معالجة اللغة العربية باستخدام تقانات الذكاء الاصطناعي

1. توطئة

تُوفّر المعلوماتية فرصاً هامّةً لخدمة الثقافة العربية، حيث تَعاظمَ هذا الدّور إقليمياً وعالمياً وتنموياً وسياسياً، وأهم أدوات الثقافة هي اللّغة، وقد حَقّقت المعلوماتية معالجاتٍ متقدّمةً للغة في تطبيق أساليب متطورة لإعطاء الآلة المهاراتِ اللّغويةَ من اشتقاقٍ وتصريفٍ واختصارٍ وفهرسةٍ وترجمةٍ آليَّة، وإذا لم تتفاعل الثقافة العربية مع المعلوماتية فستَحدُث فجوة لغوية حادة تفصل اللّغةَ العربية عن الاستخدام والتّداول [34]. لذلك لابدّ من دراسة اللغات ومعالجتها بشكل تطبيقي لخلق برامج وأنظمةٍ معلوماتية "ذكية" يكون لها دورُها في خدمة المجتمع بشكل عام، والمساعدةِ على تحسين وتطوير اللّغات للمعوقين بصرياً، وسمعياً ونُطقاً بشكل خاص.

اللغة هي وسيلة من وسائل الاتصال والتخاطب البشريين، وهي المنطلق الأساسي للتواصل مع الثقافات والحضارات الأخرى. وتحوي اللغة العربية العديد من المقومات والأسس، وتتميز بالدقة الصارمة التي تجعلها قادرة على تبوء مكانها الصحيح في هذا العصر التقني الجديد الذي سادت فيه اللغات الطبيعية وأضحت المعلومات بجميع فروعها تعالج وتخزّن وتحلّل لغوياً. ويشهد العالم اليوم انفجاراً معرفياً هائلاً وثورة علمية تكنولوجية نتجت عن طوفان عارم لثورة المعلومات والاتصالات أسهم الانسان في صنعه، مما أدى إلى بزوغ فجر جديد من التطور والنتاج المعرفي المتسارع والذي يهدف بشكل أساسي إلى حل جميع المشكلات التي تتعرّض لها الانسانية اليوم واستقراء آفاق المستقبل، ولا شك أن المعلوماتية هي قارب النجاة الذي علينا أن نصل من خلاله إلى صنع الانسان والاسهام بشكل فاعل في صياغة واقع تقني متميز للوطن العربي. بما أن الكمبيوتر هو الأداة الطليعية في هذا العصر بل هو العنصر الرئيس الذي لم نعد نستطيع الاستغناء عنه في جميع مجالات الحياة الاجتماعية والاقتصادية والعلمية، فهذا مؤشر هام على مدى انتشاره كمّاً ونوعاً بسبب وظائفه وتعدد تطبيقاته، لذلك لابدّ من استيعاب هذه التقانة وتعريبها والاهتمام بإحياء اللغة العربية الأم ومحاربة ما أصابها من زحف العاميات المحلية وتآكل مفرداتها الفصحى والركاكة والإسفاف في اختيار تعابيرها، وذلك عن طريق معالجتها آلياً بواسطة الكمبيوتر للاعتماد عليها في مختلف الدراسات والأبحاث. ومن أجل الاستخدام الأمثل للتطورات التقنية اعتماداً على الكمبيوتر، فان الاهتمام باتباع اساليب الذكاء الاصطناعي واستخدام كافة فروعه له التأثير المباشر في حل العديد من الصعوبات والعقبات

التي يمكن أن تعترضنا في التعامل مع اللغة العربية ومحاكاتها، وذلك بتصميم برامج ومنظومات قادرة على معالجة هذه اللغة.

تعمل تقانة المعلومات على إزالة الحواجز بين العلوم والفنون وبين المعارف والخبرات، فالمجتمع لا تشكله الإدارة أو الاقتصاد بقدر ما يُشكله التواصل السائد بين الأفراد والجماعات والمؤسسات، وقد جاءت تقانة المعلومات لتزيد هذا المفهوم رسوخاً، والفرصة مهيأة أكثر من أي فترة ماضية لإقامة تقانة إنسانية واجتماعية تعيد إلى المجتمع الإنساني توازنه وعقلانيته [33]، وينعكس تطور تقانة المعلومات بشكل إيجابي على التراث العربي، ولكن تبقى استفادة العرب منها ومشاركتهم فيها محدودة بسبب عدة عوامل اقتصادية واجتماعية وتنموية، وقد أشار بعضهم إلى حدوث ما يسمى الفجوة الرقمية بين الدول العربية والدول الأخرى المتقدمة [22]. 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 طُوّرت وساهمت في مجال معالجة اللغة أعمال الألسني Chomsky الذي وضع أسس الكتابة الشكلية للنحو، وهو مؤسّس النظرية التوليدية والتحويلية التي اعتمد فيها على مفاهيم أساسية وأهمّها الكفاية اللغوية والأداء الكلامي، والمقدرة الفطرية عند الإنسان لاكتساب اللغة، وتنوّع اللغات وتحديد قواعدها الكلّية، واعتماد مستويين لدراسة جمل اللغة بالتمييز بين البنية السطحية (تَتابُع الكلمات المنطوقة) والبنية العمقية (القواعد) [38]، مع كل ما يحتويه من مفردات وما يحيط بها (صور، فهارس، إلخ...)، وكل ما يتعلق بشكل الكلمات وعلامات الترقيم ومكان الجمل في النص. كما تمّ تحديد أهداف تعليم اللغة الأم والتطرّق إلى الواقع اللغوي للعربية، وعوامل الضعف اللغوي، وسبُل الإصلاح وتكوين المهارات اللغوية من خلال المناخ اللغوي الفصيح والقدوة الصالحة وتوجيه المصادر المقروءة، وفي أوائل الأشياء وأواخرها وفي الطول والقصر وفي القلة والكثرة وغيرها [8] [39].

على الرغم من كون الكمبيوتر حصيلةَ قرونٍ من التطوّر المعرفي والتقاني الذي شاركت فيه الإنسانية جمعاء، إلا أن مبتكريه لم يأخذوا بالاعتبار أثناء تصميمه من ناحية العتاد والبرمجيات، سوى حاجات أسواقهم وثقافتهم ولغتهم، ولم تكن لديهم رؤية استراتيجية لتطويره وانتشاره على مستوى العالم. لكن ما إن حدث ذلك الانتشار، ودخل استخدام الكمبيوتر في المجالات كافة، حتى أصبحت الحاجة ماسّةً لمواءمته مع اللغات الأخرى، ومنها اللغة العربية، وظهرت تحدّيات إمكانية تنفيذ ذلك على أنظمة لم تُصمَّم أصلاً للغات مختلفة، إلا أن بنيةَ الكمبيوتر الشخصي المفتوحة، وتطوُّرَه السريع، سهَّلا نسبياً إضافةَ برمجياتٍ وتقاناتٍ ملحقةٍ خاصةٍ بثقافاتٍ ولغاتٍ

أخرى، كانت اللغة العربية واحدةً منها. لكن الأنظمة التي وُجِدَت لمعالجة اللغة العربية لم تُستخدم على نطاق واسع، لأنها لا تُوفّر حلولاً مناسبةً في المجال العملي، ولا تستطيع تقديمَ ثقافةِ العالم للمستخدمين العرب، ونقلَ اللغة والثقافة العربية إلى العالم. وأهمّ ما يواجه مطوِّري تلك التقانات من صعوباتٍ؛ طبيعةُ اللغة العربية الثريّة والمرتفعة المستوى، وقلّةُ الأبحاث الأكاديمية المتعلّقة بها، ولذلك كان لا بدّ من تطوير هذه التقانات، وإجراءِ أبحاثٍ في علوم اللغويات العربية، وإيجادِ أنظمةٍ ذكيّة متطورة مُعالِجةٍ للّغة العربية، تؤدي إلى تحويل النصّ المكتوبِ إلى نصٍّ منطوق وبالعكس.

تُعدّ دراسة المقدرة اللغوية للإنسان ومحاكاتها من أهم مجالات البحث في علم الذكاء الاصطناعي، فاللغة هي وسيلة للتفكير والتعبير عن الأفكار والمعلومات ونقل الخبرات إلى الآخرين، والتواصل والتعليم والتعلّم وقضاء حاجات الإنسان الاجتماعية. وتدخل محاكاة هذه المقدرة اللغوية في معظم البرمجيات التي يضعها علماء الذكاء الاصطناعي، كالأنظمة الخبيرة وتمثيل المعرفة والتعليم بمساعدة الكمبيوتر [54]. وبما أنّ الجزء الأكبر من المعرفة الإنسانية موجود في نصوص لغوية، فلا بدّ للكمبيوتر من طريقة تمكّنه من التعامل مع هذه النصوص لاستخلاص هذه المعرفة.

2. الذكاء الإنساني

يتميز الذكاء الإنساني باكتساب المعلومات والقدرة على التعلّم والفهم من خلال الممارسة الفعلية والتطبيق العلمي والخبرة المُكتسبة، ويؤدي ذلك إلى التمييز الدقيق بين القضايا والتوصُّل من العموميات إلى الجزئيات واستبعاد المعلومات غير المناسبة. يستجيب هذا الذكاء لمختلف المواقف بمرونة تامة وسرعة مع عدم الانحياز الخاطئ، ويمتلك القدرة على اتخاذ القرارات الصحيحة بناء على الإدراك الحسي والعقلي لجوانب المشكلة والاحتمالات الواردة ونتائجها. فالذكاء الإنساني له القدرة على استنباط القوانين العامة من الأمثلة المحددة، وعلى اكتساب المعرفة واستخدامها في حل المشاكل والقضايا الصعبة، ونقل التجربة والخبرة الذاتية إلى مواقف جديدة، والقدرة على اكتشاف الأخطاء وتصحيحها وصولاً إلى تحسين الأداء في المستقبل، ويقوم على فهم وتحليل المواقف الغامضة وغير التقليدية باستخدام أسلوب الاستنتاج المنطقي.

تَبيَّن أنه ليس من السهل محاكاة هذه القدرات والخبرات على الكمبيوتر، ومن هنا ظهرت الحاجة إلى علم الذكاء الاصطناعي. وتكمن صعوبة محاكاة الذكاء البشري في عدم القدرة على وضع تصور دقيق لكيفية قيام الإنسان بالأفعال الذكية، مثل نقل المعلومات إلى المخ أو عملية استرجاع

المعلومات أو ربط الأحداث، ومن هنا نشأ علم الإدراك أو علم المعرفة وهو أحد علوم الـذكاء الاصطناعي.

3. الذكاء الاصطناعي

يبحث الذكاء الاصطناعي Artificial Intelligence (AI) في طرائق برمجة الكمبيوتر لإنجاز مهمات تحتاج إلى الذكاء البشري فيما لو أنجزها الإنسان بنفسه [56]، ويُصنَّف هذا العلم بأنـه أحد العلوم التي تهتم ببناء برمجيات ومكونات مادية تستطيع محاكاة السلوك البشري، ويهـدف إلى محاكاة بعض عمليات الإدراك والاستنتاج المنطقي التي يجيـدها الإنـسان [30]. عـرّف Gheorghe Tecuci الذكاء الاصطناعي AI أنه العلم المتعلق بالناحيتين العمليـة والنظريـة للأنظمة المتطورة التي تقوم بعرض المميزات المشتركة مع السلوك الذكي للإنسان مثل الإدراك Perception ومعالجـة اللغـات الطبيعيـة Natural Language Processing والاستنتاج Reasoning والتخطيط Planning والتعليم Learning وحل المـشكلة Problem Solving [63].

لقد ظهر الذكاء الاصطناعي إلى الوجود نتيجة للثورة التي حدثت فـي نظريتـي المعلومـات والتحكم الآلي، فقد بدأت أبحاثه بالظهور في الخمسينيات من القرن الحالي، وترمي هذه الأبحاث إلى تحقيق هدفين رئيسيين: الأول هو الوصول إلى فهم أفضل وأعمق للذكاء الإنـساني؛ عـن طريق محاولة محاكاته، أما الهدف الثاني فهو الاستثمار الأفضل للكمبيوتر واسـتغلال مجمـل إمكاناته [63]. يستخدم علماء الذكاء الاصطناعي تقانات عدة للارتقاء بمستوى ذكاء الكمبيـوتر منها النمذجة Modeling والتقليد أو المحاكاة Simulation للحصول علـى نمـاذج للـذكاء البشري. من أهداف الذكاء الاصطناعي فهم المبادئ التي تجعل الذكاء ممكناً وتطوير الأنظمـة الذكية، وتكوين المعرفة وآلية الاستدلال، وجعل التعامل مع الكمبيوتر سهلاً كالتعامل مع البـشر [61]. إن السلوك الذكي للإنسان لا ينتج عن طرائق الاستنتاج التي يتبعها بقدر ما ينـتج عـن المعرفة أو الخبرة المختزنة في ذاكرته، والتي يكتسبها طوال حياته ليـستعملها وقـت الحاجـة بالشّكل الأمثل، لذلك فقد ابتُكر الكثير من الطرائق والتقانات لتمثيل المعرفة وفق نماذج يفهمهـا الكمبيوتر، ليتسنى له الاستفادة منها واستخدامها بشكل يُحاكي الذكاء البشري قدر الإمكان.

