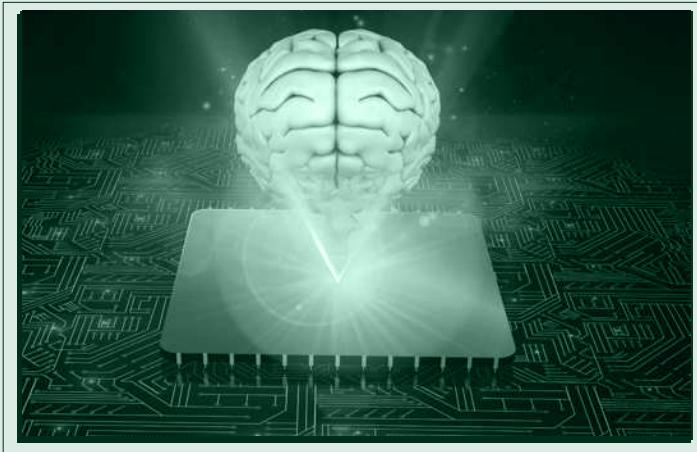




العربية والذكاء الاصطناعي



تحرير
المعتز بالله السعيد



العربية والذكاء الاصطناعي

تحرير

المعتز بالله الشعيب

المشاركون

أحمد راغب

المعتز بالله الشعيب

محمد عطيه

نعميم عبدالغنى

العربية والذكاء الاصطناعي

محمد عطيه

الرياض ، ١٤٤٥ هـ

البريد الإلكتروني: nashr@ksaa.gov.sa

ح / مجمع الملك سلمان العالمي للغة العربية ، ١٤٤٥ هـ
فهرسة مكتبة الملك فهد الوطنية أثناء النشر

..ص : ..سم

رقم الإيداع: ١٤٤٥/١٤٣٨

ردمك: ٣-٥٩-٨٤١٣-٦٠٣-٩٧٨

لا يسمح باعادة إصدار هذا الكتاب، أو نقله في أي شكل أو وسيلة ، سواء أكانت
الكترونية أم يدوية ، بما في ذلك جميع أنواع تصوير المستندات بالنسخ ، أو التسجيل
أو التخزين ، وأنظمة الاسترجاع ، دون إذن خطى من المجمع بذلك.

(صدر هذا الكتاب عن مركز الملك عبدالله للتخطيط والسياسات اللغوية، والذي
جرى دمجه في مجمع الملك سلمان العالمي للغة العربية).

◆ هذه الطبعة إهداء من المجمع، ولا يُسمح بنشرها ورقياً، أو تداولها تجاريًّا

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

أطلق مجمع الملك سلمان العالمي للغة العربية ضمن أعماله وبرامجه مشروع: (المسار البحثي العالمي المتخصص)؛ لتلبية الحاجات العلمية، وإثراء المحتوى العلمي ذي العلاقة بمجالات اهتمام المجتمع، ودعم الإنتاج العلمي المتميّز وتشجيعه، ويضم المشروع مجالات بحثية متنوعة، ومن أبرزها: (دراسات التراث اللّغوي العربي وتحقيقه، والدراسات حول المعجم، وقضايا الهوية اللّغوية، ومكانة العربية وتعزيزها، واللسانيات، والتخطيط والسياسة اللّغوية، والترجمة، والتّعريب، وتعليم اللّغة العربية للناطقين بها وبغيرها، والدراسات البيئية).

وصدر عن المشروع مجموعة من الإصدارات العلمية القيمة (جزء منها- ومن بينها هذا الكتاب - صدر عن مركز الملك عبدالله بن عبدالعزيز للتخطيط والسياسات اللّغوية والذي جرى دمجه في مجمع الملك سلمان العالمي للغة العربية). ويسعد المجمع بدعوة المختصين، والباحثين، والمؤسسات العلمية إلى المشاركة في مسار البحث والنشر العلمي، والمساهمة في إثرائه، ويمكن التواصل مع المجمع مسار البحث والنشر عبر البريد الشّبكي: (nashr@ksaa.gov.sa).

والله ولي التوفيق

هذا المشروع

مشروع تأليف سلسلة كتب في مجال (حوسبة العربية) يهدف إلى بناء تراكمٍ معرفي في مجال حيوي مهم، هو مجال (حوسبة العربية) . ويعد هذا الكتاب واحداً من سلسلة كتب صدرت في المركز.

يقع هذا المشروع ضمن سلسلة (مباحث لغوية) التي يشرف المركز على اختيار عنواناتها، وتكتلif المحررين والمؤلفين، ومتابعة التأليف حتى إصدار الكتب. وهي سلسلة يجتهد المركز أن تكون سداداً لحاجات بحثية وعلمية تحتاج إلى تنبية الباحثين عليها، أو تكشف البحث فيها.

ويعد هذا الكتاب واحداً من كتب ثلاثة متراقبة في مشروع علمي واحد متخصص في (الذكاء الاصطناعي) :

- العربية والذكاء الاصطناعي.
- تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية.
- خوارزميات الذكاء الاصطناعي في تحليل النص العربي.

مدير مشروع (العربية والذكاء الاصطناعي)

د. عبدالله بن يحيى الفيفي

فهرس الكتاب

الصفحة	الموضوع
٥	هذا المشروع
١١	كلمة المركز
١٣	مقدمة
١٧	الفصل الأول: الذكاء الاصطناعي ونمذجة اللغات الطبيعية: الطموح، الواقع، والآفاق
٢٠	١. تاريخ مختصر للذكاء الاصطناعي.
٦٥	٢. مدرستان في معاجنة المسائل في إطار الذكاء الاصطناعي.
٧٠	٣. تحديات نمذجة اللغات الطبيعية ومعاججتها آلياً.
٧٩	٤. إلى أين وصلت المعاجنة الآلية للغات الطبيعية؟
٨٢	٥. هل هناك سقف لمعالجة اللغات الطبيعية آلياً؟
٨٤	٦. هل لغة العربية خصوصية مع الذكاء الاصطناعي؟
٨٦	٧. ما هي آفاق مستقبل المعاجنة الآلية لغة العربية عبر الذكاء الاصطناعي؟

٨٨	٨. الخاتمة والتائج
٩١	الفصل الثاني: المعالجة الآلية للغة العربية المكتوبة-مقدمة في ذكاء الآلة
٩٤	١. الذكاء الاصطناعي واللغة المكتوبة.
٩٦	٢. المعالجة الآلية الكتابية [المحرفية / الجرافيمية].
١٠٢	٣. المعالجة الآلية البنوية [الصّرفية].
١٠٩	٤. المعالجة الآلية التركيبية.
١١٧	٥. المعالجة الآلية الدلالية.
١٢٧	الفصل الثالث: المعالجة الآلية للغة العربية المنطوقة
١٣٠	١. الصوت اللغوّي وعلم الأصوات.
١٣١	٢. مصطلحات أساسية.
١٣٢	٣. التحليل الحاسوبي لمكونات الصوت اللغوّي.
١٥٧	٤. من التقنيات الصوتية الحاسوبية.
١٧٥	الفصل الرابع: الذكاء الاصطناعي وتعليم اللغة العربية: نحو منصة تعليمية متكاملة
١٧٨	١. اللغة والتعلم.
١٧٩	٢. طبيعة إنتاج اللغة.
١٨١	٣. الذكاء الاصطناعي وإنتاج اللغة.
١٨٢	٤. الذكاء الاصطناعي واللغة العربية.
١٨٣	٥. الذكاء الاصطناعي وتعلم النطق.
١٨٧	٦. برامج إثراء الثروة اللغوية.
١٩٢	٧. الذكاء الاصطناعي والكتابة.

١٩٤	٨. الذكاء الاصطناعي وتعليم الإملاء.
١٩٦	٩. الذكاء الاصطناعي وتعليم النحو.
١٩٨	١٠. الذكاء الاصطناعي وإعادة بناء النصّ.
٢٠٠	١١. مشكلات الذكاء الاصطناعي في التعامل مع اللغة العربية.
٢٠١	١٢. مقترن لنصة تعليمية ذكية.
٢٠٩	الباحثون

كلمة المركز

يعمل المركز في مجال البحث العلمي ونشر الكتب مستهدفاً التركيز على المجالات البحثية التي مازالت بحاجة إلى تسلیط الضوء عليها، وتكثیف البحث فيها، ولفت أنظار الباحثين والجهات الأكاديمية إلى أهمية استثمارها بمختلف وجوه الاستثمار، وذلك مثل مجال (التخطيط اللغوي) و (العربية في العالم) و (الأدلة والمعلومات) و (تعليم العربية لأنوائها أو لغير الناطقين بها) إلى غير ذلك من المجالات، وإن من أهم مجالات البحث المستقبلية في اللغة العربية مجال (العربية والحوسبة ، والذكاء الاصطناعي) حيث إن حياة اللغات ومستقبلها مرهونة بمدى تجاوיבها مع التطورات التقنية والعالم الافتراضي، وكثافة المحتوى الإلكتروني المكتوب، وهو ما يشكل تحدياً حقيقياً أمام اللغات غير المنتجة للمعرفة أو للتقنية.

وقد عمل المركز على تسلیط الضوء على هذا المجال التخصصي؛ مستعيناً بالكفاءات القادرة من المهتمين بالتخصص البيني (بين اللغة والحاسوب) مقدراً جهودهم، وهادفةً إلى نشرها، وتعزيز مبادئها، راغباً أن يكون هذا المسار العلمي مقرراً في الجامعات في كلية العربية والحاسوب، و مجالاً بحثياً يقصده الباحثون الأكاديميون، والجهات البحثية العربية.

وقد أصدر المركز سابقاً ستة عشر كتاباً مختصاً في (حوسبة العربية) وفي الإفادة من (المدونات اللغوية) في الأبحاث العربية، ويحتفل بإصدار سبعة كتب جديدة مختصة في (حوسبة العربية والذكاء الاصطناعي)، ويقدمها للقارئ العربي، وللجهات الأكاديمية؛ للإفادة منها في مناهج التعليم والبناء عليه، وهذه الكتب السبعة هي: (العربية والذكاء الاصطناعي، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، خوارزميات الذكاء الاصطناعي في تحليل النص العربي، مقدمة في حosome اللغة العربية، الموارد اللغوية الحاسوبية، المعالجة الآلية للنصوص العربية، تطبيقات أساسية في المعالجة الآلية للغة العربية).

ويشكر المركز السادة مؤلفي الكتب، ومحرريها، لما تفضلوا به من عمل علمي رصين، وأدعو الباحثين والمؤلفين إلى التواصل مع المركز لاستكمال المسيرة، وتفتيق فضاءات المعرفة.

وفق الله الجهود وسدد الرؤى.

الأمين العام
أ. د. محمود إسماعيل صالح

مُقدمة

يُمثل «الذكاء الاصطناعي Artificial Intelligence» ميدانًا بحثيًّا رحباً ذا أبعادٍ اقتصاديَّةٍ فائقة، تظهرُ آثارُها على أصعدِ مُختلفٍ وفي ميادينٍ شتَّى، تشملُ جوانبَ البحثِ والصناعةِ والتدريس. ولا يزالُ الأملُ قائماً لدى المعنيينِ بمستقبلِ الآلةِ في الوصولِ بالذكاءِ الاصطناعيِّ إلى مُستوياتٍ متقدمةٍ، يُمكِّنُ معها توفيرُ كثيرٍ من الجهدِ والطاقاتِ البشرية، من خلالِ تنميةِ قُدراتِ الآلةِ علىِ محاكاةِ ذكاءِ الإنسان. ولعلَّ الطَّفرةِ المائِلَةِ التي نشهُدُها اليومَ تُعجلُ بتحقيقِ الآمالِ والطموحاتِ.

لقد مرَ الذكاءُ الاصطناعيُّ - بمفهومِه الحديثِ - بثلاثةِ أطوارٍ رئيسَة، تتمثلُ في: طورِ النَّشأة، وطورِ التَّجربَة، وطورِ النَّهضةِ التي نَلَمْسُها الآن. وكانتَ لكلِّ طورٍ من هذهِ الأطوارِ سماتٌ ورؤى وأفكار، تراكمُ في أطْرِها العامةَ سعياً إلى توجيهِ الآلةِ إلى فهمِ أعمقِ لقدرَاتِ العقلِ البشريِّ؛ أو بعبارةٍ أخرى، سعياً إلى محاكاةِ قدرَاتِ الدِّماغِ البشريَّةِ ومهاراتِ التَّفكيرِ العُليَا لدىِ الإنسان، كالقدرةِ علىِ صناعةِ القرارِ، والتَّفسيرِ، والاستنباطِ، والتَّفكيرِ الإبداعيِّ، وغيرِ ذلكِ.

وتتفَرَّغُ عن الذَّكاء الاصطناعيِّ مجالاتٌ معرفيةٌ عديدة، منها ذلك المجال الذي نحنُ بِصَدِّيهِ في هذا الكتاب، ونعني «المعالجة اللُّغات الطَّبَيعيَّةِ» Natural Language Processing؛ وهو مجالٌ معرفيٌّ يُسعي من خلالِهِ إلى توجيهِ الآلة إلى فهم اللُّغة الطَّبَيعيَّةِ عبرَ مُستوياتها المُتعدِّدة، ومعالجةَ وَحداتِها في هذه المُستويات تحليلًا وتوليدًا، وخلق بيئةٍ تفاعليةٍ قادرةٍ على تحقيقِ التَّوَاصُل بينَ الإنسان والآلة. وحينَ تحدثَ عن اللُّغة الطَّبَيعيَّةِ، فنحنُ معنِّيُونَ بِصُورِيَّتها الرَّئيسيَّتين، المكتوبة والمنطقية.

واللُّغةُ العربيَّةُ إحدى اللُّغات الطَّبَيعيَّةِ التي تخطى بعنانِي الباحثين في الذَّكاء الاصطناعيِّ عمومًا ومعالجة اللُّغات الطَّبَيعيَّةِ على وجهِ الْخُصُوصِ. ذلك أنها واحدةٌ من أكثر اللُّغات انتشارًا في عالمنا المُعاصر؛ حيثُ تأتي في المرتبة الرابعة من حيثُ عدد مستخدميها، بعدَ الصَّينيَّة والأرديَّة-الهندية والإنجليزية.

كذلك فإنَّها واحدةٌ من أكثر اللُّغات الطَّبَيعيَّةِ تَنَامِيًّا في العالم. ويكفي أن نعرفَ أنها انتقلت من المرتبة السادسة في مُنتصف العقدِ الأوَّل من القرن الحادِي والعشرين إلى المرتبة الرابعة في مُنتصف العقدِ الثاني، من حيثُ عددِ مستخدميها الذي يقتربُ من نصف المليار نَسْمة. ولعلَّ هذا الانتشار والتَّنَامِي يُعطينا صُورَةً استشرافيةً عَمَّا يُمْكِنُ أن تصلَ إليه العربيَّةُ في المستقبل القريب.

إنَّا نُؤمِنُ أنَّ اللُّغةَ العربيَّةَ قابلةٌ للمعالجة الآلية، ومُهِيأةٌ للوصول إلى مُستوياتٍ مُتقدِّمةٍ للغاية، إذا توافرت لها المواردُ والطَّاقاتُ من ناحية، وإذا توافر لها الدَّعمُ المُناسبُ من ناحيةٍ أخرى. ذلك أنَّ طبيعةَ العربيَّةِ قياسيةٌ في شَطَرِ كبيرٍ منها، لا سيَّما في بنيتها الصَّوتيةِ وقوانيِّنِها الصَّرْفيةِ. وفي الوقت ذاته، فالعربيَّةُ لُغَةٌ مُعرَبةٌ، وتتَّمَعُ بنظامٍ استنفافيٍّ توليدِيٍّ، وتتَّمَعُ أيضًا بنظامٍ كتابيٍّ خاصٌّ؛ وهي أمورٌ تُمثِّلُ إشكالاتٍ ليست هَيِّنةً؛ ولن تكونَ معالجةُ هذه الإشكالات يسيرةً ما لم تتوافر السُّبُلُ والأدواتُ المُؤَدِّيةُ إلى ذلك.

