

معالج صرفي تحليلي للواصق الكلمة المتصرفة في اللغة العربية باستخدام بحث العمق أولاً في البحث
في شبكة الانتقالات المعززة

م.د. علي فاضل مرهون	م.د. سلمى عبد الباقي محمود	م.د. علياء سلمان صابر
جامعة البصرة	جامعة البصرة	جامعة البصرة
كلية العلوم	كلية العلوم	كلية العلوم
قسم علوم الحاسبات	قسم علوم الحاسبات	قسم علوم الحاسبات

الملخص

يقدم هذا البحث نظاماً تحليلياً معجمياً صرفياً للكلمة المتصرفة المدخلة وذلك من خلال اعتماد طريقة بحث العمق كآلية ناجحة لاجتياز شبكة الانتقالات المعززة (Augment network (ATN transition التي استخدمت لتمثل العلاقة ما بين اللواصق (Affixes) والكلمة المتصرفة وطريقة جيدة لتحليل الكلمة المتصرفة ، حيث تجرى عملية التحليل الصرفي للكلمة المدخلة (اسم وفعل و حرف) وذلك بفصلها إلى جذع وسوابق ولواحق مع تحديد الخصائص المعجمية (الجنس والعدد والشخص) لجذع الكلمة المتصرفة ولجميع إضافاتها ، وقد مكنتنا آلية بحث العمق أولاً من التخلص من مشكلة الغموض في التحليل الصرفي وذلك من خلال رفض كل المسارات المشتركة وغير الصحيحة والإبقاء على المسار الأفضل .

الكلمات المفتاحية : التحليل الصرفي ، شبكة الانتقالات المعززة ، معالجة اللغات الطبيعية .

1 - المقدمة

يعد الانتشار الواسع في استخدام الحاسب الالكتروني في العديد من المجالات في البلاد العربية والتقدم الكبير في مجال نقل المعلومات وتبادلها بوساطة شبكة المعلومات وتوجه العلماء إلى تصميم حاسبات لها القدرة على التعامل مع الإنسان بلغته الطبيعية احد الأسباب الرئيسة لتوجه الباحثين إلى الاهتمام بمجال معالجة اللغة الطبيعية حاسوبياً بشكل عام واللغة العربية بشكل خاص ، فاللغة العربية هي اللغة الأم والرسمية لسكان البلدان العربية ، وإنها لغة القرآن الكريم -وواحدة من اللغات الست المعتمدة في الأمم المتحدة ، وهي من اللغات السامية التي تمتاز عن اللغات الاندو -أوربية بوجود الجذور التي يتم منها توليد معظم كلمات اللغة ، لذا تمتاز بكونها لغة اشتقاقية وعالية في التصريف وتعد هذه الميزة نقطة قوة للغة بكونها لغة غنية بالمفردات ولكنها تعد عائقاً في طريق المعالجة الآلية للغة لصعوبة حصر كل قوانين الاشتقاق والتصريف ، تمر الكلمة في نظام معالجة اللغة العربية بعدة

مراحل منها المعالجة المعجمية والصرفية والتركيبية والدلالية . وتعد مرحلة المعالجة الصرفية العمود الفقري في أنظمة معالجة اللغات الطبيعية وتطبيقاتها المختلفة مثل الترجمة الآلية (Mt –Machine Translation) واسترجاع المعلومات (IR-Information Retrieval) واستخلاص المعلومات (IE- Information Extraction) وتصنيف النصوص (TC- Text Classification) وتلخيص النصوص (TS –Text Summarization) ، وغيرها من التطبيقات، فنظام المعالجة الصرفية يهتم بدراسة هيكل الكلمة أو بناء كلمات اللغة أسماءً وأفعالاً، ويعد مجال المعالجة الصرفية واحداً من أغنى المجالات البحثية، إذ تعد الطبيعة الصرفية والاشتقاقية مجالاً للعديد من الباحثين الذين اهتموا بدراسة نظم صرفية حاسوبية وتصميمها ، وقد تعددت الأساليب المعتمدة في بناء أنظمة الصرف العربي ومنها [1] :

- 1- آلة الحالات المحددة (FSM) Finite State Machines
- 2- قواعد الزوائد Affix grammars
- 3- شبكة الانتقالات المحددة (ATN) Augmented transition networks

يتناول موضوع البحث الحالي تصميم نظام للمعالجة الصرفية لكلمات اللغة العربية معتمداً على شبكة الانتقالات المعززة آليةً للتحليل وذلك لما توفره من إمكانية لتحليل الكلمة المتصرفة إلى سوابق ولواحق تمكن من تطبيق آليات البحث عن الاحتمالات الممكنة للكلمة المراد تصريفها ، ومن هنا جاء اقتراح الية بحث العمق أولاً كأ ستراتيجيةً لاجتياز الشبكة من اجل الحصول على التحليل الصرفي المناسب للكلمة المدخلة ، حيث يتم تحليل الكلمة إلى جذع ولواحق (سوابق ولواحق) وكذلك يتم تعيين الخصائص الدلالية (الجنس والعدد والشخص) لكل لاصق ، ويعتمد النظام الصرفي المقترح على معجم معتمد على الجذع stem-based lexicon ويبدأ هذا النظام عمله بعد فشل البحث المعجمي من ضمن أي تطبيق من تطبيقات معالجة اللغات الطبيعية.

