برامج النطق الآلي أو ما يعرف بـ(مركبات الكلام) وعلاقتها باللغة العربية

م.م.جمانة خالد محمد/كلية التربية للبنات – جامعة بغداد

E:jkalmashhadani@gmail.com

خلاصة:

في ظل هذا التطور الكبير في عالم الحاسوب الآلي(الكمبيوتر)، ودخوله في كل مفصل من مفاصل حياتنا اليومية، كان من المنطقي أن تلتقي اللغة بالحاسوب، وذلك لسبب أساسي وبسيط، هو كون اللغة تجسيداً لما هو جوهري في الإنسان، أي نشاطه الذهني بكل تجلياته، واتجاه الحاسوب نحو محاكاة وظائف الإنسان وقدراته الذهنية، فكان من نتاج ذلك ظهور برامج النطق الآلي أو ما يعرف بـ(مركبات الكلام)، وقد دخلت هذه البرامج في نواحي الحياة المختلفة: التعليم، ومساعدة المعاقين من المكفوفين وضعاف البصر، والهواتف النقالة والحواسيب الآلية، وغيرها من المجالات ويمكن تعريف هذه البرامج بأنها[**برامج حاسوبية تقوم بتحويل أي نص مكتوب إلى كلام منطوق]**.

وهذه البرامج متوافرة في اللغات الأجنبية، أما في اللغة العربية فما زالت في بداية الطريق، ويعود سبب ذلك إلى أن النظم الحاسوبية قد صممت أصلاً للتعامل مع اللغة الانكليزية، واللغة العربية تتمتع بميزات تؤهلها للدخول في عالم الحاسوب الآلي الذي يقوم على نظام رياضي، فاللغة العربية يمكن تحويلها إلى معادلات رياضية يفهمها الحاسوب لامتيازها بنظام تقليب جذور الكلمة وبيان المستعمل منها والمهمل، وخاصية الوزن التي امتاز بها النظام الصرفي العربي.

ونحن بهذا البحث نحاول الدعوة إلى السعي الحثيث إلى ردم هذه الهوة واعتماد الدراسات الحديثة التي تجمع بين علوم اللغة وعلوم الحاسوب، وهو ما يعرف بـ(اللسانيات الحاسوبية ، أو علم اللغة الحاسوبي) .

Abstract:

In light of this great development in the world of computer (PC), and entry in every detail of the joints of our daily lives, it makes sense to meet the language surprise, and cause it to a basic and simple, is the fact that language embodiment of what is essential in humans, ie, his mental in all its manifestations, and the direction of a computer to simulate the functions of man and his abilities of mind, was a product of the emergence of software speech automated or what is known (compounds to speak), has entered these programs in different aspects of life: education, and helping the disabled of the blind and visually impaired, mobile phones and computers , and other fields can be defined as those programs **[software converts any written text to the words uttered]**.

These programs are available in foreign languages while in the Arabic language is still at the beginning of the road, and due to the fact that computer systems were designed originally to deal with the English language, and the Arabic language has features that qualify to enter the world of computers based on the system of sports, Arabic language can be converted into mathematical equations is understood by the computer to turn the franchise system, the roots of the word and statement of user and neglected, and the feature; weight that RPR morphological Arab regime.

We call this research we try to strive to bridge this gap and the adoption of recent studies that combine linguistics and computer science, a so-called (Computational Linguistics, Linguistics or computer) .

المقدمة:

ظهر الحاسوب الآلي أو(الكمبيوتر) في أربعينيات القرن الماضي بإمكانياته القابلة للتطور، مما ينبئ بعصر جديد يوازي عصر الثورة الصناعية. مما دفع المهندسين وعلماء الرياضيات إلى السعي إلى وضع نظرية تبين كيفية التعامل مع الحاسوب الآلي (الكمبيوتر) إلى أن توصل أي.شانون إلى (نظرية المعلومات) التي في أثنائِها استطاعوا التواصل مع الحاسوب الآلي من طريق تحويل الأوامر إلى أرقام يفهمها الحاسوب،ومن هنا كبرت تطلعات العلماء وأمانيهم وسعيهم إلى جعل الحاسوب الآلي (الكمبيوتر) يفهم لغة البشر[١] ((إن المتتبع اليوم للإنجازات والإبداعات في مختلف قطاعات النشاط البشري يلاحظ الدور الذي يلعبه الحاسوب والتقانة ذات العلاقة إلى درجة جعلت المعلوميات تغزو معظم الأنشطة البشرية إن لم نقل كلها))[٢].

وفي ظل هذا التطور الكبير في عالم الحاسوب الآلي (الكمبيوتر) ،ودخوله في كل مفصل من مفاصل حياتنا اليومية ،((كان من المنطقي بل من الحتمي أيضاً،أن تلتقي اللغة بالكمبيوتر، وذلك لسبب أساسي وبسيط ، هو كون اللغة تجسيداً لما هو جوهري في الإنسان ، أي نشاطه الذهني بكل تجلياته ، في الوقت نفسه الذي يتجه فيه الكمبيوتر نحو محاكاة وظائف الإنسان وقدراته الذهنية))[٣]، وللوصول إلى هذا الهدف قامت هناك دراسات في مجالات متعدِّدة، منها[٤]:

١.علوم اللسانيات: من طريق تطوير الدراسات اللغوية، وإخضاعها للمعالجة الرياضية، والمنطقية الإحصائية.

٢.تطوير تكنولوجيا المعلومات بشقيها(العتاد hardware والبرمجيات software) ، وما رافق ذلك من تطور في لغات البرمجة، وأساليب الذكاء الصناعي، والسعي إلى الارتقاء بلغات البرمجة وجعلها تتسم بالقوة والمرونة وجعلها شبيهة باللغات الإنسانية.

في ظل هذا الواقع كان لزاما على اللغة العربية مواكبة هذا العصر والدخول في عالم الحاسوب الآلي، ولكن تطبيق هذا الحلم واجه صعوبات جمَّة، سنعرض لها في بحثنا هذا وسيكون في فصلين:

الأول : تعريف ببرامج النطق الآلي للنصوص أو ما يعرف بـ(مركبات الكلام)، وسنعرض فيه إلى التعريف بهذه البرامج، وآلية عملها، والتطبيقات المستعملة فيها هذه البرامج.