ونحنُ نُقدِّمُ هذا الكتابَ (العربيَّةُ والذَّكاءُ الاصطناعيُّ) الذي نسعي من خلالِه إلى الْوُقُوف على علاقَةِ اللُّغةِ العربيَّةِ بالذَّكاءِ الاصطناعيِّ، وما انتهَتَ إليه جُهودُ معالجتها آليًّا، وما يُمْكِنُ أن يُقدِّمُ الذَّكاءُ الاصطناعيُّ للعربيَّةِ مُستقبلاً. ونُشُدُّ في عملِنا هذا

الإجابة عن أسئلة، منها: ما الذكاء الاصطناعي، وما أطواره، وما إشكالاته؟ وما علاقته باللغات الطبيعية بالذكاء الاصطناعي، وما علاقه العربية خصوصاً به؟ وإلى أي مدى يمكن توظيف الذكاء الاصطناعي في معالجة اللغة العربية المكتوبة؟ وما واقع توظيفه في معالجة اللغة العربية المنطقية؟ وما أبرز تحديات حوسبة العربية؟ وكيف يمكن الإفاده من تقنيات الذكاء الاصطناعي وتطبيقاته في تعليم العربية؟

يشتمل الكتاب على أربعة فصول؛ حيث يأتي الفصل الأول بعنوان (الذكاء الاصطناعي ونَمْذَجَةِ الْلُّغَاتِ الطَّبِيعِيَّةِ: الْطَّمُوحُ، الْوَاقِعُ، وَالْآفَاقُ)، وَيُمْثِلُ تقدِيمًا مُهُمًا وثريًا لمفهوم الذكاء الاصطناعي وإرهاصاته وواقعه وآفاقه وعلاقته باللغات الطبيعية، ويُجِبُ الفصل أيضًا عن سؤالاتٍ مطروحة بشأن طبيعة اللغة العربية وقابليتها للمعالجة الآلية. ويأتي الفصل الثاني بعنوان (المعالجة الآلية للغة العربية المكتوبة: مقدمة في ذكاء الآلة)، ويتناولُ رؤيةً منهجيةً حول مستويات اللغة العربية المكتوبة ووحداتها والتوجيه الآلي لها ووسائل المعااجلة السطحية والعميقة، مع الإشارة إلى تطبيقات معااجلتها آليًا في كُلِّ مُسْتَوِيٍ على حدة. أمّا الفصل الثالث فيأتي بعنوان (المعالجة الآلية للغة العربية المنطقية)، ويتناولُ طرحاً لمكونات الصوت اللغوي وأليات تحليله حاسوبياً، كما يعرض بعض التقنيات الصوتية الحاسوبية وأليات بنائها وتطويرها، ويركز خصوصاً على تقنية تحويل النص العربي المكتوب إلى صوت منطوق. ويأتي الفصل الرابع أخيراً بعنوان (الذكاء الاصطناعي وتعليم اللغة العربية: نحو منصة تعليمية متكاملة). ويتناول هذا الفصل رؤيةً حول كيفية الإفاده من الذكاء الاصطناعي في تعليم مهارات اللغة العربية، مع التركيز على مهارات الكتابة والتحدد، ويعرض الفصل أيضًا مقتراً لمنصة تعليمية ذكية، يمكن تطويرها واستثمارها في تعليم اللغة العربية بمستوياتها المختلفة.

وبعد؟

فإنَّا نرجو أن يكونَ هذا الكتاب فاتحةً للباحثين في ميادين الذكاء الاصطناعي، خصوصاً أولئك المعنيين بالمعالجة الآلية للغة العربية، لطرق آفاقٍ أكثرَ سعةً ورحابة من التفكير والإبداع، سعياً إلى الوقف على إشكالات حوسبة العربية والعمل على

مُعاجِّلَتها، وتقويم أدواتها وخوارزمياتها ومواردها، والانتقال بها إلى مجتمع معرفي قادرٍ على تطوير الآلة لتحقيق نهضته وبناء ملامح مستقبله.

نسأل الله تعالى أن يتقبلَّ هذا الجهد بالذكر الحسن والأجر الجزيل، وأن يجعلَه من العلم الذي ينفع أصحابه بعد مماتهم.

ربَّنا عليكَ توَكَّلنا وإليكَ أتَبَّنا وإليكَ المصير.

المحرر

د. المُعْتَزِّ بِالله السَّعِيد

الفصل الأول

الذكاء الاصطناعي ونَمْذَجَةُ اللُّغَاتِ الطَّبَيِعِيَّةِ

الطموح، الواقع، والآفاق

د. محمد عطية محمد العربي أحمد

ملخص

يملاً الحديثُ عن مصطلح «الذكاء الاصطناعي» وأخبارُ تقنياته والتبيّنُ بإنجازاته في السنوات الأخيرة وسائل الإعلام في السنوات الأخيرة، فما هو الذكاء الاصطناعي؟ من أين وكيف نشأ؟ وما تارikhه؟ ما هي عَرَاثَتُه وما هي نجاحاته؟ هذا ما يُفصّلُهُ القسمُ الأول بأجزاءه الثانية التي تمثّلُ الأساس الذي تَبْنِي عليه الأقسامُ التالية من الفصل. وينتقلُ القسمُ الثاني إلى التعرّف على كيفية معالجة المسائل المعتبرة في إطار الذكاء الاصطناعي بمدرستَيْه؛ التحليلية، والعدديّة الإحصائية.

وينصبُ التركيز في القسم الثالث على إحدى المسائل الكبرى المطروحة في مجال الذكاء الاصطناعي - إن لم تكن أكبرها على الإطلاق - ألا وهي تَمْذِيَّة اللغات الطبيعية ومعالجتها آلّياً والتحديات الأساسية التي تواجهُها. وفي القسم التالٍ له يقف القارئ على الوضع الراهن للبحث العلمي والتطور التقني للمعالجة الآلية للغات الطبيعية.

ثم يناقش القسم الخامس ما إذا كان تطور المعالجة الحاسوبية الآلية للغات الطبيعية سيمضي إلى أن يحاكي ملَكات الإنسان العاقل وربما يتجاوزه، أم أنَّ هناك سقفاً لذلك لا يمكن المروق منه. وهي ليست فقط قضية فلسفية نظرية عميقة، بل إنها محدّدة لمسار البحث والتطوير على هذا الصعيد.

وينظر القسم السادس من الفصل في حال اللغة العربية مع الذكاء الاصطناعي من حيث ما إذا كانت لها خصوصية عن غيرها من اللغات الطبيعية سواء في نمذجتها أو معالجتها الآلية، وما تلك الخصوصيات إن وُجِدَتْ؟ أما القسم السابع في مُختَتم الفصل فإنه يستشرف في ضوء ما سبق من أقسامه آفاقَ مستقبل اللغة العربية مع الذكاء الاصطناعي على رجاء أن يكون هذا الفصل الأول كُلُّه زادًا للقارئ وعُونًا له على استيعاب بقية فصول الكتاب والتفاعل الإيجابي معها.

الكلمات المفتاحية:

التعلم الحاسوبي، الذكاء الاصطناعي، اللغات الطبيعية، اللغة العربية، النَّمْذَجَة الحاسوبية، المقاربة التحليلية.

١. تاريخٌ مختصرٌ للذكاء الاصطناعيٌ.

تعلّمنا مدرسةُ التاريخ على اختلاف مواضيعها أن جذورَ النّشأة دائمًا ما تطبعُ أثراًها على الحاضر وعلى المآلات، والذكاء الاصطناعي ليس بذُراً مما دوَّنه التاريخ، ولذلك فليس فَضْلَةً ولا تَرْبِيدًا أبداً الاستفاضةُ في القسم الافتتاحي من هذا الفصل من أجل وضع القارئ في صورة نشأة الذكاء الاصطناعي وتطوره التاريخي ومبادئ آليات عمله الأساسية بتيسيرٍ غير مُخلٍّ، ومن ثم علاقته بميّكانة اللغات الطبيعية. ولعل القارئ يجني بعد مطالعته لهذا القِسْمِ الأوَّل ثمارَ ذلك أثناء دراستِه للأقسام التالية من الفصل؛ استيعابًا سلِيسًا لما سيرُدُّ فيها من اصطلاحاتٍ ومسائلٍ حosomeُ اللغات الطبيعية، وفهمًا لما إذا أخذَت بعضُ الخيارات دون غيرها في مسارِها، وربطًا بما يرد في ختام الفصل من استشرافات، ولعله يحصلُ عِلَوَةً على ذلك كله متعةً تتبعُ قصة ميلاد علمٍ جديدٍ من رحم تفاعلاتٍ مطولةً مختدمٍ بين ما كان قائماً قبلَه من فلسفةٍ وعلومٍ وتقنياتٍ معاً.

١، حُلْمٌ قديمٌ وسعيٌ مستمرٌ

ما بَرَحَ كِيَارُ الحكام والقادة والمُخْطَطِين في مختلف أنحاء العالم عبر أزمنة التاريخ قدِيمَها وحديثَها تداعبُ خواطِرِهم أحَلامُ امتلاكِ جيوشٍ كبيرة العدد من العَمَال أو المقاتلين غير البشرِين المسخَّرين تمامًا للقيام بما يقوم به البشر من وظائف العمل أو القتال ولكن دون تكُبُّد أعباء البشر الحقيقيين الذين يحتاجون للراحة ولل الطعام وللksesاء وللعلاج وللماوى وللترفيه، وللتعليم والتدرِّيب، وكذلك لمراعاة معتقداتهم ومعالجة شؤونهم المعنوية، ومع ذلك كُلِّه فقد يتمرسون وينخرجون عن السيطرة. فمن يمتلك دون خصوِّمه مثل تلك الجيوش من الآلَيَّن أشباه البشرِين تحت كامل سيطرته فلا شك أنه سيهزِّ مهُم اقتصادياً أو يسوِّدُهم عسكرياً.

قبل حوالي ثلاثة آلاف سنة في الصين يُعْتَقَدُ أن مهندسًا يُدْعَى «يان شي» قدم للإمبراطور آنذاك «تشاو ميو وانج» آلة ميكانيكية من صنعه على هيئة آدمي وبحجمه الطبيعي تستطيع الخطو وتحريك اليدين وتأتي ببعض تعبيرات الوجه، وكذلك يُدْعَر الروماني «هيرون» عالم الرياضيات والمهندس الميكانيكي الشهير الذي عاش في مدينة الإسكندرية في مصر قبل حوالي ألفي سنة والذي اشتهر بآلاته التي ميَّزَت العديد من الوظائف التي اعتاد البشر القيام بها؛ مثل العزف على آلة «الأورغن» وبرمجة المؤثرات السمعية والبصرية على المسرح، وتماثيل ميكانيكية على هيئة بشرية تصب الشراب وأخرى تقوم بفتح وغلق أبواب القصور والمعابد.

ومن بين مئات الآلات التي صممها ونفذها قبل نحو ثانية قرونٍ أحد رُوَادُ الهندسة الميكانيكية ومشاهيرها «بديع الزمان أبو العز ابن إسماعيل الجزارِي» – الذي خدم في بلاط «بني آرتوق» في «ديار بكر» ثم في البلاط الأيوبي – عشراتُ الآلات الميكانيكية على هيئة بشرية تقوم بمهامَّ متنوعةٍ؛ منها مثلاً خادم آلي يصب الشراب، وآخر يصب الماء للوضوء وغسيل اليدين، وجُوقة آلين تعزف الآلات الموسيقية تلقائياً، وغير ذلك كثير. ثم نجد على هذا الطريق في نهاية القرن الثامن عشر الميلادي عدداً من الأوروبيين نذكر من بينهم صانع الساعات البارع الفرنسي «بيير جاك درو» وابنه «هنري لوبي» الذي اشتهرت دُمَاهُما الميكانيكية ذات الهيئة البشرية المتقدة بين دُمَاهِّ تعزف الموسيقى وأخرى تخط حروفًا بالقلم ونحو ذلك، وقد تميزت هذه الدُمَاهُ الميكانيكية بوجود ذاكرة ميكانيكية تُمثِّل «برناجاً» (خطوات حركية متسلسلةٌ معرَّفةٌ سلفاً) ولا تزال ثلاثة من هذه الدُمَاهِّ معروضةً في متحف الفن والتاريخ بمدينة «نوشاتيل» في «سويسرا».

ويحفل تاريخُ مختلف الأمم عتيقه وقربيه مثلما يحفل التراث العريض لأدب الخيال العلمي بالطراائف والنفائس من أمثلة تلك الآلات التي حاولت محاكاة البشر في ملمح أو أكثر، وتشير كلها دون لبس إلى قَدَمٍ وتواصلُ الولع بحلم صناعة «أشباء البشر». وكثيراً ما يستخدم العامةً اصطلاح «الخادم الآلي robot»^(١) بدلاً من «شَيْهِيَّةِ البشر humanoid»؛ وخصوصاً في سياق التحكُّم والتسيير، والواقع أن الاصطلاحين ليسا متطابقين كما سيرد في القسم الفرعِي ١,٥ من الفصل.

١- كلمة «رُوبُوت» مشتقة من الكلمة «رُوبُوتا robota» باللغة التشيكية التي تعني «السُّخرة».

٢، ١. نهضة الرياضيات قاطرةً للفيزياء والتقنية.

على صعيد آخر وثيق الصلة؛ فإن حقل الرياضيات قد انتقل إلى حقبة جديدةٍ مع مطلع القرن السادس عشر الميلادي حيث نقل الأوروبيون معارف الرياضيات من العربية إلى اللاتينية واتسعت وقتها كثيراً قاعدة المشغلين بها اطلاقاً من الإمارات الإيطالية، ويشير العديد من مؤرخي العلوم إلى عام ١٥١٩ م - حين توصل الإيطاليون إلى صيغة الخل لإيجاد جذور «كثيرة الحدود polynomial» من الدرجة الثالثة - على أنه نقطة انتقال التراكم المعرفي في الرياضيات من تراكم رتيب بطيء قبلها إلى تراكم على منحنى متزايد التسارع بعدها فاستقر التعبير الرمزي في الرياضيات، وصيغت نظرية عامة للأعداد، وللजبر، ثم ابتكر حساب التفاضل والتكامل، والمعادلات التفاضلية ودراسة حلولها، والإحصاء وحساب الاحتمالات، والجبر الخطّي، وحساب المصفوفات، والأعداد المركبة، ثم هندسة السطوح المنحنية (الهندسة اللا إقليدية Non-Euclidean geomerty) ... وغير ذلك كثيراً مما يضيق المقام هنا عن ذكره.

