

# معالجة اللغة العربية آلياً باستخدام الذكاء الصناعي

\*سمر معطي، \*\*فاضل سكر

قسم الذكاء الصناعي، كلية الهندسة المعلوماتية، جامعة حلب، حلب – سورية

\*samarmouti@hotmail.com

\*\*fadelsuk@cns.bu.edu

الذكاء البشري فيما لو أنجزها الإنسان بنفسه، ويهدف إلى محاكاة بعض عمليات الإدراك والاستنتاج المنطقي [4] [5]. وتدخل محاكاة هذه المقدرة اللغوية في معظم البرمجيات التي يضعها علماء الذكاء الصناعي، كالأنظمة الخبيرة وتمثيل المعرفة والتعليم بمساعدة الحاسوب [6].

تتجلى أهمية برمجيات معالجة اللغات الطبيعية Natural Languages Processing (NLP) في تأمين التخاطب المباشر بين الإنسان والحاسوب كتابة أو كلاماً بلغة الإنسان الطبيعية، تُستعمل تقانات معالجة اللغات الطبيعية في الكثير من المجالات، مثل نظم الاستفاداة من بنوك المعلومات التي تُخزّن الموسوعات والكتب القيمة، ونظم الإجابة عن الأسئلة، ونظم فهم الكلام وتوليدته، ونظم قراءة النصوص وفهمها، ونظم تأليف النصوص، وغير ذلك [8] [7].

## II. المعالجة الآلية للغة العربية

تملك المعالجة الآلية لكل من الكلمة والجمله العربية في الحاسوب مستويات متعددة، فهناك المعالجة الصرفية والنحوية والدلالية والمعجمية والصوتية وغيرها، ويجب أن تُبنى هذه المعالجة على قواعد واضحة ومحددة، مما يستوجب الكشف عن دقائق بنية اللغة والإحاطة التامة بمفرداتها [9] [10]. للغة العربية خصائص ومميزات تميزها من غيرها من اللغات، ففيها خصيصة الاشتقاق الصرفي، والمرونة النحوية، واعتماد المعجم على الجذور، والصلة الوثيقة بين المبنى والمعنى، واطراد القياس في كثير من الحالات الصرفية والإعرابية والصوتية، وهي أيضاً توصف بأنها لغة جبرية، فجميع الأفعال الثلاثية مثلاً تنتظمها صيغة (فعل). فاللغة ليست نظاماً عشوائياً بل منظومة مُسّقة تقيدتها الضوابط وتحكمها القواعد المطردة، وفي المنظومة اللغوية يكمن كثير من التشابكات والتداخلات الدقيقة، التي تدين للتحليل وتخضع للتقنين والتقييد [11]. فمنظومة الكيان اللغوي Linguistic هي منظومة شاملة لعدة منظومات فرعية منها منظومة الصوتيات Phonology، وتتحكم بالطريقة التي تخرج بها الأصوات اللغوية وتلون درجاتها [12]، ومنظومة الصرف Morphology، وهي المسؤولة عن بنية الكلمات من حيث هيكل عناصرها والعلاقات التي تربط بين هذه العناصر [13] [14]، ومنظومة النحو Syntax وتتضمن عدة آليات تتحكم في ترتيب الكلمات

المخلص-- تمّت معالجات متطورة للغة العربية بالاعتماد على التقانات المعلوماتية المتقدمة، فقد استُخدم الذكاء الصناعي (AI) Artificial Intelligence الذي يهتم ببناء برمجيات ومكونات مادية تستطيع محاكاة السلوك البشري. تمّ الاعتماد على لغة البرمجة المنطقية Visual Prolog التي تُعدّ الأنسب والأقوى لمعالجة اللغات الطبيعية Natural Languages Processing (NLP) يكونها اللغة الأقرب إلى اللغة الإنسانية. صُممت منظومة آلية لمعالجة اللغة العربية، وذلك بعد تحليل هذه اللغة ودراسة قواعدها لبناء المحلل النحوي المسؤول عن التراكيب النحوية، والمحلل الصرفي المختصّ ببنية الكلمات والعلاقات التي تربط بينها، والمحلل الدلالي المرتبط بالمعنى، والمعرب الذي يشكل النصوص المُعالجة، والمحلل الصوتي الذي صُمم للتعرف على الحروف والكلمات المنطوقة وتحويلها إلى نصوص أو أوامر، والمركب الصوتي الذي يتحكم بالطريقة التي تخرج بها الأصوات اللغوية، ويقوم بتحويل النصوص المكتوبة إلى منطوقة.

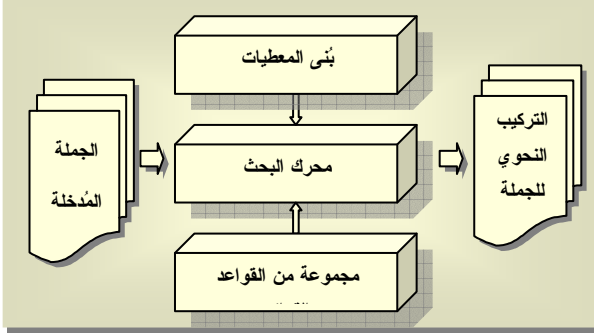
الكلمات المحورية-- الذكاء الصناعي، معالجة اللغة العربية، المحلل النحوي، المحلل الصرفي، المحلل الدلالي، المحلل الصوتي، المركب الصوتي

## I. معالجة اللغات الطبيعية

حققت المعلوماتية معالجات متقدمة للغة في تطبيق أساليب متطورة لإعطاء الآلة المهارات اللغوية من اشتقاق وتصريف واختصار وفهرسة وترجمة آليّة، وإذا لم تتفاعل الثقافة العربية مع المعلوماتية فستحدث فجوة لغوية حادة تفصل اللغة العربية عن الاستخدام والتداول [1]. تعمل تقانة المعلومات على إزالة الحواجز بين العلوم والفنون وبين المعارف والخبرات، وقد جاءت تقانة المعلومات لتزيد مفهوم التواصل بين الأفراد والجماعات والمؤسسات رسوخاً [2]. بدأ اللغويون وعلماء الألسنية بالتفكير بكيفية ترميز coding اللغة شكلاً باعتمادهم على علوم الرياضيات والمنطق الرياضي الذي وُضعت أسسه مع Boole في النصف الثاني من القرن التاسع عشر، ومن أهم الأعمال التي طوّرت وساهمت في مجال معالجة اللغة أعمال الألسني Chomsky الذي وضع أسس الكتابة الشكلية للنحو، وهو مؤسس النظرية التوليدية والتحويلية [3]. تُعدّ دراسة المقدرة اللغوية للإنسان ومحاكاتها من أهم مجالات البحث في علم الذكاء الصناعي (AI) Artificial Intelligence الذي يبحث في طرائق برمجة الحاسوب لإنجاز مهمات تحتاج إلى

### أ. تصميم المحلل النحوي

تمّ بناء محلل نحوي من أجل إيجاد التراكيب النحوية للجمل العربية، وذلك من خلال تصميم بُنى المعطيات (مجموعة الحقائق (القوالب) والأساليب والقواعد اللازمة) المناسبة، وإنشاء محرك بحث (إجراءات مُبرمجة) يقوم بعملية الاستدلال والاستنتاج للوصول إلى القالب النحوي الصحيح من مجموعة من الحلول الممكنة، وذلك اعتماداً على مجموعة من القواعد والقوانين المتعلقة باللغة العربية. تمّ إيجاد المجموعات الوظيفية (مجموعة الظروف، مجموعة حروف الجر، ...) التي تساعد محرك البحث في الوصول إلى الحل الصحيح. يُبين الشكل (1) المخطط العام للمحلل النحوي.