الثاني: بيان علاقة اللغة العربية بالحاسوب الآلي، وسنعرض فيه إلى خصائص ومزايا تؤهل اللغة العربية الدخول في عالم الحاسوب الآلي، والأسس اللغوية التي تعتمد عليها برامج النطق الآلي ، أسباب قلة مشاريع برامج النطق الآلي أو(مركبات الكلام).

وفي الخاتمة سنعرض لأهم النتائج التي توصل إليها البحث .

الفصل الأول

تعريف ببرامج النطق الآلي للنصوص أو ما يعرف بـ(مركبات الكلام) ، وآلية عملها وتطبيقاتها

- تعريف ببرامج النطق الآلي:

لما كان الكلام وسيلة التواصل الأيسر والأكثر استعمالا بين بني البشر، فهو الوسيلة الأسهل لإيصال الأفكار والمعتقدات وتوضيحها ، ومع تطور الحياة في عصرنا الحالي ودخول الحاسوب الآلي في مناحي الحياة، سعى الباحثون والمهندسون إلى جعل سبل التعامل معه أسهل وأسرع من طريق العمل على إنشاء جيل من الحواسيب الآلية الذي يتواصل كلاميا مع المستعمل(٥).

ولتحقيق هذا الهدف ظهرت برامج وتطبيقات متعدِّدة تمكن الحاسوب من نطق بعض الكلمات المسجلة سلفًا ، مثل: البرامج المستعملة في نظم الرد الآلي المعروفة ، أو ألعاب الفيديو(٦).

ولما كانت هذه الطريقة مكلفة و معقدة لم تحقق الهدف المنشود من ورائها ، فقد سعى العلماء والمتخصصون جاهدين إلى إيجاد طرائق بديلة لإنتاج الكلام، فحاولُوا منذ ستينيات القرن الماضي ((إنتاج إشارة تشبه إشارة الكلام بطرق ميكانيكية ، ثم كهربائية وأخيرا حاسوبية رقمية وذ لك اعتمادًا على فهم آلية إنتاج الكلام عند الإنسان ونمذجة هذه العملية، فظهرت مركبات الكلام المتعددة وللغات عديدة.))(٧).

واستعملت هذه التقنية في نطق النصوص الداخلة إلى الحاسوب وإخراجها على ((شكل كلام منطوق بصوت بشري وبشكل فوري))(٨) ، وهذه التقنية تزداد أهميتها يوما بعد يوم، ولا سيما في الأعمال التي تحتاج إلى تحديث بشكل مستمر، فبإمكان هذه البرامج تحويل الأخبار والطقس وحسابات البنوك، والأسهم، والبريد الإلكتروني إلى كلام منطوق بصوت بشري يوفر على المستعمل الوقت والعناء في البحث وقراءة البيانات(٩).

مما سبق يمكننا القول: إن برامج النطق الآلي أو ما يعرف .(مركبات الكلام) هي:

[برامج حاسوبية تعمل على إمكانية تحويل أي نص مكتوب إلى كلام منطوق]

- آلية عمل برامج النطق الآلي :

من التعريف الإجرائي السابق لهذه البرامج يتضح مدى صعوبة هذه العملية وتعقيدها (أي عملية تحويل النص إلى كلام)، فمن المعلوم أن عملية الكلام التي يقوم بها الإنسان بشكل مستمر هي عملية معقدة تقوم بالأساس على تحويل الهواء الخارج من الرئتين (الزفير) إلى موجة صوتية في أثناء التضييق

عليه أو إغلاق الطريق أمامه أو فتحه لإنتاج أصوات ذات صفات مختلفة تؤلف بمجموعها كلمات اللغة المنطوق بها.

ولما كان((نظام النطق عند الإنسان هو نظام ميكانيكي تحكمه مجموعة عضلات (اللسان، والحنك، والحبال الصوتية، وغيرها)، ونتيجة لعطالة هذا النظام وحركته البطيئة نسبياً يمكن عدّه ساكناً في مدّة تراوح بين ١٠ و ٢٠ ميلي ثانية[١٠]، مما يسمح بنمذجته[١١]، بعدّة موسطات في هذه المدة. تسمح هذه النمذجة عملياً من تقليد إنتاج مختلف الأصوات))[١٢].

وتختلف مركبات الكلام أو برامج النطق الآلي بعضها عن بعض اعتماداً على طريقة نمذجة نظام النطق، منها[١٣]:

١.المركبات المفصلية الموجية articulators modulation[١٤].

٢.المركبات البواني أو المشكِّلات formant synthesizers[١٥].

٣.المركبات التي تعتمد تقانة التنبؤ الخطي linear predictive coders[16].

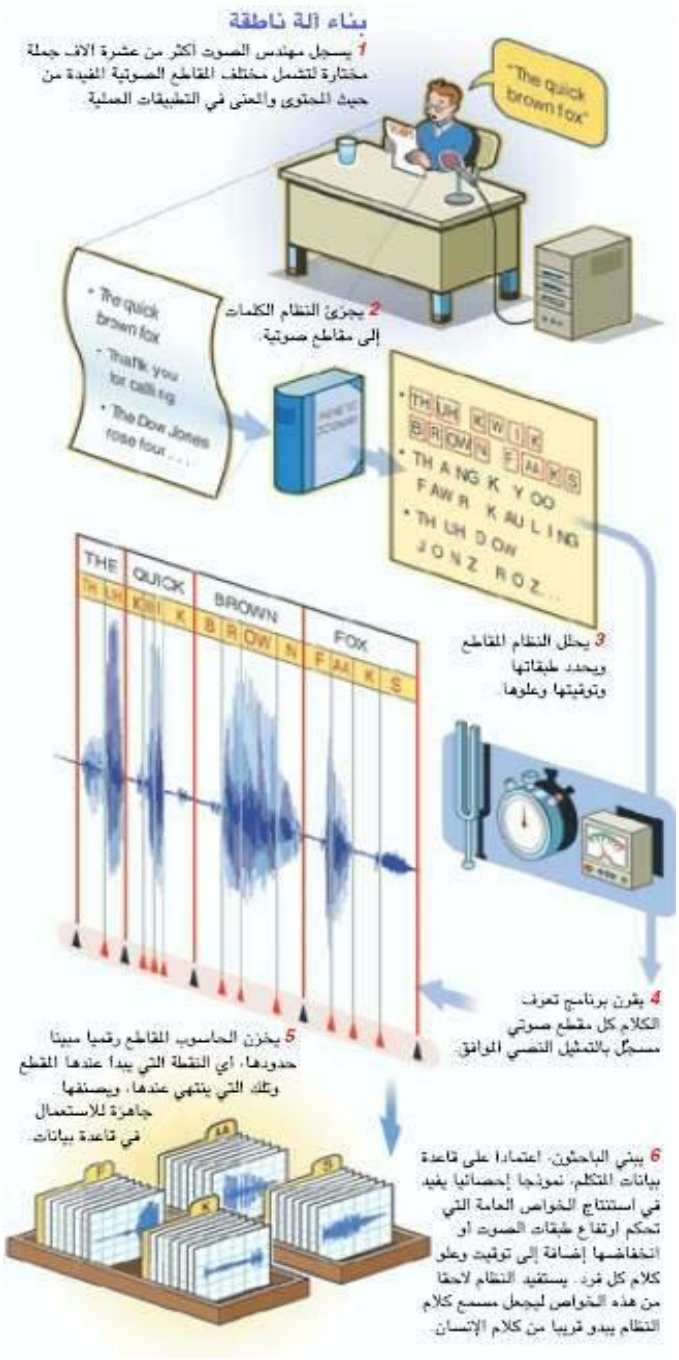
إلا أن آلية عمل هذه البرامج واحدة فهي تقوم على مراحل[١٧]:

١. يُسجِّلُ المهندسون والمبرمجون يُسجلون ذلكَ جملاً كلامية بأصوات عدد من الأشخاص الذين لا يملكون لكنة أو لهجة خاصة. في حجرة تسجيلاً، ويقوم الشخص بالقراءة بصوت مرتفع أكثر من عشرة آلاف جملة، وهو ما يعادل نحو خمس عشرة ساعة تسجيل.

٢. يحول البرنامج كلمات النص المنطوق بها إلى سلسلة من المقاطع الصوتية وذلك باستعمال قاموس ألفاظ، وهو المرجع الذي يضم المقاطع الصوتية التي تؤلف كل كلمة. يحفظُ البرنامج الملامح الخاصة ويتذكّرها بكل مقطع صوتي، مثل المقاطع التي سبقته والتي أتت بعده أكان هذا المقطع في بداية الكلمة يقعُ أم نهايتها أم في الجملة؟

٣. ما إن تتم معالجة النص حتى يحلل برنامجنا التسجيل الصوتي، وذلك بهدف قياس ثلاث صفات مميزة هي(طبقة الصوت،وتوقيته،وعلوّه). التي تسمى العروضprosody ، ومعرفة هذه الصفات لكلّ واحد من المقاطع تساعدنا على اختيار أي مثال يجب استعماله لتركيب عبارة معطاة.

٤. وبعد ذلك، وباستعمال تقنيات تعرّف الكلام، وهي برامج إملاء تترجم الكلام إلى نص. يقرن البرنامج الحاسوبي كل مقطع صوتي مسجل بجزئه النصي المقابل. وهكذا، مع وجود النص والصوت معا تكون برمجيتنا قادرة على تحليل كل تسجيل وهيَ تُعيّن كذلكَ بدقة الحدود التي يبدأ بها كُلُّ مقطعٍ أو ينتهي عندها. إن هذه وفي الشكلين (١) و (٢) مخططان تفصيليان يوضحان خطوات عمل البرنامج.


بناء آلة ناطقة
1 يسجل مهندس الصوت أكثر من عشرة آلاف جملة مختارة لتشمل مختلف المقاطع الصوتية المفيدة من حيث المحتوى والمعنى في التطبيقات العملية
"The quick brown fox"
2 يجزئ النظام الكلمات إلى مقاطع صوتية
3 يحلل النظام المقاطع ويحدد طبقاتها وتوقيتها وعلوها
4 يقرن برنامج تعرف الكلام كل مقطع صوتي مسجل بالتمثيل النصي الموافق
5 يخزن الحاسوب المقاطع رقميا مبينا حدودها، أي النقطة التي يبدأ عندها المقطع وتلك التي ينتهي عندها، ويصنفها جاهزة للاستعمال في قاعدة بيانات
6 يبني الباحثون، اعتمادا على قاعدة بيانات المتكلم، نموذجا إحصائيا يفيد في استنتاج الخواص العامة التي تحكم ارتفاع طبقات الصوت أو انخفاضها إضافة إلى توقيت وعلو كلام كل فرد. يستفيد النظام لاحقا من هذه الخواص ليجعل مسمع كلام النظام يبدو قريبا من كلام الإنسان
THE
QUICK
BROWN
FOX

مخطط رقم(١)

52°, winds 10 mph
FIFTY TWO DEGREES, WINDS TEN MILES PER HOUR
WINDS?
WINDS?
FIFTEE TOO DUH-GREEZ...
2 يختار النظام القطع الصوتي ذا الطبقة والتوقيت والعلو الأكثر ملاءمة لتركيب كل جزء من الجملة المراد نطقها.
3 بعد اختيار القاطع الصوتية ووصفها واحدا إثر آخر، يزيل النظام الانقطاعات الناجمة عند وصف القاطع المختلفة. فمثلا، عندما يختلف مقطعان متجاوران في الطبقة يُجري البرنامج تعديلات بسيطة كأن يحني الطبقة قليلا نحو الأعلى أو نحو الأسفل عند أطراف كل عينة وذلك من أجل تلاؤم جيرانها.
4 وأخيرا «يقرأ» النظام النص المعروض
"Fifty two degrees, winds ten miles per hour"

مخطط رقم (٢)

مما تقدم يظهر لنا صعوبة هذه البرامج وتعقيدها، وحاجتها إلى بذل الجهد لتحقيق انتصار تقاني وحلم إنساني، تعود بدايته ومحاولاته الأولى إلى أواخر القرن الثامن عشر حينما صنع العالم الهنغاري (ولفانج فون كمبلن) أول آلة ناطقة، استعمل فيها مجموعة معقدة من القصبات والصفارات وعُلب الطنين لتولِّد كلمات بسيطة، واستطاع إنتاج بضع كلمات وجمل قصيرة بمعالجة أكياس جلدية منفوخة ،ضخت الهواء عبر صندوق خشبي وقطعة مطاطية بشكل جرس كانت تمثِّلُ الفم[١٨].

الإجرائية أساسية لأنه ما إن تُحدَّدُ المقاطع الصوتية وتُسمَّى حتى يتمكن البرنامج من تصنيفها بدقة ضمن قاعدة بيانات[١٩] يمكن البحث فيها.