ويعني ذلك كله امتلاك العديد من الأدوات الناجعة للتمكن من الرصد والتحليل والتصميم والحساب والإدارة والتحكم ، فكانت هذه الأدوات أعمدةً قامت عليها نهضة أخرى هائلة في اكتشاف وتطور علوم الطبيعة (الفيزياء^(١))، والتي بدورها كانت أساساً آخر قامت عليه نهضة تقنية صناعيةٌ كبرى بدءاً من النصف الثاني للقرن الثامن عشر الميلادي تتبعها موجاتها لتبلغ بنهاية الحرب العالمية الثانية نضجاً وتمكيناً كبيراً، لا سيما على محاور الميكانيكا، والبصريات، والطاقة، ثم الكهرومغناطيسيات، والاتصالات، وترميز البيانات وتشغيرها.

ومن المفيد هنا الإلارج إلى أن الرياضيات قد انتقلت عبر ذلك الوقت بالتدريج من مجرد تداول ومعالجة المعادلات والمتباينات - منها كان تعقيدُها - إلى القدرة على تقديم حلولٍ متكاملة لمسائل مركبة مصوّفةً على هيئة إجراءات عديدة الخطوات لمعالجة مدخلات المسألة تجمع إلى جانب المعادلات حدوداً شرطيةً تعرّف الانتقال بين هذه الخطوات مما يُعرف ضمنياً المسارات (السيناريوهات) الممكنة لحلول

١- وتحقّق بها الكيمياء، وتطوّر المواد.

المسألة قيد المعالجة، وهو ما اصطلح على تسميته بـ «الخوارزم»^(١) وهو الصورة الرياضية التي يمكن تفديتها أو ميكانتها عبر آلة مناسبة قابلة للبرمجة.

٣. ابعاد الفلسفة وتفاعلها مع الرياضيات سعياً وراء قوانين التفكير البشري.

وعلى صعيد ثالثٍ يتكامل مع سابقيه فقد اقتفى مسار الفلسفة خطًّا مشابهًة حين بدأ انبعاثها في أوروبا خروجاً عن الإطار المعرفي الجامد الذي فرضته الكنيسة الكاثوليكية لقرون طويلة من ناحيةٍ وحين اكتمل انتقال التراث الفلسفي من العربية إلى اللاتينية من ناحيةٍ أخرى، فبلغت في أوروبا أسماءً كبيرةً في دنيا الفلسفة من القرن السادس عشر فصاعداً من أمثال «فرانسيس بيكون» إلى «رينيه ديكارت» إلى «جوتفرید لاينيتر» إلى «إيهانويل كانط» ... وغيرهم الذين طرحوا باستفاضة قضايا كبيرة تتصل بـ باهية المادة والطبيعة، والعلوم، والعقل والإدراك، وقوانين التفكير.

وفي وقت باكرٍ من تلك الحقبة الجديدة كان الاتصال لا يزال وثيقاً بين الرياضيات وعلوم الطبيعة (الفيزياء) والفلسفة، حتى أن الفيزياء كانت تسمى في وقت «إسحاق نيوتن» ومعاصريه بـ «الفلسفة الطبيعية»، ولذلك أسفرت الجدلية بين الفلسفة و المعارف الطبيعية والرياضيات متتسارعة النمو في القرون من السادس عشر إلى التاسع عشر الميلادية عن السعي الحثيث لفهم قوانين التفكير البشري وتقعیدها بل الطموح إلى ميكنة هذا التفكير، والذي أنتج بدوره إنجازاتٍ فارقةً.

ففي منتصف القرن التاسع عشر استطاع الإنجليزي «جورج جون بول» صياغة منظومة «المنطق الرياضي الثنائي Boolean Algebra» في صورة قوانين جبرية محكمة^(٢)، كما انتشرت في القرن نفسه بين الفلاسفة والمناطق الرياضيين نزعة لتأسيس

١- وذلك اعتقاداً بفضل الفلكي وعالم الرياضيات الأبرز في العصور الوسيطة «محمد بن موسى الخوارزمي» المتوفى في بغداد عام ٨٥٠م، وتحمل كذلك اسمه العملية الرياضية الشهيرة «لوغاريتم Logarithm» التي هي معكوس رفع عدد لأسٍ. وهو صاحب كتاب «الجبر والمقابلة»، وهو كذلك من أدخل الصفر في نظام العد العشري. ولا يزال علم الجبر يحمل نفس الاسم العربي في كل لغات العالم «Algebra» ولا يزال كذلك اسم الصفر في مختلف لغات العالم مشتقاً من اسمه العربي؛ فلا رياضياتٍ حديثة بدون نظام العد العشري ولا بدون الصفر ولا بدون علم الجبر، والخوارزمي رحمة الله هو الرائد المبرز في ذلك كله.

٢- وهو الأساس الذي تعمل وفقه كل الدوائر الإلكترونية التي تنفذ عمليات التحكم والحساب في كل الحواسيب الرقمية بمختلف أنواعها.

الرياضيات بالكامل على أساس مُحْكَمٍ من المنطق البَحْثُ الذي لا يتضمن سوى مجموعة محدودة من المسلَّمات مع القواعد الاستنباطية الأساسية فقط لا غير، فإذا تم ذلك فلربما صار الطريق مفتوحًا لـمِلكة عملية البرهنة الرياضية، فـكَرَّسَ الالماني «غُوتلِيبٌ فِريغِيَهُ» حياته المهنية في الرابع الأخير من نفس القرن لمحاولة تأسيس الحساب العددي بإحكام على المنطق وأصدر عمله في مجلدات ثلاثة، وقبيل طباعتها اكتشف أحد كبار المَناطقة الرياضيين الآخرين الإنجليزي «برِّتراند رِّاسِل» تناقضًا يخلخل بنية عمل «فِريغِيَهُ»، وكان «رِّاسِل» قد أخذ على عاتقه هو القيام بهذا التأسيس فنشر بين عامي ١٩١٠ و١٩١٢ م مع أستاذة الفيلسوف وعالم الرياضيات «آلفريد نورث وايت هيد» عملاً نظريًا ضخماً آخر^(١) يُؤسس للحساب العددي على أساس منطقي يَحْتِ بما يتتجنب ثغرة «فِريغِيَهُ».

وأثناء ذلك فقد كان مجتمع كبار علماء الرياضيات بعد الحرب العالمية الأولى - وقد تشعبت فروعها وتطاولَ بنائها المعرفي ليبلغ ذُرًا شاهقًا - يمور بالجدل حول ما إذا كانت أُسُسُها متينةً بما يكفي من أجل ضمان أن تتحقق منظومة الرياضيات شرطين أساسيين (كان غالب الظن وقتها أنها متحققةان تلقائياً) ألا وهم: خلوها من إمكانية تَوْلُد متناقضاتٍ، والاكتمال (أي قدرتها من حيث المبدأ على برهنة أية حقيقة رياضية يُمْكِنُ اشتقادُها من مُسلَّماتها الأولى). ولذلك دعا العالم الأَبْرُز مسموع الكلمة في ذلك الوقت الألماني «ديفيد هِيلِبِيرت» إلى مؤتمر عالميٌّ جامع للبت في هذه القضايا. وكانت ثمرة ذلك في بداية ثلاثينيات القرن العشرين هي المفاجأة الكبرى التي فجرها العبرى النمساوي^٢ (كِيرْت غِيدِل) بـمُبْرَهَتَيِّهِ البديعتين الشهيرتين عن استحالة اكتمال أية منظومة رياضياتٍ وخلوها من التناقض في الوقت ذاته، وبذلك حَدَّ بالبرهان الرياضي سقفاً للرياضيات وأثبت دون شك وجود حدود للمعرفة فيها مع تبيان أسباب ذلك. ومن رحم تلك الجدلية المحدثة بين من حاولوا تأسيس الرياضيات على المنطق الرياضي الصافي - ومن ثم إمكانية مِلكة عملية البرهنة واشتقاق الحقائق الرياضية من مسلَّماتها الأولى - وبين من أثبت رياضياً استحالة تحقيق ذلك بالكامل، تشكلت الملامح الأولى لـ«نظرية تعين الصعوبة» Complexity theory التي واصل روادُ مثل الأمريكية

١- وُشيرَ تحت عنوان «Principia Mathematica» ويعني باللاتينية «المبادئ الأساسية للرياضيات».

المحَرِّي «جون فون نُويان» والبريطاني «آلن تيورِنخ» تأسيسها والتي صارت بمثابة اللُّبُّ النظري لـ «علوم الحاسوب Computer Science»^(١).

٤ . ١ . وبدأ عصرُ الحواسيب الرقمية.

بحلول أربعينيات القرن العشرين الميلادي سواء أثناء الحرب العالمية الثانية وبُعيدَها، كانت كل الظروف قد اجتمعت من أجل بناء حواسيب رقمية ذات قدرات حسابية عالية تتصف بعمومية الاستخدام؛ بمعنى قابليتها للبرمجة كي تنفذ خوارزمات مختلفة، مع القدرة على استقبال بياناتٍ رقميةٍ كمدخلاتٍ تعمل عليها تلك الخوارزمات، وكذلك تصدير بياناتٍ رقميةٍ كمخرجاتٍ لعمل تلك الخوارزمات. وتمثلت تلك الظروف في أن:

- حلم صناعة شبيه الإنسان قائم لم يغب منذ الْقِدَم، ومكانة التفكير البشري في قلبه، ومكانة القدرات الحسابية حجر أساس فيه.
- الفِكْرُ الرياضي وأدواته صارت ناضجةً لذلك.
- التقنية الكهرومغناطيسية تُمْكِّنُ من ذلك.
- الصراع العسكري العالمي المشتعل والسعي وراء امتلاك كل أسباب القوة فيه يتوجان ويوظفان كل ما تقدَّمَ من ظروف.

ولذلك كان من الطبيعي أن تُبْنَى أثناء ذلك «الحواسيب الرقمية» الأولى التي يمكن برمجتها، وكان من طبائع الأمور أيضًا أن يتتصدر أكبرُ المتصررين في الحرب العالمية

١- هناك اعتقاد غير صحيح أن «علوم الحاسوب» لا وجود لها إلا بوجود «الحواسيب الرقمية»، والصواب أن علوم الحاسوب ب مختلف أدواتها ذات وجود سابق ومستقل عن تلك الحواسيب، ولا تفترض لعملها إلا توافر «وحدات» تقوم بإجراء العمليات الحسابية والمنطقية الأولية؛ فربما تكون هذه الوحدات آلات إلكترونية أو كهربائية أو حتى ميكانيكية بل ربما تكون بشرية؛ ومن الطريق في هذا المقام الإشارة إلى أن «وكالة الطيران والفضاء الوطنية: ناسا» الأمريكية الشهيرة كانت منذ تأسيسها في نهاية الخمسينيات من القرن الميلادي الماضي حتى متتصف ببعينياته بتوظيف كتاب من الموظفين المتخصصين في الرياضيات العددية الذين كانوا يعملون كحواسيب بشرية لمساعدة علماء الفلك وعلماء الصواريخ وعلماء الطيران حيث تقوم تلك الكتبات بتنفيذ خوارزمات للحصول على حلول عددية لمعادلات معقدة يحتاجها هؤلاء العلماء في تصاميمهم ويتعدى الحصول على حلول تامة مغلقة لها.

الثانية «الولايات المتحدة الأمريكية» ريادةً هذا المضمار^(١)، كما لم يكن مستغرباً أن تكون الاستخدامات الأولى لتلك الحواسيب استخداماتٍ عسكريةً وأن تعتبر هي نفسها أسلحةً حربيةً ضاربة^(٢).

وعلى الرغم من أن الوحدات الحسابية الأولية في تلك الحواسيب كانت ذات تقنية كهربائيةٍ ولم تتحول إلى إلكترونية بالكامل إلا بعد عقدين من الزمان فإن «مِعْمارها architecture» لا يزال في جوهره هو المتبع في بناء الغالبية العظمى للحواسيب التي نستخدمها حتى يومنا هذا. والقصةُ بعد ذلك من تحول الحواسيب للتقنية الإلكترونية بدلًا من الكهربائية ثم تضاؤل هذه الوحدات الأولية الإلكترونية وزيادة كثافتها في وحدة المساحة بنمطٍ أُسّيًّ، ومن ثم تضاعف القدرات الحسابية والتخزينية للحواسيب كل حوالي ثمانية عشر شهراً (وهو ما يعرف بقانون «موور»)، فضلاً عن التطور المطرد في وسائل إدخال وتصدير وعرض البيانات، واندماج الحواسيب مع تقنيات الاتصالات والشبكات الرقمية ... إلخ إلخ حتى نصل إلى وقتنا الحاضر هي قصةٌ معروفةٌ نعيشها ولا حاجة بنا للاستفاضة فيها في هذا المقام.

٥ . إذن، ما هو الذكاء الاصطناعي بالضبط؟ وما هو هدفه النهائي؟

بعد انتشار أخبار صناعة الحواسيب الرقمية القابلة للبرمجة وإسهامها في حل مسائل حسابية كبيرة، اشتعلت الحماسة وتجددت الآمال بين الأوساط العلمية في اقتراب تحقيق حلم العُصور الذي طال انتظاره بصناعة «أشباه للبشر» ذوي قدراتٍ ذهنية وحركية وانفعالية تشابه ما للإنسان من قدرات بما يتجاوز الدُّمَى الميكانيكية القديمة التي أشرنا إليها في القسم الفرعي ١,١ من هذا الفصل. فقد كانت الفلسفة المادية الجاحظة هي

١- يبرز في ملحمة مشوّعات بناء أول تلك الحواسيب اسم عالم الرياضيات والفيزياء النافذ «جون فون نوينان» مصمّماً أول و مدیراً لتلك المشروعات، إلا أن الأقل شهرةً والأسبق ريادةً هو المهندس الألماني «كونراد رويس» الذي كانت حاسيباته الأولى ميكانيكية كهربائية مختلطةً، ولعل هزيمة ألمانيا ومحورها في الحرب أمام الولايات المتحدة الأمريكية وحلفها هو ما أخفقت ذِكره.

٢- استُخدمت تلك الحواسيب أثناء الحرب العالمية الثانية بكثافةٍ كأدواتٍ لفك شفرة الرسائل العسكرية السرية، كما ساهمت بفعاليةٍ وقتذاك في مشروع «ماياباتن» لصناعة القنابل الذرية «الانشطارية» الأمريكية التي أتتت على «هروشبي» و «ناجازاكى» اليابانيتين في نهاية الحرب، ثم كذلك في تصميم القنابل الهيدروجينية «الاندماجية» الأشد فتكاً ضمن سباق التسلح المحتدم مع الاتحاد السوفييتي وحلفه الشيوعي والتي فَجَّرَت الولايات المتحدة الأمريكية أولها تجربياً بنجاح عام ١٩٥٢ م.

السائدة بين الأوساط العلمية في حقبة الأربعينيات والخمسينيات من القرن العشرين الميلادي، وكان مُنطَّلَقَ الحِبَاسُ أن تلك الحواسيب الرقمية يمكنها أن تحاكي دور الدماغ البشري الذي هو من وجهة نظر تلك الفلسفة المادية في النهاية مجرد آلية حسابية منطقية متطرورة، ومع زيادة القدرات الحاسوبية الرقمية تدرِّيжиًّا فإنها سوف تنافس الدماغ البشري وربما تُفُوقُه! كما أن الهندسة الميكانيكية متقدمةً (وقتذاك) بما يمكنها من حماكة الجسد بعد ربطها كهربائيًّا والتحكم فيها بواسطة الدماغ (الحواسيب الرقمية).