الشكل (1) المخطط العام للمحلل النحوي

تمّ تفسير المعرفة وتمثيلها على هيئة قواعد وحقائق، حيث تمّ تحديد التعبيرات الإعلانية للجملة الاسمية من خلال قوالبها التي تمّ تشكيلها بناءً على القواعد التوليدية والتحويلية. مثلاً يمكن إيجاد التعبير الإعلاني للقالب "مبتدأ + خبر" بالشكل التالي:

("مبتدأ", "خبر") sentn2

تُعدّ كل كلمة من كلمات الجملة بارامترراً لها.

يقوم محرك البحث بربط القواعد والحقائق المعيّنة لتكوين خطّ الاستنتاج والاستدلال، وقد تمّ بناؤه ليضمّ آليات التحكم لتمكين الفرضيات المولدة من الوصول إلى الحلّ الصحيح، ويسمح هذا المحرك بتوليد الافتراضات بناءً على المعلومات المتوفرة في قاعدة المعرفة، وقد قام محرك البحث بالعمل بطريقة بحث العمق أولاً Depth First وهي تشبه التسلسل الخلفي Backward Chaining حيث يتمّ اختيار الحال البدائية أولاً ثم الانتقال إلى الحال الثانية وهكذا حتى الوصول إلى الحال النهائية، فإذا تمّ الوصول إلى حال نهائية غير الحال المطلوبة فعندئذٍ تتمّ العودة إلى الحال السابقة مباشرة، وتكرّر هذه العملية حتى الوصول إلى الحلّ الأفضل. قام هذا المحرك بالبحث للحصول على القوالب النحوية المُمكنة لهذه الجملة، سواء أكانت اسمية أم فعلية، وذلك بعد إجراء فهرسة لأسماء التعبيرات الإعلانية، حيث يحتفظ محرك البحث بقائمة من المؤشرات لجميع القواعد والحقائق وذلك لتسهيل عملية البحث عنها، وتدلّ عملية الفهرسة هذه على أن الحقائق يمكن أن تأخذ الكثير من الترتيبات المختلفة والتي تعطي السلوكية ذاتها، ومن ثمّ يتمّ الوصول إلى الهدف المطلوب وهو القالب النحوي الصحيح للجملة، وذلك وفقاً لعدة مراحل.

داخل الجمل، مثل تلك الخاصة بالتقديم والتأخير وغير ذلك [15] [16]، والعدد المحدود من الكلمات في كل لغة هو المادة التي تمكن من خلق وتركيب عدد لا محدود من الجمل، ولكن ذلك لا يتمّ عشوائياً ومن دون ضوابط وإنما تحكمه مجموعة من القوانين تسمى القواعد النحوية [17] [18]، ومنظومة الدلالة Semantic، وتختص بالقضايا المتعلقة بالمعنى وهي تفرّض قيودها على منظومة النحو بحيث تمنع توليد جمل سليمة نحويّاً وغير مقبولة دلاليّاً [19]. إن دراسة الكلام المنطوق المسموع مقدّمة لا بدّ منها لدراسة الأنظمة (القواعد) اللغوية (اللغة نفسها) وأصبح علم الأصوات تمهيداً بالملاحظة الحسية لإنشاء علم الصوتيات الذي هو تخطيط عقليّ لقواعد الأصوات بناءً على هذه الملاحظة الحسية [20].

### III. تصميم منظومة آلية لمعالجة اللغة العربية

تمّت معالجة النصوص العربية على كافة المستويات النحوية والصرفية والدلالية والصوتية من أجل تصميم منظومة آلية لمعالجة اللغة العربية، لذلك صُممت المحللات النحوية والصرفية والدلالية والصوتية بالإضافة إلى إنشاء المركب الصوتي، كما سيُبين في المقاطع التالية.

#### 1.III. المحلل النحوي

يقوم المحلل النحوي بتقسيم الجملة وتحديد قوالبها، ويُجزّئ الكلمات فيها لإيجاد العلاقة النحوية فيما بينها، ومن ثمّ يقوم باختيار القالب النحوي المناسب للجملة العربية بعد سلسلة من عمليات الاستدلال والاستنتاج التي يقوم بها محرك البحث (مجموعة من الإجراءات) للوصول إلى الحلّ الأفضل بالاعتماد على القواعد النحوية في اللغة العربية؛ حيث أن الجملة والقالب مصطلحان مختلفان ولكنهما متّصلان اتّصلاً وثيقاً، فمثلاً "أكل الولد التفاحة" جملة، ولكن (فعل + فاعل + مفعول به) قالب. لكلّ قالب عددٌ لانتهائي من الجمل، ومن بعض القوالب الشّهيرة في اللغة العربية:

1. فعل + فاعل + مفعول به	2. فعل + فاعل + مفعول مطلق
كتبَ + الولدُ + الدرسَ	نامَ + الولدُ + نوماً
3. فعل + فاعل	4. مبتدأ + خبر
جاءَ + زيدٌ	التفاحةُ + حلوةٌ

تمّ تحديد أقسام الجملة العربية وقوالبها والمجموعات الوظيفية (مجموعة حروف الجر، مجموعة الظروف، ...)، وقد ارتكز محرك البحث على بُنى المعطيات (قوالب الجمل العربية) التي تمّ إنشاؤها وفقاً للقواعد التوليدية والتحويلية معتمداً على طريقة بحث العمق أولاً Depth First للحصول على القوالب النحوية المُمكنة للجملة البسيطة المُدخلة إلى المحلل النحوي، ومن ثمّ الوصول إلى الهدف المطلوب وهو القالب النحوي الصحيح للجملة.