٥.حين إدخال النص ليقوم برنامج النطق الآلي بإخراجه على هيئة صوت بشري يحول البرنامج النص المكتوب إلى مقاطع صوتية، وهذه مهمة أصعب مما تبدو عليه أول وهلة، فالجملة تحوي علامات ترقيم واختصارات يجب لفظها، إذ يجب البدء بترجمة النص إلى سلسلة الكلمات التي يجب على مركب الكلام أن يلفظها.

٦. وبعد تعيين سلسلة الكلمات التي يجب قراءتها، يجب أن يكتشف النظام الأسلوب الذي يجب أن يتبعه حين القراءة. إذ تعتمد قراءة بعض الكلمات على موقعها في الكلام[٢٠].

٧. في هذه المرحلة يصبح نظام تركيب الكلام جاهزاً لتحويل الكلمات إلى مقاطع صوتية.ثم بعد اختيار المقاطع الصوتية، ورصفها واحداً بعد الآخر، يزيل النظام الانقطاعات الناتجة عن رصف المقاطع المختلفة جنبا إلى جنب، فمثلا حينما يختلف مقطعان متجاوران في الطبقة يجري البرنامج تعديلات بسيطة، كأن يحني الطبقة قليلا نحو الأعلى أو نحو الأسفل عند أطراف كل عيّنة؛ وذلك لتلائم جيرانها من المقاطع.

٨. ثم ينطقُ البرنامج أخيراً النصَّ المطلوب.

وتطورت المحاولات، وتتالت على مرِّ السنين حتى وصلت إلى إنتاج برامج النطق الآلي أو ما يعرف بـ(مركبات الكلام) التي تعتمد التقنية الرقمية في إصدارها للصوت.

- تطبيقاتها:

بعد هذا العرض المبسط لآلية عمل برامج النطق الآلي او ما يعرف بـ(مركبات الكلام) صار لزاماً علينا بيان مجالات تطبيق هذه البرامج ليتضح للقارئ الكريم أهميتها وتأثيرها في جميع جوانب الحياة ، فهي تدخل في مجالات متعدِّدة، منها:

١.في مجال التعليم إذ توفر هذه البرامج فرصة لتعليم اللغات من طريق ادخال الكلمات والجمل وسماعها، وتعلم كيفية نطقها، وأنها تستعمل في القواميس الناطقة التي تظهر معنى الكلمة مع امكانية تعلم كيفية نطقها ببرنامج النطق الآلي.

٢.في مجال مساعدة المعاقين: وذلكَ من طريق إمكانية تحويل الدروس والمحاضرات إلى تسجيلات صوتية تفيد فئة المكفوفين، ويمكن استعمال هذه التقنية لمساعدة ذوي الاحتياجات الخاصة من الصم والبكم في تمكنهم من التواصل مع المجتمع حولهم في أثناءِ كتابة ما يريد باستعمال هذه البرامج ويقوم البرنامج بالنطق.

٣.في الهواتف النقالة والحواسيب الآلية: إذ يمكن بهذه التقنية تمكين الهاتف من نطق اسم المتصل مما يوفر جهدا على كبار السن وضعاف البصر، وتستعمل هذه التقنية في قراءة التقارير والمعلومات لرجال الأعمال ومن يحتاج عملهم إلى متابعة كل المستجدات.

٤.في مجال الخدمات: الاستعمال في أجهزة الرد الآلي وفي المطارات وفي أجهزة الطيران إذ تعلم قائد الطائرة بكل تغير يطرأ على لوحة قياس السرعة وغيرها.

نخلصُ مِمَّا سبق ظهور مدى أثر هذه البرامج في حياتنا غير أنه من المؤسف أن هذه التطبيقات أكثرها متوافر باللغة الانكليزية وغيرها من اللغات الأخرى، أما اللغة العربية فإن المحاولات ما زالت في أول الطريق وهي لا تعدو أن تكونَ محاولات فردية تعود لشركات خاصة مما يجعلها باهظة الثمن وغير متاحة للأفراد من أمثال شركة صخر للبرمجيات، وشركة IBM، ولا تعدو محاولات بعض الدول العربية إنتاج مثل هذه البرامج محاولات معدودة محصورة في داخل المراكز البحثية، وغير متاحة لمن يرغب فيها.

الفصل الثاني

اللغة العربية وعلاقتها بالحاسوب الآلي

لمَّا دخل الحاسوب الآلي في عصرنا الحالي جميع مناحي الحياة، ولأن اللغة ليست بمعزل عن محيطها، فقد ظهرت دراسات تحاول الجمع بين العلوم اللغوية وعلوم الحاسوب الآلي، ومن المعلوم أن إخضاع اللغة (أي لغة) للتحليل والمعالجة من طريق الحاسوب الآلي ليس أمرا هيناً سهلاً، ولا طريقاً مفروشاً بالورود والرياحين، وإنما هو طريق صعب وشاق تكتنفه عقبات كثيرة وهو أصعب وأشق على اللغة العربية أكثر من أي لغة أخرى.

ويعود السبب في ذلك إلى أن النظم الحاسوبية قد صممت أصلا للتعامل مع اللغة الانكليزية، و((إن استيعاب اللغة العربية في نطاق الانكليزية هي عملية خاطئة))[٢١] ، ويرجع الخطأ لسببين رئيسينِ[٢٢].

أحدُهُما: اختلاف اللغة العربية عن اللغة الانكليزية، لأنَّ العربيَّةَ أعقد اللغات السامية، والإنكليزيَّة أسهل لغات الفصيلة الهندو أوربية، إذن فهما على طرفي نقيض.

والآخر: اختلاف نظام الكتابة في كلا اللغتين فاللغة العربية في معظم أمورها(نحوها، وصرفها وكتابتها) أعقد من اللغة الانكليزية.

لذلك كانت((عملية استيعاب اللغة في إطار الانكليزية بمثابة منطق معكوس أو محاولة مستحيلة لاستيعاب الأعقد في نطاق الأبسط))[٢٣].

وللنهوض بواقع اللغة العربية، واعتمادها أساساً في البرامج الحاسوبية، سنحاول في هذا الفصل، بشيء من الاختصار بيان نقاط تتعلق بعلاقة اللغة العربية بالحاسوب الآلي، من مثل:

- خصائص ومزايا تؤهل اللغة العربية للدخول في عالم الحاسوب الآلي.
- الأسس اللغوية التي تعتمد عليها برامج النطق الآلي أو (مركبات الكلام)
- أسباب قلة مشاريع برامج النطق الآلي أو (مركبات الكلام) في اللغة العربية.