بعد وقت ليس بالطويل من المحاولة والبحث المعمق للمسألة استبانت الصعوبة البالغة لصناعة «شبيهٍ للبشر humanoid» حتى ولو بقدرات رضيع حديث الولادة؛ فالحاوسيات الرقمية كانت وقتها تختل أجنحةً كاملةً من مبني كبيرة، وتستهلك طاقةً كهربائيةً هائلةً، ومع ذلك فإن قدراتها الحسابية عند توظيفها لتنفيذ خوارزمات «البحث الشجري tree search»^(١) من أجل المفاضلة بين البدائل والتخاذل القرار يجعل تشغيل تلك الخوارزمات بطريقًا جدًّا؛ وبالتالي فإن استعجابة ذلك «الشبيه» لم يُشَرِّخَارجيًّا ما لن تبدي إلا بعد ساعاتٍ من حصوله. كما كان الجمع بين متانة الجسد واستدامته ومردنته وسلامة حركته، والحفاظ على اتزانه، وتوفير استقلاله من حيث الطاقة وتجددها ... إلخ كلها أمورًا متعددةً تماماً على الصعيد الميكانيكي. وكذلك كانت وسائل اكتساب البيانات من البيئة المحيطة ضوئيًّا أو صوتياً أو تلُّمِسًا أضعفَ كثيراً من أن تشابه البصر أو السمع أو الإحساس عند البشر، ناهيك عن تفسيرها وتحليلها للخلوص إلى معلومةٍ مفيدة. وعلاوةً على ذلك كله فلم تكن هناك آليات يُعوَّلُ عليها لاكتساب الخبرات وللتعلم التلقائي وهو ملمح جوهريٌّ يميّز البشر عن الآلات.

شم ماذا عن القدرات اللغوية لذلك «الشبيه البشري» وهي وسيلة التواصل التي لا بد منها كي تُقْنَع من سيسخدمه أو يتعامل معه أنه « بشري »؟ لم يكن لدى أحد وقتها أية فكرة عن ميكنة اللغة تتجاوز عملية الترميز^(٢).

نجمَت عن إدراك هذه العقبات كلها في ذلك الوقت الباكر عدُّ استجاباتٍ تهمنا أحدها في سياق هذا الفصل؛ ذلك أن العلماء المهتمين أيقنوا أن معالجة تحدي صناعة

١- سوف نأتي في القسم الفرعي التالي من هذا الفصل بشيء من التفصيل على شرح البحث الشجري.

٢- وذلك استلهاماً من إنجازات «نظرية الاتصال المعلوماتي Information theory» التي كانت قد تأسست على يد الأمريكي «كلود شانون» ثم استوت على عودها وقتذاك قبل حوالي عقد من الزمان.

«شبيه البشر» جُملةً واحدةً يكاد يقع في مقام الاستحالات، ولذلك يُؤُوا إلى «المعالجة بالتجزئة divide and conquer» وهي الاستراتيجية المُجربة نجاعتها مع أمثال تلك التحديات الكبرى. ومن وقتها استقر تقسيم العمل في تطوير أشباه البشر الآلين إلى الجناحين التاليين:

- «الذكاء الاصطناعي Artificial Intelligence» ويُشتهر كذلك بالاختصار «**AI**» ويعني إجمالاً بمحاكاة وظائف «المُخ» في رأس هذا الآلي.
- «الروبوتيات Robotics» وتعني إجمالاً بمحاكاة وظائف «باقي الجسم» في هذا الآلي.

ويطبق الجدول التالي هذا التقسيم على المسائل التي شكلت تحديات كبرى أمام صناعة «أشباه البشر» مما ذكرناه آنفًا في هذا القسم الفرعي:

مسائل الروبوتيات (في باقي الجسم)	مسائل الذكاء الاصطناعي (في الرأس)
اكتساب البيانات من البيئة المحيطة تائماً صوتيًّا	اكتساب البيانات من البيئة المحيطة ضوئياً أو صوتيًّا
الموازنة بين متانة الجسم وخففة حركته ومرونته	تحليل البيانات المكتسبة للخلوص إلى معلوماتٍ مفيدة
صيانة الجسم واستدامته	المفاضلة بين البدائل والتخاذل القرار
تخطيط حركة الأطراف وتحييئها لحظةً بلحظة	إدارة الذاكرة والاستدعاء الموضوعي لمحفوظاتها
سلامة حركة الجسم والحفظ على اتزانه	تحقيق التراكم المعرفي
استقلال الجسم الآلي في اكتساب الطاقة وتجددها	التعلم التلقائي
الدفاع عن الجسم في مواجهة الأخطار الخارجية	القدرات اللغوية تحليلًا وبناءً
مسائل وتحديات أخرى كثيرة ...	مسائل وتحديات أخرى كثيرة ...

الجدول ١: تصنيف أولٍ لمسائل صناعة «شبيه البشر» بين حقولي الذكاء الاصطناعي والروبوتيات وفيها نستأنف رحلتنا في هذا الفصل مع الذكاء الاصطناعي مُضيّقين بؤرة اهتمامنا أكثر لتركيز على معالجة الذكاء الاصطناعي لـ اللغات الطبيعية وخصوصاً العربية، فمن

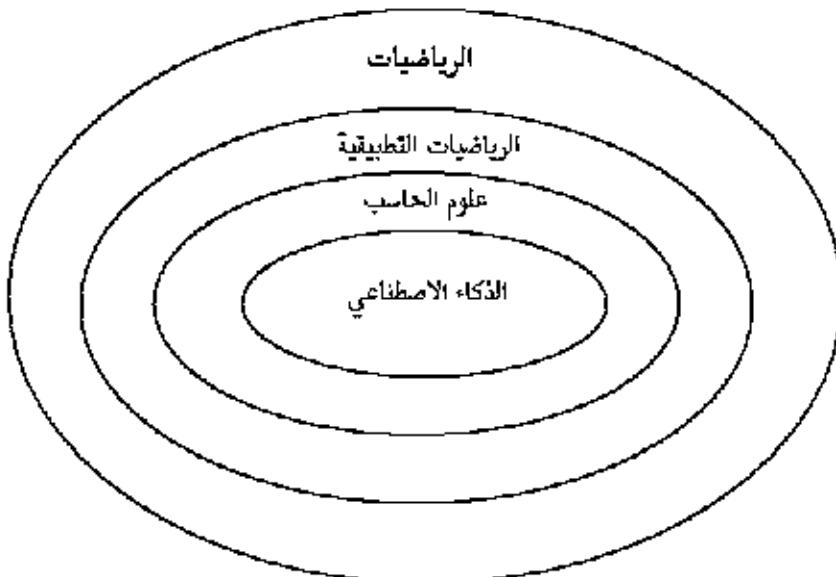
المفيد ونحن نودع «الروبوتات» أن نُلْمِح إلى أنها قد حققت الآن بعض النجاحات المتقدمة من حيث المواد التي تُصنَع منها وتحقيق المثانة وخفقة الوزن معاً، وكذلك من حيث تخطيط وتنفيذ الحركة؛ فلم تعد «الروبوتات» في زمننا هذا متقيدة حَصْرًا بالشكل البشري للجسد؛ فهناك الآن الكثير والكثير منها ترَكَ على تحقيق كفاءة الوظائف الحركية حتى ولو استلهمت ميكانيكيَّة وهيئة كائنات أخرى كالحشرات مثلاً، أو كأجزاء فقط من الجسد كالذراع مثلاً، أو حتى ككائنات مجهرية دقيقة.

ورغم أن كثيَّراً من المسائل في «الروبوتات» ما زالت تمثل تحدياتٍ كبيرةً أمام البحث العلمي والتطوير الهندسي، إلا أن «الروبوتات» قد استقرت الآن كمكون فائق التطور في بنية أنشطة وصناعاتٍ أساسية مثل صناعة السيارات والطائرات والإلكترونيات والطب واستكشاف الفضاء الخارجي والتعدين وزنَ الألغام، وبجدوى اقتصادية مرتفعة للغاية. وقبل مواصلة رحلتنا مع «الذكاء الاصطناعي» لا بد لنا من تقديم تعريفٍ لهذا المصطلح الذي سُكَّ للمرة الأولى عام ١٩٥٦ م واستقر من وقتها في مختلف الأديبيات العلمية والإعلامية والأدبية، وفيها يلي محاولة لصياغة مثل هذا التعريف:

- الذكاء الاصطناعي هو العلم الذي يشتغل بابتكار وتطوير خوارزمات مفيدة تسهم في المحاكاة الآلية لِقدْراتِ الدِماغ البشري؛ من إدراكٍ للبيئة المحيطة، والاستجابة المناسبة لمثيراتها، وتعلمٍ، وتخطيطٍ، وإيجاد حلولٍ للمسائل المستجدة، والتواصل اللغوي، وإدارةٍ للتراكم المعرفي، ... إلخ (ويُطْلِقُ البعض على هذه القدْراتِ وأمثالها «المَلَكاتُ الْعُلِيَا» لِلإِنْسَان) وينجُّ من هذا التعريف المسائل العلومُ لها تعريفٌ رياضيٌّ مُحْكَمٌ والمعلومُ لها حلولٌ مفيدةٌ مُبِرَّهَةٌ رياضيًّا.
- من أجل أن يكون لأي حلٍ / خوارزم وجود مفيد في علوم الحاسوب، لا بد ألا تتخطى درجة «صعوبته complexity» على مقياس «نظرية تعين الصعوبة Complexity theory» عتبةً معينةً؛ إلا فلن يكون مفيداً حيث لا يمكن عندئذٍ تفريغه على الحواسيب الرقمية. فعلى سبيل المثال؛ لو تزايدت العمليات الحسابية و/أو الذاكرة المطلوبة لتنفيذ خوارزم ما (التقديم خدمةً ما لعدد من المشتركين فيها مثلاً) بنمطٍ أُسَيًّا مع زيادة حجم المسألة (عدد المشتركين) فإن درجة صعوبته الأُسَيَّةَ تلك تجعل تفريغه حاسوبيًّا أمراً مستحيلاً مع نمو عدد المشتركين وبالتالي تendum إفادته.

ومن المهم للغاية ملاحظة أن هذا التعريف **متحركٌ في الزمن**؛ حيث يمكن وفقاً له أن تقع مسألة ما ضمن اهتمام الذكاء الاصطناعي في وقتٍ ما ثم تخرج منه لاحقاً إذا اكتُشفَ لها تعريفٌ رياضيٌّ مُحَكَّمٌ وحُلَّ مفهُودٌ مِنْهُ رياضيًّا. ومن الأمثلة على ذلك تصميم «لغاتٍ عالية المستوى High-Level Languages» لترجمة الحواسيب الرقمية وبناء «مترجماتٍ compilers» آلية لها.^(١)

ومن المهم كذلك في ختام هذا القسم الفرعي أن نعيّنَ موضع «الذكاء الاصطناعي» على خارطة العلوم بالتعرف على سلسلة نسبٍ فيها - انظر (الشكل ١) أدناه - فالذكاء الاصطناعي يتتمي لحقل «علوم الحاسوب» وهو فرع متقدمٌ من «الرياضيات التطبيقية» وهي بدورها جزءٌ من «الرياضيات». وبطبيعة الحال فإن هذا النسبَ لا يمنع الذكاء الاصطناعي من استلهام بعض الأفكار أو المبادئ من علومٍ أخرى مثل «الفيزياء»، أو «علم النفس الإدراكي cognitive psychology» أو «العلوم التربوية».



الشكل ١: نسبُ «الذكاء الاصطناعي» بين العلوم

١- وبسبب ارتباطها الوثيق بتطور «نمذجة اللغات الطبيعية Natural Language Modeling» فسوف توقف مع هذه المسألة المضروب بها المثال هنا بمزيدٍ من التفصيل في القسمين الثاني والثالث من هذا الفصل.

٦. نجاحاتٌ مبكرةٌ وتفاؤلٌ عريض

كما ألمحنا في نهاية القسم الفرعي ١,٣ من هذا الفصل فقد تزعم عالم الرياضيات الألماني النافذ «ريفييد هيلبرت» في القسم الثاني من عشرينيات القرن العشرين وبدايات ثلاثينياته الدعوة إلى الإجابة على التساؤل المركزي «هل يمكن تأسيس منظومة الرياضيات بجميع فروعها وأكملها على أساس من المنطق الرياضي المُمحكم من أجل ضمان اكتمالها وخلوّها من التناقض؟»، وجاءت الإجابة ما بين منتصف الثلاثينيات إلى منتصف الأربعينيات مفاجئةً ويمكن تلخيصها في جُزَّأَيْنِ:

أ. برهنَ «كيرت غيدل» رياضيًّا على استحالة جمع أية منظومة رياضياتٍ لخاصَّيَّتي «الاكتمال» و«عدم التناقض» معًا، وبذلك يَبيَّنُ أنَّ هناك حدودًا لا يمكن للمنطق الرياضي أن يخترقها.

ب. في إطار هذه الحدود يمكن - من حيث المبدأ - مِيَكَنَةً أي نَسْقٍ من المنطق الرياضي، وقد جاء ذلك نتيجةً لعمل «آلان تيورنُغ» بنموذجه الحاسوبي النظري «آلة تيورنُغ Turing Machine» وعمل «الأنزو تشيرش» بنموذجه البرامجي النظري «حساب لامدا Lambda Calculus».

وكان الجزء الثاني من الإجابة على سؤال «هيلبرت» وقدًا مُسْعِلًا لِحِمَاسِ كِبِيرٍ بعد انتهاء الحرب العالمية الثانية؛ ففي النصف الثاني من أربعينيات القرن العشرين والنصف الأول من خمسينياته بدأ عدد متزايد من علماء الرياضيات، والهندسة الكهربية، والميكانيكية، والاقتصاد، وعلم النفس، والعلوم السياسية في نقاشات مكثفة حول إمكانية صناعة دماغ آلي. وكان من أمثلة فعاليات هذا النشاط ونتائجها ما يلي:

- بعد أن أظهرت أحدث مُكَشَّفات تُشْرِيغ المنظومة العصبية في الحيوانات وفي البشر - في ذلك الوقت - أن المخ يتكون من شبكة كثيفة الترابط من الخلايا العصبية، دفعت هذه المكتشفات بعض الباحثين - مثل «مارفين مينسكي» أحد الآباء المؤسسين للذكاء الاصطناعي - إلى محاولة بناء شبكة عصبية اصطناعية كهربائيةٍ عسى أن تظهر بعض قدرات المخ الطبيعي، واستطاعت هذه الشبكة العصبية بعد عمل مُطَوَّلٍ في تصويرها إظهار بعض التصرفات التي تكافيء العمليات المنطقية الثنائية البسيطة.