فالحدس Common Sense هو دروس وخبرات يتعلمها الإنسان في حياته اليومية، وتُعد هذه الدروس والخبرات أساساً لنوع من الاستنتاج يُدعى بالاستنتاج الحدسي، أما المنطق فهو يتناول طرائق محاكاة الكمبيوتر للإنسان في استنتاجه لشيء معين بالاستناد إلى مجموعـة حقـائق أو مقدمات، ويُستعمَل المنطق في الكثير من تطبيقات الذكاء الاصـطناعي، فقـد ازداد الاهتمـام

العالمي بتطوير البنى الذكية Intelligent Structures [51]، وأُجري الكثير من الأبحاث التي استخدمت تطبيقات الذكاء الاصطناعي، كتعليم الآلة Machine Learning من أجل بناء نظام Meta-Reasoning لاختيار الخوارزمية الآلية Automatic Algorithm [55]، كما تمّ تصميم آلات ذكية لتحليل نشاطات وحركات الإنسان والتعرّف عليها [47].

ذكرت Martha Palmer أن بناء آلات ذكية يخضع إلى الخطوات التالية: إيجاد تمثيل للمسألة كي تتعامل مع الأعداد أو النصوص أو المنطق، وتتمّ المعرفة الكاملة للمشكلة، والتوسع في البرامج لمعالجة الحالات التي تكون فيها المعرفة ناقصة [60]. ويَستَخدم الذكاء الاصطناعي المعالجة الرمزية Symbolic Processing من خلال المعرفة والأسلوبية، كما تعتمد بعض برامج الذكاء الاصطناعي على ما يُسمى بالقواعد التجريبية Rules of Thumb لتوجيه البحث عن الحل الأمثل للمشكلة إلى الهدف الصحيح، والوصول إليه بأقل وقت وأقصر طريق ممكن [62]. يتميز الذكاء الاصطناعي بأنه قادرعلى التعامل مع الحالات المعقدة والمحيّرة، وعلى التعلم أو الفهم من خلال الخبرة، واكتساب وتطبيق المعرفة والإبداع والخيال، لذلك أمكن تطبيقه في الكثير من المجالات منها علم الادراك Cognitive Science والانسان الآلي Robotics وعلم الكمبيوتر Computer Science. كذلك استُخدمت تقانات الذكاء الاصطناعي في التخطيط Planning عبر اختيار تسلسُل العمليات اللازمة للوصول إلى الهدف المطلوب في الكثير من المجالات، مثل أنظمة استنباط مخطط حركة الروبوت اللازم لجعلها تقوم بمهمة نقل أجسام معينة من مكان إلى آخر. وطُبّقت هذه التقانات أيضاً في البرمجة العصبية وNeural Computing ومعالجة اللغات الطبيعية Natural Languages Processing؛ وتعني المقدرة على محاكاة الكمبيوتر باللغة الطبيعية، وتشمل التعرّف على الكلام Speech Recognition؛ أي التعرّف على الكلمات المنطوقة، وفهم الكلام Speech Understanding؛ أي فهم معنى الكلام وتركيب الكلام Speech Synthesis وهذا يعني إنتاج الكمبيوتر لصوت شبيه بالصوت البشري.

1.3. لغات الذكاء الاصطناعي

تُعدّ لغة الـ Visual Prolog 6.1 الأنسب والأقوى لمعالجة اللغات الطبيعية؛ وهي لغة البرمجة المنطقية غرضية التوَّجه، المُكوَّنة من عدة عناصر والتي تملك بيئة تفاعلية متطورة ، وقد صُمِّمت هذه البيئة لتؤمِّن تعاملاً سهلاً وسريعاً ومتطوراً، وتسمح هذه اللغة للمبرمج بتمثيل العلاقات بين الأشياء وتجميعها وتنظيمها، حتى يتمكَّن من الوصول إلى استنتاج منطقي من الحقائق التي تُمثِّلها تلك العلاقات. تُنجز الجمل المبنية على القواعد Rule Based Systems

في هذه اللّغة بشكل مختلف عن اللّغات التقليدية، حيث تُعدُّ بنية التحكم Control Structure (آلية الاستدلال) التي يتمتّع بها المترجم مُهمّةً وعملية من أجل الاستخدام الأفضل للقواعد والحقائق، فقد تعطي هذه البنية الترتيبَ للسؤال أولاً، ومن ثم للقواعد والحقائق بالأهمية نفسها، ثم لاستخدام قواعد البيانات، ويُعرف هذا بالتسلسل الخلفي Backward Chaining، وقد يكون البدء من الحقائق والاستدلالات حتى الوصول إلى الاستنتاجات والذي يُعرف بالتسلسل الأمامي Forward Chaining، ويمكن تركيب بنى التحكم المختلفة للحصول على بنى التحكم الهجينة Hybrid Control Structures ويُستخدم فيها غالباً مزيج من التسلسل الأمامي والتسلسل الخلفي للحصول على مزايا مشتركة، ويمكن للغة البرولوغ القيام بتحويل القواعد من شكل إلى آخر؛ أي من التسلسل الأمامي إلى الهجين، ويتم ذلك بوساطة بعض المُعلنات الخاصة، ويقوم المترجم بفهرسة التعابير الإعلانية وذلك لمساعدة التسلسل الخلفي، أي أن المترجم يحتفظ بقائمة من المؤشرات لجميع القواعد وذلك لتسهيل عملية البحث عنها.

4. معالجة اللغات الطبيعية

تتجلى أهمية برمجيات معالجة اللغات الطبيعية Natural Languages Processing (NLP) في جعل الحوار بين الإنسان والآلة ممكناً بلغة أقرب ما تكون إلى اللغات الطبيعية التي يستخدمها في الحوار مع أقرانه، وذلك باستخدام الذكاء الاصطناعي. فاللغة هي منظومة من الرموز والمعاني والأصوات والقواعد التي يؤدي اندماجها إلى التخاطب مع الإنسان [64]، حيث يُستخدم تعبير اللغات الطبيعية للإشارة إلى اللغات الإنسانية؛ كاللغة العربية والإنكليزية والروسية وغيرها لتمييزها من اللغات غير الطبيعية؛ كلغات برمجة الكمبيوتر والشفرات السرية وغيرها. تُستخدم جميع هذه اللغات للتفاهم والاتصال سواء بين الإنسان والإنسان أو بين الإنسان والآلة، كما أن لكل منها مفرداتها ونحوها وقواعدها اللغوية التي تحدد كيفية بناء الجمل ومعناها الدلالي. يهدف هذا المجال إلى تأمين التخاطب المباشر بين الإنسان والكمبيوتر؛ كتابة أو كلاماً بلغة الإنسان الطبيعية، بدلاً من لغة البرمجة الجامدة والصعبة الفهم، وهذا يُسهِّل على الإنسان العادي استثمار الكمبيوتر.

تمّ إنجاز تقدمٌ ملحوظ في مجال معالجة اللغات الطبيعية في الفترة الأخيرة، حيث تُستعمَل تقانات معالجة اللغات الطبيعية في الكثير من المجالات، مثل نظُم الاستفادة من بنوك المعلومات التي تُخزّن الموسوعات والكتب القيمة، ونظم الإجابة عن الأسئلة، ونظم فهم الكلام وتوليده، ونظم قراءة النصوص وفهمها، ونظم تأليف النصوص، وغير ذلك.

1.4. بُنى معالجة اللغات الطبيعية

المعالجة الآلية للغات الطبيعية تعني معالجة اللغة الطبيعية في أشكالها كافة؛ مكتوبةً (نصوص) أو مطبوعةً (وثائق) أو مسموعةً (إملاء) أو مقروءةً (توليد الكلام).

1.1.4. البنية العامة

تُقسم معالجة اللغات الطبيعية إلى قسمين: معالجة النصوص المكتوبة؛ باستخدام المعطيات المعجمية والنحوية والدلالية [49]، ومعالجة النصوص المنطوقة؛ باستخدام المعطيات المذكورة نفسها بالإضافة إلى مجموعة المعطيات الصوتية [50]. تتألّف عملية معالجة اللغة الطبيعية من المحلّلات التالية:

- محلّل نحوي Syntactic Analyser لإظهار العلاقة بين كلمة وأخرى من الناحية الإعرابية [9].
- محلّل صرفي Morphological Analyser لتحليل الكلمات المفردة إلى مركباتها وتصريفها [40].
- محلّل دلالي Semantic Analyser لخلق البنى الصحيحة لكي تتلاءم الكلمات المرتبطة مع بعضها بعضاً من حيث المعنى.
- محلّل مقامي Discourse and Pragmatic Analyser ويشمل مرحلة اختيار المعاني التي تتّفق وسياق موضوع الكلام والتأثيرات البيئية الثقافية والاجتماعية [57].
- محلّل صوتي Phonological Analyser لتحليل الصوت إلى وحدات صوتية تُعرف بالفونيمات، وتشكّل كل مجموعة منها كلمة من كلمات اللغة.
- مركّب صوتي Phonological Synthesiser لتحويل الكلمات إلى التوصيف الصوتي الخاص بكل كلمة وفقاً لمعايير صوتية خاصة، ثمّ يتمّ توليد الكلام وفقاً لهذه المعايير لتحويل النصوص إلى كلام منطوق.

2.1.4. البنية الكلامية- النصيّة

تُعدُّ تقانة تحويل النص المكتوب إلى كلام Text-To-Speech (TTS) من أهم التقانات التي تُتيح معالجةَ النُّصوص العربية وتشكيلَها، ومن ثمّ تحويلَها إلى كلامٍ منطوق بصوتٍ شبيهٍ بصوت الإنسان، إذ ترتكز هذه التّقانة على ثلاث وحدات؛ وهي الوحدةُ اللُّغوية المُمثّلة بالنصِّ العربي، والوحدةُ الصوتية التي يتمُّ من خلالها تحويل الكلمات إلى التوصيف الصوتي الخاص بكلِّ كلمةٍ وفقاً لمعاييرَ صوتيّةٍ خاصّةٍ، ثمّ الوحدةُ السّمعية التي تقوم بتوليد الكلام وفقاً لهذه المعايير. يتمّ في

هذه التقانة استخدام حزمة برمجية ضمن الكمبيوتر وكرْت ومكبّرات للصوت لخلق كلام بـديل لشاشة الكمبيوتر؛ تُحلِّل البرمجيات المعطيات في ذاكرة الفيديو Video وتحول كل مـا يظهـر على الشاشة إلى مركبات صوتية Phonetic Component تتألف البرمجية من قاعدة بيانـات صوتية تضمّ مجموعة الكلمات أو الحروف المسجلة. عندما يتمّ تنفيذ أمر ما تقـوم البرمجيـة بإرسال المعطيات إلى كرت الصوت ليتمّ نطقها [2] [44]. قد تحتوي هذه المعالجة على المحلّل النحوي والصرفي والدلالي والمركّب الكلامي.

أما تقانة تحويل الكلام إلى نص مكتوب Speech-To-Text (STT) فيتمّ فيها التـدرُّبُ علـى نماذجَ من الكلام موصوفةٍ بمجموعةٍ من المميِّزات الطيفيَّة، ومن ثم إدراكُ الكلام عن طريـق التعرّفِ على النَّماذجِ المدرَّبِ عليها والجديدةِ. ويتمّ فيها الاعتماد على عدة تجارب ترتكز على إجراء اختبارات نموذجية وِفقاً لنطق مجموعة من الكلمات أو الجمل أو الأصوات، والطلب من المستمعين إليها إعادة كتابتها أو تكرارها، وذلك لإحصاء نسبة الكلمات التي تمّ تمييزها بـشكل صحيح، وهذا هو مقياس درجة فهم الكلام. تعتمد هذه النتيجة على الظـروف التـي يجـري الاختبار فيها وعلى مفردات الاختبار (كلمة أو جملة أو صوت كلامي) وتخزينها ضمن قائمـة تحوي الكثير من الأمثلة، فهناك نوعان من التجارب لهما دور فعّـال فـي إدراك الأصـوات الكلامية وهما اختباران يتعلق أحدهما بالكلام الاصطناعي (الصادر من الكمبيـوتر) والآخـر بالكلام الطبيعي.[15]. قد تحتوي هذه المعالجة على المحلِّل الـصّوتي والنّحـوي والـصّرفي والدّلالي.