– خصائص ومزايا تؤهل اللغة العربية الدخول في عالم الحاسوب الآلي :

قامت أولى الدراسات اللغوية على أساس رياضي فقد وضع الخليل بن احمد الفراهيدي (ت١٧٥هـ) مفردات معجم العين على أساس رياضي ويتضح ذلك من قوله ((اعلم أن الكلمة الثنائيَّةَ تتصرَّف على وجهين نحو : قَدْ، دَقْ، .... والكلمة الثلاثيَّةُ تتصرَّف على ستة أوجه، وتُسمَّى مَسدُوسة، وهي نحو : ضرب ضبر، برض بضر، رضب ربض،. والكلمة الرباعية تتصرَّف على أربعة وعشرين وجهاً وذلك أن

حروفها وهي أربعةُ أحرف تُضرَب في وجوه الثلاثيِّ الصحيح وهي ستَّةُ أوجه فتصير أربعة وعشرين وجها ، يُكتب مُستَعمَلُها . ويُلغى مُهمَلُها، وذلك نحو عبقر نقول منه . عقرب، عبرق، عقبر،....

والكلمة الخماسية تتصرَّف على مئة وعشرين وجها ، وذلك أن حروفها، وهي خمسة أحرُف تُضرَب في وُجُوه الرُّباعيِّ، وهي أربعة وعشرون حرفا فَتصير مئة وعشرينَ وَجْها يُستَعمَل أقَلُّه ويُلغى أكثره . وهي نحو: سَفَرجل، سفرلج، سَفجرل، سجفرل،.... وهكذا))[٢٤]، وقد أعطى هذا الأساس الرياضي للغة قدرة هائلة لا متناهية لإنتاج المفردات والمصطلحات والجمل مما جعل منها لغة ((عالية الطوعية للاستعمال العلمي ووضع مصطلحه))[٢٥]، فهناك أوجه تشابه كثيرة بين اللغة، وعلم الرياضيات فـ ((الرياضيات ذات طابع عقلي رمزي تجريدي كما هو معلوم . واللغة مبنية على رموز . وفيها من التجريد الذهني ما هو بيّن واضح لكل باحث . ثم إن اطراد اللغة يشبه اطراد النظر الرياضي))[٢٦]، لكن هذا لا يعني عدم وجود اختلاف بينهما فالاختلاف قائم على أمور متعدِّدة أهمها؛ ((أن اللغة رموز تجريدية ذات مرجعية موجودة ملموسة أو مفهومة . يعرفها كل ناطق باللغة . ولذلكَ يجد كل ناطق باللغة أمره ميسوراً، وهو يريد أن يتحدث عن أي أمر، في حين أن الرياضيات تتحدث عن قوانين تضبط مسائل في الحياة، ولا يعرف الصلة بين هذه القوانين والمسائل التي تنضبط بمقتضاها إلا المتخصص))[٢٧]، هذا الأمر مكن اللغة العربية من الدخول في مجال الحاسوب الآلي، بوصفها ((لغة خوارزمية من حيث النحو والصرف والاشتقاق والكتابة،... . بمعنى أنه يمكن وضع قواعد رياضية تحدد السلوك اللساني في العربية))[٢٨]، وفيما يأتي بيان لبعض خصائص اللغة العربية وميزاتها:

*من المعلوم أن اللغة العربية من أقدم اللغات المنطوق بها في عصرنا الحاضر، وهي تعد من اللغات الأصيلة بسبب كون((أصولها تجذَّرت ونبتت كلها في أرض الجزيرة العربية وبعض تخومها))[٢٩]، فكان هذا سببا في تميزها بـ ((قواعد ثابتة تقريباً في نطق أصوات حروفها وكتابة ألفاظها، عدا شواذ معروفة لعدد محدود جداً،... .ولندرة هذه الشواذ يمكن حفظها من دون صعوبة تذكر))[٣٠].

*قِدَمُ اللغة جعلها غنية بالمعاني والمباني والمترادفات، مما يجعلها لغة لها طواعية لتشكيل العبارة العلمية وصياغة المصطلحات وتوليدها للتعبير عن المفاهيم العلمية الجديدة[٣١].

*يمتاز النظام الصرفي العربي بخاصية الوزن الذي يمكن به اشتقاق عدد كبير من الكلمات بصيغٍ مختلفة من الجذر نفسه لكلٍّ منها معنى مختلف ، فمن معرفة (الجذر + الوزن) للكلمة العربية يمكن تحديد أو إدراك معنى الكلمة أو إدراكها، فهي طريقة يمكن بها وضع قواعد للغة العربية يمكن استعمالها في حوسبة اللغة ، مثال ذلك الوزن (استفعل) يمكن التعبير عنه بمجموعة الأوامر الآتية :

أضف السابقة (استَ) إلى الجذر (ف،ع،ل).
أضف فتحة إلى عين الفعل.

أضف فتحة للام الفعل (هذه قاعدة غير لازمة لأنها خاصة ببناء الماضي وهي من ميدان النحو). ويمكن إضافة قاعدة أخرى تربط الوزن بما يدل عليه من المعنى[٣٢].

*يمتاز النظام الصوتي للغة العربية بوجود قواعد ثابتة تحكمه، فلكل حرف في العربية رمز صوتي لا يشذ عنه إلا بوجود قانون محدد، وهو ما كتب عنه اللغويون العرب منذ عصر الخليل بن احمد الفراهيدي(ت١٧٥هـ) إلى عصرنا الحاضر، إذ يمكن ترجمة هذه القواعد إلى قوانين يمكن استعمالها في حوسبة اللغة العربية[٣٣].

هذه هي أهم مميزات اللغة العربية وخصائصها التي تؤهلها للدخول في عالم الحاسوب الآلي .