2- صعوبات المعالجة الآلية للغة العربية والصرف العربي

إن أنظمة معالجة اللغة العربية ألياً لا تزال في مراحلها الابتدائية مقارنة بالأعمال والإنجازات التي قدمت في مجال معالجة اللغات الأخرى وأهمها اللغة الإنكليزية وبعض اللغات الأوربية وذلك من خلال أنظمة استرجاع المعلومات والترجمة الآلية وتصنيف المعلومات واستخلاصها وغيرها من التطبيقات ، فهناك بعض الجوانب التي تعيق مسيرة تقدم العمل في مجال معالجة اللغة العربية ألياً بشكل عام والصرف العربي بشكل خاص [2] :-

- إنّ اللغة العربية هي لغة عالية التصريف والاشتقاق مما يجعل عملية معالجة الصرف العربي عملية غير بسيطة .
- غياب عملية التشكيل (الحركات) من أغلب النصوص المكتوبة وهذا يؤدي إلى صعوبة حل مسألة اللبس أو الغموض الصرفي والنحوي وبالتالي أظهرت الحاجة الى قواعد تحليل صرفية وتركيبية محكمة من أجل حل مسألة الغموض التي تعيق مسألة الفهم لوجود تفسيرات مختلفة للكلمات أو القواعد [10].
- تغير شكل كثير من الحروف بتغير موقعها من الكلمة وعدم تصنيف الحروف في الكتابة إلى حروف كبيرة وحروف صغيرة كما هي الحال في اللغة الإنكليزية لتميز أسماء العلم عن غيره من الكلمات مما يضيف جهداً إضافياً في التحليل المعجمي والصرفي والقواعدي .
- تعقد عمليات الإبدال والإعلال وتداخلها .
- عدم توفر البيانات المعجمية المنهجية عن الإنتاجية الصرفية، أي تلك التي تربط بين الجذر والصيغ الصرفية المنطبقة عليه وبين الكلمات المشتقة ومعانيها الصرفية ، وتصنيفاتها النحوية الفرعية ، وأطرها الدلالية التي تحدد علاقات إسنادها.
- فضلاً عما تقدّم هنالك قلة في الوثائق العربية القياسية ذات الموضوعات المختلفة وذلك لاستخدامها مصادر للمعلومات أو لأغراض التدريب أو الفحص ، كذلك قلة المعاجم الالكترونية وبرامج التحليل القواعدي الآلية التي تعدّ مراحل أساسية في سبيل التقدم البحثي في جميع مجالات معالجة اللغة العربية آلياً .

3- الأعمال ذات العلاقة

- الصرف أو "المورفولوجي Morphologie" هو فرع اللسانيات الذي يتعامل مع البنية الداخلية للكلمات ، من حيث تكوين عناصرها الأولية والتفاعل بين هذه العناصر [الحمداوي علي، 2003] ، ويعد النظام الصرفي العمود الفقري للأنظمة معالجة اللغات الطبيعية ، فلا يوجد تطبيق يمكن ان يعد ذا فائدة إذا لم يكن مدعوماً بنظام تحليل صرفي جيد ، وهذا ما دفع العديد من الباحثين للعمل في هذا المجال الذي يعد الأكثر غزارة من حيث الجهد البحثي ، ومن تلك الجهود :
- قدم الباحثان [4] خوارزمية صرفية تعمل على إيجاد الجذور الثلاثية والصيغة الصرفية للكلمة المدخلة ، حيث اقترح الباحثان استخدام هذا النظام في أنظمة استرجاع المعلومات .
 - وقدم الباحثان [1] نظاماً صرفياً تحليلياً للكلمات العربية المتصرفة ، حيث تم استخدامات آلية شبكة الانتقالات المعززة (ATN) التي اقترحها وود [5] ، إذ اقترح الباحثان طريقة البحث

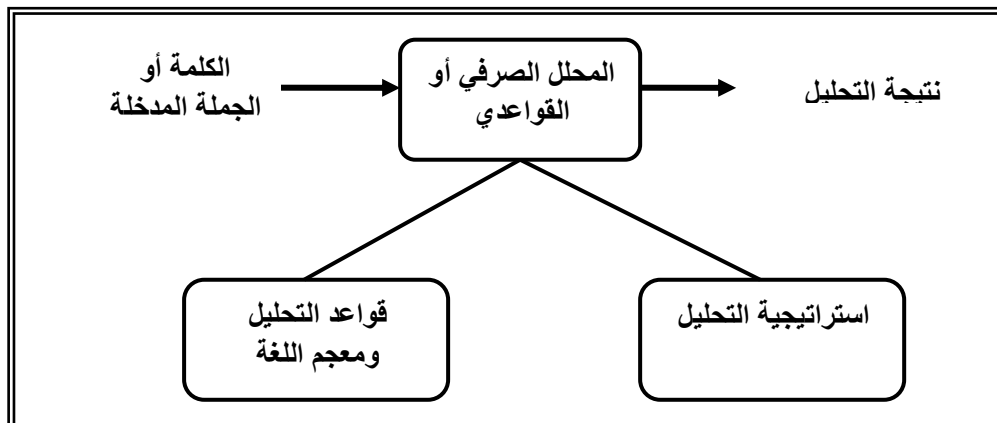
الشامل للشبكة غير المحددة من اجل اجتياز الشبكة للحصول على كل الاحتمالات الممكنة للكلمة المدخلة ، فكان لابد من وجود آلية لحل اللبس الناتج من تعدد الحلول ، وفي نظامنا المقترح اقتراحت آلية بحث العمق أولاً من اجل الحصول إلى حل واحد (التصريف الصحيح) للكلمة المدخلة وتم ذلك من خلال تصميم شبكة انتقالات محددة للمعالجة الصرفية لكل من الأسماء والأفعال في اللغة العربية .