- على صعيد «المنطق الرمزي» ابتكر كل من «آلن تيوبيل» و «هيربرت سايمون»^(١) في عام ١٩٥٥ م برنامجاً حاسوبياً سمياه «المُنَظَّرُ المَنْطَقِي» Logic Theorist «Logic Theorist» والذي استطاع في النهاية الوصول آلياً بدءاً بال المسلمات الأولية إلى برهان ٣٨ من أصل ٥٢ «مبرهنة theorems» وَرَدَتْ في عمل «برتراند راسل» و «آلفريد نورث وايت هيد» - الذي أشرنا إليه في القسم الفرعى ١,٣ من هذا الفصل - وهو النجاح الذي جعل «هيربرت سايمون» يصرّح بأن صناع هذا البرنامج قد توصلوا لحل اللغز القائم منذ القدام عن كيفية اكتساب آلة مادية لعقلٍ مفكّرٍ !
- في عام ١٩٥١ م أنسجز «كريستوف سترايتشي» في جامعة «مانشستر» ببريطانيا برنامجاً حاسوبياً يستطيع أن يمارس لعبة «الضَّامَة/ الدَّاما Checkers»، وفي السنة نفسها أنسجز «ديتريش برينز» - الذي كان وقتها يتعاون مع جامعة «مانشستر» أيضاً - برنامجاً يستطيع أن يمارس لعبة الشطرنج. وبالوصول إلى منتصف الخمسينيات من القرن العشرين كان الأميركي «آرثر لي صامويل»^(٢) قد أنسجز برنامجاً يستطيع أن يتحدى لاعباً هاوياً ذا مهارة معتبرة في «الضَّامَة».^(٣)
- اقترح «آلن تيورنگ»^(٤) في عام ١٩٥١ م اختباراً للآلات الحاسوبية المفكرة، وُيُعرَف هذا الاختبار باسم «اختبار تيورنگ Turing test»؛ حيث تخضع آلة آلة يزعم صانعوها أنها تمتلك تفكيراً ذاتياً إلى عدد من المخترين (وهم أشخاص عاديون) يدخل كُلّ منهم في حوارٍ تصيّر عن بُعد مع هذه الآلة بتناوبٍ عشوائيٍّ مع محاورٍ بشريٍّ ويُطلّب من المختبر كُلّ فترةً أثناء الحوار أن يقرر ما إذا كان

١- وهو الحاصل على جائزة نوبل في الاقتصاد عام ١٩٧٨ م.

٢- وهو من سُكَّ مصطلح «machine learning» وهو ما يمكن ترجمته إلى «التعلم التلقائي» أو «التعلم الحاسوبي» وقد اخترنا في هذا الفصل من الكتاب الترجمة الثانية، وسوف يلعب هذا المصطلح دوراً مهماً في ما يلي من هذا الفصل.

٣- ومن وقتها صارت القدرات التنافسية للبرامج التي تمارس ألعاب الذكاء (مثل الشطرنج) أمام منافسيها من البشر مقيماً لما وصل إليه مستوى التطور في مجال الذكاء الاصطناعي؛ فقد كان - على سبيل المثال - تغلب الحاسوب «Deep Blue» الذي طورته شركة «آي بي إم» في شهر مايو من عام ١٩٩٧ م على الروسي الذي افرد ببطولة العالم في الشطرنج عشرة أعوام متتالية «غاري كاسباروف» بـ «ثلاث جولات ونصف مقابل جولتين ونصف» علامةً مهمة على تطور حقل الذكاء الاصطناعي كله.

٤- وهو أحد الآباء المؤسسين لكلٍّ من «علوم الحاسوب» و «الذكاء الاصطناعي»، والذي ورد ذكره آنفًا أكثر من مرة في هذا الفصل.

محاوره الآن بشرّيًّا أم آليًّا، وبناءً على التحليل الإحصائي لإجابات المختربين تُعطى كُلُّ آلةٍ درجةً لتقييم ذكائتها. ورغم الانتقادات اللاحقة التي وُجّهت إليه، فلا يزال يُعدُّ «اختبار تيورنج» أول طرح جاد في «فلسفة الذكاء الاصطناعي».

مع تزايد هذا النشاط البحثي وترافق إنجازاته كان من المهم أن تتداعى الشخصيات العلمية البارزة وفتقها في هذا الخضم إلى تدارس هذا الحقل الجديد الوعاد تدريًّا مُطولاًً معمقاً، فكان أنْ عُقدَ مؤتمر «دارمُوث»^(١) في صيف ١٩٥٦ م الذي يُعدُّ بمثابة الميلاد الأكاديمي للذكاء الاصطناعي حيث سُكَّ المُضطَّلَّ.

وفي الحقبة التي تلت هذا المؤتمر ارتفعت موجة البحث والتطوير مدعاةً بمنَح ماليةً من «وكالة مشروعات الأبحاث الدفاعية المتقدمة (داربَا) DARPA» في الولايات المتحدة الأمريكية لعدد من كبرى الجامعات الأمريكية؛ مثل «معهد ماساتشوسيتس للتكنولوجيا إم آي تي MIT» وكذلك «جامعة ستانفورد» و «جامعة كالينيغي ميلون» مما مكّنها من إنشاء برامج أو أقسام أو حتى معاهد مكرسة لتخصص «الذكاء الاصطناعي»،

١- هو ورشة عمل طويلة استمرت لمدة شهر ونصف في جامعة «كلية دارتموث» في مدينة «هانوفر» بولاية «نيوهامشير» بالولايات المتحدة الأمريكية، وفيه اجتمع أحد عشر من علماء الرياضيات وعلوم الحاسوب للقيام بالنقاش والعرض الذهني المطّول. وقد صارَّ تُصُّنُّ المقترن الأصلي لورشة العمل وثيقةً تاريخيةً يستحق أن تُورّده في المليون بلي لفائدة موضوع هذا الفصل وقارئه:

“We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer”.

وترجمةً للعربية هي:

«نقترح العكوف على دراسة للذكاء الاصطناعي تمت لشهرين عشرة مشاركين تُعقدُ في صيف ١٩٥٦ م في كلية دارتموث في مدينة هانوفر بولاية نيوهامشير. وسوف تسير الدراسة على أساس الافتراض القائل بأن أي ملْمح من ملامح عملية التعلم أو أي ملْمح آخر للذكاء يمكن من حيث المبدأ وصُفُّه بدقةٍ تكفي لجعل الحاسوب قادرًا على محاكاته. كما سنقوم بمحاولة لاستكشاف كيفية جعل الحواسيب قادرةً على استعمال اللغة، وتكوين المفاهيم المجردة، وحل المسائل التي لا يقدر على معالجتها الآن إلا البشر، وتطوير الحواسيب لدواتها بذواتها. ونظن أنه يمكن إنجاز تقدم معتَبر في واحدةٍ أو أكثر من هذه المسائل إذا عكفت مجموعة من العلماء المتَّقِّنون بعنايةٍ على العمل عليها معًا طيلة الصيف»

كما يُذكَر «مختبر الذكاء الاصطناعي» بجامعة «أُدِنِرِه» في «اسكتلندا» في ذلك الوقت كمثال على مراكز الذكاء الاصطناعي الهامة خارج الولايات المتحدة الأمريكية.^(١)

ولذلك يطلق العديد من مؤرّخي العلوم اسم «السنوات الذهبية» للذكاء الاصطناعي على تلك الحقبة الممتدة من عام ١٩٥٦م إلى عام ١٩٧٤م، وفيما يلي سوف نحاول وضع القارئ في بعض أجواء ما ساد في تلك الفترة الحافلة من أساليب معالجة^(٢) المسائل المطروحة وبعض الإنجازات التي تحققت أثناء تلك «السنوات الذهبية»:

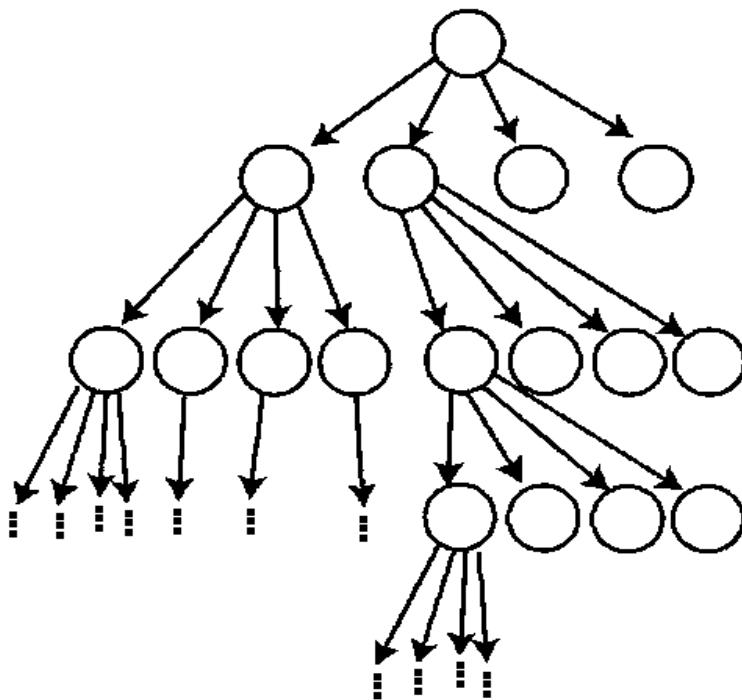
- صياغة ومعالجة المسائل كعملية بَحْثٍ شَجَرِيًّا: حيث يعتمد هذا الأسلوب على رُكْنَيْنِ أو لهما هو صياغة المسألة محل الدراسة على هيئة «حالات states» بدءاً بـ «حالة ابتدائية initial state» مصحوبةً بآلية لتوليد الحالات التي يمكن أن تقود إليها كل حالة ومع تعريف الحالات التي تُعَدُّ حَلًّا مقبولاً للمسألة. أما الركن الثاني فهو إجراء عملية «البحث الشجري tree search» التي تُفْتَسِح بالحالة الابتدائية للمسألة والتي إن لم تمثل حَلًّا مقبولاً يتم توليد الحالات التي تقود إليها الحالة الابتدائية، ومن ثم يجري اختبار كل حالة مولدة من حيث كونها حَلًّا مقبولاً أم لا، ثم يتواتي هذا التوليد والاختبار للحالات جُولَةً وراء أخرى حتى يتم الحصول على أحد الحلول المقبولة (أو حتى تحصيل عددٍ معَرَّفٍ ابتداءً من تلك الحلول) أو حتى يتم توليد عددٍ تراكميًّا أقصى معَرَّفٍ ابتداءً من الحالات وحينها يتم الإعلان عن عدم التوصل إلى حل للمسألة.^(٣)

ويصور (الشكل ٢) المجاور مثلاً لشجرة الحالات الناتجة عن عملية البحث الشجري؛ حيث تمثل كُلُّ حَالَةٍ على هيئة كرةٍ صغيرةٍ، وترسمُ الحالة الابتدائية (جذر الشجرة) عادةً في أعلى الشجرة، وتمثّل العلاقة بين كل حالة والحالات التي يمكن أن تتولد منها بأسهم في اتجاه هذا التوليد، وتُسمّى الحالات التي لا تخرج منها أية أسمهم «حالاتٍ طَرْفِيَّةٍ terminal states».

١- بالتوالزي مع ذلك كان اليابانيون في تلك الحقبة وهم يعذبون بناء الاقتصاد والصناعة في بلادهم بعد دمار الحرب العالمية الثانية يركزون جهودهم على تطوير الروبوتات الصناعية.

٢- والتي لا بدَّ للقارئ أن يُلْمِمَ بها من أجل استيعاب الموضع التي تستعرضها الأقسام التالية من هذا الفصل - وخصوصاً قسمُه الثاني - وقد اجتهدنا قدر الإمكان من أجل عرضها بوضوح وتبسيط غير مُحِلٍّ للقارئ غير المتخصص.

٣- يتم أيضًا الإعلان عن عدم التوصل إلى حلٍ للمسألة عند عدم توليد أية حالات جديدة.



الشكل ٢: مثال على «شجرة البحث» .
Search tree

وقد تفنن باحثو الذكاء الاصطناعي وتوسعوا في ابتكار خوارزمات البحث الشجري بهدف التوصل إلى حل / حلول مقبولة (إن وجدت) للمسألة بأسرع ما يمكن عبر الاقتصاد في توليد حالات المسألة قدر الإمكان بتوجيهه أولاً ناحية المسارات التي يتوسم فيها تحقيق أقصى اقترابٍ من الحل^(١)؛ فتجاوزوا أساليب البحث العميم مثل «البحث الشامل» exhaustive search وابتكرت خوارزمات بارعة؛ مثل «مونت كارلو» Monte Carlo و كذلك «A* search».

كذلك ابتكرت خوارزمات بحث شجري لاختيار أفضل اختيارٍ / نقلةٍ خاصةٍ بالألعاب التي يتنافس فيها الحاسوب مع خصم آخر - سواء كان بشرياً أو برنامجاً جاسوبياً آخر - بالتناوب (مثل الشطرنج) ومن أبرزها خوارزم «ألفا-بيتا» Alpha-Beta. الواقع أنه مع بدء سبعينيات القرن العشرين الميلادي صار

1- وهو ما يُعرف بـ «blind search» في مقابل البحث غير الموجه (الأعمى) الذي يُعرف بـ «heuristic search».

البحث الشجري وخوارزماته رُكناً أساسياً كبيراً من مباحث الذكاء الاصطناعي مؤسساً على صياغة رياضية مُحَكَّمة.

وبالاعتماد على خوارزمات البحث الشجري هذه طُورَتْ العديدُ من البرمجيات ذات القدرات المُبَهِّرة «آنذاك»؛ فاستطاع بعضُها أن يواحدَ - ولو بأداءٍ أوَّلِيًّا - خصائصَ الألعاب التناوبية، كما تمكنَتْ أخرى من حلَّ بعض المسائل في الجبر والهندسة الإقليدية^(١)، وأخرى تستطيع إجراء «التكاملات بالرموز symbolic integrals^(٢)»، وأخرى لخطيط حركة الروبوتات^(٣).

• ميكانة المنطق الرياضي والتَّوسيع في حوسبيته: بينما شهدت العقود الأخيرة من القرن التاسع عشر والعقود الأولى من القرن العشرين الميلادي تَنظِيرَ المنطق تَنظِيرًا رياضيًّا عميقًا فاستُكشِفَتْ حدودُه وصار من الفروع النظامية الأساسية للرياضيات - كما أشرنا في القسم ١,٣ آنفًا - فإن «السنوات الذهبية» قد شهدت حوسبية هذا المنطق الرياضي^(٤) بتوسيعٍ ليصير أحد الأدوات الرئيسية التي يرتكز عليها حقل «الذكاء الاصطناعي» ويوظفها بكثافةٍ في تطبيقاته التي دَعَّمت الثقة في القدرات الوعادة لهذا الحقل الناشئ؛ حيث ظهرت برمجيات تَمكَنَتْ ابتداءً من «المسلمات الأولى axioms» من الوصول إلى إثباتات بعض «مبرهنات الهندسة الإقليدية والجبر الأوَّلِيّ وحلَّ بعض مسائلِها theorems»

١- ومن أمثلتها «Geometry Theorem Prover» من تطوير «هيربرت غيلبرت» أستاذ «علوم الحاسوب» بجامعة ستوني بروك في «نيويورك» في الولايات المتحدة الأمريكية عام ١٩٥٨ م.

٢- ويعرف أحد أمثلتها الباكرة باسم «Symbolic Automatic INTegrator (SAINT).

٣- ومن تلك البرمجيات «STRIPS» الذي جرى تطويره في جامعة ستانفورد الشهيرة في الولايات المتحدة الأمريكية عام ١٩٦١ م والذي جرى توظيفه كاداةٍ للتحكم في سلوك الروبوت «شاكى Shakey» الذي كانت قد طرطره الجامعة وقتذاك.

٤- المنطق الرياضي المقصود هنا لا يشمل فقط «المنطق الأُرْسْطِيّ» الذي من أمثلته: المقدّمان <<كل إنسان فان>> و <<سالم إنسان>> فتَسْتَعِي ذلك التَّبيَّنة <<سالم فان>>، فهذا المنطق المعروف أيضًا باسم «المنطق اللُّفْظِيّ / الكلامي propositional logic» يُعَدُّ في منظومة المنطق الرياضي الآن بمثابة «الدرجة الصُّفرية first order zeroth order» وهو ذو قدرة محدودة في التعريف والاستنتاج. أما المنطق الرياضي من «الدرجة الأولى first order» المعروف أيضًا باسم «Predicate Calculus» فهو أقوى كثيراً في قدرته على التعريف والاستنتاج والاستنتاج والتعليم بسبب اشتتماله على دوالٍ متغيراتٍ وهو الأَكْثَر شيوعاً في البرمجة المنطقية الرياضية الآن، كما أن هناك المنطق الرياضي من درجاتٍ أعلى (ثانيةً فما فوق) الذي يسمح بأن تكون بآيَّمْرَاتٍ دَوَالَةً دَوَالَةً في حد ذاتها.