3.1.4. البنية الكلامية

- **فهم الكلام**

يدلّ فهم الكلام على فهم الكلام المنطوق بوساطة الكمبيوتر، ويعتمد عمل منظومته على تحليـل الإشارة الصوتية (الكلام) [53] [65]، فتمييز الكلام يشتمل على تحويله من موجات صوتية إلى نماذج للكلام Speech Models، ولكي يستطيع الكمبيوتر تمييز الكلام، ثمـة مـدخلان لهـذا الموضوع؛ المدخل الأول هو التمييز المُعتمد على المتكلم الذي يُصمَّم ليميّـز الكـلام لـشخص معين، والمدخل الثاني التمييز غير المُعتمد على المتكلم، ويُصمّم لتمييـز الكـلام لأي مـتكلم. والمشـاريع الأكثر شـيوعاً في فهم الكـلام هي نظام Hearsay الذي اعتمد على المتكلم في معدّل التمييز، ونظام Harpy الذي لم يعتمد على المتكلم [50]. يتمّ الاعتمـاد علـى التحليـل الصوتي للوحدات الصوتيَّة في عملية تمييز أو إدراك الصوت، حيث تُستَخلص مجموعـة مـن المميّزات لكل وحدةٍ صوتيَّةٍ مثل Formant Frequency، أما إدراك النماذج فلا يحتاج إلـى

معرفةٍ عميقةٍ في ميزات الوحدات الصوتيَّة؛ بل يتمُّ التدريب على نماذج من الكلام موصوفةٍ بمجموعةٍ من المميزات الطيفيَّة، ومن ثمّ إدراك الكلام عن طريق المقارنة بين النماذج المدرَّب عليها والجديدة، والطرائق الشائعة المُستخدمة في التعرُّف على النماذج هي منابع ماركوف المخفية Hidden Markov Model (HMM)، والشبكات العصبية الاصطناعية (ANNs) Artificial Neural Networks [18]. فقد تمّ بناء شبكة عصبية للتعرّف على الكلام المنطوق، حيث استُخدمت عدة معايير وطرائق وخوارزميات رياضية تقوم باستخلاص المعلومات من إشارة الكلام وتدريب الشبكة على هذه المعطيات [21].

- **توليد الكلام**

تهدف عملية توليد الكلام Speech Production إلى تطوير برامج لها القدرة على إنتاج لغة حية باستخدام الكمبيوتر، أي أن هذه البرامج قادرة على إنتاج الكلام الطبيعي لجمل وكلمات مكتوبة بلغة حية. وتوجد طريقتان لتوليد الكلام في الكمبيوتر: الطريقة الأولى التي يتمّ فيها تسجيل عدد محدود من الكلمات ينطق بها الكمبيوتر عندما يُطلَب منه ذلك، وتتميز هذه الطريقة بأنها تنطق الكلمات المخزّنة بسهولة ووضوح. أما الطريقة الثانية فهي تتضمن إنتاج الكلمات عن طريق استخدام المقاطع الصوتية أو الفونيمات المسجّلة للغة ما من اللغات.

يرتكز توليد الكلام على ثلاث وحدات: لغوية وصوتية وسمعية حيث تُشكل الوحدة اللغوية النص، ثم تحول الكلمات إلى التوصيف الصوتي الخاص بكل كلمة. وتعمل الوحدة الصوتية على حساب المعايير الصوتية المطلوبة لنطق الأصوات وفقاً للوصف الصوتي الذي تقدمه الوحدة اللغوية، بينما تُولِّد الوحدة السمعية الكلام وفقاً لهذه المعايير المحددة. يوجد العديد من الأبحاث في مجال تركيب الكلام لمختلف اللغات الطبيعية، لاستخدام هذه التقانات في مجالات مختلفة مثل القراءة الآلية للنصوص ومساعدة ذوي الاحتياجات الخاصة في التواصل مع المحيط من خلال الآلات الناطقة Talking Machine [48] [58].

2.4. خصائص اللغة العربية

تتسّم اللغة العربية بتوازن في معظم خصائصها اللغوية يضعها في منطقة متوسطة بين كثير من المحاور التي تُحدِّد مجالات التنوع اللغوي؛ فأبجدية الكتابة العربية ليست فونيمية صرفة كالأسبانية، كما أنها ليست مقطعية Syllabic كاليابانية، فالأبجدية العربية رغم كونها فونيمية أساساً، فإنها تتضمن حروفاً ذات طبيعة مقطعية (لأ)، ويفرض التوسط اللغوي ضرورة الاهتمام بالدراسات اللغوية المقارنة Comparative والتقابلية Contrastive [32].

تُستخدم اللغة العربية على مجمل المستويات العلمية والتقانية كباقي اللغات؛ فالعربية صالحة لأن تكون لغة للمُدخَلات، كما أنها صالحة أيضاً لأن تكون لغة للمُخرَجات. عُولِجت اللغة العربية في نظم استرجاع المعلومات؛ وهي قضية لغوية كمبيوترية معقدة تتجاوز أبعادها علم المكتبات والمعلومات، فتلك القضية ذات جوانب متعددة تشكل جوانبها نسيجاً متداخلاً من علم المعلومات والكمبيوتر واللغويات وعلم المكتبات والإحصاء وغيرها [26]. تُقسم الكلمة إلى اسم وفعل وحرف؛ فالاسم هو ما وُضِعَ ليدلّ على معنى مستقل بالفهم لا يتعلق بالزمن (رجل، كتاب)، والفعل ما وُضِعَ ليدلّ على معنى مستقل بالفهم ويتعلق بالزمن (كتب، يقرأ)، والحرف ما وُضِعَ ليدلّ على غير معنى مستقل بالفهم (هل، في) [14]. تمّ تحديد الضوابط التركيبية والتحويلية للفعل، فلقد قسّم النحاة الأقدمون الفعل إلى اتجاهات تختلف باختلاف المعايير المعوّل عليها لضبط الفعل ضبطاً يُتوَخّى إبرازه في حيزه اللساني بوصفه عنصراً له سماته التي تميزه مما عداه من العناصر اللسانية الأخرى، تبين القواعد التفريعية للفعل نوعين من السمات التركيبية: السمات الذاتية؛ ترتد إلى طبيعة الفعل في ذاته بمعزل عن السياق، والسمات الانتقائية؛ وهي سمات سياقية تُلحَق بالمداخل المعجمية [12]. وتتميز اللغة العربية بتقابل الأصوات والمعاني في تركيب الألفاظ، وبأثر الحروف في تقوية المعنى أو إضعافه، وبالانسجام بين أصوات الحروف التي تتركب منها الألفاظ ودلالاتها، وأكثر ألفاظها يتكون في أصوله الاشتقاقية من ثلاثة حروف صوتية، من دون حساب الحركات، وهذه الحروف الثلاثة ثابتة لا تتغير [41]. فالحركة الإعرابية لها دورٌ لا يقلّ في أهميته عن دور أي حرف من حروف الكلمة في الوصول إلى المعنى الدلالي للجملة [36]، وتُعدّ الحروف العربية الحروف الأساسية لعدد آخر من اللغات، وتُستخدم مع حروف أخرى لكتابة لغات مختلفة منها الفارسية والكردية والأوردو [1].

تمّ تصنيف جذور مفردات اللغة العربية الواردة في معجم الصحاح في صورة جداول تُوضِّح تكرار الحروف في المواقع المختلفة من الكلمة، وتَتَابُعها واختلافها من حرف لآخر، وقد تبين بعد توزيع كلمات اللغة العربية بالنسبة إلى عدد حروفها أن غالبيتها 85% مكون من ثلاثة حروف، وأغلب الباقي مكون من أربعة حروف، بنسبة 13.6 %، وهذا يؤكد أن اللغة العربية تعتبر لغة ذات أصل ثلاثي [13]. يوجد فرق بين بناء الجملة أو إنتاجها وبين فهمها، فبناء الجملة يقوم على إحياء كم غير منظم من المفردات بترشيح مجموعة من العلاقات الأساسية بينها، أما فهم الجملة فمعناه إدراك مجموعة العلاقات الأساسية التي تربط بين مفرداتها المتفرقة [5]، فعلم الدلالة لا ينفصل عن بقية العلوم الأخرى التي تساعده على القيام بمهمته [43].

3.4. معالجة اللغة العربية

أُجرِيت دراسات عدة في مجال معالجة اللغة العربية بوساطة الكمبيوتر، وتمَّ إيجاد نظام لتشكيل النصوص العربية، من خلال بناء محلّل صرفي نحوي، يتمّ من خلاله تركيب وتحليل الجمل في اللغة العربية، ووُضِع معجم للمعالجة الآلية لهذه اللغة [59]، كما قُدِّمت رسالة دكتوراه أخرى في هذا المجال حول تصميم وتنفيذ نظام تحليل صرفي نحوي للغة العربية، واعتُمِد فيها على قاعدة بيانات معجمية [52]. وقد حُدِّدَت قواعد اللغة وتمَّ بناء أُنموذج لآليتها عن طريق النظرية التوليدية والتحويلية التي تُقَدِّم صورةً مكتملة للنظرية الألسنية التي تتناول قضايا اللغة وتُفسِّرها، إذ تُقَدِّم نظرةً واضحة عن بنية اللغة وميزاتها الإنسانية واكتسابها وعلاقتها بالفكر الإنساني [25]. تتضمن القواعد التحويلية والتوليدية لائحة من التراكيب الشكلية التي تُكَوِّن جمل اللغة، شرط أن لا تحتوي هذه اللائحة على جمل ليست من جمل اللغة، وتُشَكِّل أصول اللغة في إطار هذه النظرية تنظيماً يربط بين الأصوات والمعاني، وتتألف من ثلاثة أقسام متماسكة يشتمل كلٌّ منها على تنظيم قواعدي، وهذه الأقسام هي: المكوَّن الفونولوجي، والمكوَّن التركيبي، والمكوَّن الدلالي [23]. عالَج محمد علي الخولي النظرية التحويلية وقوانينها الأساسية والمفرداتية والتحويلية والمورفيمية الصوتية وطبّق هذه القوانين على اللغة العربية، كما قام بإجراء دراسات تحليلية بهدف التعرّف على الصيغ الصرفية والوظائف النحوية الشائعة فيها، وعرض بعض التراكيب اللغوية التي قد يحدث فيها لبس نحوي من حيث المعنى وبيّن أسباب هذا الالتباس [17]. يمكن تحديد عمل القواعد التوليدية والتحويلية بقدرتها على توليد جمل اللغة الأصولية والتمييز بينها وبين غير الأصولية، وتحليل العلاقات القائمة بين عناصر الجملة [24]. تنظر البرمجيات إلى اللغة بأنها مجموعة من الجمل، تتكون كل منها من رمز واحد أو أكثر من مفردات اللغة، ويتكون نحو اللغة من تحديد رياضي لهذه المجموعة من الجمل، ولكتابة برنامج يستطيع التحكم بما أُدْخِل إليه إذا كان جملة في لغة أو لا، لابدّ أن يحتوي على نحو هذه اللغة، ويتكون هذا النحو من رباعية Quadruple وهي مجموعة منتهية Finite Set تضم المفردات الموجودة في اللغة، ومجموعة منتهية أخرى تضم أسماء المركبات النحوية (مركب اسمي، جار ومجرور،..)، ورمزاً للإشارة إلى الجملة A start symbol لمجموعة المركبات النحوية، ومجموعة من القواعد التوليدية، ويستطيع هذا النحو توليد عدد لا نهائي من الجمل [35]. ولا يمكن أن تكون اللغة نظاماً من المعاني التي لا مباني لها، لأنَّ المباني رموز المعاني ولا غِنى عن الرمز في نظام كاللغة هو في أساسه نظام رمزي؛ أي أن المعاني هي وظائف تؤديها المباني مثال على ذلك كلمة "ضَرَبَ" معناها الفعلية والماضي، أما مبناها فهو صيغة

"فَعَلَ" [30]. تمّ بناء نظام الاشتقاق والتصريف للغة العربية، الذي قسّم المعالجة الآلية للكلمة العربية إلى قسمين أساسيين: الاشتقاق والتحليل، ففي الاشتقاق (التوليد) يتمّ الانتقال من جذر الكلمة إلى مجمل الكلمات المشتقة منه، وفي التحليل يتم الانتقال من الكلمة إلى جذرها، وتحديد وضعها الصرفي والنحوي والدلالي. اعتُمِد في ذلك على معارف لغوية مُعجمية وقواعدية احتوى عليها المعجم المخزّن في الكمبيوتر [6] [7]. كما أَمكن تحليل الجملة ومعرفة أجزائها باعتبارها مكوَّنة من جزأين، وكل جزء مكوَّن بدوره من جزأين أيضاً حتى الوصول إلى الكلمة المفردة، وصُنِّفَت الكلمات بطريقتين إحداهما صرفية والأخرى نحوية (وظيفية) [16]، وقسّم بعضهم الجملة العربية إلى قسمين: القسم الأول جملة كبرى؛ وهي مكونة من جملتين أو أكثر إحداهما مبتدأ أو فاعل أو خبر أو مفعول ثان لفعل ناسخ، مثل "سواءٌ عليَّ أيَّ شيءٍ فعلتَ"، والقسم الثاني جملة صغرى؛ التي تكوّن جزءاً مُتمِّماً للجملة الكبرى؛ قد تكون مبتدأ أو فاعلاً أو خبراً أو مفعولاً ثانياً مثل "أيّ شيء فعلتَ" [37]. والصرف في اللغة هو التقليب والتغيير بينما في الاصطلاح فهو علم بأصول تُعرف به أحوال أبنية الكلم التي ليست بإعراب، وينقسم علم الصرف إلى قسمين القسم الأول يمثل مجمل أنواع المباني؛ الأسماء والأفعال والمشتقات والجمع، بينما يمثل القسم الثاني الحالات الطارئة التي تحدث في بنية الكلمة لتغييرها من بنيتها الأصلية إلى بنية أخرى [27].