- قدم الباحث [6] طريقة جديدة للتحليل الصرفي التي تختلف بشكل كبير عن كل الطرائق التقليدية التي تعتمد على القواعد المشتقة من قواعد التوليد والتحليل الصرفي ، بظهور ما يسمى المحلل الصرفي ذو المستويين (Two-level morphology) ، إذ استخدمت آلة الحالات محدودة الانتقالات FST في عملية التحليل الصرفي .
- قام الباحثان [7] بتطوير نظام صرفي PLPNET اعتمد على تطبيق طريقة التحليل الصرفي ثنائي المستوى على اللغة العربية.

4- نظام التحليل الصرفي المقترح

إنّ أي عملية تحليل صرفي أو قواعدي في أنظمة معالجة اللغات الطبيعية لابد أن تعتمد على نوعين أساسيين من المكونات [8] والموضحة بالشكل (1) :-

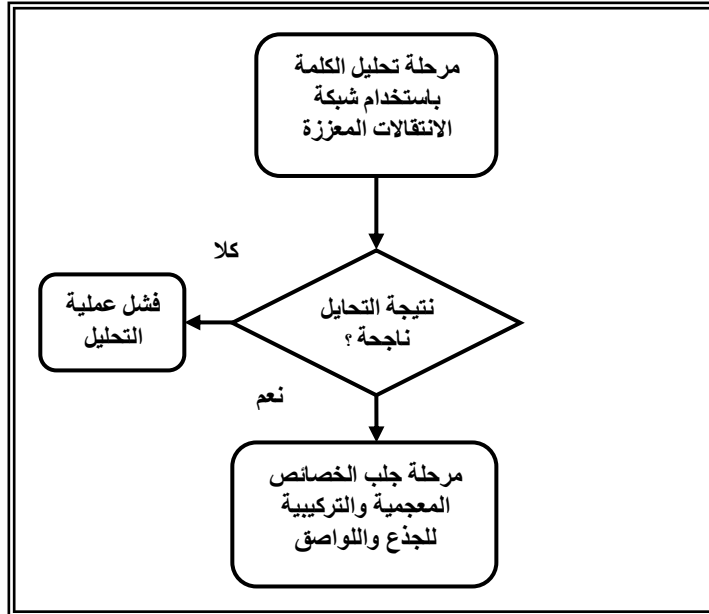
- 1- المكونات الوصفية Declarative Component : التي تقابل المعرفة اللغوية المتمثلة بالقواعد المستخدمة في الية التحليل والمعجم اللغوي للغة المراد تحليلها .
- 2- المكونات الإجرائية Procedural Component : التي تمثل آلية أو إستراتيجية التحليل المتمثلة بالخوارزميات التي تحدد بالتفصيل كل عملية في آلية التحليل.



الشكل (1) المكونات الأساسية لآلية التحليل

لذا فان المكونات الوصفية في نظامنا تتمثل أولاً في المعجم، إذ صمّم معجم معتمد على الجذع stem based lexicon ، ليضم أصناف الكلمات المكونة للغة العربية وهي (الاسم و الفعل و الحرف)، فالجذع يعد اصغر جزء ذي معنى من الكلمة ، مع مجموعة من الصفات المعجمية والتركيبية لكل صنف ، وثانياً استخدامات فكرة شبكة الانتقالات المعززة التي اقترحها وود (wood) عام 1970 ، آلية ناجحة لتحليل الكلمة المتصرفة المدخلة الى سوابق ولواحق وجذع ، حيث صممت شبكة انتقالات معززة وهو موضح في الشكل (3) لتمثيل المعرفة حساسة السياق حول العلاقة ما بين الجذع واللواحق الصرفية . أما المكونات الإجرائية فتمثل باقتراح آلية بحث العمق أولاً كطريقة اجتياز الشبكة من اجل الوصول إلى الحل الأمثل المتمثل بتحليل صرفي صحيح واحد للكلمة المتصرفة المدخلة(اسم وفعل وحرف) غير المشكولة أو مشكولة جزئياً . إن آلية عمل النظام المقترح تمر بمرحلتين وهو مبين في الشكل (2) ، ففي مرحلة التحليل التي تبدأ بعد فشل مرحلة البحث المعجمي ، تجرى عملية تحليل والكلمة المتصرفة وفصلها الى جذع وسوابق ولواحق سواء كانت اللواحق ملتصقة (مثل حرف العطف "ف" او حروف الجر "ب ، ل ، ك" و تاء التانيث "ة") او منفصلة عن الجذع مثل (واو العطف ، و "ون" الجمع) ، ولهذه العملية أهمية كبيرة في تحديد طبيعية الكلمة لمرحل التحليل اللاحقة مثل التحليل القواعدي ، إذ يمكن تمثيل نتيجة تحليل الكلمة المتصرفة بالصيغة الآتية :