### 2.III. المحلل الصرفي

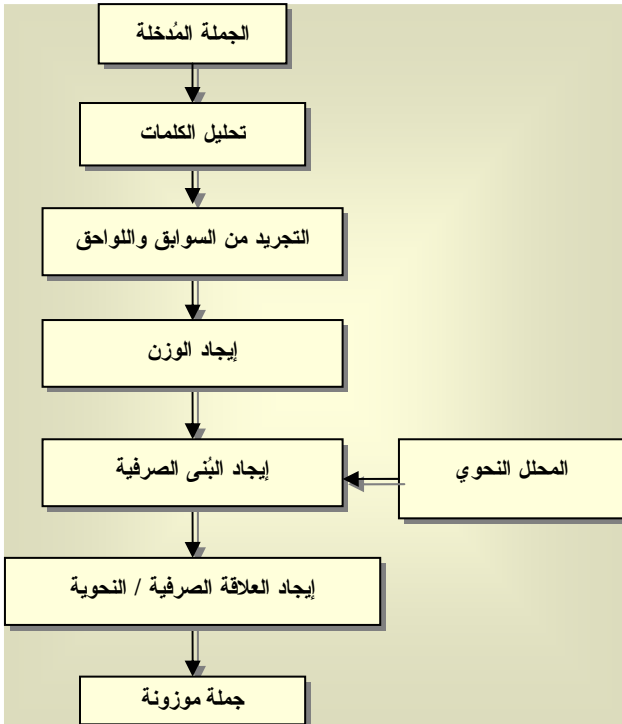
يقوم المحلل الصرفي بتحديد البنية الصرفية النحوية للجملة العربية، وذلك بعد تمثيل القواعد الصرفية لهذه الأوزان على شكل جمل منطقية. اعتمد المحلل على تجريد الكلمات من السوابق واللواحق وتحديد أوزانها المقابلة عن طريق استدعاء الهيكل البنوي الملائم لكل كلمة من بُنى المعطيات (أوزان الأسماء والأفعال)، وإيجاد العلاقة الصرفية النحوية من أجل التوصل إلى القالب النحوي الموزون للجملة المدخلة، وهو يُمثل القالب الصرفي الموزون لها.

اعتمدت خوارزمية المحلل الصرفي على تحليل الكلمة وعلى إيجاد البنى الصرفية والعلاقة الصرفية النحوية بين مركبات الجملة، وذلك بعد تمثيل القواعد الصرفية لهذه الأوزان على شكل جمل منطقية، وتم بناء الهياكل الصرفية التي قامت بُنى التحكم بفهرسة لأسماء المُعلّات المكوّنة لها. كما تمّ بناء وحدة ملاءمة وربط بين بُنى المعطيات والمحلل النحوي، وقام محرك البحث بعمليات التسلسل الخلفي المتعاقب والاستدلال، واستنتاج البنى الصرفية النحوية للجملة.

#### أ. تصميم المحلل الصرفي

يقوم المحلل الصرفي بتحليل كلمات الجملة المدخلة وتحديد بياناتها الصرفية، وذلك بعد إجراء عدة خطوات:

- تجريد الكلمة من السوابق واللواحق مثل "ال" التعريف والضمائر المتصلة وغير ذلك.
- إيجاد الوزن المُقابل لكل كلمة من كلمات الجملة، وذلك من خلال بُنى المعطيات.
- إيجاد العلاقة الصرفية النحوية بين مركبات الجملة وذلك طبقاً للقواعد النحوية التي تُحدّد الموقع الصرفي لكل وزن، ويُبيّن الشكل (2) المخطط العام للمحلل الصرفي.



### ب. الاختبار والأداء للمحلل النحوي

على سبيل المثال في حال الجملة الفعلية "استيقظ صباحاً" وبعد خضوعها لعملية التحليل النحوي، تمّ تقطيعها إلى قائمة من الكلمات: ["استيقظ", "صباحاً"], List = [ ]، تمّت معرفة عدد بارامترات هذه الجملة (عدد كلماتها): عدد الكلمات في الجملة = 2، والحلول الممكنة لإيجاد القالب النحوي لهذه الجملة تظهر بالشكل التالي:

1. فعل لازم + فاعل
  2. فعل لازم + حال
  3. فعل لازم + ظرف
  4. فعل متعدّد مع فاعل ضمير متصل + مفعول به
  5. فعل مبني للمجهول + نائب فاعل
- يقوم محرك البحث بتطبيق مجموعة من القواعد الأساسية على كل كلمة من كلمات الجملة لاختيار القالب النحوي الصحيح لها:

- الكلمة الأولى "استيقظ" تأخذ كل الحلول.
- بسبب انتماء الكلمة الثانية "صباحاً" إلى مجموعة الظروف يتم اختيار القالب 3 أي تمّ الوصول إلى حلّ وحيد وهو القالب النحوي الصحيح للجملة:

استيقظ صباحاً = فعل لازم + ظرف = استيقظ صباحاً

أما في حال الجملة "حضر الطالب"، يتمّ استدعاء نفس القوالب النحوية الممكنة السابقة (لأن هذه الجملة مكونة من كلمتين أيضاً)، يقوم محرك البحث بتطبيق مجموعة من القواعد على كل كلمة من كلمات الجملة:

- الكلمة الأولى "حضر" تأخذ كل الحلول.
- الكلمة الثانية "الطالب" بسبب عدم انتماء هذه الكلمة إلى مجموعة الظروف يتمّ استبعاد القالب 3 أي يتمّ الوصول إلى الحلّ الممكنة التالية:

1. فعل لازم + فاعل
2. فعل لازم + حال
3. فعل متعدّد مع فاعل ضمير متصل + مفعول به
4. فعل مبني للمجهول + نائب فاعل

يتمّ الوصول إلى حلّ وحيد بعد خضوع الجملة للمحلل الصرفي والدلالي. تمّ الحصول على الكثير من النتائج التي أثبتت فعالية النظام المصمّم وفائدة الاختبارات التي خضع لها، وقد تميّزت بالدقة وسرعة استجابة النظام، وطبّق هذا النظام على نماذج متعدّدة من الجمل العربية غير المشكّلة، ونَتج عن ذلك إيجاد القوالب النحوية للجملة بعد عرض مُختلف الحلول الممكنة، ومن ثمّ اختيار القالب النحوي المناسب بعد سلسلة من عمليات الاستدلال والاستنتاج التي قام بها محرك البحث للوصول إلى الحلّ الأفضل بالاعتماد على القواعد النحوية في اللغة العربية.

يُفاس مستوى أداء النظام بالقدرة على الاستنتاج والاستدلال وإعطاء النتائج الصحيحة، كما تلعب بُنى المعطيات ونوعية المعرفة التي تحتويها دوراً هاماً، وتتضمّن مجموعة الحقائق أو التّعابير الإعلانية التي تصف العلاقة المنطقية بين العناصر والمفاهيم، وطرائق حلّ المشكلات، حيث تمّ استخلاص المعرفة والخبرة من أسس اللغة العربية.



حي / +/- فاعل إنسان +/- مفعول حي، وغير ذلك، وتُعنون هذه الحقول بأسماء معينة تميز كل حقل من الحقول الدلالية. تمّ تخزين هذه الحقول على شكل تعابير إعلانية في بنى الحقول الدلالية التي يتمّ استدعاؤها إلى البرنامج عند اللجوء إلى المحلل الدلالي.