تمّ تصنيف كلمات اللغة العربية وفقاً لمعانيها في أبواب كثيرة، ومنها ما يتعلق بصفات الإنسان الخَلقية والخُلقية، والكلام والكتابة والأصوات وفي مشي الإنسان وانتقاله وسكونه وغيرها [29]. تمّ تقسيم الكلام العربي إلى قسمين رئيسين: القسم الأول الخبر؛ وهو كلام يحتمل الصدق والكذب لذاته، وهو ما يتحقق مدلوله في الخارج من دون النطق به مثل "العلمُ مفيدٌ". والقسم الثاني الإنشاء وهو كلام لا يحتمل صدقاً ولا كذباً، وهو ما لا يحصل مضمونه ولا يتحقق إلا إذا تلفظت به مثل "اقرأ الدرسَ" [42].

1.3.4. تطبيقات معالجة اللغة العربية

تتجلى أهمية المعالجة الآلية للغة العربية في الكثير من التطبيقات مثل التعرّف على الكلام وتركيبه، والقراءة الآلية للنصوص المكتوبة (TTS)، والكتابة الآلية للنصوص المنطوقة (STT) ، وقارئ المستندات؛ الذي يتمّ فيه تحويل المستندات التي تمّ مسحها ضوئياً إلى نصّ مقروء، ثم تُقرأ المحتويات بوساطة محرك النطق الآلي للنصوص (TTS)، ومعالجة الوثائق آلياً Image Processing بحيث تَسهل قراءتها وتحليلها والوصول إليها أو إلى أي جزء منها

بطريقة مباشرة أو غير مباشرة، وفي تطـوير تطبيقـات الوسـائط المتعـددة Multimedia Authoring، وفي التعرّف على النماذج Pattern Recognition مثل الحـروف وغيرهـا، والترجمة الآلية والمُصحّح الإملائي؛ أي اكتشاف الأخطاء اللغوية في النصوص وتـصحيحها، وفي تقانات نُظُم استخلاص المعلومات Retrieval Information Systems للوصول بسهولة ودقة إلى المواد المتعلقة بموضوع معين، والتحاور مع الآلة باللغة العربية، والفهرسـة الآليـة للنصوص، وضغط النصوص واسترجاعها، وإمكانية تعلّم هذه اللغة وتعليمها للنـاطقين بهـا، وتشكيل النصوص غير المشكّلة أو المشكّلة جزئياً.

2.3.4. بناء منظومة صوتية للغة العربية

تملك المعالجة الآلية لكل من الكلمة والجملة العربية في الكمبيوتر مـستويات متعـددة، فهنـاك المعالجة الصرفية والنحوية والدلالية والمعجمية والصوتية وغيرها، ويجب أن تُبنى هذه المعالجة على قواعد واضحة ومحددّة، مما يستوجب الكشف عن دقائق بنيـة اللغـة والإحاطـة التامـة بمفرداتها؛ فالإحاطة بالمفردات أمر ممكن، لكن الإحاطة بالتراكيب والجمل بأنماطها المختلفـة أمر غير ممكن وذلك لأنها غير محدودة؛ فاللغة هي استعمال لا محدود لوسائل محدودة.

للغة العربية خصائص ومميزات تميزها من غيرها من اللغـات، ففيهـا خصيـصة الاشـتقاق الصرفي، والمرونة النحوية، واعتماد المعجم على الجذور، والصلة الوثيقة بين المبنى والمعنى، واطراد القياس في كثير من الحالات الصرفية والإعرابية والصوتية، وهي أيضاً توصف بأنهـا لغة جبرية، فجميع الأفعال الثلاثية مثلاً تنتظمها صيغة (فَعُل).

فاللغة ليست نظاماً عشوائياً بل منظومة متَّسقة تقيدها الضوابط وتحكمها القواعد المطردة، وفي المنظومة اللغوية يكمن كثير من التشابكات والتداخلات الدقيقة، التي تـدين للتحليـل وتخـضع للتقنين والتقعيد [46]. فمنظومة الكيان اللغوي Linguistic هي منظومة شاملة لعدة منظومات فرعية منها منظومة الصوتيات Phonology، وتتحكم بالطريقة التي تخـرج بهـا الأصـوات اللغوية وتلوُّن درجاتها [4]، ومنظومة الصرف Morphology، وهي المـسؤولة عـن بنيـة الكلمات من حيث هيكل عناصرها والعلاقات التي تـربط بـين هـذه العناصـر [20] [19]، ومنظومة النحو Syntax وتتضمن عدة آليات تتحكم في ترتيب الكلمات داخل الجمل، مثل تلك الخاصة بالتقديم والتأخير وغير ذلك [3] [28]، والعدد المحدود من الكلمات في كل لغـة هـو المادة التي تمكّن من خلق وتركيب عدد لا محدود من الجمل، ولكن ذلك لا يتمّ عشوائياً ومـن دون ضوابط وإنما تحكمه مجموعة من القوانين تسمى القواعد النحوية [10] [45]، ومنظومـة

الدلالة Semantic، وتختص بالقضايا المتعلقة بالمعنى وهي تفرض قيودها على منظومة النحو بحيث تمنع توليد جمل سليمة نحوياً وغير مقبولة دلالياً [31].

إنَّ دراسة الكلام المنطوق المسموع مُقدّمة لا بدّ منها لدراسة الأنظمة (القواعد) اللغوية (اللغة نفسها) وأصبح علم الأصوات تمهيداً بالملاحظة الحسيّة لإنشاء علم الصوتيات الذي هو تخطيطٌ عقليّ لقواعد الأصوات بناءً على هذه الملاحظة الحسيّة [11].

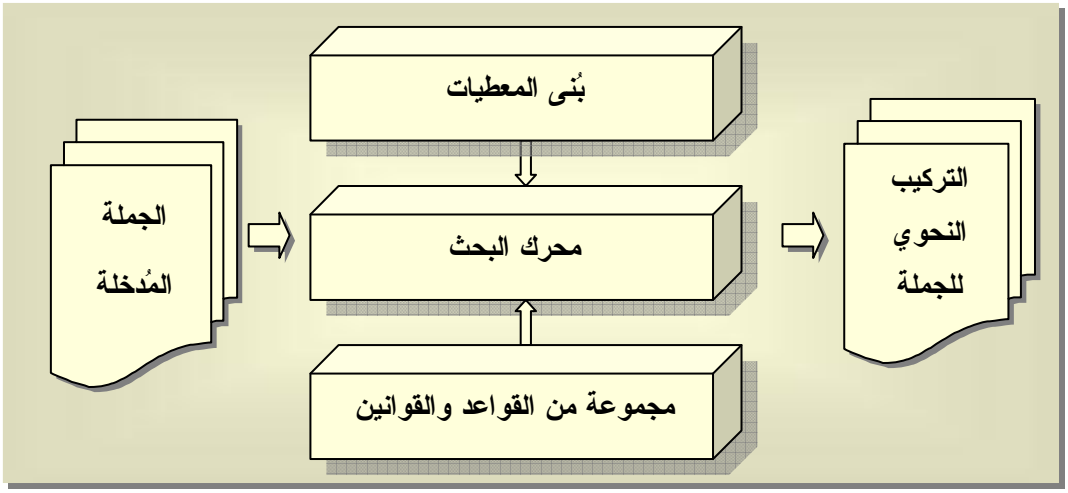
1.2.3.4. المحلّل النحوي

يقوم المحلِّل النحوي بتقسيم الجملة وتحديد قوالبها، ويُجزِّىء الكلمات فيها لإيجاد العلاقة النّحوية فيما بينها. حيث أن الجملة والقالب مصطلحان مختلفان ولكنهما متَّصلان اتِّصالاً وثيقاً، فمثلاً "أكلَ الولدُ التفّاحةَ" جملة، ولكن (فعل + فاعل + مفعول به) قالب. لكلِّ قالب عددٌ لانهائي من الجمل، ومن بعض القوالب الشَّهيرة في اللُّغة 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

- **تصميم المحلّل النحّوي**

تمَّ بناء محلّل نحوي من أجل ايجاد التراكيب النحوية للجمل العربية، وذلك من خلال تصميم بُنى المعطيات (مجموعة الحقائق والأساليب والقواعد اللازمة لإيجاد القالب النّحوي للجمل) المناسبة، وإنشاء محرك بحث (إجراءات مُبرمجة) يقوم بعملية الاستدلال والاستنتاج للوصول إلى القالب النّحوي الصحيح من مجموعةٍ من الحلول الممكنة، وذلك اعتماداً على مجموعة من القواعد والقوانين المتعلقة باللغة العربية. يُبيّن الشكل (1) المخطط العام للمحلّل النحوي.



الشكل (1) المخطط العام للمحلّل النحوي

تمّ تشفير المعرفة وتمثيلُها على هيئة قواعدَ وحقائق، حيث تمَّ تحديد التعابير الإعلانيـــة للجملـــة الاسمية من خلال قوالبها التي تمّ تشكيلها بناءً على القواعد التوليدية والتحويلية. مثلاً يمكن إيجاد التعبير الإعلاني للقالب "مبتدأ +خبر" بالشكل التالي:

sentn2 ("مبتدأ"،"خبر").

تُعدُّ كل كلمة من كلمات الجملة بارامتراً لها.

يقوم محرّك البحث بربط القواعد والحقائق المعيَّنة لتكوين خطِّ الاستنتاج والاستدلال، وقـــد تـــمّ بناؤه ليضمّ آلياتِ التحكُّم لتمكين الفرضيات المولَّدة من الوصول إلى الحلِّ الصحيح، ويسمح هذا المحرّك بتوليد الافتراضات بناء على المعلومات المتوفِّرة في قاعدة المعرفة، وقد قـــام محـــرّك 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 فعندئذ تتمّ العودة إلـــى الحال السابقة مباشرة، وتتكرَّر هذه العملية حتى الوصول إلى الحلِّ الأفضل. قام هذا المحـــرّك بالبحث للحصول على القوالب النّحوية المُمكنة لهذه الجملة، سواء أكانت اسميةً أم فعليةً، وذلـــك بعد إجراء فهرسةٍ لأسماء التعابير الإعلانية، حيث يَحتفِظ محرّك البحث بقائمة من المؤشـــرات لجميع القواعد والحقائق وذلك لتسهيل عملية البحث عنها، وتدلّ عملية الفهرسة هـــذه علـــى أن الحقائق يمكن أن تأخذ الكثيرَ من التَّرتيبات المختلفة والتي تعطي السُّلوكية ذاتَها، ومن ثـــمّ يـــتمّ الوصول إلى الهدف المطلوب وهو القالب النّحوي الصحيح للجملة، وذلك وفقاً لعدّة مراحل.

الاختبار والأداء

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

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

يقوم محرِّك البحث بتطبيق مجموعة من القواعد الأساسية على كل كلمة من كلمات الجملة لاختيار القالب النّحوي الصحيح لها:

- الكلمة الأولى "استيقظ" تأخذ كل الحلول.
- بسبب انتماء الكلمة الثانية "صباحا" إلى مجموعة الظروف يتمّ اختيار القالب 3 أي تمَّ الوصول إلى حلٍّ وحيد وهو القالب النّحوي الصحيح للجملة:

استيقظ صباحا = فعل لازم + ظرف= استيقظَ صباحاً

أما في حال الجملة "حضر الطالب"، يتمّ استدعاء نفس القوالب النحوية الممكنة السابقة (لأن هذه الجملة مكونة من كلمتين أيضاً)، يقوم محرِّك 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 الى الحلول الممكنة التالية:

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

يتمّ الوصول الى حل وحيد بعد خضوع الجملة للمحلّل الصرفي والدلالي. تمّ الحصول على الكثير من النتائج التي أثبتت فعاليّةَ النّظام المصمَّم وفائدةَ الاختبارات التي خَضَع لها، وقد

تميّزت بالدقّة وسرعة استجابة النظام، وطُبِّق هذا النظام على نماذجَ متعدِّدةٍ من الجمل العربيـــة غيرِ المشكّلة، ونَتج عن ذلك إيجادُ القوالب النّحوية للجملة بعد عرضِ مُختلَف الحلول الممكنـــة، ومن ثمّ اختيارُ القالب النّحوي المناسب بعد سلسلةٍ من عملياتِ الاستدلال والاستنتاج التي قام بها محرّك البحثِ للوصولِ إلى الحلِّ الأفضلِ بالاعتماد على القواعد النّحوية في اللّغة العربية.
يُقاس مستوى أداء النظام بالقدرة على الاستنتاج والاستدلال وإعطاءِ النتائج الصحيحة، كما تلعب بُنى المعطيات ونوعيةُ المعرفة التي تحتويها دوراً هاماً، وتتضمّن مجموعةَ الحقائق أو التَّعـــابير الإعلانية التي تصِف العلاقة المنطقية بين العناصر والمفاهيم، وطرائقَ حلِّ المشكلات، حيث تمّ استخلاص المعرفة والخبرة من أسس اللّغة العربية.