- الأسس اللغوية التي تعتمد عليها برامج النطق الآلي :

لما كانت برامج النطق الآلي أو (مركبات الكلام) تسعى إلى تحويل النص المكتوب إلى صوت منطوق يفهمه السامع، فإن هذه البرامج تعتمد اعتماداً مباشراً على القواعد الصوتية للغة التي ينطق بها البرنامج، هذه القواعد هي :

١.تحديد واضح لأصوات اللغة المنطوق بها، ونقصد بقولنا الأصوات دون الحروف، ذلك أن الحرف الواحد أو الرمز الكتابي، قد ينطق بصور مختلفة تبعا لموقعه في الكلمة، لذلك وجب إحصاء الأصوات المنطوق بها، وقد أحصى علماء اللغة القدامى أصوات اللغة العربية المنطوق بها إذ بلغت عندهم اثنان وأربعون صوتاً[٣٤]. ومن المعلوم أن الحروف في اللغة العربية عددها تسعة وعشرون حرفاً، أي أن هناك أحد عشر صوتاً منطوقا بها ليس لها رمز كتابي محدد .

٢.وصف كل صوت من هذه الأصوات بشكل مفصل من حيث المخرج، والصفة، والآلية التي يحدث بها الصوت، وهو ما يعرف بـ (ميكانيكية النطق)[٣٥]، إذ في أثناءِ هذه الدراسة وصف شكل الموجة الصوتية للصوت المنطوق به من معرفة التردد[٣٦]، والشدة[٣٧]، والزمن[٣٨]، وتعتمد هذه المرحلة على المختبرات الصوتية وأجهزة قياس الصوت الحديثة[٣٩].

٣.وضع رمز كتابي لكل صوت من هذه الأصوات وهو ما يعرف بـ (الكتابة الصوتية)[٤٠]، وهي شبيهة بالكتابة العروضية التي ابدعها الخليل بن أحمد الفراهيدي .

٤.اعتماد نظام الصوت العربي المعتمد في تركيب الكلمات ، من أمثال عدم السماح بورود أكثر من صامتين متتاليين وهو ما يعرف عند اللغويين العرب بعدم التقاء الساكنين، وهذا يعني حتمية وجود صائت لكل صامت إما سابق له وإما لاحق به، وغالباً ما يقع الصامت بين صائتين، فالصوائت منتشرة بكثرة في الكلام العربي على الرغم من قلة عددها[٤١].

٥.اعتماد التراكيب المقطعية الخاصة باللغة العربية وهي تراكيب مقطعية محدودة وذات سمة مميزة، فكلها تبدأ بصامت متبوع بصائت . وهذا عكس ما هو معروف في اللغة الانكليزية، مثلا، إذ يمكن فيها أن يبدأ المقطع أو ينتهي بصامت أو صامتين أو ثلاثة صوامت. هذا الأمر جعل من عملية توليد الأصوات أو التعرف الآلي عليها سهلاً؛ لأن تراكيب المقاطع محدودة[٤٢].

هذه هي أهم الأسس اللغوية التي تعتمد عليها برامج النطق الآلي أو مركبات الكلام في بناء نظامها الخاص، وهي أسس يسهل التعامل معها حاسوبياً إذا ما قورنت مع اللغات الأخرى كالإنكليزية والسويدية وغيرها.

- أسباب قلة مشاريع إنتاج برامج النطق الآلي أو (مركبات الكلام) في اللغة العربية:

تعود أسباب قلة مشاريع إنتاج برامج النطق الآلي أو(مركبات الكلام) إلى عدة أسباب منها:

أ. أسباب ترتبط بالجامعات وخططها ، وهو خارج نطاق بحثنا.

ب. أسباب ترتبط بالبحث اللغوي العربي ، وهو ما يمس موضوع بحثنا ومن بين هذه الأسباب:

١. أن فكرة إخضاع أصوات اللغة العربية لدراسة معملية في مختبرات صوتية يمكن من طريقها بيان مخرج الصوت وصفته، تعد من الدراسات الحديثة غير الشائعة في جامعاتنا العربية، ولم تتعدَّ حدود الدراسات الأكاديمية، ومشاريع البحوث التي قام الدكتور منصور الغامدي[٤٣]، في معهد بحوث الحاسوب الآلي والالكترونيات في مدينة الملك عبد العزيز للعلوم التطبيقية، وكتاب الدكتور سمير استيتية (الأصوات اللغوية– رؤية عضوية ونطقية وفيزيائية).

٢. اختلاف علماء اللغة المحدثين في تعريفهم للمصطلحات الحديثة ذات الصلة بالصوت من أمثلة: الفونيم، والمورفيم، والالفون، والصويته، والصوتيم، وغيرها، ويرجع السبب في ذلك إلى أن هذه المصطلحات ظهرت في الدراسات اللغوية الحديثة على يد الغربيين، وقد حاول علماء اللغة المحدثون، اعتمادها في أثناء ترجمتها، ومحاولة إخضاع القواعد الصوتية العربية بما يتوافق وهذه المصطلحات.

٣. اعتماد الألف بائية العالمية(IPA) ووصفها للأصوات، وضعف الاهتمام بوصف الأصوات العربية بأسلوب الألف بائية العالمية، وعدم اعتماد آلية صحيحة للقيام بذلك، فلا يتعدى الأمر البحوث والدراسات الأكاديمية حبيسة المكتبات[٤٤].

٤. تعد فكرة دمج الدراسات اللغوية والحاسوب الآلي من الأمور المستحدثة في الدراسات اللغوية العربية، ولم تدخل بعد بشكل واسع في الجامعات العربية.

كل هذا وغيره من الأسباب وراء قلة مشاريع حوسبة اللغة العربية، ومن بينها مشاريع إنتاج برامج النطق الآلي أو (مركبات الكلام).

الخاتمة

بعد هذا العرض المبسط لبرامج النطق الآلي أو(مركبات الكلام) وعلاقتها باللغة العربية، نتوصل إلى النقاط الآتية:

١. إن هذه البرامج لها أثر بالغ في الحياة العملية في عصرنا الحاضر، في أثناء تطبيقاتها التي تدخل في جميع مناحي الحياة.

٢. إن اللغة العربية لغة أصيلة قوية، جذوروها ثابتة يمكنها الدخول في مجال علوم الحاسوب الآلي بشكل قوي لما تمتاز به من نقاط قوة تميزها من غيرها من اللغات.

٣. السعي إلى فتح المجال أمام فروع الدراسات الحديثة التي تعتمد على الجمع بين أكثر من تخصص، من أمثلة ذلك تخصص اللسانيات الحاسوبية، أو علم اللغة الحاسوبي، والتي تجمع بين الدراسات اللغوية وعلوم الحاسوب الآلي، والسعي إلى اعتماد هذه التخصصات في الجامعات العربية للنهوض بواقع اللغة العربية.