#### • الحقول الدلالية

تمّ إيجاد الروابط الدلالية بين المجموعات المصنّفة وذلك اعتماداً على مراتب الروابط النحوية التي يتمّ الحصول عليها من خلال المحلل النحوي الذي تخضع له الجملة، حيث أن لكل رابط دلالي مرتبة أي رقماً يدلّ على الرابط النحوي، فهناك رابط الفاعلية والمفعولية والجار والمجرور وغير ذلك. ويتمّ الحكم على الجملة من حيث الصحة الدلالية أو عدمها من خلال مفاهيم ومدلولات الروابط. يقوم محرك البحث بالكثير من الإجراءات من أجل الوصول إلى الحلّ المطلوب من خلال إيجاد الروابط المثلى لبُنَى الحقول الدلالية لتكوين الاستدلال والاستنتاج المطلوب، وذلك وفقاً للخطوات التالية:

#### • الفهرسة الضمنية

تمّت عملية فهرسة ضمنية للحقول الموجودة ضمن البُنَى الدلالية، فبعد إيجاد الحقول الدلالية للأسماء والأفعال، تتمّ عملية فهرسة لهذه الحقول من أجل تسهيل عملية الوصول إليها، حيث يقوم المحلل النحوي باستدعاء هذه الحقول عند فحص كلمات الجملة من أجل إزالة اللبس الناجم عن إعراب بعض الكلمات، ومن أجل الحكم على صحة الجملة دلاليّاً أو عدم صحتها.

#### • ارتباط الحقول مع المحللات

تمّ ربط الحقول الدلالية مع بعضها على أساس الاستعمال والموقع النحوي، حيث يضمّ كل حقل مجموعة من الكلمات ذات السمات المشتركة التي توّهلها لاحتلال موقع معين من الجملة، وتتم هذه العملية بناء على السمات الدلالية الانتقائية التي يتميز بها كل حقل دلالي، فمثلاً الحقل الذي يحمل معنى الأفعال العامة يتميز بالسمات التالية [ +فاعل حي إنسان، + مفعول به غير حي ] والحقل الذي يحمل معنى الفاعلية يتميز بالسمات التالية [ +جسم حي، +إنسان عاقل ] لذلك يتمّ ربط هذين الحقلين.

#### • إيجاد مرتبة الرابط الدلالي

تمّ إيجاد مرتبة الروابط الدلالية لكلمات الجملة المُدخلة، وذلك بعد إيجاد مرتبة الروابط النحوية لهذه الكلمات عن طريق المحلل النحوي، الذي يقوم بإيجاد الموقع النحوي لكل كلمة من كلمات الجملة بالاعتماد على المحلل الصرفي.

#### ب. الاختيار والأداء للمحلل الدلالي

في حال الجملة "حضر الطالب" يقوم محرك البحث بتطبيق مجموعة من القواعد الأساسية على كل كلمة من كلمات الجملة لاختيار المقبولة منها دلاليّاً.

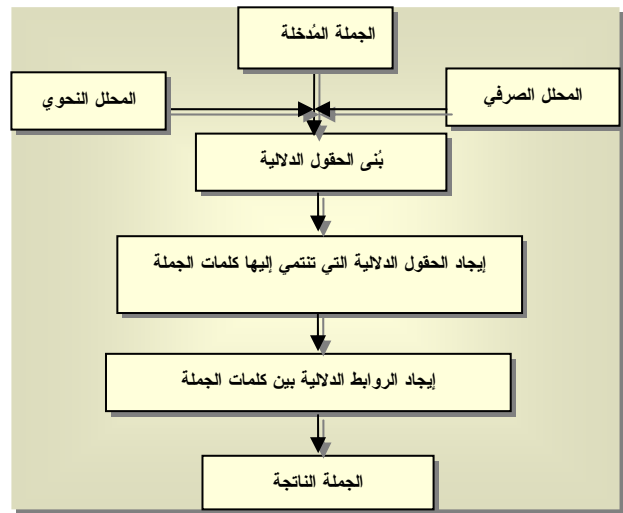
تمّ الوصول إلى الحلّ الممكنة التالية بعد خضوع الجملة للمحلّين النحوي والصرفي:

فَعَلَ + فَاعِل ← حَضَرَ الطالِبُ A  
فَعَلَ + فَاعِل ← حَضِرَ الطالِبُ B

حسب المجالات التي تتناولها الكلمة وموقعها الوظيفي في الجملة، وإيجاد الروابط والمرتبات الدلالية بينها. قام محرك البحث بفهرسة ضمنية للحقول الموجودة ضمن البُنَى الدلالية، ومن ثمّ تمّ ربط هذه الحقول مع المحللات على أساس الاستعمال والموقع النحوي، وإيجاد مرتبة الرابط الدلالي بالاعتماد على الروابط النحوية، لتحديد الجمل المقبولة نحويّاً ودلاليّاً.

#### أ. تصميم المحلل الدلالي

تمّ إيجاد الحقول الدلالية التي تنتمي إليها كل كلمة من كلمات الجملة، وإيجاد الروابط الدلالية ومراتبها بين كلمات الجملة مع مراعاة الروابط النحوية، ويبيّن الشكل (3) المخطط العام للمحلل الدلالي.



الشكل (3) المخطط العام للمحلل الدلالي

تمّ إنشاء بُنَى للحقول الدلالية الخاصة بالأسماء والأفعال، والتي تضم حقولاً من اللغة العربية. تمّ تصميم هذه البُنَى من خلال المراحل التالية:

#### • هياكل المفردات

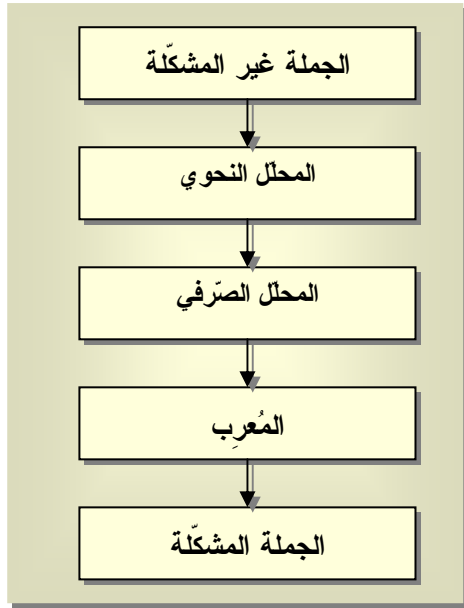
تمّ بناء قوائم بمفردات اللغة ذات السمات الدلالية الذاتية والانتقائية والتي تأخذ دوراً مهماً في تكوين الجملة وذلك بعد تركيبها بالاعتماد على القواعد الأساسية للغة العربية، وقد تكون هذه المفردات أسماء ذات سمات تقرّيبية دلالية مثل +/- جسم / +/- حي / +/- إنسان / +/- رجل / +/- عاقل / أو غير ذلك، وقد تكون أفعالاً ذات سمات دلالية ذاتية مثل: +/- حالة / +/- حركة / +/- موقف / +/- عمل، أو ذات سمات دلالية انتقائية مثل: +/- فاعل حي / +/- فاعل إنسان / +/- مفعول حي..