Inflection word = { Pred1 + Pred2 + pred3 }+ stem + {Succ1 + Succ2 + Succ3+ Suff4}



الشكل (2) مراحل عمل النظام المقترح

وهو مبين في الصيغة السابقة كل من السوابق واللواحق تكون اختيارية ، وتختلف ما قد تعنيه قيم السوابق واللواحق للأسماء عنه للأفعال ، ولكن السابقة Pred1 تعدّ مشتركة بين كل أقسام الكلام لأنها تمثل حروف العطف وقد تم تمثيلها بـقيد تخزن فيه نتيجة التحليل . والجدول (1) يبين السوابق واللواحق التي يمكن ان تلتصق بالكلمة .

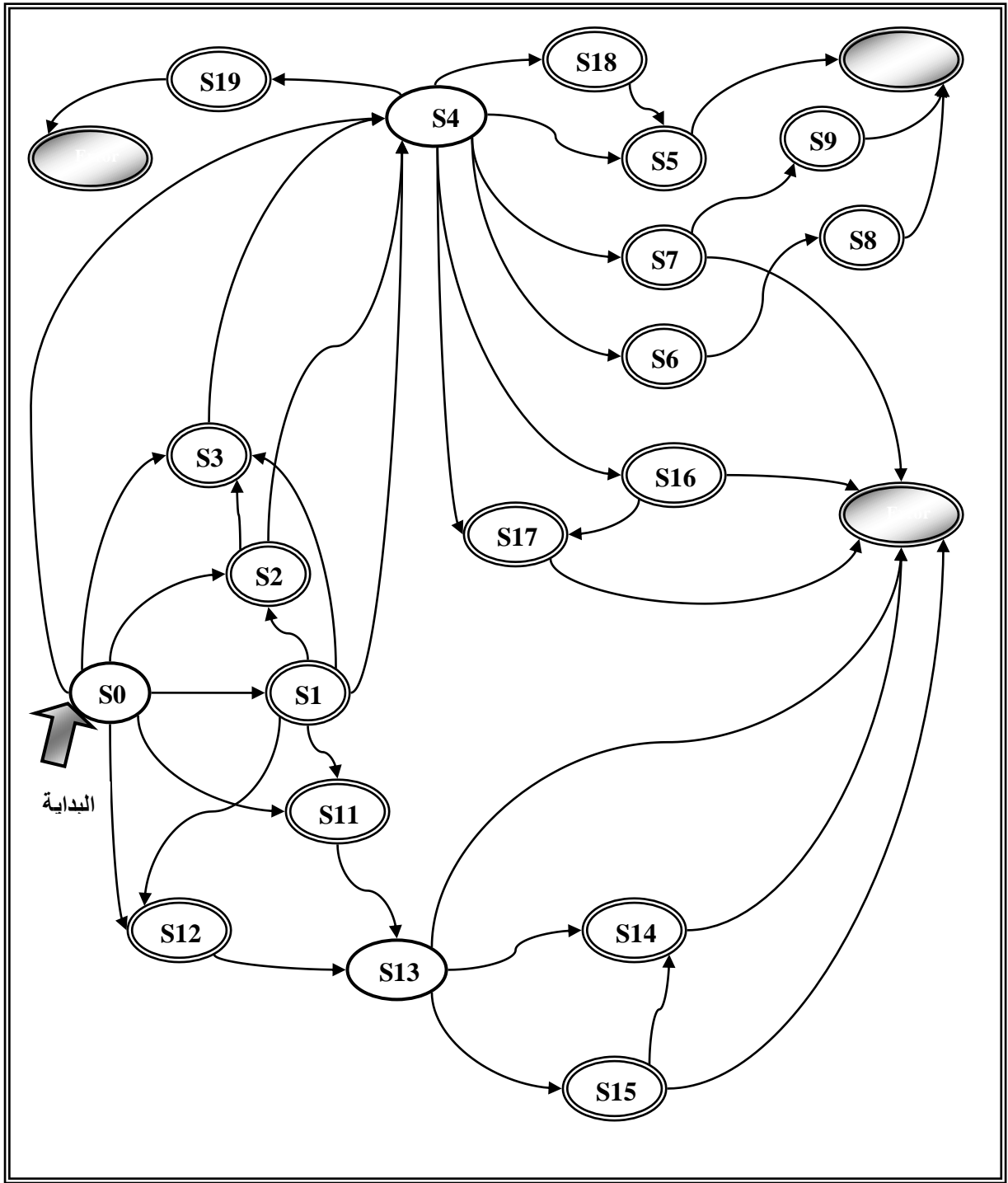
اسم الحقل	نوع الكلمة	نوع اللاصق	اللاصق
Pred1	مشترك	حروف العطف	" و " ، " ف "
Pred2	الاسم	حروف الجر المتصلة	" ل " ، " ب " ، " ك "
Pred2	الفعل	أحرف المضارع وأحرف الاستقبال	" ا " ، " ت " ، " ي " ، " ن " ، " سا " ، " ست " ، " سي " ، " سن "
Pred3	الاسم	التعريف	" ال " التعريف
Succ1	الاسم	التأنيث	" ة "
Succ1	الفعل	-	-
Succ2	الاسم	حروف المثني والجمع	" ان " ، " ون " ، " ات " ، " ين "
Succ2	الفعل	ضمائر الرفع المتصلة بالفعل الماضي والمضارع	" ا " ، " وا " ، " ت " ، " تا " ، " ن " ، " تما " ، " تم " ، " تن " ، " نا " ، " ان " ، " ون " ، " ين "
Succ3	الاسم	ضمائر المفعول به وضمائر المخاطب ونا المتكلمين و ي النسب	" ه " ، " ها " ، " هما " ، " هم " ، " هن " ، " نا " ، " ني " ، " ك " ، " كما " ، " كم " ، " كن " ، " ي " ، " أ "
Succ3	الفعل	ضمائر المفعول به والغائب وضمائر المخاطب ونا المتكلمين	" ه " ، " ها " ، " هما " ، " هم " ، " هن " ، " نا " ، " ني " ، " ك " ، " كما " ، " كم " ، " كن "
Succ4	الاسم	" ي " المتكلم والنسب وتنوين الفتح "ي"	" ي " ، " يّ " ، " آ "