هوامش البحث

(١) ينظر: مقدمة في نظرية المعلومات:١٤

(٢) استخدام اللغة العربية في المعلوماتية:١١٨

(٣) العرب وعصر المعلومات:٣٤٥

(٤) ينظر:العرب وعصر المعلومات:٣٤٥

(٥) ينظر: تحليل الكلام وتركيبه- الموسوعة العربية:المجلد ٦ ص ١٤٣.

(٦) ينظر: حواسيب محاورة –العلوم - عدد يوليو(اغسطس) ٢٠٠٩ المجلد٢١

(٧) اللغة العربية والحاسوب – المنظمة العربية للتربية والثقافة والعلوم: ٤

(٨) الناطق العربي الآلي :٣

(٩) ينظر: بحث الناطق الآلي:٣

(١٠) الملي ثانية =١/١٠٠٠ من الثانية.

(١١) النمذجة (sampling):وهي طريقة أخذ العينات وتعد الأساس لتحديد معدل أخذ العينات المناسبة للاشارة المعينة، ولها أهمية كبيرة في معالجة الإشارات ونظرية الاتصال .(ينظر: MODERN DIGITAL AND ANALOG COMMUNICATION SYSTEMS.p251.

(١٢) تحليل الكلام وتركيبه – الموسوعة العربية:المجلد ٦ ص١٤٣.

(١٣) ينظر:تحليل الكلام وتركيبه- الموسوعة العربية:المجلد٦ ص١٤٣.

(١٤) المركبات المفصلية الموجية: وهي تعتمد على محاكاة ميكانيكية للسبيل الصوتي وحركة العضلات المشاركة في عملية إنتاج الكلام. يُقرّب شكل السبيل الصوتي بسلسلة نحو عشرين مقطعاً اسطوانياً منتظماً، ويمكن الحصول على مساحة كل من هذه المقاطع بالاستعانة بالتصوير الشعاعي، أو حسابها من مواقع المفاصل (العضلات) المعنية وهي: اللسان والشفاه والحنك . أو من تحليل الإشارة الكلامية الناتجة باستعمال خوارزميات متطورة. إلا أن هذه النمذجة لا تعطي كلاماً مركباً جيداً.(ينظر:المصدر السابق نفسه)

(١٥) مركبات البواني: وهي طريقة كهربائية تعتمد على نمذجة السبيل الصوتي بمجموعة مرشحات تمرير حزمة سهلة من الدرجة الثانية.في الحقيقة ، تتميز الاستجابة الترددية للسبيل الصوتي بوجود ترددات تتمركز عندها الطاقة نسمي هذه الترددات بالبواني. توافق ترددات الطنين لمرشحات تمرير الحزمة ترددات هذه البواني. ومع أن هذه الطريقة طبيعية؛ لأنها ترتبط مباشرة بموسطات طبيعية هي ترددات الطنين للجهاز الصوتي إلا أن عمليات تحليل الأصوات لاستنتاج هذه الموسطات شاقة وطويلة وتتطلب الكثير من الوقت والخبرة.(ينظر:المصدر السابق نفسه.)

(١٦) مركبات التنبؤ الخطي: وهي أيضاً كهربائية تعتمد نمذجة السيبل الصوتي بمرشح وحيد له أقطاب متعددة. يجري حساب ثوابت هذا المرشح باستعمال تقانة التنبؤ الخطي، والتي تعتمد على إمكان التنبؤ بقيمة كل عينة من إشارة الكلام كتركيب خطي لعينات متعددة سابقة (بدقة مقبولة). تتميز هذه الطريقة بكونها قابلة للأتمته في مرحلتي التحليل والتركيب ، ويمكن تحسين جودة الكلام المركب بنمذجة إشارة التحريض بنبضات متعددة في دور اهتزاز واحد للحبال الصوتية ، فتسمى تقانة التنبؤ الخطي المتعدد النبضات MPLPC أو-multi pluse linear prediction coders أو بإشارة أكثر تعقيداً كما في عملية التنبؤ الخطي المحرض برماز CELP أو code excited linear prediction.(ينظر:المصدر السابق نفسه)

(١٧) ينظر:مجلة العلوم:المجلد٢١ عدد يوليو(أغسطس)٢٠٠٥.

(١٨) ينظر: مجلة العلوم: المجلد٢١،يوليو (اغسطس) ٢٠٠٥.

(١٩) قاعدة البيانات: هو برنامج وسيط يحوي عينات مسجلة لكل واحد من المقاطع الصوتية للغة التي ينطق بها البرنامج، وتتم الاستفادة من قاعدة البيانات في إنشاء نموذج إحصائي يفيد في استقراء تلقائي للخواص العامة التي تحكم ارتفاع طبقة الصوت، أو انخفاضها. وتفيد في تحديد مدَّةِ كلام الفرد وعلوّه. ثم تجري الاستفادة من هذا النموذج لاحقاً عند تطبيق هذه الخواص لجعل مَسْمَع كلام النظام أقرب إلى كلام الانسان.(ينظر:مجلة العلوم: المجلد٢١ عدد يوليو(أغسطس)٢٠٠٥.

(٢٠) لذلك يجب أن يقترن البرنامج ببرنامج التشكيل الآلي للنصوص الخاص باللغة العربية، بالاضافة إلى برنامج المحلل الصرفي.(ينظر:بحث الناطق العربي الالي).

(٢١) العرب وعصر المعلومات-نبيل علي:ص٣٤٨.

(٢٢) ينظر: المصدر السابق:ص٣٤٩-٣٤٩.

(٢٣) المصدر السابق:ص٣٤٩.

(٢٤) العين : ج ١ ص ٥٩

(٢٥) بعض ما يميز العربية في اصطلاحها للعلم ومصطلحه – د.جميل الملائكة ، مجلة مجمع اللغة العربية في الأردن العدد ٦٢ ص٢٤١

(٢٦) اللسانيات – د.سمير استيتيه :٥٦٣

(٢٧) المصدر السابق ٥٦٣-٥٦٤

(٢٨) قوانين الفونولوجية العربية :٣

(٢٩) بعض ما يميز العربية ص ٢٤١

(٣٠) المصدر السابق نفسه : ص ٢٤١

(٣١) ينظر المصدر السابق ٢٤٥ - ٢٤٩

(٣٢) ينظر اللسانيات العامة، مصطفى حركات:ص ٥٩ وما بعدها.

