالا للتوضيح، فإن مهمةَ هذ المحول ستكون التعرفَ على اللاصقة (تْ =t) في المستوى السفلي(المدخل) وربطها بالمعلومات الصرفية في المستوى العلوي (+3Per_Fem_Sig) والتي تمثل الشخص والتأنيث والعدد، وقد رُوعي في عملية الإلصاق الخطي بالنسبة للمضارع المطابقة بين السوابق واللواحق من حيث الشخص، فالسابقة (تَ= ta) قد تدل على المخاطب أو الغائبة، والذي يميز ذلك هو اللاحقة التصريفية، كما رُوعي كذلك اختلاف حركاتِ تلك السوابق باختلاف صيغةِ البناء للفاعل أو غير الفاعل، وكذلك كان لا بد هنا من التمييز بين اللاصقة التي تدل على البناء للفاعل (تَـ) واللاصقة التي تدل على البناء لغير الفاعل (تُ) باعتبار كل واحدة منهما لاصقة مستقلة.

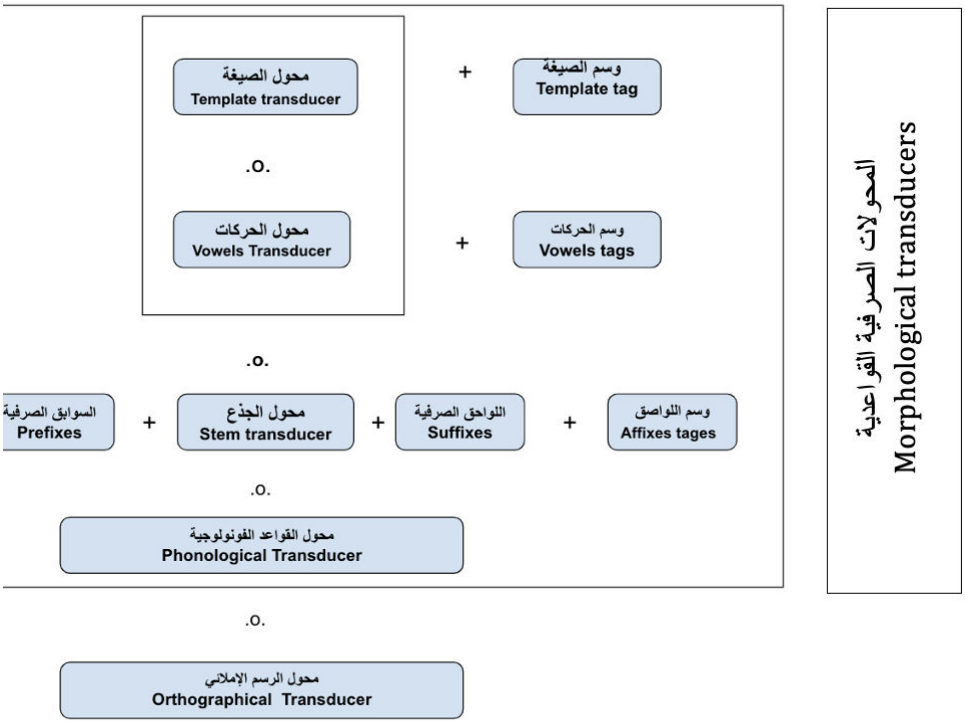
ونظرًا لاعتمادنا على الأوزان الأصلية للفعل العربي سنكون مضطرين لإضافة محول آخر رابع خاصٍ بالقواعد الفونولوجية، كقواعد الحذف والإبدال الإضافة وغيرها، ومهمة هذه المحول قراءة البنية السطحية للكلمة بدلا من أصلها الافتراضي الذي يسمى بالبنية العميقة. فمن المعلوم أن البنية العميقة للفعل يعتريها بعض التغيرات الصوتية المحكومة بقوانين فونولوجية يجب أخذها بعين الاعتبار؛ فشكل الكلمة الذي قد ينتجُ عن دمج المحولات السابقة يمثل ما يسمى في الفونولوجيا التوليدية بالبنية العميقة، وهذا يستدعي القيام بالتحويلات اللازمة للوصول إلى البنية السطحية أو ما يمكن أن نسميه بالشكل المنطوق. وإضافة هذا المحول تمكّنُنا من التعامل مع الكلمة على أنها تتكون من مستويين: أحدهما يمثل البنية العميقة والآخر يمثل المستوى السطحي المنطوق.