جدول (1) جدول اللواحق

لقد استخدمت شبكة الانتقالات المعززة آلياً للتحليل ، إذ تعد شبكة الانتقالات المعززة واحدة من أهم الطرائق المستخدمة للتمثيل والوصف الشكلي لقواعد أنظمة فهم اللغات الطبيعية من خلال استخدامها طريقةً شكلية لكتابة قواعد الإعراب المستخدمة بكثرة في مجال الذكاء الاصطناعي (Artificial intelligence) واللغويات الحاسوبية (computational linguistics) ويمكن تلخيص حسنات استخدام شبكة الانتقالات

المعززة: [Shapiro , S ,1982] بالوضوح أو السهولة ، والقوة التوليدية ، وكفاءة في التمثيل ، والقدرة على إظهار أو تمييز العموميات والشواذ أو تمييزها ، فضلاً عن الكفاءة في العمل .

إنّ سر النجاح في استخدام قواعد شبكة الانتقالات المعززة في أنظمة معالجة اللغات الطبيعية كونها أداة تفسير وتحليل ناجحة ، إذ استخدمت في النظام المقترح أداةً تحليلية ناجحة للعلاقة بين الجذع واللواصق . تتكون شبكة الانتقالات المعززة من مجموعة من العقد Nodes أو الحالات state ، وروابط بين تلك العقد Arcs ، وهو بيّن في الشكل (3) والعقد على أنواع منها عقدة البداية والعقد الوسيطة وعقدة أو عقد النهاية وعقد حالات الخطأ التي يتم الانتقال إليها عند فشل عملية المطابقة ، في حين يمثل الرابط القاعدة rule الواجب تحقق شرطها condition للانتقال إلى العقدة اللاحقة بعد تنفيذ حدث action مرتبط بتحقق الشرط ، أي أنّ قواعد التحليل الصرفي مثلت من خلال مجموعة من قواعد الإنتاج production rules وكالاتي ، (If Condition Then Action) ، تمثل الشروط بدوال منطقية Boolean function والفعاليات بأنهجة معالجة Procedures كتبت بلغة دلفي، أما الحالات فتم تمثيلها بنهجة معالجة ، فالحالات النهائية تجري فيها عملية المطابقة المعجمية لجذع الكلمة المتصرفة مع المعجم وعندها تنتهي عملية التحليل، ليتم بعدها بدء مرحلة جلب الخصائص المعجمية والتركيبية للجذع ولجميع اللواصق المرتبطة معها ، فحالة البداية (S0) وهو مبين في الشكل (3) يرتبط بها عدد ما الروابط أو مسارات التحليل التي تمثل قواعد فصل السوابق (Pred1 , Pred2) التي تلتصق ببداية الكلمة (اسم و فعل) ، إذ إنّ احد هذه المسارات تمثل التحليل الصحيح للكلمة المدخلة ، وفي حالة تحقق أكثر من شرط (أي وجود أكثر من مسار تحليل) ، وجود حالة لبس صرفي يتم عندها فحص تلك المسارات من الأعلى إلى الأسفل.