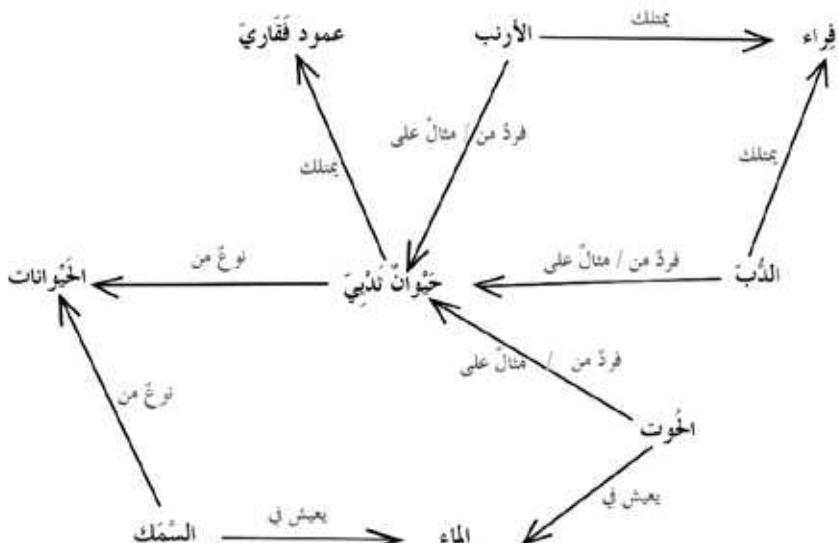
كذلك بالتزاوج مع خوارزمات البحث الشجري كما أشرنا في الفقرة السابقة، وتوّج ذلك مع نهايات تلك الحقبة - وتحديداً عام ١٩٧٢ م - بميلاد لغة برمجة الحواسيب المعروفة باسم «Prolog» - اختصاراً لعبارة «البرمجة عبر المنطق» Logic Programming - التي تيسّر لمطوري البرمجيات استدعاء وتوظيف أدوات المنطق الرياضي في برمجياتهم بسلاسة.

- المحاولات الأولى لمعالجة اللغات الطبيعية: فيما كان طبيعياً أثناء تلك السنوات الذهنية للذكاء الاصطناعي أن تكون ميكنة تحليل وتوليد «اللغات الطبيعية natural languages» في بؤرة اهتمام الباحثين، فقد كان لافتاً وقتها ذلك النضج المبكر التمثيل في تجاوز المستويات السطحية الشكليّة لِللغة الطبيعية والولوج مباشرةً إلى تناول مستوياتها العميق وتحديداً المستوى «الدلالي semantic» و«الأُنطولوجي ontological».

فليّاً كانت «نظرية الاتصال المعلوماتي information theory» التي تأسست واستقرت قبل تلك الحقبة بعُقدين وحققت نجاحات هائلة في هندسة الاتصالات وعلوم الحاسوب قد عالجت مسائل الترميز على مستوى متقدم وطورت في ذلك السياق آلياتٍ رياضيةً رفيعةً، فإن الباحثين نظروا إلى المستويات الشكليّة من اللغة الطبيعية (مثل «النَّسق الْكَتَابِي orthography»، و «النَّسق الْصَّرْفِي morphology») على أنها مسائل ترميزية تسهل معالجتها بنفس آلياتِ نظرية الاتصال المعلوماتي. وبالتالي مع ذلك كان العمل في ستينيات ذلك القرن جارياً على قدمٍ وساقٍ في تطوير لُغاتٍ برمجيةً للحواسيب الرقمية تشبه إلى حدٍ كبيرٍ لُغةً توصيف الخوارزمات الرياضية لتسير عملية تطوير البرمجيات وإبعادها عن تفصيلات «العتاد hardware» الإلكتروني الذي تتكون منه الوحدات الحسابية المنطقية ووحدات التحكم ووحدات الذاكرة في مختلف تلك الحواسيب.^(١) وكان من لوازם ذلك العمل تطوير تقييداتٍ

١- وذلك من أجل توسيع قاعدة البرمجين، ورفع إنتاجية عملية تطوير البرمجيات، وتقليل الأخطاء فيها، وكذلك - وربما كان هذا هو الأهم اقتصادياً - جعل نفس البرمجيات قابلة للاستخدام على حواسيب رقمية مختلفة في معيار عَتَادها الإلكتروني وهي الخاصية المعروفة باسم «portability». وستتوقف مجدداً مع لغات البرمجة هذه في القسمين الثاني والثالث من هذا الفصل.

وأدواتٍ رياضيةٍ رفيعةٍ لمعالجةِ العديد من الجوانب المعجمية والنحوية من اللغة - إضافةً إلى الجوانب الترميزية التي أشرنا إليها أعلاه - وهو ما افترض باحثو الذكاء الاصطناعي - بتفاؤلٍ مُفرطٍ - وقتذاك أنها قابلةٌ أيضًا للتطبيق على اللغات الطبيعية.



الشكل ٣: مثالٌ على شَذْرَةٍ من شبَكةٍ دلَاليةٍ

ولذلك انصبت أبحاثُ الذكاء الاصطناعي في معالجةِ اللغات الطبيعية على مستواها الدلالي حيث تجتمع الوحدات الترميزية لِلغة («فرداتها» على سبيل المثال) تحت «مفاهيم concepts» ثم تُدرَس وتوصَّف العلاقات بين هذه المفاهيم بروابطٍ دلَاليةٍ فيكون حاصلها ما استقرت تسميته بـ «الشبكة الدلالية semantic network»، ويصوَّر (الشكل ٣) أعلاه شذْرَةً صغيرةً مبسطةً من مثل تلك الشبكة الدلالية. وينسبُ ابتكار الشبكة الدلالية المحوسبة إلى «ريتشارد ريتشنز» عام ١٩٥٦ من جامعة «كامبريدج» الإنجليزية لتكون جسراً بين اللغات المختلفة في الترجمة الآلية.

أما الدرس الأونطولوجي فيعني بوصف العالم الواقعي حاسوبياً وربطِ وحدات اللغة بهذا النموذج الحاسوبي للعالم، وقد كان ذلك - ولا يزال - تحدياً كبيراً حتى لو

لم يكن هذا العالم الواقعي المطلوب وصفه حاسوبيًّا هو العالم الحقيقي بأكمله بل فئةً جزئيةً محدودةً منه.^(١)

ومن أمثلة التطبيقات التي أظهرت وقتها تعاوًلاً لافتًا مع اللغة الطبيعية برنامج حاسوبيٌّ سُميَّ «الطالب STUDENT» طوره «جورج بُوبُرو» ضمن أطروحته للدكتوراه في «معهد ماساتشوسيتس للتقنية MIT» عام ١٩٦٤ م بقدرته على حل بعض المسائل الكلامية في الجبر من مستوى المدارس الثانوية، وبرنامج حاسوبيٌّ طَرِيفٌ آخرٌ سُميَّ «إليزا ELIZA» طُور في الفترة بين ١٩٦٤ م و ١٩٦٦ م أيضًا في «معهد ماساتشوسيتس للتقنية» يستطيع إجراء محاوراتٍ «نصيَّةً» مع مستخدمين بشريِّين ظنَّ أحدهُم أنَّ محاورهم بشريٌّ وليس آلةً حيث بدا أداء «إليزا» طبيعياً للغاية وإن كان ذلك نتيجةً استخدام بعض الحيل البسيطة كالآجوبة المعلبة العامةً وكذلك تكرار أسئلة المستخدمين في الردود عليهم.^(٢)

• التوجُّه لدراسة الذكاء الاصطناعي في ظُروفٍ مَعْمَلِيَّة مقيَّدةٍ: وهو المَلْمَحُ الأخيرُ الذي نوهَ إليه هنا من تلك الملامح التي صبغت أجواء تلك «السنوات الذهبية» للذكاء الاصطناعي؛ حيث نادت في نهاية السبعينيات من القرن العشرين الميلادي أصواتٍ مسمومةٍ في حقل الذكاء الاصطناعي من أبرزها «مارفين مينسكي»^(٣) و«سايمور بايرت» وكلًاهما من «معهد ماساتشوسيتس للتقنية MIT» إلى الانكباب على تطوير خوارزميات الذكاء الاصطناعي - وهو ما زال حقلًا ناشئًا - على مسائل كائنةٍ في بيئةٍ مثالىٍّ (افتراضية)^(٤) إلى أن تُفهم

١- هناك بالطبع مستوىً آخر للغة وهو «اللغة المنطقية speech» استباقاً ونُفَقَّاً، وقد اخذت معالجة اللغات الطبيعية وقذفها على هذا المستوى - لحسن الحظ - مساراً مختلفاً بعيداً عن الذكاء الاصطناعي، وقد لعب ذلك المسار دوراً مهماً فيها بعد، سوف نأتي على ذكره في الأقسام التالية من هذا الفصل وخصوصاً من الثاني للرابع.

٢- وهذا مثالٌ على التقدِّم الوجَّه لاختبار «تيورينغ» لتعيين درجة ذكاء الآلات المفكرة المشار إليه آنفًا في هذا القسم الفرعى.
٣- وهو أحد الاثنين الذين دعَاوا إلى مؤتمر «دارتموث» الذي أشرنا إليه آنفًا في هذا القسم الفرعى، وهو كذلك أحد مؤسسي مختبر الذكاء الاصطناعي في «معهد ماساتشوسيتس للتقنية MIT»، وينظر إليه الآن على أنه أحد الآباء المؤسسين لهذا الحقل.

٤- وكانت الحجة القوية لأصحاب هذه الدعوة أن هذه المقاربة قد أتَتَتَ بنجاحٍ في العلوم الأخرى؛ فالفيزياء النظرية على سبيل المثال قد استخدمت مفاهيم مثالىَّة كالغازات المثالىَّة، والسوائل المثالىَّة، والسطوح المثالىَّة، والجسم الأسود، ... إلخ لتطوير قوانينها الأساسية، ثم استُخدِّمت هذه القوانين فيما بعد بنجاحٍ عريضٍ في ما يصعب حصره من التطبيقات الهندسية بعد الأخذ في الحسبان تأثيرَ الفوارق بين الافتراض المثالى والواقعُ الحقيقى.

ديناميكياتها بعمق وإلى أن تنجذب كفاءة وموثوقية هذه الخوارزمات وحينئذٍ يمكن تطبيقها على تلك المسائل في بيئاتها الحقيقة بتفاصيلها المعقدة. وكانت البيئة المثالية المُجرَّدة المفترحة هي كُتلٌ من أشكال وأحجام وألوانٍ مختلفةٍ موزَّعةٍ على سطحٍ مُستَوٍ، وبالفعل انصبَّ الكثير من النشاط البحثي على معالجة المسائل في إطار هذه البيئة المثالية. واستطاع «مينسكي» و«بابيرت» الداعيان لهذه المقاربة تطبيقها لبناء ذراعٍ رُوبوتيةٍ تمكنَت بنجاحٍ من التقاط ورصف هذه الكتل في تشكييلاتٍ مُعَقَّدةٍ نسبيًّا.

وربما كان النجاح الذي توجَّ كل ما سردناه آنفًا من ملامح تلك المرحلة هو البرنامج الحاسوبي «SHRDLU» الذي طوره بين عامي ١٩٦٨ و ١٩٧٠ م «تيري آلين وينوغراد» أستاذ علوم الحاسوب في «جامعة ستانفورد» الأمريكية والذي كان يستطيع التواصل بالإنجليزية الاعتيادية مع مُستخدِّميه ويستقبل طلباتهم بأداءً مهامًّا معينةً لمناولة الكتل في البيئة المثالية الموصوفة آنفًا، ومن ثم يقوم بتحطيم خطوات تحقيق المناولة المطلوبة وتنفيذها على شاشة رسومية.

وفي خضمٍ تلك الحقبة كان التفاؤل بمستقبل الذكاء الاصطناعي مفرطًا ويكفي ذكر بعض اقتباسات من تصريحات كبار الباحثين وقتذاك كي يلمس القارئ مدى الاندفاع في ذلك التفاؤل:

- ذكر «هيربرت سايمون ألكساندر»^(١) و «آلين نيويل»^(٢) في ورقةٍ نُشرت لها في دورية «بحوث العمليات Operations Research» عام ١٩٥٨ م أنه «في غضون عشر سنواتٍ سوف يكون بطل العالم في الشطرنج حاسوبًا رقميًّا، وفي غضون عشر سنواتٍ أيضًا سوف يكتشف حاسوبٌ رقميٌّ مُبرهنٌ رياضيًّا رئيسيةً جديدةً وثبتت صحتها».
- أكد «هيربرت سايمون ألكساندر» في مؤلفه «طبيعة ميكانة الأعمال من أجل البشر والإدارة» المنشور عام ١٩٦٥ م أنه «في غضون عقدين من الزمان سوف تصير الآلات قادرةً على أداء أي عمل يستطيع الإنسان أداءه».

١- وكان أستاذًا بجامعة «كارنيجي ميلون» الأمريكية، وهو أحد الباحثين الرواد والمفكرين النافذين في حقل الذكاء الاصطناعي في حقبته الأولى، وهو أيضًا الحاصل على جائزة نوبل في الاقتصاد عام ١٩٧٨ م، وقد سبق ذكره في هذا الفصل.

٢- وكان أستاذًا في «علوم الحاسوب» و«علم النفس الإدراكي» بجامعة «كارنيجي ميلون»، وقد سبق ذكره في هذا الفصل.

- أكد كذلك «مارفين مينسكي» في مؤلفه «الحوسبة: آلات محدودة ولا نهائية» المشور عام ١٩٦٧ م أنه «خلال جيل واحد فقط ... فإن مسألة تخلق ذكاءً اصطناعيًّا سوف تكون قد حلّت إلى حدٍ بعيد».
- وفي مقابلة له عام ١٩٧٠ م مع «مجلة الحياة» Life Magazine الأمريكية الشعبية النافذة وقتذاك صرّح «مارفين مينسكي» أنه «في غضون من ثلاثة إلى ثماني سنوات فإننا سنتمكّن من بناء آلية ذات مستوى ذكاءً عاماً مساوٍ لذكاء الإنسان العادي».

فهل صَدَّقَت الأحداث التالية ذلك التفاؤل الكبير؟

٧. السنوات العجافُ، ونَقْدُ السنوات الذهبية.