ومن المتعارف عليه عند معالجة اللغة العربية المكتوبة ضـــرورةُ أخذ طبيعة الرســـم الإملائي في الاعتبار لما قد يترتب على إهماله من تعثر أي معالجة آلية للنصـــوص المكتوبة. وقد عُولجت هذه المشـكلة بصـورة تشـبه الطريقة التي عوجلت بها الظواهر الفونولوجية، وذلك ببناء محول خامس يتم فيه التعامل مع الرسم الإملائي على أنه يتكون من مستويين من العلاقة أحدهما المستوى الافتراضي الإملائي والآخر يمثل المستوى النهائي الذي يمثل الصورة المكتوبة الخاضعة لقوانين الرسم الإملائي. وهذه الخطوة ســـتطبق على أنها المرحلة النهائية لبناء المحلل الصـــرفي بحيث تأتي تالية لكل المراحل السـابقة. ولعل من أوضـح الأمثلة على هذه القوانين الإملائية التي يجب أخذها في الاعتبار مسألةَ الكتابةِ من دون الحركات القصـــيرة/التشـــكيل. وتجاهل ذلك أثناء بناء المحللات الصـــرفية العربية سـيؤدي إلى فشـلها في إدراك وتحليل الأفعال غير المشـكولة، وقد عالجنا هذه المشـكلة باسـتخدام قاعدة إعادة الكتابة الآتية: 0 (→) Vowels التي تقوم بعملية تحويل الحركة إلى لا حركة تحويلا اختياريًا؛ ليتمكن من قراءة الكلمة المشكولة وغيرها؛ ولتحقيق ذلك استخدمنا الأقواس () الدالة في خوارزميات تقنية الحالات المتناهية على الاختيار، وهي إحدى الخصـائص التي تتيحها هذه التقنية. وهناك الكثير من الأمثلة على القواعد الإملائية التي لا يتسع المقام والمقال لذكرها.

1.2. الخطوة الثالثة: دمج المحولات في محول واحد

شكل 3

هذا الشكل يوضح عملية بناء المحلل الصرفي من عدة محولات وكيفية دمج هذه المحولات في محول واحد. تشير علامة (.o.) إلى عملية الدمج/التوليف، بينما تشير علامة (+) على عملية الإلصاق الخطي.



بعد الانتهاء من بناء محولات صرفية مستقلة، كما وضحته الخطوة السابقة، تأتي الخطوة الثالثة والأخيرة التي تتمثل في عملية دمج تلك المحولات في محول واحد يتكون من مستويين: المستوى السفلي (Lower) ويمثل مستوى الكلمة المدخلة والمستوى العلوي (Upper) ويمثل مستوى التحليل الصرفي المقابل لهذه الكلمة، وهو مستوى يحتوي على المعلومات الصرفية الخاصة بالجذر والصيغة والبناء للفاعل أو غير الفاعل والشخص والجنس والعدد ونوع ضمير المفعول به المتصل، وغير ذلك من المعلومات المشفرة في بنية الكلمة الداخلية أو المشفرة من خلال اللواصق الصرفية. وفي هذه الخطوة يتم الدمج بشكل تدريجي تراكمي يبدأ بعملية دمج محول الصيغة في محول الحركات القصيرة باستخدام عملية الدمج التوليفي المعروفة في خوارزميات هذه المحولات بـ(composition) التي ينوب عنها الرمز (.o.) في الشكل [3]. وأثناء عملية الدمج بين المحولين تتم عملية إضافة الجذر