الشكل (3) شبكة الانتقالات المعززة المقترحة

فعند فشل احد المسارات يتم العودة للخلف backtrack لتجربة المسار المتحقق الآخر وهكذا إلى أن يتم الوصول إلى المسار الصحيح ، وفي حالة العثور على الحل تتوقف عملية التحليل وتهمل كل المسارات الأخرى . أما الحالات الوسيطة مثل الحالة (S4, S13) ، فهما تتعاملان بالطريقة نفسها التي تعاملت بها حالة البداية مع تعدد المسارات الخارجة منها وحالة اللبس الصرفي ، ولكنها تختلف عنها

بوجود ارتباطات داخلية إليها وهذا يعني انه يتم الانتقال إليها من حالات سابقة لها ولكن هنا لا تتم عملية المطابقة وجلب الخصائص كما هي في الحالات النهائية ، فمثلا في الحالة S4 تتم عملية الانتقال إلى نهاية الكلمة المدخلة لبدء بفصل اللواحق (Succ1 , Succ2 , Succ3) التي من الممكن أن تلتصق بنهاية الكلمة (اسم ، فعل). ويمكن تقسيم حالات الشبكة إلى ثلاث مجاميع ، حالات لتحليل الفعل المضارع ، وحالات لمعالجة الفعل الماضي ، وأخرى للأسماء ، ولكن هذا لا يعني بأنها حالات مفصولة ولكنها تشترك في مواضع وتفترق في مواضع أخرى ، والجدول (2) بين هذا التقسيم :

S4 ، S1 ، S0	الحالات المشتركة
S15 ، S14، S13 ، S12 ، S11	حالات الفعل المضارع
S17 ، S16	حالات الفعل الماضي
S19 ، S18 ، S9، S8 ، S7 ، S6 ، S5 ، S3 ، S2	حالات الأسماء والحروف

الجدول (2) جدول الحالات المشتركة

أما الشروط والفعاليات المقابلة لها أي قوانين التحليل الصرفي المحتملة بشكل قوانين الإنتاج لربط الحالات المبينة في الشكل (3) يمكن تلخيصها في الجدول (3)

المصدر	الهدف	الشروط	الحدث	ملاحظات
S0	S1	Pred1_cond	Omit_pred1	
S0	S2	Npred2_cond	Omit_Npred2	
S0	S3	Npred3_cond	Omit_Npred2	
S0	S4	Nil_Cond	Go to word End	
S0	S11	Vpred21_cond	Omit_Vpred21	أحرف المضارع (ا ، ي ، ن ، ت)
S0	S12	Vpred22_Cond	Omit_Vpred22	أحرف الاستقبال
S1	S11	Vpred21_cond	Omit_Vpred21	
S1	S12	Vpred22_Cond	Omit_Vpred22	
S1	S3	Vpred3_Cond	Omit_Npred2	
S1	S4	Nil_Cond	Go to word End	
S2	S4	Nil_Cond	Go to word End	
S3	S4	Nil_Cond	Go to word End	
S4	S5	Nsuff1_Cond	Omit_Nsuff1	
S4	S7	Nsuff3_Cond	Omit_Nsuff3	
S4	S6	Nsuff2_Cond	Omit_NSuff2	

	Omit_VSuff3	VSuff3_Cond	S16	S4
	Omit_VSuff2	VSuff2_Cond	S17	S4
حالة انتهاء الاسم بـ " ي المتكلم أو ي النسب"	Omit_NSuff31	NSuff31_Cond	S18	S4
حالة انتهاء الاسم بـ " أ "	Omit_NSuff32	NSuff32_Cond	S19	S4
إضافة "ة" المحذوفة عند جمع التأنيث	Omit_Add_Nsuff2	Nsuff2_cond & Nsuff2=' ات '	S8	S6
	Omit_Nsuff2	Nsuff2_Cond	S9	S7
	Omit_Vsuff2	Vsuff2_Cond	S17	S16
	Omit_Nsuff1	Nsuff1_Cond	S5	S18
	Go to the End	Not Match	S13	S11
	Go to the End	Not Match	S13	S12
	Omit_VSuff3	VSuff3_Cond	S14	S13
	Omit_VSuff2	VSuff2_Cond	S15	S13
	Omit_VSuff3	VSuff3_Cond	S14	S15
	End processing	Not match	Error	S19,S9, S8,S17, S16,S14, S15

الجدول (3) شروط وقوانين الإنتاج وفعاليتها

من الجدول نلاحظ انه توجد ثلاثة أنواع من الشروط :

النوع الأول : يتحقق عندما تطابق بداية أو نهاية الكلمة مع مجموعة اللواصق المرتبطة به مثلا الشرط pred1_Cond يتحقق عندما تكون الكلمة مبتدئة بأحد حروف العطف .

النوع الثاني : Nil_Cond إن هذا الشرط يتحقق عندما تكون الكلمة غير مسبوقه بسوابق ما عدا حروف العطف .

النوع الثالث : Not match إن هذا الشرط يتحقق عند فشل عملية المطابقة ، والانتقال إلى نهاية الكلمة لفحص وجود اللواحق .

أما الفعاليات أو العمليات فتقسم أيضا إلى ثلاثة أنواع :

النوع الأول : عملية الحذف مثل Omit_pred1 حيث يتم حذف حرف العطف وخرنه في الحقل pred1 في ضمن قيد نتيجة التحليل للكلمة المتصرفه وخرن الباقي من الكلمة في حقل الجذع Stem .