#### • الحقول الوظيفية

تمّ تصنيف المفردات حسب المجالات أو المفاهيم التي تتناولها وحسب الموقع الوظيفي الذي تشغله في الجملة، وذلك بالاعتماد على السمات الدلالية الانتقائية (السياقية) التي تحدّد الوضع الوظيفي للكلمة في الجملة مثل +/- فاعل

## 4.III. المُعرب

المُعرب عبارة عن مجموعة من قواعد اللغة العربية، يقوم بإعراب الجملة العربية سواء كانت فعلية أو اسمية بعد مرورها على مجموعة من القواعد الإعرابية الأساسية، حيث يقوم باختيار القالب المناسب للجملة، وذلك لتشكيلها بالعلامات المناسبة. يتم إعراب الجملة المدخلة؛ أي تشكيلها بعد مرورها على المحلل النحوي، الذي يحلل الجملة إلى كلمات، ويحدد طولها ويعمل على إيجاد قوالب هذه الجملة؛ أي الطرائق الممكنة لإعرابها. ثم تخضع إلى المحلل الصرفي، الذي يحلل الكلمة إلى حروف منفصلة، وذلك لتحديد ظهور أو تقدير الحركات على أواخر الكلمات. يبين الشكل (5) مراحل خوارزمية الإعراب، التي اعتمدت على المحللين الصرفي والنحوي وعلى المُعرب، شكّلت بُنى المعطيات من قواعد الإعراب، وتمت تجزئتها ضمن هياكل، كما تمّ تصميم وحدة ملائمة وربط بين المُعرب والمحلل الصرفي والنحوي، وتمت فهرسة وحفظ البنى من أجل استدعائها لتطبيق القاعدة المناسبة على كل كلمة من كلمات الجملة من قِبَل محرك البحث الذي اعتمد على طريقة بحث العمق أولاً، وعلى قواعد الإعراب المُخزّنة في بُنى المعطيات، وذلك بعد تطبيق القواعد الصرفية على الكلمات من أجل ظهور الحركة أو تقديرها.

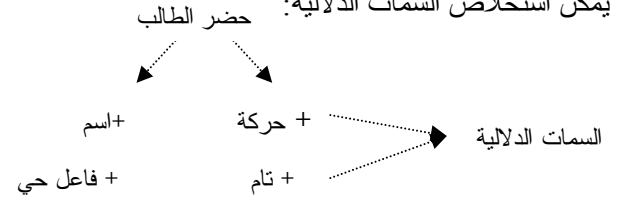


الشكل (5) خوارزمية الإعراب

أ. تصميم المُعرب

تمّ تصميم بُنى المعطيات المناسبة، وهي قواعد الإعراب المشفرة، وتُجرى عملية تجزئة لهذه البنى؛ أي يتمّ وضع البنى المشتركة ببعض الخصائص مع بعضها بعضاً ضمن هيكل واحد Slot، فتمّ وضع القواعد الخاصة بالأسماء والأفعال المرفوعة ضمن هيكل المرفوعات، والقواعد الخاصة بالأسماء والأفعال المنصوبة ضمن هيكل المنصوبات، أما هيكل المجرورات فيضمّ القواعد الخاصة

قام المحلل الدلالي بفحص كلمات هذه الجملة من حيث انتمائها إلى الحقول الدلالية مع إيجاد الروابط الدلالية مع هذه الحقول، فالكلمة الأولى تنتمي إلى الحقل الدلالي G23 (حقل الأفعال اللازمة)، أما الكلمة الثانية فهي تنتمي إلى الحقل الدلالي G21 (حقل الأسماء المتعلقة بالإنسان)، حيث يُمكن استخلاص السمات الدلالية:

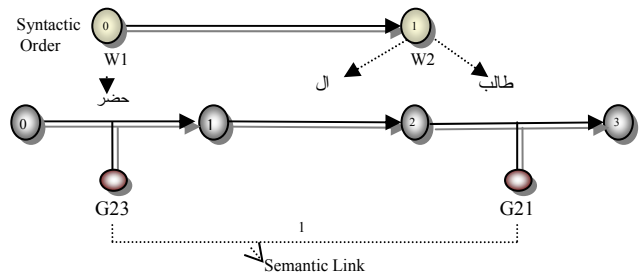


تمثل مراتب الروابط النحوية للجملة في الشكل:  
 $S = W1 + W2$   
 حيث أن  $W_i$  رمز للكلمة  $i$  في الجملة  
 $S$

الرابط بين الحقلين G21 و G23 هو رابط الفاعلية، لأن الفعل "حضر" ينتمي إلى مجموعة الأفعال الدالة على عمل حركي، وهي المجموعة التي تقتضي فاعلاً حياً، والاسم "الطالب" ينتمي إلى المجموعة التي تحمل السمات [+جسم حي، +إنسان عاقل]، ومرتببة الرابط الدلالي بين هذين الحقلين هي 1:



يُمكن إظهار المراتب النحوية والدلالية للجملة في الشكل (4).



الشكل (4) الروابط النحوية والدلالية لجملة "حضر الطالب"

يتم اختيار الجملة A التي تمثل المعنى الدلالي الصحيح للجملة وهي "حضر الطالب". تميّزت الاختيارات التي خضع لها بمعلومات دقيقة واستجابة سريعة، وتطبيقها على نماذج متعددة من الجمل تم التوصل إلى تحديد الجمل المقبولة أو غير المقبولة دلاليًا بعد عمليات استدلال واستنتاج قام بها محرك البحث بالاعتماد على بُنى الحقول الدلالية للوصول إلى الحل الأفضل.

فالفاعل الماضي يأخذ الفتحة، لأنه مبني دائماً على الفتحة إذا صُرّف مع الضمائر المنفصلة مثل "هو".

- من أجل الكلمة الثانية "فاعل" يتم استدعاء هيكل المرفوعات الذي يضم قاعدة الفاعل، وتأخذ الضمة الظاهرة، لأن الفاعل اسم مرفوع يأتي بعد فعل تام معلوم.

تكون النتيجة "حضر الطالب". أثبتت النتائج فعالية النظام المصمم، والاختبارات التي خضع لها تميّزت بالدقة وسرعة الاستجابة، إذ أمكن تطبيقها على نماذج متعددة من الجمل غير المشكّلة، والتوصّل إلى التشكيل والإعراب الصحيحين للجمل بعد عمليات استدلال واستنتاج قام بها محرك البحث للوصول إلى الحلّ الأفضل بالاعتماد على القواعد النحوية في اللغة العربية. تمّ التوصل من خلال البرمجة التي تمّ بناؤها إلى إعراب الجملة العربية سواء أكانت فعلية أم اسمية. تمّ تطبيق هذه البرمجة على مجموعة من الجمل وأعطت الإعراب الصحيح لهذه الجمل، وذلك بعد مرورها على مجموعة من القواعد الأساسية.

### 5.III. المحلل الصوتي

يقوم المحلل الصوتي اعتماداً على بُنى المعطيات الصوتية بالتعرّف على الحروف المنطوقة، وتطبيق طرائق المعالجة المختلفة للإشارة الكلامية الناتجة، ومن ثمّ التعرف على الكلمات المنفصلة، وذلك من خلال بناء البرمجيات وتصميم شبكة عصبية ذات انتشار خلفي. تمّ التوصل إلى أنه يمكن فصل الحركة عن الحرف المتحرك، وذلك بعد إجراء معالجة أولية للإشارة الصوتية من حذف للضجيج وتقطيع وترشيح، حيث تمّ أخذ بارامترات الحروف المتحركة والساكنة، والحركات الثلاث (الفتحة والضمّة والكسرة) في مستوي الزمن وفي مستوي التردد لتحليلها والتمييز بينها. بُنيت شبكة عصبية ذات انتشار خلفي، ودُرّبت على الحروف المتحركة والساكنة المُسجّلة. تمّ اختبار الشبكة على هذه الحروف وكانت نسبة التعرف 100%، وصُنِّقت الحروف الساكنة والمتحركة غير المُدْرَبَة مع الحروف المُماثلة لها والمتناسبة معها.