والصيغة في المستوى العلوي لتكون من ضمن المعلومات الصرفية التي يزودنا بها التحليل. أما في المستوى السفلي فيتكون لدينا ما يسمى بالجذع. بعد ذلك، يتم دمج محول الجذع بشكل خطي مع محول اللواصق السابقة و/أو اللاحقة باستخدام عملية الإلصاق (concatenation) التي يمثلها الرمز (+) في الشكل [3]، وفي هذه المرحلة يتم إضافة المعلومات/الواسمات المتعلقة بالجهة والبناء للفاعل/ غير الفاعل والشخص والعدد وغيرها من اللواصق في المستوى العلوي. ولأن الصورة الناتجة عن عملية دمج الجذر في الصيغة والحركات قد تؤدي إلى إدراك الصورة العميقة للكلمة فقط، تأتي عملية دمج صورة الكلمة (في بنيتها العميقة) مع محول القواعد الفونولوجية الذي سيمكن هذا المحول من إدراك الكلمة في صورتها السطحية/الظاهرة. وأخيرًا، تأتي عملية دمج المحول الناتج عن عملية الدمج في الخطوات السابقة مع محول الرسم الإملائي للتغلب على مشكلات الكتابة ومن أهمها مشكلة الكتابة من دون حركات. والجدول [2] يقدم أمثلة لعملية التحليل الصرفي باستخدام هذا النظام. وهنا نلاحظ أن لدينا مستويين: المستوى السفلي (Lower) الذي تظهر فيه الكلمة المدخلة، والمستوى العلوي (upper) الذي تظهر من خلاله المعلومات/الواسمات الصرفية.

جدول2

بعض الأمثلة لمخرجات النظام حيث يمثل(UPPER) مستوى التحليل الصرفي بينما يمثل (LOWER) مستوى الكلمة المدخلة.

UPPER: LOWER:	ktb+CVCVC+Perf+Act+3Per_Fem_Sig katabat (كَتَبَتْ)
UPPER: LOWER:	fhm+CVCVC+Perf+Act+1Per_Pl fahimnA (فَهِمْنا)
UPPER: LOWER:	qbl+staCCVC+Perf+Act+1Per_Sig {istaqobalotu
UPPER: LOWER:	qbl+staCCVC+Perf+Act+3Per_Fem_Sig {istaqobalat(اسْتَقْبَلَتْ)
UPPER: LOWER:	CVCCVC+Imperf+Act+snb+3Per_Masc_Sig yusan~ibu(يُسَبِّبُ)
UPPER: LOWER:	Hfl+CtVCVC+Perf+Act+3Per_Fem_Sig {iHotafalat(احْتَفَلَتْ)
UPPER: LOWER:	Hfl+CtVCVC+Perf+Act+3Per_Fem_Pl {iHotafalona(احْتَفَلْنَ)

2.النتائج والتقييم

تختلف طرائق تقييمِ المحللات الصـــرفية باختلاف الأهداف منها وباختلاف مســـتوى التحليل الذي يهدف إليه هذا المحلل أو ذلك، وقد درجت الدراســـات على اتباع طرائق آلية وأخرى يدوية يتم من خلالها تقييمُ أداءِ ودقةِ المحلل. إحدى هذه الطرائق الألية تتمثل في مقارنة أداء المحلل الصرفي قيد الدراســـة بما يســـمى (Gold Standard)، وهي بيانات تم تجهيزها وتحليلها يدويًا لتكون مادةً لاختبار أداء ذلك المحلل واختبار دقته. ومن طرائق التقييم الأخرى مقارنةُ أداء المحلل الصـــرفي بمحللاتٍ أخرى ســـابقة. وهنا قد تكون المقارنةُ من وجهٍ من عدةِ وجوه كأن تكون المقارنة على مســـتوى قدرةِ المحلل على اســـترجاعِ كافة المعلومات الصـــرفية بدقة أو من حيث قدرته على تحليل الكلمة باختلاف أشكالها الإملائية (بالتشكيل أو بدونه) أو قدرته على تحليل كلماتٍ جديدة، إلى غير ذلك من المعايير التي قد تتخذ للمقارنة والتقييم.

ولأن الهدفَ من المحلل الذي بين أيدينا هو التغلبُ على مشـــكلةِ اعتمادِ المحللات الصـــرفية السابقة على وجود قاعدة بيانات معجمية، فإن تقييمنا لأداء هذا المحلل سوف يقتصر فقط على مقارنة نتائج هذا المحلل بنتائج المحللات الصـــرفية الأخرى من حيث قدرته أولا على تحليل كلماتٍ عربيةٍ قديمةٍ وكلماتٍ معاصـــرةٍ (عربيةٍ أو مُعرَّبةٍ) وثانيًا من حيث قدرته على تحليل الكلمة المدخلة ســـواء أكانت مشـــكولة أم لا. وقد قُيِّم أداء هذا المحلل الصـــرفي غير المعجمي بمقارنته باثنين من المحللات المتاحة على شـــبكة (الإنترنت)، وهما محلل Araflex ومحلل مجموعة كامل (camel_analyzer)؛ ولأن محلل Araflex لا يتوفر إلا على (الإنترنت) وهو ما يصعُب معه المقارنة آليا، ســـنعتمد في تقييمنا على التقييم اليدوي وذلك باختبار 100 فعل تشـــتمل على أفعال قديمة وأفعال معاصرة عربية أو مُعرَّبة، وقد اختِيرتْ هذه الأفعال بشكل عشوائي مع مراعاة شمولية الأوزان الصرفية وكذلك مراعاة اختلاف اللواصق التصريفية. والجدول رقم [3] يوضح نتائج هذه المقارنة