(٣٣) لقد وضع الباحثون في مدينة الملك عبد العزيز للعلوم والتقنية وجامعة الملك فهد للبترول والمعادن، قوانين للفونولوجيا العربية يمكن استعمالها في برامج الحاسوب الآلي . ينظر:قوانين الفونولوجيا العربية، منصور الغامدي، حسني المحتسب، مصطفى الشافعي، ورقة بحث للمشروع رقم : أت ١٨ – ١٢ .

(٣٤) ينظر كتاب سيبويه : ج٤ ص٤٣٢، المقتضب:ج١ص٣٢٨.

(٣٥) ينظر: اللسانيات د.سمير استيتية :ص٢٩، صوتيات العربية منصور الغامدي :ص ١٠٣ وما بعدها.

(٣٦) التردد : هو عدد الذبذبات في الثانية الواحدة ويقاس بالهيرتزHertz أو مايعرف اختصارا بـ Hz (ينظر الصوتيات العربية : ص ١٠٥) .

(٣٧) الشدة : وتعني مدى شدة الموجة (أي سعتها) وتقاس بالديسبيل decibel أو ما يعرف اختصارا بـ db. (ينظر المصدر السابق ص ١٠٥).

(٣٨) الزمن : هو المقياس الزمني للموجات الصوتية ويقاس بالمليثانية (millisecond) وهي جزء من الألف من الثانية . (ينظر المصدر السابق ص ١٠٥) .

(٣٩) ينظر: علم الأصوات اللغوية : ص ٢١ وما بعدها، دراسة الصوت اللغوي ص٥٥ وما بعدها .

(٤٠) وهي نوع من الكتابة تعتمد الرموز اللاتينية، تظهر فيها كيفية نطق الكلمة من تبيين الخصائص الدقيقة للألوفونات. (ينظر معجم علم الأصوات – محمد الخولي ص ١٣٧).

(٤١) ينظر : وحدات صوتية لتوليد الكلام العربي آليا : ص٥.

(٤٢) ينظر : وحدات صوتية : ص ٦ .

(٤٣) له كتاب (الصوتيات العربية) ، وبحوث متعددة في هذا المجال من بينها بحث (البصمة الصوتية)، المجلة العربية للدراسات الأمنية.(يراجع موقع الدكتور منصور الغامدي على الانترنيت).

(٤٤) من مثل بحوث الدكتور منصور الغامدي: الألف بائية الصوتية العربية العالمية، ألف بائية صوتية عالمية بالحرف العربي، نظام ترميز اللغة العربية الجديد.

المصادر

- استخدام اللغة العربية في المعلوماتية، المنظمة العربية للتربية والثقافة والعلوم،تونس،١٩٩٦.
- بعض ما يميز العربية في صلاحها للعلم ومصطلحه، د.جميل الملائكة، مجلة مجمع اللغة العربية في الأردن، العدد٦٢.
- تحليل الكلام وتركيبه، الموسوعة العربية،دمشق، دار الفكر.
- تحليل اللغة العربية بواسطة الحاسوب، د.سعد القحطاني، مجلة مجمع اللغة العربية في الأردن، العدد ٦٨.
- حواسيب محاورة، مجلة العلوم،الترجمة العربية لمجلة ساينتفك أمريكان، تصدرعن مؤسسة الكويت للتقدم العلمي، عدد يوليو(أغسطس)٢٠٠٥.
- دراسة الصوت اللغوي، أحمد مختار عمر، عالم الكتب،١٩٩٧.
- الصوتيات العربية، منصور الغامدي، دار التوبة، الرياض.
- العرب وعصر المعلومات، نبيل علي، عالم المعرفة، أبريل ١٩٩٤، العدد١٨٤.
- علم الأصوات اللغوية، مناف مهدي الموسوي، دار الكتب العلمية، بغداد ، ط٣،٢٠٠٧.
- العين، الخليل بن احمد الفراهيدي(١٧٥هـ)، تحقيق مهدي المخزومي، ابراهيم السامرائي، سلسة المعاجم والفهارس.

- قوانين الفونولوجية العربية، منصور الغامدي، حسني المحتسب، مصطفى الشافعي، مجلة جامعة الملك سعود،علوم الحاسوب والمعلومات (١٤٢٤هـ)، العدد١٦.
- الكتاب، سيبويه، ابي بشر عمر بن عثمان بن قنبر، تحقيق عبد السلام هارون، مكتبة الخانجي بالقاهرة، دار الرفاعي بالرياض،ط١٩٨٢،٢.
- اللسانيات العامة وقضايا العربية، مصطفى حركات، المكتبة العصرية، ط١٩٩٨،١.
- اللسانيات (المجال والوظيفة والمنهج)، سمير شريف استيتية، عالم الكتب الحديث، جدار للكتاب العلمي، ط٢٠٠٨،٢.
- اللغة العربية والحاسوب، د.ندى غنيم، د.أميمة دكاك،المنظمة العربية للتربية والثقافة والعلوم ، الاجتماع الثاني لخبراء المعجم الحاسوبي للغة العربية، مايو٢٠٠٨.
- معجم علم الأصوات، محمد الخولي ، ط١٩٨٢،١.
- المقتضب، المبرد، أبي العباس محمد بن يزيد(٢٨٥هـ)، تحقيق: محمد عبد الخالق عظيمة، جمهورة مصر العربية، وزارة الأوقاف،المجلس الأعلى للشؤون الإسلامية، لجنة إحياء التراث الإسلامي،١٩٩٤.
- مقدمة في نظرية المعلومات(الرموز ،الإشارات،والضجيج)، جون ر .بيرس، ترجمة:المهندس فايز فوق العادة ، منشورات وزارة الثقافة في الجمهورية العربية السورية، دمشق،١٩٩١.
- الناطق الآلي، بحث مقدم في جامعة الملك عبد العزيز للعلوم التطبيقية، بإشراف: عبدالملك سلمان السلمان.
- وحدات صوتية لتوليد الكلام العربية آلياً، منصور الغامدي،حسني المحتسب، مصطفى الشافعي، ورشة العمل الرابعة حول علوم الحاسوب والمعلومات، جامعة الملك فهد للبترول والمعادن(١٤٢٣هـ).
-MODERN DIGITAL AND ANALOG COMMUNICATION SYSTEMS، BY: B.P.LATHI.