2.2.3.4. المحلّل الصرفي

يقوم المحلّل الصرفي بتحديد البُنية الصّرفية النّحوية للجملة العربية، وذلك بعد تمثيـــل القواعـــد الصّرفية لهذه الأوزان على شكلِ جُملٍ منطقية. اعتمد المحلّل على تجريد الكلمات من السوابق واللواحق وتحديد أوزانها المُقابلة عن طريق استدعاء الهيكل البنيوي المُلائم لكل كلمة من بُنـــى المعطيات، وإيجاد العلاقة الصّرفية النحوية من أجل التوصّل إلى القالـــب النحـــوي المـــوزون للجملة المُدخلة، وهو يُمثّل القالبَ الصّرفيَّ الموزونَ لها.

- **تصميم المحلّل الصّرفي**

يقوم المحلّل الصرفي بتحليل كلمات الجملة المُدخلة وتحديد بياناتها الصّرفية، وذلك بعد إجـــراء عدة خطوات:

- تجريد الكلمة من السوابق واللّواحق مثل "ال" التعريف والضمائر المتّصلة وغير ذلك.
- إيجاد الوزن المُقابل لكل كلمةٍ من كلماتِ الجملة، وذلك من خلال بُنى المعطيات.
- إيجاد العلاقة الصّرفية النحويةً بين مُركّبات الجملة وذلك طبقاً للقواعد النحوية التي تُحدِّد الموقعَ الصرفي لكل وزن، ويُبيّن الشكل (2) المخطط العام للمحلّل الصرفي.

الشكل (2) المخطط العام للمحلّل الصرفي

تمّ تصميم بُنى المعطيات للمحلّل الصرفي التي تضمّنت أوزان الأسماء والأفعال في اللغة العربية، حيث تمّ تمثيل المعلومات والقواعد الصرفية لهذه الأوزان بشكل مُعلنَات، وتمّ وضع التَّعابير الإعلانية ذات السمات المُشتركة ضمن هياكل Slots، فتمّ إيجاد هيكلِ أوزانِ الاسم الثلاثي المجرّد، وهيكلِ الاسم الرباعي المجرّد، وهيكلِ الفعل الثلاثي المجرّد، وهيكلِ الفعل الثلاثي المزيد وغير ذلك، وقامت بُنى التحكم بفهرسةٍ لأسماء المعلنات المكوّنة لهذه الهياكل، وذلك لمساعدة التسلسل الخلفي من أجل عملية الوصول إليها، ومن الهياكل التي تمّ بناؤها في بُنى المعطيات:

هيكل الأسماء الثلاثية المجرّدة:

w_Anoun3 ("فَعْلٌ"،"فَعِلٌ"،"فَعَلٌ"،"فُعُلٌ"،"فُعْلٌ"،"فُعَلٌ"،"فَعُلٌ"،"فِعْلٌ"،"فَعِلٌ"،"فِعَلٌ"،"فِعِلٌ").

هيكل الأسماء الرباعية المجرّدة:

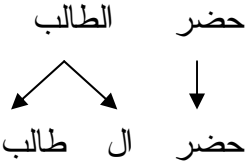
w_Anoun4 ("فَعْلَلٌ"،"فُعْلَلٌ"،"فُعْلُلٌ"،"فِعْلَلٌ"،"فِعْلِلٌ"،"فِعَلٌّ").

يتم تصميم وحدة ملاءمة وربط بين بُنى المعطيات الصرفية والمحلّل النحوي. يُـربط المحلـل النحوي بمجموعة الحقائق والقواعد الصّرفية اللازمة في بُنى المعطيات الصّرفية عـن طريـق وحدة ملاءمة تؤمّن تبادل المعطيات من وإلى محرّك البحث، وتتعامل مع سلاسلَ من الكلمـات التي تُمثّل أوزان الأسماء والأفعال في اللغة العربية، وتعمل كمُوزّع للمعطيات، وتقوم بعمليـة عزل سهلة ومرنة بين المعطيات الصّرفية ومحرّك البحث.

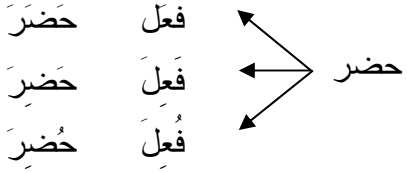
يقوم محرك البحث باستدلال واستنتاج القالب الموزون للجملة المُدخلة، حيث فرضت بُنى التحكم تسلسلاً خلفيّاً متعاقباً على كل حالة معتمِدة على طريقة بحث العمق أولاً Depth First، ويسمح هذا المحرك بتوليد الافتراضات بناءً على الأوزان المخزّنة ضمن الهياكل، وقد تمّ بناء المحرّك ليضمّ آليات التحكُّم لتجريد الكلمة من السوابق واللواحق، ولإيجاد الوزن المُقابل لكل كلمة مـن كلمات الجملة وذلك بعد استدعاء الهيكل المُناسب من بنى المعطيات، ولإيجاد العلاقة الـصّرفية النحوية بين مُركَّبات الجملة، وذلك تبعاً للقوالب النحوية الناتجة من المحلّل النحوي الـذي يـتمّ استرجاعه من أجل مقارنة الُبنى الصرفية الناتجة عن توزين كل كلمة مع القوالب الناتجة مـن المحلل النحوي، فيتمّ تقليصُ الحلول الممكنة حتى الوصول إلى الحل الصحيح.

الاختبار والأداء

في حال الجملة "حضر الطالب"، وبعد خضوعها لعملية التحليل الصرفي، تمَّ تقطيع كلماتها إلى قائمة من الحروف، ومن ثم تجريدها من السوابق واللواحق:



تمّ تحليل الكلمة الأولى "حضر"، وعُرِف عدد حروفها وهو 3 حروف، وحُكم عليها بأنها تنتمي إلى زمرة الأفعال بسبب وجود "الـ" التعريف في الكلمة الثانية، فتمّ استدعاء هيكل الأفعال الثلاثية المجرّدة الماضية w_Averb3، واستُنتجت ثلاثة أوزان لهذه الكلمة وهي:



أما الكلمة الثانية "طالب" فيستنتج أنها اسم، لأنه لا يُمكن ورود فعلين متتابعين مباشـــرةً، وبعـــد معرفة عدد حروفها وهي 4 حروف، تمّ استدعاء هيكل الأســـماء الثلاثيـــة المزيـــدة بحـــرف w_Enoun3 الذي يحوي 26 وزناً وهي:

("أفعَـــلٌ"، "مُفعَـــلُ"، "فَـــاعِـــلُ" ، "فَيْـــعَـــلُ"، "فُعـــالُ"، "فَعِـــيلُ"، "فُعـــلَى"، "فُعْلَـــةٌ"، "فُعَـــالٌ"، "فِعَالٌ"، "فَعُولٌ"، "مِفْعَلٌ"، "فَوْعَلٌ"، "فُيْعَلٌ"، "فَعْوَلٌ"، "فَيْعَلٌ"، "فُعْوَلٌ"، "فُنْعَلٌ"، "فُعْلُلٌ"، "فِعِلّمٌ"، "فِعِلِلٌ" ، "فِعِيْلٌ"، "فِعْوَلٌ"، "فِعَلٌّ"، "فَعْلَةٌ"، "فِعْلَةٌ")، لكن بسبب وجود الحرف "ا" كحرفٍ ثانٍ يـــتمّ اختيـــار الوزن "فَاعِل"، فتنتج البُنى الصرفية التالية لهذه الجملة:

1. فَعَلَ + فَاعِل
2. فَعِلَ + فَاعِل
3. فُعِلَ + فَاعِل

بعد خضوعها للمحلّل النّحوي تم استنتاج الحلول الممكنة التالية:

1. فعل لازم + فاعل
2. فعل لازم + حال
3. فعل مبني للمجهول + نائب فاعل

بعد إيجاد العلاقة الصرفية النحوية بين أوزان الكلمات يتمّ استبعاد البنية الصرفية 3، وذلك بسبب كون وزن الكلمة الثانية "فَاعِل، وهذا الوزن لا يأتي بعد "فُعِل" لذا لا يُمكن أن تكون الكلمة الأولى فعلاً مبنياً للمجهول، فتَنتج البُنى الصرفية النحوية التالية:

فَعَلَ + فَاعِل ⟵ حَضَرَ الطَالِبُ A

فَعِلَ + فَاعِل ⟵ حَضِرَ الطَالِبُ B

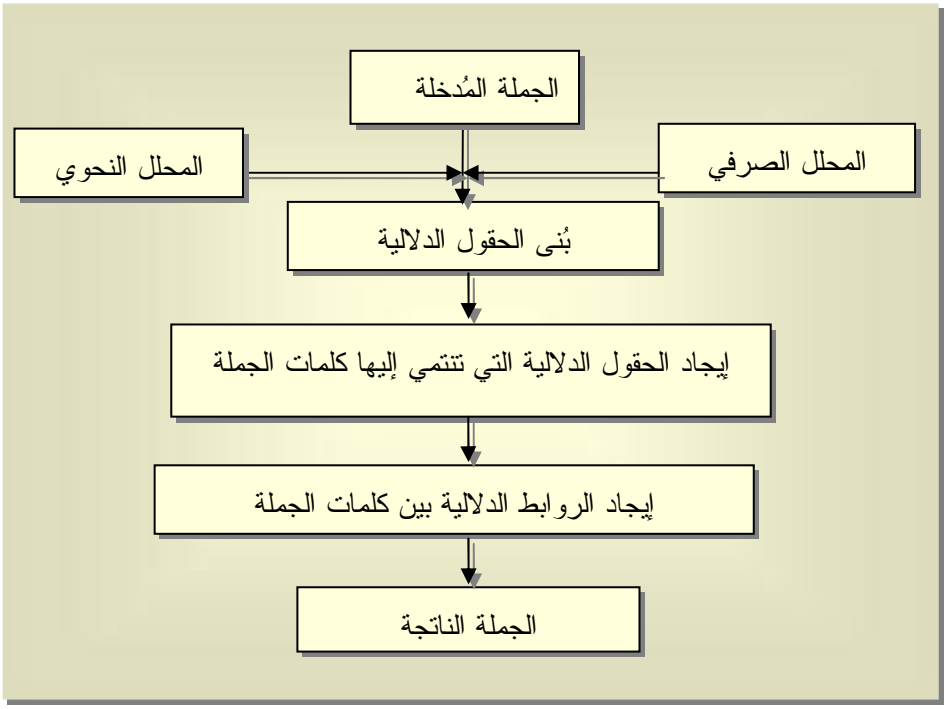
بالاستعانة بالمحلّل الدلالي يتمّ الوصول إلى قالب صرفي نحوي وحيد من بين A وB. أثبتـــت التجارب فعّالية المحلّل الصرفي، حيث كان لهذا المحلل دورٌ أساسيٌّ في تقليص وتـــسهيل آليـــة البحث عن الحلِّ الأمثل، وتميّزت النتائج بالدقّة والاعتماد على عدد قليل من الحلول الممكنـــة، وبسرعةِ الوصول إلى القالب النّحوي المُناسب والوحيد للجملة، ومن ثم استنتاج القالب الصرفي الصحيح.

3.2.3.4. المحلّلَ الدّلالي

يقوم المحلّل الدلالي بتفسير الجملة من الناحية الدلالية، ومدى قبولها أو عدم قبولها دلاليّاً وذلك بعد القيام بعدة خطوات، واختيار الإعراب الصحيح لها انطلاقاً من دراسة القواعد الكلية ومعرفة أنواع الدّلالات للكلمة وأنواع المعاني.

- **تصميم المحلّل الدلالي**

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



الشكل (3) المخطط العام للمحلّل الدلالي

تمّ إنشاء بُنى للحقول الدلالية الخاصة بالأسماء والأفعال، والتي تضم حقولاً من اللغة العربية. تمّ بناء هذه البُنى 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 تفريعية 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**

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

- **مراتب الحقول الدلالية**

تمَّ 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 مرتبةً أي رقماً يدلّ على الرابط النحوي، فهناك رابط الفاعلية والمفعولية والجار والمجرور.وغير ذلك. ويتمّ الحكم على الجملة من حيث الصحة الدّلالية أو عدمها من خلال مفاهيم ومدلولات الروابط.

يقوم محرك البحث بالكثير من الإجراءات من أجل الوصول إلى الحلّ المطلوب من خلال إيجاد الروابط المُثلى لبُنى الحقول الدلالية لتكوين الاستدلال والاستنتاج المطلوب، وذلك وفقاً للخطوات التالية:

1. **الفهرسة الضمنية**

تمّت عملية فهرسة ضمنية للحقول الموجودة ضمن البُنى 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 اللُّبس الناجم عن إعراب بعض الكلمات، ومن أجل الحكم على صحة الجملة دلالياً أو عدم صحتها.