جدول 3

نتائج تقييم أداء المحلل الصرفي الحالي مقارنة بـ (Araflex) و (Camel_analyzer)

	نوع الأفعال	عدد الأفعال	مقارنة النتائج		
			AraFlex	Camel_ Analyzer	المحلل الحالي
أفعال مضبوطة بالشكل	قديمة	10	٪ 60	٪70	٪90
	معاصرة	30	٪73.33	٪83.33	٪93.33
	مُعَرّبة	10	٪ 50	٪60	٪100
أفعال غير مضبوطة بالشكل	قديمة	10	٪ 70	٪70	٪80
	معاصرة	30	٪75	٪86.7	٪90
	مُعَرّبة	10	٪50	٪50	٪100
	المجموع	100	-	-	-
	متوسط النسبة		**٪63**	**٪70**	**٪92.2**

وتظهر النتائج الموضحة في الجدول [3] تفوقَ المحلل الصرفي الحالي في أدائه على المحللين الصرفيين الآخرين، وهذا غير مستغرب لأنه قد تخلص من مشكلة الاعتماد على المعجم، وهي المشكلة المصاحبة لغيره من المحللات الصرفية والمسؤولة عن تدني أدائها. والسؤال الذي قد يتبادر إلى ذهن القارئ هنا هو سؤال عن السبب الكامن خلف تحقيق المحلل الحالي تلك النسبة (92.2٪) مع أننا قد توقعنا أنه باعتماده على الأوزان الصرفية واستغنائه عن المعجم يستطيع أن يحلل أي مُدْخَلٍ فعلي. وبالنظر في المخرجات وتحديدًا الأفعال التي لم يستطع تحليلها تبين أن السبب الرئيسي هو عدم استيعاب المكون الفونولوجي لكل القواعد الفونولوجية التي تقوم بالربط بين البنية العميقة/الأصلية والبنية السطحية/الظاهرة للكلمة. فالفعل (يصْطَدِمُون)، على سبيل المثال، هو أحد الأفعال التي لم يستطع تحليلها، والسبب في ذلك راجع إلى اختلاف صورة هذه الكلمة عن غيرها من الكلمات التي تأتي على وزن (افْتَعل)، فالتغيرات الفونولوجية التي حدثت للبنية الأصلية (يصتدمون) والمحكومة بقاعدة المماثلة، قاعدة قلب التاء طاء لمماثلة الصاد في صفة التفخيم، لم يتم إدراجها في المكون الفونولوجي المصاحب لهذه المحلل الصرفي، وهو ما أدى إلى عدم قدرة هذه المحلل الصرفي على التعرف على هذه الكلمة.

ولأن هذا النظام المقترح في هذه الدراسة ليس إلا نموذجًا تجريبيًا لمحلل صرفي غير معجمي للأفعال العربية، فسيكون العمل في المستقبل منصبًا على تطوير قدرته التحليلية، وذلك بالتركيز

على المكون الفونولوجي بحيث يشتمل على جميع القوانين الفونولوجية التي تحكم بنية الفعل العربي، وبعدها سيكون العمل منصبًا على اختباره على عدد كبير من الأفعال واختبار دقة مخرجاته ومقارنتها مع مخرجات غيره.