#### أ. تصميم المحلل الصوتي

تمت المعالجة الأولية للإشارات الصوتية بعد تسجيل الصوت عبر المايكروفون فتحوّل التغيّرات في الضغط الجوي إلى تغيّرات في التيار (الجهد) الكهربائي، ثمّ تُحوّل الإشارة التشابهيّة إلى رقميّة عن طريق النمذجة؛ وهي أخذ عيّات من الإشارة (أي قراءات الضغط الجوّي) في فترات زمنيّة متساوية. يتمّ تسجيل الحروف العربية بوساطة المايكروفون المتصل بالحاسوب، ويؤخذ تردد نمذجة [24000 Hz] في عملية التسجيل؛ هذا التردد يتبع لكثرت الصوت، وأكبر تردد يمكن نمذجته هو نصف تردد النمذجة للملف الصوتي [48000 Hz]، وعدد القنوات واحد Mono وطول القناة 16 Bits.

تمّ حذف الضجيج وتقطيع الإشارة في هذه المرحلة، حيث يوجد عدة عوامل تؤثر في تشويش الإشارة الصوتية عند تسجيل الحرف ومنها الوسط المحيط، والترددات العالية

بالأسماء المجرورة، وهيكل المجزومات يضمّ القواعد الخاصة بالأفعال المجزومة، وهيكل المبنيات يضمّ القواعد الخاصة بالأسماء والأفعال المبنية.

تُستخدّم الهيكل لتجميع القواعد الإعرابية المُشتركة، وتتمّ عملية فهرسة لأسماء الهياكل وذلك لمساعدة التسلسل الخلفي، أي يتمّ حفظ قائمة من المؤشرات لمجمل الهياكل، وذلك لتسهيل عملية البحث عنها واسترجاعها.

إن تصميم آلية ملائمة وربط تهدف إلى إيجاد وحدة ربط بين المُعرب والمحلّلين الصرفي والنحوي، حيث لا بدّ من ربط مجموعة بُنى المعطيات مع المحلل النحوي والصرفي، وذلك لاستدلال واستنتاج الحلّ الأفضل. تقوم وحدة الملائمة والربط بتبادل البيانات من المُعرب إلى المحلل النحوي وبالعكس، حيث يتمّ تطبيق القواعد الإعرابية على الكلمة وذلك بعد معرفة موقعها النحوي من المحلل النحوي، ويتمّ تطبيق آلية المحلل الصرفي على هذه الكلمة للحكم على الحركة من حيث الظهور والتقدير، وعملية التبادل هذه ثنائية الاتجاه.

يقوم محرك البحث بضمّ آليات التحدّم التي يفرضها عليه المحلل النحوي من خلال سلسلة من العمليات لتحليل الجملة ومعالجتها، وإيجاد الإعراب المُمكن لها، ثم يُطبّق مهام المحلل الصرفي الذي يحلّل الكلمة ويُعالجها من حيث السوابق والواحد وإيجاد الوزن الموافق لها، ويقوم المُعرب بربط بُنى المعطيات لتكوين خط الاستنتاج والاستدلال لتشكيل الجملة بالحركات المناسبة بعد اختيار الإعراب الصحيح لها.

يقوم محرك البحث بتوليد الآليات المناسبة اعتماداً على قواعد الإعراب المُخزّنة في قاعدة المعرفة، واعتمدت بُنى التحكم تسلسلاً متعاقباً على كل حال مُتبعة طريقة بحث العمق أولاً Depth First، حيث يتمّ اختيار الكلمة الأولى من الجملة أولاً ويتمّ إيجاد حالات الإعراب الممكن لها، ثم الانتقال إلى الكلمة الثانية وهكذا حتى الوصول إلى آخر كلمة، فإذا تمّ الوصول إلى حال نهائية غير الحال المطلوبة، فعندئذٍ تتمّ العودة إلى الحال السابقة مباشرة، وتكرّر هذه العملية حتى الوصول إلى الحلّ الصحيح، ومن ثمّ يتمّ الوصول إلى الهدف المطلوب وهو الإعراب والتشكيل الصحيحان للجملة.

#### ب. الاختبار والأداء للمعرب

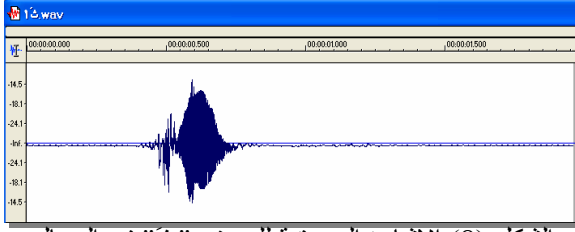
في حال الجملة الفعلية "حضر الطالب" وبعد خضوعها لعملية التحليل النحوي تمّ الحصول على الطرائق الممكنة لإعراب هذه الجملة، يقوم محرك البحث بتطبيق آلية المحلل النحوي لاختيار الإعراب الصحيح لها، وتمّ الوصول إلى حل وحيد وهو الإعراب الصحيح للجملة:

فعل لازم + فاعل

يتمّ استدعاء القواعد المناسبة لتحليل الكلمات، وذلك للحكم على الحركة من حيث الظهور والتقدير، وتبيّن أن الحركة في كلتا الكلمتين ستكون ظاهرة، ويقوم المُعرب باستدلال واستنتاج الحركات المناسبة لكل كلمة من كلمات الجملة:

- من أجل الكلمة الأولى "فعل لازم" يتمّ استدعاء هيكل المبنيات الذي يضم قاعدة الجملة الفعلية للفاعل الماضي،

ويوضّح الشكل (8) شكل الإشارة الصوتية للحرف "ث" في هذا المجال.



الشكل (8) الإشارة الصوتية للحرف "ث" في المجال الزمني

تمّ تسجيل حروف اللغة العربية المتحركة والساكنة بالإضافة إلى بعض الكلمات، ودُرِبَت الشبكة العصبية على هذه النماذج. تمّ اختبار الشبكة على هذه الحروف والكلمات، وكانت نسبة التعرف 100%، وصُفِّت الحروف الساكنة والمتحركة غير المُدربة مع الحروف المُماثلة لها والمتناسبة معها. لقد أثبتت النتائج فعالية الشبكة المُصممة، وذلك بعد تطبيق طرائق أفرزت معلومات دقيقة، وتميّزت بنسبة نجاح عالية من خلال برمجة تمّ فيها تقليل الأخطاء إلى أدنى حد مع اختزال زمن التدريب.