لا رَيْبَ أن القارئ الذي صحّبنا حتى هذه النقطة من هذا الفصل قد لاحظ أن مَهْدَ الذكاء الاصطناعي وَمَرْكَزُ الثقل الأَكْبَر لأنشطة البحث والتطوير فيه من خمسينيات إلى سبعينيات القرن العشرين الميلادي هو الولايات المتحدة الأمريكية^(١) حيث تحصل الأبحاث الجامعية على غالب تمويلها من المُجَمَع الصناعي العسكري. ولذلك فإن استمرار تمويل هذه الأبحاث مشروطٌ بتحقيق هذه الأبحاث لمستهدفات تطوير التقنيات الصناعية و/أو العسكرية التي وَعَدَتْ بها وَقْتَ حصولها على تلك التمويلات؛ فعلى سبيل المثال كانت «وكالة مشروعات الأبحاث الداعية المتقدمة (دارِيَا) DARPA» الأمريكية - وهي إحدى كُبُريات الجهات الممولة لهذه الأبحاث - تحت ضغط السياسيين كي تَقْصُرْ تمويلتها على المشروعات ذات المهام العَمَلَية المحدَّدة بوضوح والتي تتلزم بتنفيذها بدِقَّةٍ لا سيما بعد تمرير التعديل القانوني في هذا الاتجاه المنسوب للسيناتور الأمريكي «مانسفيلد» سنة ١٩٦٩ م، وبطبيعة الحال فإن أكبر الضحايا جَرَاءَ ذلك هي الأبحاث الاستكشافية المفتوحة التي تُعَدُّ وفق هذا التَّوْجِهِ أنشطةً «نظريَّةً بَحْثَةً»!

١- تليها بمسافةٍ غير قصيرة حلينتها العسكرية والاقتصادية الأولى في تلك الفترة «المملكة المتحدة».

وفي حين أخذت «السنوات الذهبية»^(١) تقترب من نهايتها، فإن أجواءً متصاعدةً من الإحباط بل السخط كانت تصاعدُ في أوساط الجهات الممولة لأبحاث الذكاء الاصطناعي^(٢)، وبالوصول إلى عام ١٩٧٤ انتهت تلك الجهات الممولة إلى قناعةً بأن حقل الذكاء الاصطناعي قد فشل فشلاً ذريعاً في تحقيق الآمال العريضة المعلقة عليه قياساً على التوقعات ذات السقوف العالية التي كانت تروّجها بشقة كبار الشخصيات المتقدمة للمشهد البحثي - مثل تلك التي وردت في ختام القسم الفرعي السابق. ومنذ ذلك الوقت ولعدة سنواتٍ لاحقةٍ توقفت المنح التمويلية لأية مقرراتٍ بحثيةٍ في إطار «الذكاء الاصطناعي» توقفاً شبه كاملاً استمر حتى سنة ١٩٨٠ م.

بالرغم من الإنجازات العديدة المبهرة التي حققها هذا الحقل الناشئ في تلك الحقبة - مما أشرنا إليه في القسم الفرعي السابق - والتي روجت لقدراته في المستقبل على صنع المعجزات، فإن إبهارها ذلك قد نبع من مقارنتها بلا شيءٍ قبلها، أما بالمقارنة مع التوقعات المُسِرِّفة في التفاؤل التي كانت تُطلُّق آنذاك بالوصول إلى حلولٍ ناجعةٍ في وقت قريبٍ للمسائل الكبرى المتصلة بمحاكاة الذكاء البشري فإن تلك «الإنجازات» تُعدُّ بمثابة «ألعاب الأطفال» على حد وصف تقارير رسمية صدرت في نهاية ستينيات وبداية سبعينيات القرن العشرين الميلادي لرفع واقع أبحاث الذكاء الاصطناعي على وجه العموم أو بعض مسائله على وجه الخصوص؛ ومن أمثلتها المؤثرة: «تقدير آبلاك» (Automatic Lighthill) عام ١٩٧٣ م لتقييم كامل أوسع أبحاث الذكاء الاصطناعي في بريطانيا. لم تقتصر تلك الانتقادات القاسية وفتئِد على التقارير الرسمية بل إن الساحة العلمية والفكرية حفلت بالعديد من الفلاسفة والكثير من كبار المختصين في الرياضيات وعلوم الحاسوب والاتصالات وعلم النفس الإدراكي ... - بل وحتى باحثين في الذكاء الاصطناعي

١ - وهي الفترة بين أعوام ١٩٥٦ م و ١٩٧٤ م كما أسلفنا في القسم الفرعي السابق.

٢ - ومن أهمها في تلك الفترة بالإضافة إلى «داريا» الأمريكية: «مجلس الأبحاث الوطني» (NRC) في الولايات المتحدة الأمريكية، و«مجلس الوزراء البريطاني» في المملكة المتحدة.

نفسه - من تصدّوا النقد آليات الذكاء الاصطناعي والتبنّي في ضوء هذا النقد لمزاعم بعض نجوم الذكاء الاصطناعي بقدرته خلال وقت قريب على محاكاة الذكاء البشري.

وبينما أدت تلك الحملة المكثفة التي سُطّرت فيها أكواًم من المقالات والأوراق البحثية والكتب إلى انتقاصٍ كبيرٍ من مصداقية حقل الذكاء الاصطناعي استمر قائمًا لعدة سنوات تالية، فإن ما أنتجه من تشخيصٍ تفصيليًّا لمشاكل الأسلوب المتّبع وقيّد في هذا الحقل - والذي استقر من وقتها في أدبيات علوم الحاسوب عمومًا والذكاء الاصطناعي خصوصًا - كان معلّمًا فارقاً في ترشيد مسيرة هذا الحقل وتطويره كثيرًا فيما بعد؛ بل استمر أثره حتى وقتنا الحاضر. وفيما يلي نوجز أهم تلك العقبات التي كان على الذكاء الاصطناعي أن يضعها بعين الاعتبار ومن ثم يحاول تجاوزها حتى يواصل مسيرته نحو تحقيق إنجازاتٍ صناعيةٍ وخدّاميةٍ ملموسةٍ ذات آثارٍ اقتصاديةٍ ومعرفيّةٍ ملموسة في العالم الواقعيٍّ: ^(١)

- كان «البحث الشّجيريُّ» السلاح الأساسيّ الضارب الذي وظّفه الباحثون آنذاك لتوسيف وحلّ الكثير من - أو بالأصح «أغلب» - المسائل التي تصدّروا لها، وفي حين أن هذا الأسلوب يتمتع بميزات: «البساطة كنمودج رياضيٌّ حاسوبيٌّ»، و «سهولة التنفيذ براجبيًّا»، و «العمومية» ^(٢)، ولكنه في الوقت ذاته يعاني من عيوبٍ ثلاثةٍ (على الأقل) يطعن كل منها بشدّةٍ في فعاليته:

الأول: التضخمُ الأُسْيُّ للعدد الذي يجري توليدِه من حالات المسألة في شجرة البحث؛ فإذا رُمِّزَ لمتوسط عدد ^(٣) الحالات التي تولّدها كل حالةٍ في شجرة البحث بالرمز M وإلى العمق المطلوب استكشافه حتى الانتهاء من المسألة

١- وتحصّ جمّع هذه العقبات خصوصيّةً وثيقّةً المعالجة الحاسوبية لـ«اللغات الطبيعية»، ولا مفر من فهمها بعمق ومن ثم تقادها أو ترويضها لمن أراد أن يخوض بجدية مضمار هذه المعالجة من باب البحث والتطوير. ولذلك تُبرّر هذه العقبات عَزْرَ تقدّمَ موجز لبعض الآليات الأساسية التي ارتكز عليها هذا الحقل في تلك الحقبة، وقد قدمنا لبعض هذه الآليات يالجاري في القسم الفرعي السابق من هذا الفصل، وسوف نعاود في هذا الفصل ولا سيما قسميه الثاني والثالث التعرّض لثلاث العقبات والتعامل معها.

٢- بمعنى أن خوارزم البحث الشجيري ذاته يعاد استخدامه مع مسائل مختلفة ويجري فقط تغيير «الإجراء» *routine* المسؤول عن توليد حالاته، وكذلك الإجراء المسؤول عن اختيار كل حالةٍ من حيث كونها حالًّا مقبولاً للمسألة أم لا.

٣- المقصود هنا ليس «المتوسط الحسابي arithmetic mean» ولكن المقصود هو «المتوسط الهندسي geometric mean» H الذي نحصل عليه لعددين d_1, d_2, \dots, d_n وفق الصيغة الحسابية: $H = (d_1 \times d_2 \times \dots \times d_n)^{\frac{1}{n}}$ ، وعلى نحو العموم فإن المتوسط الهندسي لمجموعة أعداد $\{d_1, d_2, \dots, d_n\}$ يُحسب وفق الصيغة: $H = (d_1 \times d_2 \times \dots \times d_n)^{\frac{1}{n}}$.

بالرезультат فإن إجمالي عدد الحالات المطلوب توليدها واختبارها أثناء البحث = $(م - 1) \div (م - 1)$ ^(١) فإذا لم تكن المسألة «ضحلة» العمق، فإن أي حاسوب ستتقاضر قدراته سريعاً عن التوغل في عمق هذا البحث مما عَظَمْت تلك القدرات الحاسوبية. ^(٢)

الثاني: في أغلب المسائل غير التافهة المُنْدَجَة في صيغة البحث الشجري فإن حجم وعمق شجرة البحث حتى العثور على حلٌ للمسألة (أو استيفاء حلولها) ليس معلوماً قبل إجراء عملية البحث، ولذلك فإن التوقف عن البحث عند حد معين دون العثور على حلول بسبب نفاد الزمن المحدد أو الموارد الحاسوبية يُنْتَج إجابةً «غير يقينية undecided» بين وجود أو عدم وجود حلول للمسألة.

الثالث: لا يمكن نَمْدَجَة جميع المسائل ذات الأهمية في هذا الحقل في صيغة البحث الشجري.

• كذلك كانت «حوسبة المنطق الرياضي» ^(٣) سلاحاً ضارباً أساسياً آخر في ترسانة الذكاء الاصطناعي آنذاك، ورغم النجاحات المبكرة التي حققها بالتوليد الآلي لإثباتات بعض «المبرهنات الرياضية theorems» - مما أغري أسماءً كبيرةً في هذا الحقل بالزعم أن ذلك يمثل شوطاً كبيراً في رحلة محاكاة الذكاء البشري بل

١- إذا افترضنا أنه في مسألة ما - على سبيل المثال - كانت $M = 3$ فإنه عند عمق $U = 5$ تكون $J = 121$ حالة، وعند عمق $U = 10$ تكون $J = 29524$ حالة، أما عند عمق $U = 15$ تتجاوز قيمة J السبعة ملايين حالة، وبالتالي الوصول إلى عمق $U = 20$ تتجاوز قيمة J ألف وسبعمائة وأربعين مليون حالة! وهذا فإن أي حاسوب مهما عظمت قدراته سوف يعجز سريعاً مع تزايد عمق البحث، ولن يسعفنا حتى «قانون مور Moore's law» (الذي ألمحنا إليه قرب نهاية

القسم الفرعي ٤، من هذا الفصل) الذي يصف نمط زيادة قدرات الحواسيب مع مرور السنين.

٢- تُثْبِتُ أَفْضَلُ حُوازِّمَاتِ الْبَحْثِ الْعَدِيدَةِ (التي ألمحنا إلى بعضها آنفًا بعد الشكل رقم ٢) لترويض الانفجار الأُسْيَّ الملازم للبحث الشجري في تقليل ذلك العدد الإجمالي بما يك足 تقليل معامل توليد الحالات M ، لكن نمو هذا الإجمالي يظل ذات نمط أُسْيَّ مما يؤجل قليلاً استنفاد القدرات الحاسوبية مع زيادة عمق البحث لكنه لا يغير نمط النمو إلى نمط آخر أهداً.

٣- وتحديداً «المنطق الرياضي من الدرجة الأولى Calculus of Predicates» - وفي باطنه «المنطق الرياضي من الدرجة الصفرية» الذي يُعرَفُ أيضاً بـ «المنطق الكلامي» أو المنطق الأُرْسُطِي Propositional Logic «الحالات خاصة بسيطة» منه - حيث يَؤُولُ التحليل المنطقي الرياضي المحوسَبُ إلى عملية تُسمَى «النَّفْنِيد refutation» التي تكون بدورها من شقَّنَ يُسَرَّانَ آنِيَّاً؛ ففي الشق الأول تتولد شجرة بدائل «التبيرات المنطقية الرياضية predicates» ومساراتها، وفي الشق الثاني يتواصل البحث في هذه الشجرة أولاً بأول حتى الوصول إلى «العبارة المنطقية الحالية empty clause» حيث تنتهي عملية النَّفْنِيد حيثَ بَأنَ العبارة المنطقية الرياضية المطلوب إثباتها يصبح ثبوتها منطقياً من المسلمات الابتدائية.

أسلفنا - فإن هذا المنطق الرياضي المحوسّب يعاني أيضًا من ثلاثة عيوب (على الأقل) يطعن كل منها بشدّةٍ في نجاعته:

الأول: معضلة «التضخم الأسّي exponential growth» في توليد شجرة التعبيرات المنطقية الرياضية، وهي المعضلة نفسُها التي ناقشناها في نقد البحث الشجري آنفًا.

الثاني: إذا انتهت عملية «التحليل المنطقي» - المسماة أيضًا بـ «التفنيد refutation» - بالعثور على «العبارة المنطقية الخالية empty clause» فقد ثبّتت العبارة المنطقية المطلوب إثباتها وحُلّت المسألة^(١)، ولكن إذا لم يُعثّر على تلك العبارة المنطقية الفارغة فلا سبيل إلى تقرير ما إذا كانت العبارة المطلوب إثباتها صحيحة أم خاطئة استنادًا إلى المسلمات الابتدائية. وبالتالي فلا ضمانةً أن هذا الأسلوب سيتهي دومًا بالبُلْت في صحة المسألة المطلوبة أم عدمها.

الثالث: لا يمكن تَمْذَجَةُ جميع المسائل ذات الأهمية المعتبرة في هذا المضمار على أنها تعبيرٌ منطقيةٌ رياضيةٌ يُراؤ إثباتها عبر آلية المنطق الرياضي المحوسّب.

وينبغي عند هذه النقطة التنويه إلى أن هناك آلياتٍ وخوارزماتٍ أخرى لا يتسع المقام لسردها جيًعاً وظفها الذكاء الاصطناعي في محاولة حل المسائل فيها سُمّيَ بسنواته الذهبية من منتصف خمسينيات إلى مطلع سبعينيات القرن العشرين الميلادي، وهي آلياتٍ تشتَرك مع «البحث الشجري» و «المنطق الرياضي المحوسّب» في العيوب الثلاثة الأساسية التي ذكرناها، ولكننا مثَلنا بهاتين العائلتين من الآليات الحاسوبية لكونهما الأبرز والأعم والأكثر شيوعًا فيما استخدمه الذكاء الاصطناعي في تلك الفترة. كما ينبعي التأكيد أيضًا على أن كل هذه الآليات الحاسوبية لا تزال مهمًّا ومفيدةً حتى وقتنا هذا؛ حيث يجري تلافي عيوبها بالتقيد بالشروط الحاسوبية المناسبة للمسائل التي تعالجها، ويدمجها مع آليات حاسوبية أخرى سوف نتعرّض لبعضها لاحقًا في هذا الفصل.

١- انظر المامش السابق.