2. **ارتباط الحقول مع المحللات**

تمّ ربط الحقول الدلالية مع بعضها على أساس الاستعمال والموقع النحوي، حيث يضمّ كل حقل مجموعةً من الكلمات ذات السمات المشتركة التي تؤهلها لاحتلال موقع معين من الجملة، وتتم هذه العملية بناء على السمات الدلالية الانتقائية التي يتميز بها كل حقل دلالي، فمثلاً الحقل G1 الذي يحمل معنى الأفعال العامة يتميز بالسمات التالية [+فاعل حي إنسان + مفعول به غير

حي] والحقل G7 الذي يحمل معنى الفاعلية يتميز بالسمات التالية [+جسم +حي +إنسان +عاقل] لذلك يتمّ ربط هذين الحقلين.

3. 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.

الاختبار والأداء

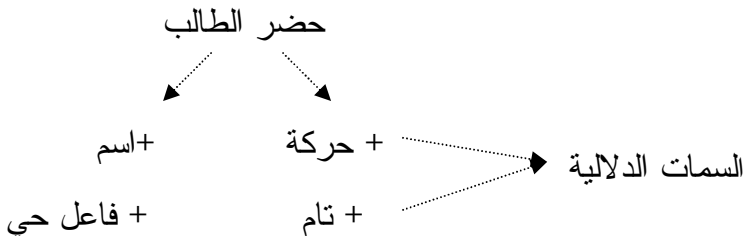
في حال الجملة "حضر الطالب" يقوم محرك البحث بتطبيق مجموعة من القواعد الأساسية على كل كلمة من كلمات الجملة لاختيار المقبولة منها دلاليّاً.

تمّ الوصول الى الحلول الممكنة التالية بعد خضوع الجملة للمحلّلين النحوي والصرفي:

فَعَلَ + فَاعِل ⟵ حَضَرَ الطَّالِبُ A

فَعِلَ + فَاعِل ⟵ حَضِرَ الطَّالِبُ B

قام المحلّل الدلالي بفحص كلمات هذه الجملة من حيث انتماؤها 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:



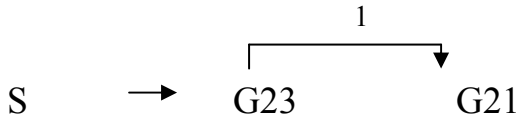
تُمثّل مراتب الروابط النحوية للجملة في الشكل:

S = W1 + W2 حيث أن Wi رمز للكلمة i في الجملة S

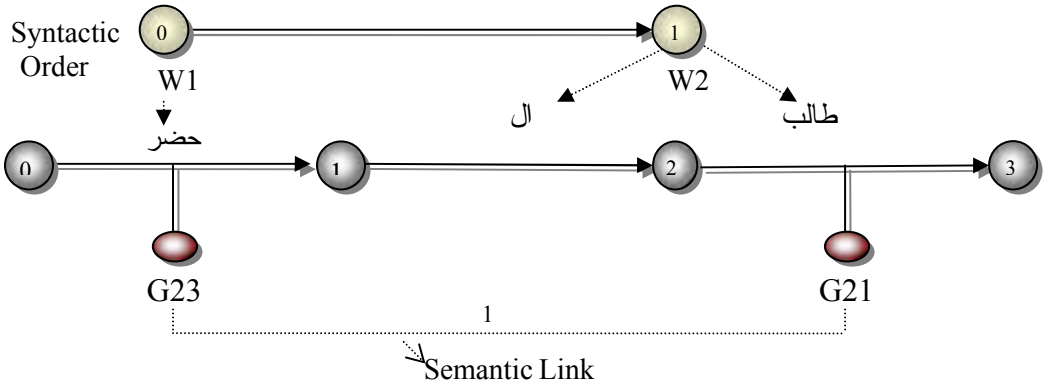
0 1

الرابط بين الحقلين G23 و G21 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]، ومرتبة الرابط الدلالي بين هذين الحقلين هي 1:



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



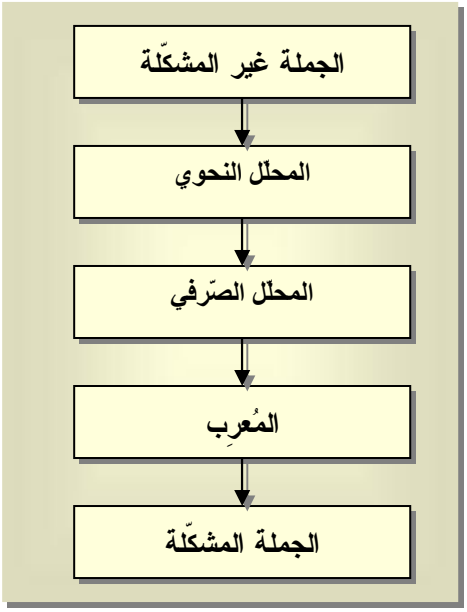
الشكل (4) الروابط النحوية والدلالية لجملة "حضر الطالب"

يتمّ اختيار الجملة A التي تمثل المعنى الدّلالي الصحيح للجملة وهي " حَضَرَ الطَالِبُ ". تميّزت الاختبارات التي خضع لها بمعلومات دقيقةٍ واستجابةٍ سريعة، وبتطبيقها على نماذج متعددة من الجمل تم التوصّل إلى تحديد الجمل المقبولة أو غير المقبولة دلالياً بعد عمليات استدلال واستنتاج قام بها محرّك البحث بالاعتماد على بُنى الحقول الدلالية للوصول إلى الحلِّ الأفضل.

4.2.3.4. المُعرِب

المُعرِب عبارة عن مجموعة من قواعد اللغة العربية، يقوم باختيار القالب المُناسب للجملة، وذلك لتشكيلها بالعلامات المُناسبة. يتمّ إعراب الجملة المُدخلة؛ أي تشكيلها بعد مرورها على المحلّل النحوي، الذي يحلّل الجملة إلى كلمات، ويحدّد طولها ويعمل على إيجاد قوالب هذه الجملة؛ أي الطرائق الممكنة لإعرابها. ثم تخضع إلى المحلّل الصرفي، الذي يحلّل الكلمة إلى

حروف منفصلة، وذلك لتحديد ظهور أو تقدير الحركات على أواخر الكلمات. يُبيّن الشكل (5) مراحل خوارزمية الإعراب.



الشكل (5) خوارزمية الإعراب

- **تصميم المُعرِب**

يتم تصميم بُنى المعطيات المناسبة، حيث تشفّر قواعد الإعراب وتمثّل بشكل بُنى معطيات، وتُجرى عملية تجزئة لهذه البُنى؛ أي يتمّ وضم البُنى 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 بالأسماء المجرورة، وهيكل المجزومات يضمّ القواعد الخاصة بالأفعال المجزومة، وهيكل المبنيات يضمّ القواعد الخاصة بالأسماء والأفعال المبنية.

تُستَخدم الهياكل 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 ملاءمة وربط تهدف إلى إيجاد وحدة ربطٍ بين المُعرِب والمحلِّلين الصرفي والنحوي، حيث لا بدّ من ربط مجموعة بُنى المعطيات مع المحلّل النحوي والصرفي، وذلك لاستدلال واستنتاج الحل الأفضل. تقوم وحدة الملاءمة والرّبط بتبادل البيانات من المُعْرِب إلى المحلل النحوي وبالعكس، حيث يتمّ تطبيق القواعد الإعرابية على الكلمة وذلك بعد معرفة

موقعها النحوي من المحلّل النحوي، ويتمّ تطبيق آلية المحلّل الصرفي على هذه الكلمة للحكم على الحركة من حيث الظهور والتقدير، وعملية التبادل هذه ثنائية الاتجاه.

يقوم محرّك البحث بضمّ آليات التحكُّم التي يفرضها عليه المحلل النحوي من خلال سلسلة من العمليات لتحليل الجملة ومعالجتها، وإيجاد الإعراب المُمكن لها، ثم يُطبِّق مهام المحلل الصرفي الذي يحلّل الكلمة ويُعالجها من حيث السوابق واللواحق وإيجاد الوزن الموافق لها، ويقوم المُعْرِب بربط بُنى المعطيات لتكوين خط الاستنتاج والاستدلال لتشكيل الجملة بالحركات المُناسبة بعد اختيار الإعراب الصحيح لها.

يقوم محرك البحث بتوليد الآليات المُناسبة اعتماداً على قواعد الإعراب المُخزنة في قاعدة المعرفة، واعتَمدت بُنى التحكم تسلسلاً متعاقباً على كل حال متّبِعةً 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، فعندئذ تتمّ العودة إلى الحال السابقة مباشرة، وتتكرّر هذه العملية حتى الوصول إلى الحلّ الصحيح، ومن ثمّ يتمّ الوصول إلى الهدف المطلوب وهو الإعراب والتشكيل الصحيحان للجملة.

الاختبار والأداء

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، وتمَّ 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، وتبيَّن أن الحركة في كلتا الكلمتين ستكون ظاهرة، ويقوم المُعرِب باستدلال واستنتاج الحركات المُناسبة لكل كلمةٍ من كلمات الجملة:

- من أجل الكلمة الأولى "فعل لازم" يتمّ استدعاء هيكل المبنيات الذي يضم قاعدة الجملة الفعلية للفعل الماضي، فالفعل الماضي يأخذ الفتحة، لأنه مبني دائماً على الفتحة إذا صُرّف مع الضمائر المنفصلة مثل "هو".
- من أجل الكلمة الثانية "فاعل" يتمّ استدعاء هيكل المرفوعات الذي يضم قاعدة الفاعل، وتأخذ الضمّة الظاهرة، لأن الفاعل اسم مرفوع يأتي بعد فعل تام معلوم.

تكون النتيجة "حضرَ الطالبُ". أثبتت النتائج فعاليّة النظام المصمّم، والاختبارات التي خضع لها تميّزت بالدقة وسرعة الاستجابة، إذ أمكن تطبيقها على نماذج متعددة من الجمل غير المشكّلة، والتوصّل إلى التشكيل والإعراب الصحيحين للجملة بعد عمليات استدلال واستنتاج قام بها محرّك البحث للوصول إلى الحلِّ الأفضلِ بالاعتماد على القواعد النّحوية في اللغة العربية. تمَّ التوَصُّل من خلال البرمجية التي تمَّ بناؤها إلى إعراب الجملة العربية سواء أكانت فعليةً أم اسميةً. تمَّ تطبيق هذه البرمجية على مجموعة من الجمل وأعطت الإعراب الصحيح لهذه الجمل، وذلك بعد مرورها على مجموعة من القواعد الأساسية.

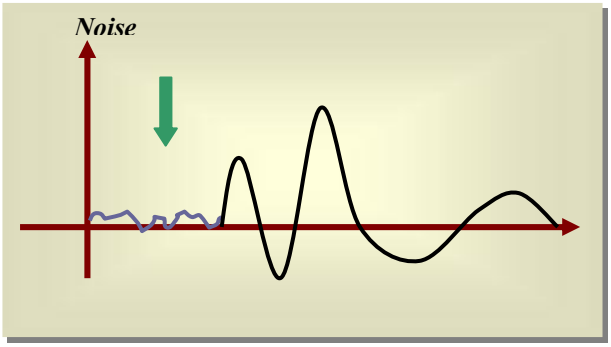
5.2.3.4. المحلّل الصوتي

يقوم المحلِّل الصوتي بالتعرّف على الحروف المنطوقة، اعتماداً على بُنى المعطيات الصوتية، وتطبيق طرائق المعالجة المختلفة للإشارة الكلامية الناتجة، ومن ثمّ التعرّف على الكلمات المنفصلة، وذلك من خلال بناء البرمجيات وتصميم شبكة عصبية ذات انتشار خلفي.

- **تصميم المحلّل الصوتي**

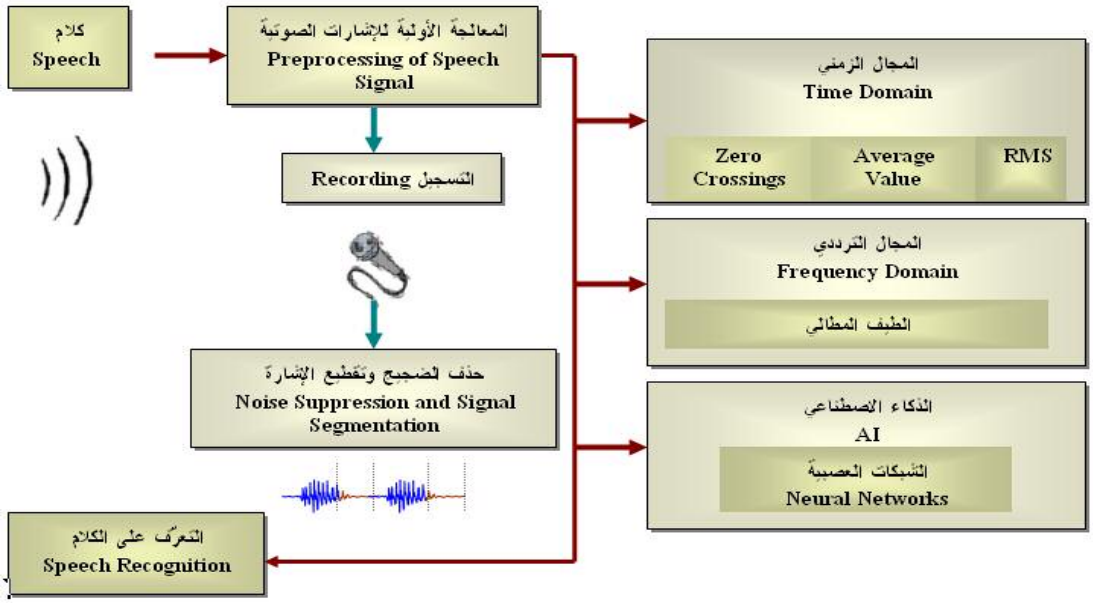
تتم المعالجة الأولية للإشارات الصوتية بعد تسجيل الصوت عبر المايكروفون فتتحوّل التغيّرات في الضغط الجوي إلى تغيُّرات في التيار (الجهد) الكهربائي، ثمّ تُحوّل الإشارة التشابهيَّة إلى رقميّة عن طريق النمذجة؛ وهي أخذ عيّنات من الإشارة (أي قراءات الضغط الجوّيّ) في فتراتٍ زمنيَّةٍ متساويةٍ. يتمّ تسجيل الحروف العربية بوساطة المايكروفون المتّصل بالكمبيوتر، ويؤخذ تردد نمذجة [Hz] 24000 في عملية التسجيل؛ هذا التردد يتبع لكرت الصوت، وأكبر تردد يمكن نمذجته هو نصف تردد النمذجة للملف الصوتي [Hz] 48000، وعدد القنوات واحد Mono وطول القناة 16 Bits.