3. الخاتمة

تناولت هذه الورقة بالشرح والتفصيل طريقةَ بناءِ محللٍ صرفي غير معجمي للأفعال العربية. وقد كان الدافع وراء هذا المقترح هو التغلب على إحدى المشكلات التي تواجهُها أشهرُ المحللات الصرفية العربية، أعني مشكلةَ الاعتمادِ على المعجم التي جعلت هذه المحللات غيرَ قادرةٍ على التعامل مع الكلمات التي لا تحتويها قواعدُ بياناتها المعجمية. وقد اعتمدنا في تطويرنا لهذا المحلل على أمرين: أحدهما الأوزانُ الفعليةُ لما تتسم به من انتظام يمكن الإفادة منه، والآخر هو تقنية الحالات المتناهية (Finite state technology) وتحديدًا خوارزميات المحولاتِ الصرفيةِ لما تتمتع به من سرعة وقدرة على التعامل مع الكلمة العربية وتحليلها على عدة مستويات وهو ما يناسبُ طبيعتَها الصرفية المعقدة.

وقد بُني هذا المحلل ابتداء من خمسة محولات صرفية مستقلة لكل منه منها أهدافه الخاصة المتعلق بمستوى من مستويات تمثيل بنية الفعل العربي، ثم دُمجت تدريجيًا وبشكل تراكمي في خطوة لاحقة مع بعضها لتنتج لنا محولا صرفيًا واحدًا يتعامل مع الكلمة المدخلة في مستويين فقط: المستوى السفلي (Lower) ويمثل مستوى الكلمة المدخلة والمستوى العلوي (Upper) ويمثل مستوى التحليل الصرفي المقابل لهذه الكلمة، وهو مستوى يحتوي على المعلومات الصرفية الخاصة بالجذر والصيغة والبناء للفاعل/غير الفاعل والشخص والجنس والعدد ونوع ضمير المفعول به المتصل، وغير ذلك من المعلومات المشفرة في بينة الكلمة الداخلية أو المشفرة من خلال اللواصق الصرفية. وقد أشرنا في موضع سابق أن هناك خاصيتين من خصائص الحالات المتناهية ساعدت في هذا الدمج: الأولى منهما ما يعرف بخاصية الدمج التوليفي (composition)، وقد أفادتنا في التعامل مع التداخل بين الجذر والصيغة والتعامل مع القواعد الفونولوجية والإملائية، والأخرى هي خاصية الإلصاق

الخطي(concatenation) التي مكنتنا من دمج الجذع بشكل خطي مع محول اللواصق السابقة و/أو اللاحقة.

وقد اتضحتْ من نتائج تقييم هذا المحلل قدرتُه على تحليل أفعال لم تستطع تحليلَها المحللاتُ الصرفيةُ المختارة للمقارنة في هذه الدراسة لعدم وجود أصولها المعجمية في قواعد البيانات/ المعاجم المصاحبة لتلك المحللات، وهو ما لا يتطلبه هذا المحلل الصرفي، فقد حقق نسبة 92.2 % مقارنة بنسبة 63 % و 70 % للمحللات الأخرى، وهذه النتائج الواعدة هي الهدف الذي قامتْ عليه هذه الدراسة. وحتى يمكن الاستفادة من هذا النموذج الصرفي وتحقيق دقة أعلى فإن العمل سيكون منصبًا في دراسات مستقبلية على تطوير قدرتِه، خصوصًا فيما يتعلق بمحول القواعد الفونولوجية، وذلك بإضافة بعض القواعد الفونولوجية التي لم يشتمل عليها هذا المحول، ومن ثم اختباره آليًا على مدونات نصية ذات نطاق أوسع.

References

Alkhairy, M., Jafri, A., & Smith, D. A. (2020). Finite state machine pattern-root Arabic morphological generator, analyzer and diacritizer. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*.

Alothman, A., & Alsalman, A. (2020). Arabic Morphological Analysis Techniques. *International Journal of Advanced Computer Science and Applications*, *11*(2). https://doi.org/10.14569/ijacsa.2020.0110229

Altantawy, M., Habash, N., & Rambow, O. (2011). Fast yet rich morphological analysis. *FSMNLP 2011 - Proceedings of the 9th International Workshop Finite State Methods and Natural Language Processing*.

Altantawy, M., Habash, N., Rambow, O., & Saleh, I. (2010). Morphological analysis and generation of Arabic Nouns: A morphemic functional approach. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*.

Attia, M., Pecina, P., Toral, A., Tounsi, L., & van Genabith, J. (2011). An open-source finite state morphological transducer for modern standard Arabic. *FSMNLP 2011 - Proceedings of the 9th International Workshop Finite State Methods and Natural Language Processing*, 125–133.