النوع الثاني: عملية الانتقال إلى نهاية الكلمة (Go to The End) ، في هذه العملية لا تجرى أي تعديلات على قيد نتيجة التحليل ، ولكن يتم الإعداد لفحص مؤخرة الكلمة لمقارنتها مع اللواحق.

النوع الثالث: عملية الحذف والإضافة (Omit_Add_op) إن هذه العملية هي عملية مزدوجة إذ يتم حذف بعض اللواحق وإضافة بعض الأحرف المحذوفة عند عملية تكوين الكلمة ، مثل كلمة "سيارة" فعند الجمع تحذف "ة" ويضاف "ات" ، لذلك عند عملية التحليل الصرفي لا بد من إضافة "ة" بعد عملية حذف "ات" من نهاية الكلمة .

إما عملية الانتقال إلى الحالات الخطأ فتتم عندما تفشل كل مسارات التحليل المتحققة أو الفعالة من إيجاد التحليل الصحيح للكلمة المدخلة .

3 - الأمثلة والمناقشة

ُطبق المحلل الصرفي على كلمات مفردة وعلى نص متكون من العديد من الكلمات، وفيما يأتي عرض بعض الأمثلة لعملية التحليل للكلمات مفردة مرة وللنص مرة أخرى .

المثال الأول:الجدول (4) يبين نتيجة مرحلة التحليل للمجموعة من الكلمات

الملاحظات	Suff4	Suff3	Suff2	Suff1	Stem	Pred3	Pred2	Pred1	الكلمة
	-	هم	-	-	فجر	-	سن	-	سنفجرهم
	-	-	ان	-	فجر	-	ي	-	يفجران
	-	هم	نا	-	فجر	-	-	-	فجرناهم
	-	-	-	-	تفجير	ال	-	-	التفجير
	-	-	-	ة	ناسف	-	-	-	ناسفة
اعادة اللـ"ة" المحذوفة	-	-	ات	-	سيارة	ال	ب	-	بالسيارات
	أ	-	-	-	شخص	-	-	-	شخصاً

الجدول (4) نتائج مرحلة التحليل

والجدول (5) يبين نتائج مرحلة جلب الخصائص المعجمية والتركيبية من جداول قاعدة بيانات المعجم الساندة حيث يتم تحديد نوع الكلمة (اسم وفعل وحرف) وكذلك تحديد الخصائص المعجمية مثل (العدد ، الجنس ، الشخص) لجذع ، وكل لواحق الكلمة ، إضافة إلى تحديد جزء الكلام وزمن الفعل في حالة الأفعال.

الكلمة	جزء الكلام	عدد الجذع	جنس الجذع	شخص الجذع	عدد ضمير الفاعل	جنس ضمير الفاعل	شخص ضمير الفاعل	عدد ضمير المفعول به	جنس ضمير المفعول به	شخص ضمير المفعول به

سفنجرهم	فعل	مفرد	مذكر	غائب	-	-	-	جمع	مذكر	غائب
يفجران	فعل	مفرد	مذكر	غائب	مثنى	مذكر	غائب	-	-	-
فجرناهم	فعل	مفرد	مذكر	غائب	جمع	مذكر- مؤنث	غائب	جمع	مذكر	غائب
التفجير	اسم	مفرد	مذكر	غائب	-	-	-	-	-	-
ناسفة	اسم	مفرد	مؤنث	غائب	-	-	-	-	-	-
السيارات	اسم	جمع	مؤنث	غائب	-	-	-	-	-	-

الجدول (5) نتائج مرحلة جلب الخصائص

إنّ نتائج مرحلتي التحليل وجلب النتائج تكون ذات أهمية كبيرة في تحديد جزء الكلام وحل حالات اللبس القواعدي من خلال وجود بعض السوابق واللواحق ، فمثلا لحل اللبس بين الفعل الماضي "قتل" والاسم "قتل" يمكن أن يحل إذا سبق بحرف جر ، أو "ال" التعريف ، خلافا لذلك تترك عملية حل اللبس لمعالج التحليل القواعدي ، كما تتم الاستفادة من الخصائص المعجمية في عملية التحليل القواعدي والدلالي .

المثال الثاني : الجدول (6) يبين نتيجة مرحلة التحليل للنص " انفجرت سيارتان مفخختان في بغداد بالقرب من مدخل المنطقة الخضراء ، وقد أدى الانفجار إلى إصابة عدد من المواطنين بجروح " ، مع ملاحظة أن الكلمات التي ليس لها تحليل صرفي يعني أنها اجتازت مرحلة البحث المعجمي.