### 6.III. المركب الصوتي

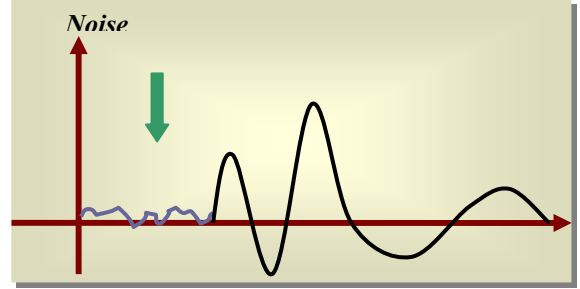
يقوم المركب الصوتي اعتماداً على البنى الصوتية المخزّنة بتركيب الحروف المتحركة من حروف ساكنة مع الحركات، وتركيب الكلمات من الحروف المنفصلة. اعتمدت خوارزمية هذا المركب على القواعد الصوتية، وتبيّنت مواضع النبر والتنغيم، وقامت بالتقسيم المقطعي مع تحديد المقاطع القصيرة، ومن ثمّ تمّ تركيب الحروف صوتياً اعتماداً على البنى الصوتية المخزّنة والتي يبلغ عددها 31 ملفاً صوتياً لتركيب الحروف المتحركة من حروف ساكنة مع الحركات، وتركيب الكلمات من الحروف المنفصلة. تمّ تركيب الحرف الساكن مع الحركة الموافقة له ومع الحركة الموافقة لحرف آخر، مع إجراء دراسة وتحليل لبارامترات الإشارة الصوتية الناتجة في المجال الزمني، ولوحظ تشابه بين الأطياف الناتجة في المجال الترددي وتمثال في الألفاظ الصوتية، وتمّ تركيب الحروف المتحركة من حروف ساكنة مع الحركات بإجراء معالجة الإشارة الصوتية للحرف المركب أو الكلمة المركبة وفقاً لمعادلات رياضية خطية أو غير خطية، أو من خلال تطبيق توابع النافذة Blackman و Gaussian و Hamming وعمليات التنعيم الملائمة.

#### أ. تصميم المركب الصوتي

يتمّ توليد النصوص المنطوقة باتباع الخطوات التالية:

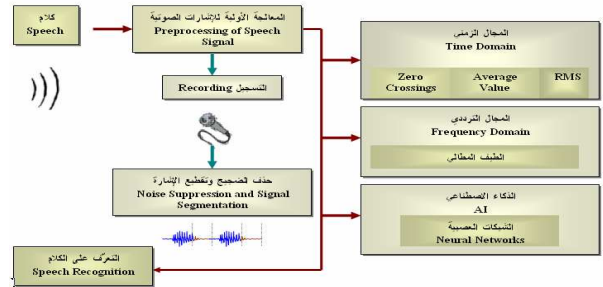
- إدخال النص المكتوب والمشكول المراد نطقه آلياً.
- تحويل النص المكتوب إلى نص صوتي باستخدام القواعد الصوتية.
- تحديد مواضع النبر ونوع أسلوب الجملة؛ أي التنعيم.

ضمن الحاسوب والميكروفون وكرت الصوت، والاختلاف الكبير لأصوات الإشارة الصوتية لكل حرف وغيرها. للحصول على إشارة صوتية صافية للحرف يتم فصل الضجيج عن الإشارة الصوتية الأساسية باستخدام ثلاثة معايير رياضية هي معدّل الانزياح ومعدّل المرور الصفري ومعدّل المطال، ويوضّح الشكل (6) الضجيج الذي يرافق الإشارة الصوتية.



الشكل (6) ضجيج الإشارة الصوتية

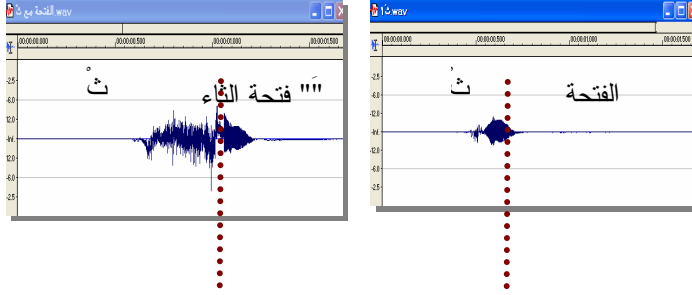
تمرّ مرحلة تنقية الكلام من الضجيج بعدة خطوات حيث يتم تسجيل ضجيج الوسط المحيط بشكل منفرد دون وجود إشارة صوتية، وحساب معدّل الانزياح ومعدّل المطال ومعدّل المرور الصفري، ومن ثمّ تسجيل جميع الحروف العربية بشكل مستقل، وحساب المعايير الرياضية السابقة لكل حرف، ويتمّ بعد ذلك تقطيع الحرف إلى إطارات ذات حجم ثابت، وحساب معدلي المطال والمرور الصفري لكل إطار على حده، وأخيراً يجب مقارنة القيم الناتجة في حساب هذين المعدلين للضجيج مع معدلي المطال والمرور الصفري لكل الإطارات، فإذا كانت القيم مُتقاربة فلا وجود لأي كلام في الإطار ويتم حذفه، والحفاظ على الإطارات ذات المعدلات المطالية والمرورات الصفرية الأكبر من معدلي المطال والمرور الصفري للضجيج، يُبيّن الشكل (7) كيفية معالجة الإشارة الصوتية للتعرف عليها.



الشكل (7) معالجة الإشارة الصوتية

يمكن ملاحظة بارامترات المجال الزمني مباشرة من خلال البرمجيات الصوتية Cool أو Sound Forge 7.0 أو Edit، أو من البيئات البرمجية Matlab أو من خلال برنامج خاص صُمم لهذا الهدف (Visual Prolog)، والبارامترات هي: Zero Crossings و Average Value و Maximum Sample Value و RMS و Minimum Sample Value. من الضروري ملاحظة أن هذه البارامترات تتغير حسب الضجيج والميكروفون والحاسوب وتُعد المتكلم عن الميكروفون أثناء تسجيل الحروف والحال الصحية والنفسية للإنسان وغير ذلك،

الصوتية المخزنة 31 ملفاً، وتمت معالجة الإشارة الصوتية للحروف العربية والكلمات في المجال الزمني. تم تركيب الحروف الساكنة مع الحركات الثلاث Short Vowels الموافقة لها، وتمت المقارنة مع الحرف المتحرك، كمثال على ذلك الشكل (10) يُوضّح الإشارة الصوتية للحرف المركب من إشارة فتحة التاء "ت" مع الحرف الساكن "ث"، والإشارة الصوتية للحرف المفتوح "ت".