Ayed, R., Bounhas, I., Elayeb, B., Evrard, F., & Saoud, N. B. Ben. (2012). Arabic morphological analysis and disambiguation using a possibilistic classifier. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7390 LNAI*. https://doi.org/10.1007/978-3-642-31576-3_36

Beesley, K. R. (1996). *Arabic finite-state morphological analysis and generation*. 89. https://doi.org/10.3115/992628.992647

Beesley, K. R. (1998). *Arabic morphology using only finite-state operations*. 50. https://doi.org/10.3115/1621753.1621763

Boudchiche, M., Mazroui, A., Ould Abdallahi Ould Bebah, M., Lakhouaja, A., & Boudlal, A. (2017). AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer. *Journal of King Saud University - Computer and Information Sciences*, *29*(2). https://doi.org/10.1016/j.jksuci.2016.05.002

Boudlal, A., Lakhouaja, A., Mazroui, A., Meziane, A., Ould Abdallahi Ould Bebah, M., & Shoul, M. (2010). Alkhalil Morpho SYS1: A Morphosyntactic Analysis System for Arabic Texts. *international Arab Conference on information technology*.

Buckwalter, T. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0. *Linguistic Data Consortium, University of Pennsylvania, LDC Catalog No.: LDC2002L49*.

Buckwalter, T. (2007). Issues in Arabic Morphological Analysis. في *Arabic Computational*

Morphology. https://doi.org/10.1007/978-1-4020-6046-5_3

Ghembaza, M. I. E. K., Smai, A. H., & Aloufi, K. S. (2018). Arabic Solid-Stems for an Efficient Morphological Analysis. *Arabian Journal for Science and Engineering*, *43*(12). https://doi.org/10.1007/s13369-017-2938-8

Habash, N., Rambow, O., & Kiraz, G. (2005). Morphological analysis and generation for Arabic dialects. *Computational Approaches to Semitic Languages Workshop Proceedings, SEMITIC@ACL 2005*, *June*, 17–24. https://doi.org/10.3115/1621787.1621791

Habash, N., Rambow, O., & Roth, R. (2009). MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization. *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, *41*.

Hulden, M. (2009). Revisiting multi-tape automata for Semitic morphological analysis and generation. *Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages, SEMITIC@EACL 2009*, 19–26. https://doi.org/10.3115/1621774.1621779

Hulden, M. (2018). Finite-State Technology. في *The Oxford Handbook of Computational Linguistics 2nd edition*. https://doi.org/10.1093/oxfordhb/9780199573691.013.39

Iazzi, S., Iazzi, A., Laaroussi, S., & Yousfi, A. (2021). The Use of the Relational Concept in the Arabic Morphological Analysis. *International Journal of Advanced Computer Science and Applications*, *12*(11). https://doi.org/10.14569/IJACSA.2021.0121165

Maamouri, M., Graff, D., Bouziri, B., Krouna, S., & Kulick, S. (2010). LDC Standard Arabic morphological analyzer (SAMA) v. 3.1. *LDC Catalog No. LDC2010L01. ISBN*.

Mahyoob, M. (2018). Deterministic Finite State Automaton of Arabic Verb System: A Morphological Study. *International Journal of Computational Linguistics (IJCL)*, *Volume (9)*(9).

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMeL tools: An open source python toolkit for arabic natural language processing. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*.

Pasha, A., Al-Badrashiny, M., Diab, M., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. M. (2014). MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*.

Said, I., Abdellah, Y., Mostafa, B., & Driss, A. (2018). Arabic Morphological Analysis Based on Graphs and Correspondence tables between Affixes and Root. *9th*

International Symposium on Signal, Image, Video and Communications, ISIVC 2018 - Proceedings. https://doi.org/10.1109/ISIVC.2018.8709237

Sawalha, M., Atwell, E., & Abushariah, M. A. M. (2013). SALMA: Standard arabic language morphological analysis. *2013 1st International Conference on Communications, Signal Processing and Their Applications, ICCSPA 2013*. https://doi.org/10.1109/ICCSPA.2013.6487311

Zalmout, N., & Habash, N. (2017). Don't throw those morphological analyzers away just yet: Neural morphological disambiguation for Arabic. *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*. https://doi.org/10.18653/v1/d17-1073