الملاحظات	Suff4	Suff2	Suff1	Stem	Pred3	Pred2	Pred1	الكلمة
	-	ت	-	انفجر	-	-	-	انفجرت
	-	ان	-	سيارة	-	-	-	سيارتان
	-	ان	-	مفخخة	-	-	-	مفختان
عدم وجد تحليل صرفي	-	-	-	في	-	-	-	في
عدم وجد تحليل صرفي	-	-	-	بغداد	-	-	-	بغداد
	-	-	-	قرب	ال	ب	-	بالقرب
عدم وجد تحليل صرفي	-	-	-	من	-	-	-	من
عدم وجد تحليل صرفي	-	-	-	مدخل	-	-	-	مدخل
	-	-	-	منطقة	ال	-	-	المنطقة
	-	-	-	خضراء	ال	-	-	الخضراء
	-	-	-	قد	-	-	و	وقد
عدم وجد تحليل صرفي	-	-	-	أدى	-	-	-	أدى
	-	-	-	انفجار	ال	-	-	الانفجار
عدم وجد تحليل صرفي	-	-	-	إلى	-	-	-	إلى
	-	-	ة	إصابة	-	-	-	إصابة
عدم وجد تحليل صرفي	-	-	-	عدد	-	-	-	عدد
عدم وجد تحليل صرفي	-	-	-	من	-	-	-	من
	-	ين	-	مواطن	ال	-	-	المواطنين
	-	-	-	جروح	-	ب	-	بجروح

الجدول (6) نتائج مرحلة التحليل للنص

وكما هو في المثال الأول يتم جلب الخصائص الدلالية والتركيبية للكلمات المحللة صرفيا

4- الاستنتاجات

1- يعد الصرف في اللغة العربية مدخلا أساسيا في فهم الأداء الشامل للغة ، هذا من الجانب اللغوي ، أما على الصعيد التقني ، فتعد معالجة الصرف العربي أليا مطلبيا أساسيا لمكننة عمليات تحليل النصوص ، المكتوبة والمنطوقة ، وفهمها وتوليدها بشكل آلي ، فضلا عن كونه جزء لا غنى عنه لمكننة المعاجم ، واسترجاع المعلومات ، وتحليل مضمون النصوص ونظم استخراج المعلومات وتعد ظاهرة اللبس الصرفي الناتجة عن الخطأ في عملية تحليل الكلمة إلى جذع ولواصق واحدة من أهم مشاكل المعالجة الآلية للصرف العربي .

- 2- استخدام شبكة الانتقالات المعززة آلياً لتحليل الكلمة العربية وفصل اللواحق عن جذوع الكلمات كان استخداماً ناجحاً لقدرة الشبكة على شمول معظم حالات اللواحق .
- 3- ساعدت آلية بحث العمق أولاً على التخلص من بعض حالات اللبس الصرفي الناتجة من تعدد المهام لنفس اللاحقة فمثلاً (حرف الـ "ب" في الكلمة "بالسيارة" هو حرف جر في حين يعد ضمن حروف الجذع في كلمة "بريد" .
- 4- تم تطبيق المحلل الصرفي المقترح على مجموعة من النصوص في مجال الإرهاب بنجاح ، وللارتقاء بعمل هذا المحلل يمكن إضافة شروط وفعاليات إضافية من أجل معالجة حالات الإبدال والإعلان .

5- المصادر

- [1] A. and Shaalan , Rafea , Lexical Analysis of Inflected Arabic words using exhaustive search of an augmented transition network ,Software Practice and Experience , Vol.23(6),pp.567-588 , John Wiley & sons ,U.K .
- [2] 1988 . مصر ، القاهرة ، دار التعريب للنشر ، اللغة العربية والحاسوب ، نبيل علي
- [3] قم ، مؤسسة انوار الهدى للطباعة والنشر والتوزيع ، شذا العرف في فن الصرف ، علي الحمداوي ، 2003 ، ايران ،
- [4] R. AL-Shalabi , and M. Evens , ,A Computational Morphology System for Arabic. Workshop on Semitic Language Processing . In the proceedings of the COLING-ACL-98 , University of Montreal ,PQ , Canada ,pp. 66-72 ,1998 .
- [5] W .Woods , Transition network grammar for natural language analysis , Comm . ACM , Vol. 10 ,pp. 591-606 , 1970 .
- [6] K. Koskenniemi , ,Two-level model for morphological analysis . In proceedings of IJCAI ,pp. 683-685 , 1983 .
- [7] K. Beesley , Arabic morphology using only finite-state operations . Proceedings of the Workshop , pp. 50-57 , 1998 .

- [8] R. Ouersighni , A major offshoot of the DIINAR-MBC Project: AraParse , A morpho-Syntactic analyzer for unvowelled Arabic texts ,Claude Bernard-Lyon University , France ,www.univ-lyon2.fr/langues/promodiinar/accueil.htm.
- [9] S. Shapiro , Generalized Augmented Transition Network Grammars for Generation From Semantic Networks , American journal of computational linguistics , Vol. 8, No. 1, pp.1-25, 1982.
- [10] "تصميم نظام استخلاص المعلومات من بعض فقرات النصوص العربية" , علياء الموزاني , جامعة البصرة , كلية العلوم , 2007 .

Abstract

This paper presents a lexical and morphological analyser for inflected Arabic words by use the deep first search method as successful technique to passing the augmented transition network(ATN) which was used to represent the relation between the morphological affix and the inflected word as a good way to analysis the inflected word , by apply the morphological analysis to the input word (name , verb , character) by separating the word into stem , affix and suffix and also determine the lexical characters (sex , number , person) to the stem of inflected word and all its suffixes .the deep first search technique give us the ability to solve the ambiguity problem in the morphological analysis by rejection all the common and not correct paths and keep on just the correct path .

Keywords : Morphological Analysis , augmented transition network ,
Natural language processing .