يتم حذف الضجيج وتقطيع الإشارة في هذه المرحلة، حيث يوجد عدة عوامل تؤثر في تشويش الإشارة الصوتية عند تسجيل الحرف ومنها الوسط المحيط، والترددات العالية ضمن الكمبيوتر والمايكروفون وكرت الصوت، والاختلاف الكبير لأصوات الإشارة الصوتية لكل حرف وغيرها. للحصول على إشارة صوتية صافية للحرف يتمّ فصل الضجيج عن الإشارة الصوتية الأساسية باستخدام ثلاثة معايير رياضية هي معدّل الانزياح ومعدّل المرور الصفري ومعدّل المطال، ويُوضّح الشكل (6) الضجيج الذي يرافق الإشارة الصوتية.



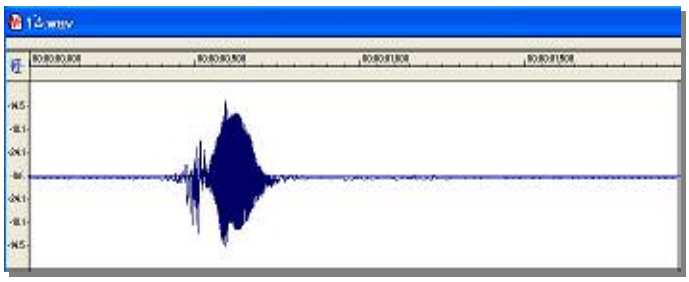
الشكل (6) ضجيج الإشارة الصوتية

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



الشكل (7) معالجة الإشارة الصوتية

يمكن ملاحظة بارامترات المجال الزمني مباشرة من خلال البرمجيات الصوتية Sound Forge 7.0 أو Cool Edit، وهـــي Zero Crossings وAverage Value وRMS و Sample Maximum Value وMinimum Sample Value، أو من البيئـــات البرمجيـــة Matlab أو Visual Prolog من خلال برنامج خاص لهذا الهدف، ومن الضروري ملاحظة أن هـــذه البارامترات تتغير حسب الضجيج والمايكروفون والكمبيوتر وبُعد المتكلم عن المايكروفون أثناء تسجيل الحروف والحال الصحية والنفسية للإنسان وغير ذلك، ويوضّـــح الـــشكل (8) شـــكل الإشارة الصوتية للحرف "ثَ" في هذا المجال.



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

تمّ التوصل الى أنه يمكن فصل الحركة عن الحرف، وذلك بعد إجراء معالجةٍ أوليـــةٍ للإشـــارة الصوتية من حذفٍ للضّجيج وتقطيعٍ وترشيح، تمّ اختبار الشبكة العصبية على الحروف الساكنة والمتحركة والحركات وحروف المدّ وعلى الكلمات، وكانت نسبة التعرّف 100%، وصُـــنِّفَت الحروف الساكنة والمتحركة والحركات وحروف المدّ غير المُدربة مع الحروف المُماثلةِ لهـــا والمتناسبةِ معها. لقد أثبتت النتائج فعّالية الشبكة المُصممة، وذلك بعد تطبيق طرائـــق أفـــرزت معلوماتٍ دقيقةً، وتميّزت بنسبة نجاحٍ عالية من خلال برمجيةٍ تمّ فيها تقليل الأخطاء إلى أدنـــى حد مع اختزال زمن التدريب.

6.2.3.4. المركّب الصّوتي

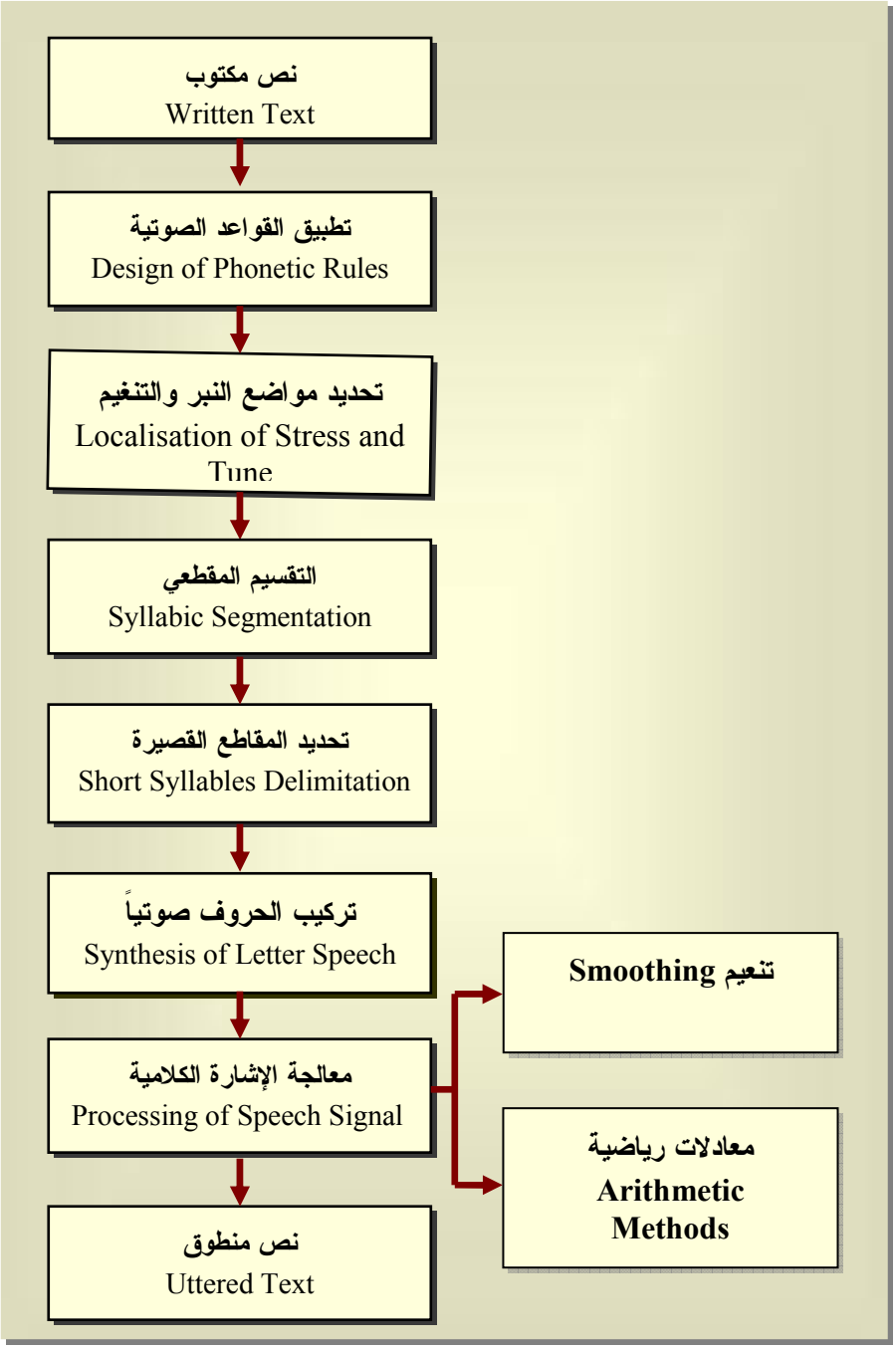
يقوم المركّب الصوتي اعتماداً على البُنى الصّوتية المخزَّنة لتركيب الحروف المتحرِّكـــة مـــن حروفٍ ساكنةٍ مع الحركات، وتركيبِ الكلمات من الحروف المنفصلة.

- **تصميم المركب الصوتي**

يتمّ توليد النصوص المنطوقة بإتباع الخطوات التالية:

- إدخال النص المكتوب والمشكول المُراد نطقه آلياً.
- تحويل النص المكتوب إلى نص صوتي باستخدام القواعد الصوتية.
- تحديد مواضع النبر ونوع أسلوب الجملة؛ أي التنغيم.
- 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 تنعيم واستخدام النوافذ مثل Hamming وGaussian وBlackman أو من خلال إيجاد معادلات رياضية خطية أو غير خطية.
- توليد النص المنطوق.

تمّ توضيح خوارزمية المركّب الصوتي اللغوي في الشكل (9).



الشكل (9) خوارزمية المركّب الصوتي

يتم تصميم بُنى المعطيات الصوتية، حيث تمّ فصل حركة الفتحة عن حرف متحـــرك مفتـــوح، وحركة الضمة عن حرف متحرك مضموم، وحركة الكسرة عن حرف متحرك مكـــسور، وتـــمّ التوصل إلى أن ألف المد عبارة عن فتحة طويلة، وواو المد عبارة عن ضمة طويلة، وياء المد عبارة عن كسرة طويلة، لذا تمّ اختصار حجم بنى المعطيات الصوتية، حيث تمّ تخزين الإشارات

الصوتية للحروف الساكنة Quiescent Letters والحركات الثلاث Short Vowels الفتحة والضمة والكسرة فقط، وكان عدد الملفات الصوتية المخزّنة 31 ملفاً، وتمّت معالجة الإشارة الصوتية للحروف العربية والكلمات في المجال الزمني.

تمّ تركيب الحروف الساكنة مع الحركات الثلاث Short Vowels الموافقة لها، وتمت المقارنة مع الحرف المتحرك، كمثال على ذلك الشكل (10) يُوضّح الإشارة الصوتية للحرف المركّب من إشارة فتحة الثاء " َ" مع الحرف الساكن "ثْ"، والإشارة الصوتية للحرف المفتوح "ثَ".

الإشارة الصوتية للحرف المفتوح "ثَ" الإشارة الصوتية لـ "ثْ" مع فتحة الثاء "َ"

الشكل (10) تركيب الفتحة من "ثَ" مع الحرف الساكن "ثْ"

بيّنت النتائج دقّةَ تركيبِ الحروف المتحرِّكة من حروفٍ ساكنةٍ مع الحركات، وتركيبِ الكلماتِ من الحروفِ المنفصلة، وذلك بعد تطبيق طرائقَ أثبتت فعاليّتَها في عمليةِ تقريبِ الإشارة الصّوتية المركّبة الناتجة من الإشارة الصوتية المتّصلة، وتميّزتْ بنسبةِ نجاحٍ عاليةٍ، وبأقلِّ خطأ مُمكنٍ وبكفاءةٍ مُتَميزةٍ، كما أنّ الاختبارات التي طبِّقت في هذا المجال كانت ناجحةً وأدّت إلى نتائج جيِّدةٍ.

5. خاتمة

من خلال الدّراسة الشّاملة لمعالجة النُّصوصِ العربية، بالاعتماد على بناء المحلّلات النّحوية والصّرفيّة والدّلالية والصّوتية، وإنشاء المركّب الصّوتي الذي يقوم بتحويل النّصوص المكتوبة إلى منطوقة، تُبيّن النّتائج مدى فعالية الأنظمة المصمّمة وسرعة تنفيذها وإمكان تطبيقها على نماذجَ عدّةٍ من الجُمل العربيّة غير المُشكّلة للوصول إلى التشكيل والإعراب الصّحيحين لها، من خلال الدّور الهام الذي يقوم به محرّك البحث في عمليّة الاستدلال والاستنتاج اعتماداً على القواعد النّحوية النّاظمة للّغة العربيّة.