الإشارة الصوتية للحرف المركب "ت" مع فتحة التاء "ت"

الإشارة الصوتية للحرف المفتوح "ت"

الشكل (10) تركيب الفتحة من "ت" مع الحرف الساكن "ث"

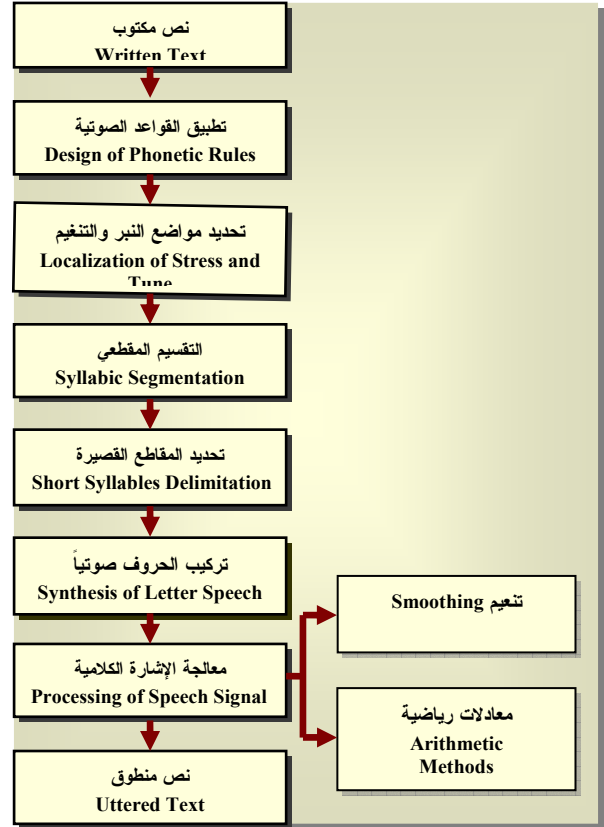
بيّنت النتائج دقة تركيب الحروف المتحركة من حروف ساكنة مع الحركات، وتركيب الكلمات من الحروف المنفصلة، وذلك بعد تطبيق طرائق أثبتت فعاليتها في عملية تقريب الإشارة الصوتية المركبة الناتجة من الإشارة الصوتية المتصلة، وتميّرت بنسبة نجاح عالية، وبأقل خطأ ممكن وكفاءة متميزة، كما أنّ الاختبارات التي طبقت في هذا المجال كانت ناجحة وأدت إلى نتائج جيّدة.

#### الخاتمة

من خلال الدراسة الشاملة لمعالجة التّصوُّص العربية، وبناء المحلّلات التّحوية والصّرفيّة والدّالّية والصّوتية، وإنشاء المركّب الصّوتي الذي يقوم بتحويل التّصوُّص المكتوبة إلى منطوقة، تُبيّن النتائج مدى فعالية الأنظمة المصمّمة وسرعة تنفيذها وإمكانية تطبيقها على نماذج عدّة من الجُمْل العربيّة غير المُشكّلة للوصول إلى التّشكيل والإعراب الصّحيحين لها، من خلال الدّور المهم الذي يقوم به محرّك البحث في عمليّة الاستدلال والاستنتاج اعتماداً على القواعد التّحوية النّاطمة للغة العربيّة.

يُقترح الاستمرار في ترقية وتطوير التقانات في مجال معالجة اللغة العربية بكافة أشكالها؛ مكتوبة أو مطبوعة أو مسموعة أو مقروءة، واستخدامها في تطبيقات الذكاء الصناعي، ووضع الخطط والبرامج اللازمة للتكامل بين البحوث المعلوماتية واللغوية.

- عملية التقسيم المقطعي، التي تتم بناءً على مجموعة من القواعد لتحديد البنية المقطعية.
- تحديد المقاطع القصيرة، وإيجاد مكافئاتها؛ أي مُقابلة كل حرف متحرك بحرف ساكن مع الحركة الموافقة له.
- تركيب الحروف صوتياً، ويتم ذلك بناءً على قاعدة بيانات صوتية، ومن ثم تركيب الحروف لتوليد الكلمات، وتركيب الكلمات لتوليد الجمل، وتركيب الجمل لتوليد النص.
- معالجة الإشارة الكلامية الناتجة، وذلك بتطبيق عمليات تعميم واستخدام النوافذ مثل Hamming و Gaussian و Blackman أو من خلال إيجاد معادلات رياضية خطية أو غير خطية.
- توليد النص المنطوق. تم توضيح خوارزمية المركب الصوتي اللغوي في الشكل (9).



الشكل (9) خوارزمية المركب الصوتي

تم تصميم بُنى المعطيات الصوتية، حيث تم فصل حركة الفتحة عن حرف متحرك مفتوح، وحركة الضمة عن حرف متحرك مضموم، وحركة الكسرة عن حرف متحرك مكسور، وتم التوصل إلى أن ألف المد عبارة عن فتحة طويلة، وواو المد عبارة عن ضمة طويلة، وياء المد عبارة عن كسرة طويلة، لذا تم اختصار حجم بُنى المعطيات الصوتية، حيث تم تخزين الإشارات الصوتية للحروف الساكنة Quiescent Letters والحركات الثلاث Short Vowels الفتحة والضمة والكسرة فقط، وكان عدد الملفات

- [13] عتيق عبد العزيز، 1985، علم المعاني، دار النهضة العربية، بيروت.
- [14] حسان تمام، 1973، اللغة العربية (معناها ومبناها)، الهيئة المصرية العامة للكتاب
- [15] Kasravi Kas, 1997, Introduction to Artificial Intelligence, Electronic Data Systems, Troy, Michigan USA.
- [16] Grenager Teg, 2003, A Quick Romp Through Probabilistic Relational Models, NLP Lunch.
- [17] Wiley John and other, 2002, Thought and Language A G E, Inc.
- [18] Manning Christopher, 2000, Information pragmatics, A Natural Language Processing Approach, CSLI IAP meeting.
- [19] Wu Jun, 2001, Maximum Entropy Language Modeling with Syntactic, Semantic and Collocational Dependencies, Center for Language and Speech Processing, Department of Computer Science, Johns Hopkins University.
- [20] Ghenima.M., 1998, Un Système de Voyellation de Textes Arabes. Université Lumière Lyon2, France.

### المراجع

- [1] علي نبيل، 2001، الثقافة العربية وعصر المعلومات/ رؤية لمستقبل الخطاب الثقافي العربي، المجلس الوطني للثقافة والفنون والآداب، سلسلة عالم المعرفة، الكويت.
- [2] علي نبيل، 1994، العرب وعصر المعلومات، سلسلة عالم المعرفة.
- [3] قباوة فخر الدين، 1997، تحليل النص اللغوي، دار الفكر المعاصر، بيروت ودار الفكر، دمشق.
- [4] طلبه محمد فهمي وآخرون، 1994، الحاسب والنكاه الصناعي، مجموعة كُتب دلنا.
- [5] وافي علي عبد الواحد، 1972، علم اللغة، دار النهضة مصر.
- [6] الأنطاكي محمد، المحيط في أصوات العربية ونحوها وصرها، دار الشرق العربي، بيروت.
- [7] الدحداح أنطوان، 1996، معجم تصريف الأفعال العربية، مكتبة لبنان، ناشرون.
- [8] الراجحي عبده، 1974، التطبيق الصرفي، دار النهضة العربية للطباعة والنشر، بيروت.
- [9] الأفغاني سعيد، 1957، في أصول النحو، مطبعة الجامعة السورية.
- [10] الصابوني عبد الوهاب، اللباب في النحو، دار مكتبة الشرق، لبنان.
- [11] جطل مصطفى، 1979، نظام الجملة، مديرية الكتب والمطبوعات الجامعية، حلب.
- [12] هارون عبد السلام، 1959، الأساليب الإنشائية في النحو العربي، مكتبة المتنبي، بغداد ومؤسسة الخانجي، مصر.