شهدت التسعينيات أول تعامل مع الأنظمة الناطقة بعد ما بات في وقت من الأوقات حلم مُبرمجي الكمبيوتر الوصول إلى طريقة تُمكنهم من التخاطب مع الكمبيوتر، وكان لا بدّ من مواجهة ثلاث قضايا رئيسية؛ هي القضايا المتعلقة باللغة، والمتعلقة بالنطق والكلام، وقضايا الصوت. وكانت هذه الطريقة وسيلة الكفيفين لإدخال النصوص وتنفيذ الأوامر والتحكم في النوافذ، ومن ثمّ سماع كل ما هو معروض على الشاشة، بالإضافة إلى ذلك تُستخدم لمساعدة من يواجهون صعوبة في استخدام لوحة المفاتيح أو الماوس ليتعاملوا مع الكمبيوتر، لذلك تمّ بناء نظام تشغيل صوتي للكفيفين باستخدام تقانات الذكاء الاصطناعي، وكان النظام ناطقاً باللغة العربية، لأن منظومة اللغة العربية تتميز بالانتظام الصوتي، فالقواعد الصوتية العربية تتصف بالاطراد، ويتسم نظام مقاطعها الصوتية ونبرها بالبساطة، وهذه الخصيصة مهمة في توليد الكلام Speech Synthesis وتمييزه Speech Recognition آلياً.

6. توصيات

- الاستمرار في ترقية وتطوير التقانات في مجال معالجة اللغة العربية بكافة أشكالها؛ مكتوبةً أو مطبوعةً أو مسموعةً أو مقروءةً، واستخدامها في تطبيقات الذكاء الاصطناعي.
- تطوير محلِّل دلالي للفهم الآلي لمضمون الجمل والسياق، فضلاً عن الحكم على الجمل بالصحة الدلالية أو عدمها، وذلك بتطوير الحقول الدلالية التي يعتمد عليها المحلل الدلالي.
- تطوير محلِّل صرفي متكامل للغة العربية بحيث يتمُّ التشكيل الكامل للكلمة بدون اللُّجوء إلى معجم حاسوبي، وذلك بإيجاد قوالب صرفية نحوية تولّد جملاً صحيحة وسليمة نحوياً وصرفياً.
- تطوير النظام الآلي لقراءة النصوص العربية، وذلك لإكساب الكمبيوتر مهارة القراءة الصحيحة للمحارف سواء منها المطبوعة أو المخطوطة، ومعالجة الوثائق العربية آلياً، وذلك بربط هذه المعلومات المكتوبة أو المسجلة بالنظام الآلي، بحيث تسهل قراءتها وتحليلها والوصول إليها أو إلى أي جزء منها بطريقة مباشرة أو غير مباشرة.
- تصميم وحدة ملاءمة وربط بين اللغة العربية والبرمجيات العالمية والقياسية، من أجل تواصل اللغة العربية مع مصادر العلوم المعلوماتية، والتوسع في إنتاج التقانات التي تخدم الثقافة العربية، بما يدعم آداب اللغة العربية وفنونها الرفيعة حفاظاً على الهوية العربية وشخصيتها المتميزة.
- تطوير تطبيقات وسائط العرض المتعددة Multimedia من حيث معالجة الوثائق العربية والصور والأصوات مكتوبة أو مطبوعة أو مرسومة أو مسموعة، وذلك لإنشاء بعض المكتبات الالكترونية العربية على الإنترنت.

- تطبيق تقانات معالجة اللغة العربية على المحمولات الإلكترونية، وذلك بتخفيض سعات التخزين المطلوبة، وتخفيض قوة المعالجات الدقيقة.
- تسخير برمجيات اللغة في خدمة ذوي الاحتياجات الخاصة، كإيجاد قارئ وكاتب لغوي متكامل للكفيفين، كذلك تطوير أنظمة تُخَدِّم الصم والبكم من خلال بناء أنظمة ذكية لتصحيح النطق وفهمه وتحليل الصوت والسمع.
- وضع الخطط والبرامج اللاّزمة للتكامل بين البحوث المعلوماتية واللغوية.

7. المراجع

1. أبو الهيجاء أحمد، 2003، تقييس استخدام اللغة العربية في الحاسوب، جامعة العلوم والتكنولوجيا الأردنية، الأردن.
2. أحمد عبد الله، 2000، مرجع في صيانة الحواسيب الشخصية، سلسلة الرضا للمعلومات.
3. الأفغاني سعيد، 1957، في أصول النحو، مطبعة الجامعة السورية.
4. الأنطاكي محمد، المحيط في أصوات العربية ونحوها وصرفها، دار الشرق العربي، بيروت.
5. بحيري سعيد حسن، 1988، نظرية التبعية في التحليل النحوي، مكتبة الأنجلو المصرية.
6. البواب مروان والطيان محمد حسان، 2000، الجلسات العملية لقسم اللغة العربية، منشورات جامعة دمشق.
7. البواب مروان والطيان محمد حسان، أسلوب معالجة اللغة العربية في المعلوماتية (الكلمة – الجملة)، المعهد العالي للعلوم التطبيقية والتكنولوجيا، دمشق.
8. الثعالبي أبو منصور، 1938، فقه اللغة وسر العربية، شركة مكتبة ومطبعة مصطفى البابي الحلبي وأولاده، مصر.
9. جارم علي وأمين مصطفى، 1962، النحو الواضح (للمدارس الثانوية)، دار المعارف، مصر.
10. جطل مصطفى، 1979، نظام الجملة، مديرية الكتب والمطبوعات الجامعية، حلب.
11. حسان تمام، 1973، اللغة العربية (معناها ومبناها)، الهيئة المصرية العامة للكتاب.
12. حساني أحمد، 1993، السمات التفريعية للفعل في البنية التركيبية، ديوان المطبوعات الجامعية، الجزائر.

13. حلمي موسى علي، 1973، دراسة إحصائية لجذور معجم الصحاح (باستخدام الكمبيوتر)، جامعة الكويت.
14. الحملاوي أحمد، 1965، شذا العَرْف في فن الصرف، شركة مكتبة ومطبعة مصطفى البابي الحلبي وأولاده، مصر.
15. حميدي محي الدين، 1991، المنظومة الكلامية (دراسة في فيزياء وبيولوجيا اللغات الشفهية)، معهد الإنماء العربي، بيروت.
16. الخولي محمد علي، 1986، أساليب تدريس اللغة العربية، مطابع الفرزدق التجارية، الرياض.
17. الخولي محمد علي، 1998، دراسات لغوية، دار الفلاح للنشر والتوزيع، الأردن.
18. خياطة محمد أبو اليمن، 2003، إدراك الصوت باستخدام الشبكات العصبيّة، مشروع دبلوم حاسبات، كلية الهندسة الكهربائية والإلكترونيّة، جامعة حلب.
19. الدحداح أنطوان، 1996، معجم تصريف الأفعال العربية، مكتبة لبنان، ناشرون.
20. الراجحي عبده، 1974، التطبيق الصرفي، دار النهضة العربية للطباعة والنشر، بيروت.
21. رضا عبد الوهاب وآخرون، 1999، التعرُّف على الصوت باستخدام الشبكات العصبيَّة، أطروحة تخرُّج، كلية الهندسة الكهربائيَّة والإلكترونيَّة، قسم الحاسبات، جامعة حلب، سوريا.
22. زكار معتصم، 2003، استعمال تكنولوجيا المعلومات في استكشاف ونشر التراث العربي، ورقة عمل مقدمة إلى الندوة الإقليمية حول توظيف تقنيات المعلومات والاتصالات في التعليم برعاية الاتحاد الدولي للاتصالات، دمشق.
23. زكريا ميشال، 1983، الألسنية التوليدية والتحويلية وقواعد اللغة العربية (الجملة البسيطة)، المؤسسة الجامعية للدراسات والنشر والتوزيع، لبنان.
24. زكريا ميشال، 1982، الألسنية التوليدية والتحويلية وقواعد اللغة العربية (النظرية الألسنية)، المؤسسة الجامعية للدراسات والنشر والتوزيع، لبنان.
25. زكريا ميشال، 1984، مباحث في النظرية الألسنية وتعليم اللغة، المؤسسة الجامعية للدراسات والنشر والتوزيع، لبنان.
26. سالم محمد، 2003، النظم المحسبة للاسترجاع الموضوعي باللغة الطبيعية، رسالة دكتوراه، http://www.cybrarians.info/news/salem.htmk.

27. السيد عبد الحميد مصطفى، 1998، المغني في علم الصرف، دار الصفاء، عمان.
28. الصابوني عبد الوهاب، اللباب في النحو، دار مكتبة الشرق، لبنان.
29. الصعيدي عبد الفتاح ويوسف محمد السيد، 1929، الإفصاح في فقه اللغة، مطبعة دار الكتب المصرية، القاهرة.
30. طلبه محمد فهمي وآخرون، 1994، الحاسب والذكاء الاصطناعي، مجموعة كتُب دلتا.
31. عتيق عبد العزيز، 1985، علم المعاني، دار النهضة العربية، بيروت.
32. عطبة عبد الرحمن، 1989، في رحاب اللغة العربية، دار الأوزاعي، لبنان.
33. علي نبيل، 1994، العرب وعصر المعلومات، سلسلة عالم المعرفة.
34. علي نبيل، 2001، الثقافة العربية وعصر المعلومات/ رؤية لمستقبل الخطاب الثقافي العربي، المجلس الوطني للثقافة والفنون والآداب، سلسلة عالم المعرفة، الكويت.
35. علي نبيل، 1987، عالم الفكر (مجلة الحاسوب) مجلد 18، عدد 3، وزارة الإعلام، الكويت.
36. عمايرة خليل أحمد، 1984، في نحو اللغة وتراكيبها عالم المعرفة، جدة.
37. قباوة فخر الدين، 1986، إعراب الجمل وأشباه الجمل، دار الأوزاعي، لبنان.
38. قباوة فخر الدين، 1997، تحليل النص اللغوي، دار الفكر المعاصر، بيروت ودار الفكر، دمشق.
39. قباوة فخر الدين، 1999، المهارات اللغوية وعروبة اللسان، دار الفكر المعاصر، بيروت ودار الفكر، دمشق.
40. قباوة فخر الدين، 1998، تصريف الأسماء والأفعال، مكتبة المعارف، بيروت.
41. مبارك محمد، 2004، فقه اللغة وخصائص العربية، رابطة أدباء الشام.
42. مجموعة اللغة العربية، 1985، دراسة أولية تتناول أنماط الجمل في اللغة العربية، تقرير داخلي رقم 10543، المعهد العالي للعلوم التطبيقية والتكنولوجيا، دمشق، سوريا.
43. مختار عمر أحمد، 1998، علم الدلالة، عالم الكتب، القاهرة.
44. نشاوي محمد أسعد، 1995، موسوعة مبرمجي الحواسيب الشخصية، دار المعاجم.
45. هارون عبد السلام، 1959، الأساليب الإنشائية في النحو العربي، مكتبة المثنى، بغداد ومؤسسة الخانجي، مصر.

46. وافي علي عبد الواحد، 1972، علـــم اللغـــة، دار النهضـــة، مصـــر.
47. Billard Aude, 2002, Genetic Algorithms For Evolving Robot Controllers, Mechatronics Programming Humanoid Robots.
48. Byron Donna, 2002 ،Artificial Intelligence at OSU, Grad Orientation.
49. Caley Richard, Simple Text-to-Speech using Festival, http://search.cpan.org/~rcaley/speech_pm_1.0/Speech/Festival/Synthesiser.pm.
50. Doyle Patrick, 1997, Natural Language, AI Qual.
51. Garcia Carlos Reyes, 2003, Implementing Fuzzy Expert System for Intelligent Buildings.
52. Ghenima.M., 1998, Un Système de Voyellation de Textes Arabes. Université Lumière Lyon2, France.
53. Goldman Jean-Philippe, 2004, Speech Analysis Tools.
54. Grenager Teg, 2003, A Quick Romp Through Probabilistic Relational Models, NLP Lunch.
55. Haipeng Guo, 2003, Algorithm Selection for Sorting and Probabilistic Inference: A Machine Learning-Based Approach, PhD. Thesis Defense Department of Computing and Information Sciences College of Engineering, Kansas State University.
56. Kasravi Kas, 1997, Introduction to Artificial Intelligence, Electronic Data Systems, Troy, Michigan USA.
57. Manning Christopher, 2000, Information pragmatics, A Natural Language Processing Approach, CSLI IAP meeting.
58. Maxey H. David, Speech Synthesis Technology, http://www. mindspring.com/~ssshp/ssshp_cd/ss_spsyn.htm
59. Ouersighni.R., 2002, La Conception et la Réalisation d' un Système d' Analyse Morpho-Syntaxique Robuste pour l'Arabe :Utilisation pour la Détection et le Dignostic des Fautes d' Accord. Université Lumière Lyon2, France.
60. Palmer Martha, 2002, Introduction to Artificial Intelligence.
61. Schmidt Cecil P, 2001, Case-Based Reasoning and Real Time Applications, Department of Computing and Information Sciences, KSU CIS 890: Special Topics in Real-Time AI, Kansas State University.
62. Sook Hee and others, 2002, A Telecommand Verification of Kompsat-2 using Rule-based and Case-based Reasoning, The Space Ops Conference, Oct. 9, 12, 2002.
63. Tecuci Gheorghe, 2003, CS 580 Introduction to Artificial Intelligence, Learning Agents Laboratory, Department of Computer Science, George Mason University.
64. Wiley John and other, 2002, Thought and Language A G E, Inc.

65. 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.

سمر معطي