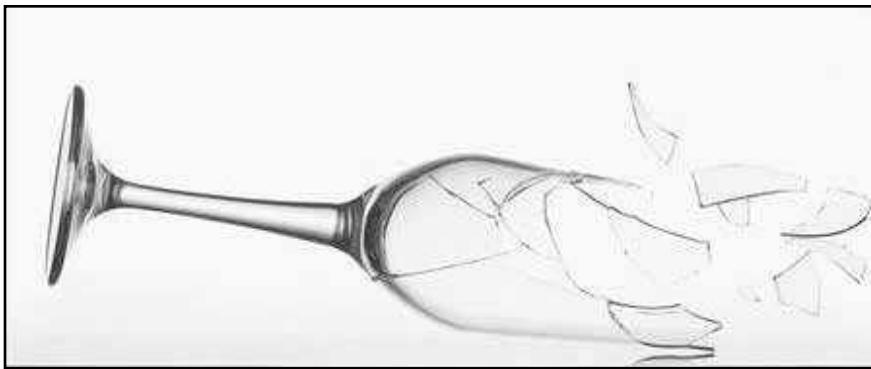
• بالوصول إلى متصف سبعينيات القرن العشرين الميلادي اقتنعت جمهرة الباحثين بعد تجارب كثيرة أن إنجاز تقدم ملموس نحو محاكاة المهام الإدراكية لدى الإنسان - مثل السمع: بمعنى تفسير الدماغ للذبذبات الصوتية الواردة إليه من العصب السمعي، أو البصر: بمعنى تفسير الدماغ للإشارات الضوئية الواردة إليه من العصب البصري، أو اللمس - ونحو اتخاذ الإنسان للقرار بناءً على هذا الإدراك يعتمد ضمن أمور أساسية أخرى على معرفة البيئة الخارجية المحيطة به.

وربما كانت «مفاوضات مورافيك»⁽¹⁾ من أكثر ما شاع في أدبيات الذكاء الاصطناعي تعبيرًا عن ذلك ومؤدًّا لها أنه «على عكس الاعتقاد المُسبَّق الشائع أن عمليات التحليل والبرهنة المنطقية الرياضية تتطلب قدرات حاسوبية أعلى من تلك التي تتطلبها العمليات الإدراكية الأساسية لدى الإنسان والحيوان، فإن الواقع هو أن محاكاة عملية إدراكية بسيطة - كالتعرف على وجه أو كالسير عبر غرفة دون ارتطام بأشيائها - تحتاج أضعافًا أضعاف الطاقات الحاسوبية التي تحتاجها عمليات التحليل والبرهنة المنطقية الرياضية الآلية»، ومَرَد ذلك الصعوبة معضلتان أساسيتان:

الأولى: الضخامة الهائلة للبيئة الخارجية؛ حتى لو كانت جزءًا محدودًا من العالم الواقعي، ويكتفي لتصور تلك الضخامة إجراء محاولةٍ يائسةٍ لحصر ما يستوعبه دماغ طفل صحيح في الثالثة من عمره عن العالم الخارجي، ناهيك عن عدم تجانس تلك المعرفة.

الثانية: تمثيلُ أغلب معرفة الإنسان عن العالم الخارجي في ذهنه تمثيلًا «كيفيًّا» qualitative «كميًّا» quantitative تمثيلًا «كميًّا» qualitative، وكذلك استخدام آليات كيفية لا كمية في معالجة هذه المعرفة لإدراك متغيرات هذا العالم واتخاذ القرار، وتسمّيَان «المعرفة الفطرية commonsense knowledge» و«التحليل الفطري reasoning» على الترتيب، وما زالت حوسِبُتهما حتى الآن معضلةً عسيرةً (انظر الشكل ٤ أدناه والهامش المتصل به للمزيد من التمثيل والتوضيح).

١- الباحث الكندي المُساوي «هائز بير مورافيك» هو أستاذ بمعهد الروبوتات في «جامعة كارنيجي ميلون» في «بنسلفانيا» في الولايات المتحدة الأمريكية، وهو من ارتبطت أسماؤهم بشدة بهذه القضية الأساسية في حقل الذكاء الاصطناعي.



الشكل ٤: مثالٌ على المعرفة الفيزيائية الفطرية: كأس زجاجية تسقط فترطم بأرضية صلبة فتهشم^(١)

• اتَّسَمَتِ الأَسَالِيْبُ الْأَسَاسِيَّةُ لِمُعَالَجَةِ الْمَسَائِلِ إِبَّانَ «السَّنَوَاتِ الْذَّهَبِيَّةِ» بِكُونِهَا «قَطْعِيَّةً/ تَحْدِيدِيَّةً deterministic» بَيْنَمَا الْبَيَانُ الْمُدَخَّلَةُ إِلَى تِلْكَ الْمَسَائِلِ مِنِ الْبَيَّنَةِ الْمُحِيطَةِ ذَاتُ مُرَكَّبَاتٍ «عَشَوَائِيَّةً stochastic» لَا يَمْكُنُ إِغْفَالُهُ، بَلْ إِنَّهُ قَدْ اسْتَقَرَّ فِي الْرِّيَاضِيَّاتِ وَالْعِلُومِ الْتِجْرِيَّةِ فِي الْعَقُودِ الْلَّاحِقَةِ أَنَّ الْعَشَوَائِيَّةَ عَامِلٌ مُتَأَصِّلٌ فِي الْعَمَلِيَّاتِ وَالظُّواهِرِ الْطَّبِيعِيَّةِ؛ فَيَا يُعْرَفُ بِنَظَرِيَّةِ «الشَّوَّاشِ chaos».

١- تَصَوَّرُ حَوَارًا يَقُولُ فِيهِ أَحَدُهُمْ لِصَاحِبِهِ الَّذِي يَحْمِلُ كَاسًا زَجاجِيًّا «احْتَرِ أَنْ تَسْقُطَ مِنْ يَدِكَ الْكَاسِ.»، ثُمَّ تَصَوَّرُ مَا الَّذِي تَسْتَدِعِيهِ هَذِهِ الْعِبَارَةُ الْبَيْسِيَّةُ لِحَامِلِ الْكَاسِ الزَّجاجِيِّ؟

إِنَّهَا تَسْتَدِعِي: أَنَّ (١) هَنَاكَ مَجَالٌ جَاذِبٌ نَحْوَ الْأَرْضِ، وَأَنَّ (٢) الْكَاسُ جُرمٌ ذُو كُتْلَةٍ إِذَا فَلَتْ مِنْ يَدِهِ سُوفَ يَسْتَارُعُ نَحْوَ الْأَرْضِ حَتَّى يَرْتَطِمُ بِأَقْرَبِ سَطْحٍ (الْأَرْضِيَّةِ)، وَكَذَلِكَ أَنَّ (٣) الْكَاسُ مَصْنُوعٌ مِنْ زَجاجٍ وَهُوَ مَادَةٌ هَشَّةٌ إِذَا اصْطَدَمَتْ مَسْرَعَةً بِسَطْحٍ صَلِبٍ فَإِنَّهَا عَلَى الْأَغْلِبِ سَتْهَشُمْ، (٤) إِذَا تَهْشَمَتْ فَإِنَّ أَجْزَاءَهَا سَتَكُونُ ذَاتَ حَوَافَّ مُدَبِّبَةٍ، وَأَنَّ (٥) هَذِهِ الْأَجْزَاءُ سَتَتَّنَاثِرُ بِسَرْعَةٍ فِي الْجَاهَاتِ عَشَوَائِيَّةٍ، وَأَنَّ (٦) أَيْ جُزْءٌ مُدَبِّبٌ مِنْهَا إِذَا اصْطَدَمَ مَسْرَعًا بِبَشَّرَةِ الْإِنْسَانِ فَمِنْ السَّهْلِ أَنْ يَمْدُشَ أَوْ يَقْطَعَ الْجَلْدَ وَاللَّحْمَ وَهُمَا مَادَتَانِ لِيَتَانَ، (٧) وَفِي الْغَالِبِ فَإِنَّ ذَلِكَ سُوفَ يَسْبِبُ جَرَحًا يَسْتَبِعُ نَزِيفًا، وَأَنَّ (٨) التَّزِيفُ رِبَّا ... إِلَخ.

يُلَاحِظُ بِوضُوحٍ غَزَارةُ الْمَعْرِفَةِ الْفَطَرِيَّةِ الَّتِي اسْتَدَعَتْهَا تِلْكَ الْعِبَارَةَ الْبَيْسِيَّةَ عَنِ الْعَالَمِ الْخَارِجِيِّ الْوَاقِعِيِّ، كَمَا يُلَاحِظُ بِوضُوحٍ أَيْضًا تَسْلِسَلُ التَّحْلِيلِ الْمُنْطَقِيِّ بِسَلَاسِلِ الْمُحُصُولِ إِلَى الْمَخَاطِرِ الْمُحْتَمَلَةِ، وَكَذَلِكَ يُلَاحِظُ أَنَّ هَذِهِ الْمَعْرِفَةُ وَتَسْلِسَلُ تَحْلِيلِهَا الْمُنْطَقِيِّ تَوْعِيَانَ لَمْ يُسْتَنَدْ فِيهَا قَوَانِينِ الْحَرْكَةِ لِتَكُونَ مَعَادِلَاتِ تَفَاضِلِيَّةٍ مُثَلًا يَتَوَجَّبُ حَلَّهَا لِلْمُحُصُولِ عَلَى مَسَارِ الْجَسْمِ السَّاقِطِ وَسَرْعَةِ وَمَكَانِ اصْطِدَامِهِ ثُمَّ حَرْكَةِ كُلِّ جُزْءٍ مِنْ أَجْزَاءِهِ الْمُتَهَشِّمَةِ ... إِلَخُ مُثَلِّي يَعَالِجُ عَلَمَ «الْفِيَزِيَّاءِ» عَادَةً مَسَائِلِ الْحَرْكَةِ مَعَالِجَةً رِيَاضِيَّةً رَصِينَةً، وَكَذَلِكَ يُلَاحِظُ السَّرْعَةَ الْلَّاحِظِيَّةَ لِاسْتَدَعَاءِ هَذِهِ الْمَعْرِفَةِ الْفَطَرِيَّةِ وَإِجْرَاءِ تِلْكَ الْمَعَالَجَةِ الْفَطَرِيَّةِ وَمِنْ ثُمَّ اتَّخَادِ حَامِلِ الْكَاسِ قَرَارًا بِالْحِرْصِ عَلَى عَدْمِ سَقْطِ الْكَاسِ مِنْ يَدِهِ وَتَنْفِذِ هَذِهِ الْقَرَارِ حَرَكَيًّا لِتَفَادِيِ الْمَخَاطِرِ الْمُرْتَبَةِ عَلَى سَقْطِ الْكَاسِ مِنْ يَدِهِ.

وَحَتَّى وَقْتَنَا هَذَا لَمْ يَتَمْ تَطْوِيرُ أَدَوَاتِ رِيَاضِيَّةٍ نَاضِجَةٍ لَحْوِيَّةٍ مُثَلُّ هَذِهِ الْمَعْرِفَةِ الْفَطَرِيَّةِ عَنِ الْعَالَمِ الْوَاقِعِيِّ (الَّتِي تُسَمَّى أَيْضًا «أُونْتُرُولُوْجِيَّا ontology» الْعَالَمِ الْوَاقِعِيِّ) أَوْ لِإِجْرَاءِ مُثَلِّ هَذِهِ التَّحْلِيلِ الْفَطَرِيِّ لِهَا بِهَذِهِ الْكَفَاءَةِ الرَّفِيعَةِ؛ مَا يَتَرَكُ نَقْصًا فَادِحًا فِي تِرْسَانَةِ الذَّكَاءِ الْأَصْطَنَاعِيِّ عَمَّا مُعَالِجَةُ الْلُّغَاتِ الْطَّبِيعِيَّةِ عَلَى وَجْهِ الْخُصُوصِ.

• في خمسينيات القرن العشرين الميلادي اكتشفت الدراسة التشريحية للمخ والأعصاب في البشر - وكذلك في الثدييات العليا/ الراقيمة كالقردة - أن المنظومة العصبية لتلك الكائنات المعقّدة هي شبكة ذات عدد هائل (يقدر بآلاف الملايين) من الخلايا العصبية يتصل كل منها بالعديد من الخلايا العصبية الأخرى في الشبكة. وفي العقود التالية تأكّد هذا الاكتشاف وأضيف إلىه تحديد مواضع المراكز العصبية على هذه الشبكة كثيفة الترابط حيث يتعاظم النشاط الكهربائي لكل مركز مع أداء طائفية معينة من الوظائف الذهنية أو الإدراكيّة للكائن الحي.

شَجَّعَتْ هذه الاكتشافات باحثي الذكاء الاصطناعي على أن يحاولوا محاكاة هذه الشبكات العصبية محاكاة حاسوبية مبسطة^(١) رجاء الحصول على محاكاة للأداء الذهني أو الإدراكي للكائنات الحية المعقّدة، فجرّت محاولات أولية في حدود ما سمحت به القدرات الحاسوبية آنذاك واستطاعت إبداء تصرُّف مشابه لبعض العمليات «المنطقية الثنائية binary logic»، لكن برامج الأبحاث في هذا الاتجاه توقفت بنهاية السبعينيات من القرن نفسه؛ ليس فقط بسبب محدودية القدرات الحاسوبية ولكن أيضًا بسبب موجة عالية من الانتقادات للافاق التي يمكن أن تصل إليها هذه المحاكيات^(٢). الأهم من كل ما سبق هو أن ذلك هو كُلُّ

- سوف نعود للنهاج الحاسوبية للشبكات العصبية في القسم الثاني من هذا الفصل، كما ستناولها مرة أخرى بتفصيل أكبر وبرسم توضيحي في القسم الثالث منه كمثال على الفارق بين واقع الظاهر الطبيعية ونمودجها الرياضي في العلوم التطبيقية.
- كان «فرانك روزينبلات» عالم النفس الأميركي في «جامعة كورنيل» في «نيويورك» بالولايات المتحدة الأمريكية هو رائد الدعوة لمحاكاة الشبكات العصبية حاسوبياً في عام ١٩٥٨ م فانطلق برنامج أبحاث نشط لإجراء هذه المحاكاة وتطبيقاتها طوال عقد السبعينيات، ولكن كل شيء توقف عام ١٩٧٩ م حين نشر كل من «مارفين مينسكي» و «سايمور بايرت» كتابهما الشهير «خوارزميات الشبكات العصبية Perceptrons» اللذين قدما فيه حججًا رياضية وفلسفية قوية تبين انسداد الأفق أمام النهاج الحاسوبية للشبكات العصبية كسلاح فَعَالٍ في جمعة الذكاء الاصطناعي.
- استمر هذا التوقف ما يقارب عقدين من الزمان حتى استأنف العمل البحثي في هذا الاتجاه ثانيةً في عقد الثمانينيات حيث ستبين رياضياً أن هذه النهاج الحاسوبية للشبكات العصبية أدوات مفيدة للغاية في «التمثيل التقريري للبيانات» عديدة المتغيرات multivariate function approximation «بمجرد معرفة نهاج غزيرة من مدخلاتها ومحركاتها ومن ثم تنصير أحد الوسائل الأساسية في «العلم الحاسوبي machine learning» كسياسي لاحقًا في هذا الفصل.
- وفضلاً عن أن «مينسكي» و «بايرت» اسمان ذوا نفوذ كبير في هذا الحقل كائنين من الآباء المؤسسين له، فإن المفارقة هي أنها كانوا أول من نَهَّى نموذجًا حاسوبيًّا لهذه الشبكات العصبية - وقد أوردنا ذلك في بدايات القسم الفرعي ١.٥ من هذا الفصل - كما أن هناك مفارقةً شخصيةً طريةً أيضًا تمثل في أن «مارفين مينسكي» كان زميل دراسة لـ «فرانك روزينبلات» في المرحلة الثانوية ثم كان سبباً في تحطيم برنامجه البحثي الذي لم يستأنف إلا بعد موته حيث توفي «روزينبلات» في عام ١٩٧١ م.

ما نعرفه - حتى الآن - عن كيفية عمل الدماغ البشري والجهاز العصبي من حيث معالجته للمسائل الذهنية والإدراكية، وبالتالي فهو غير قادرًّا أبدًا لمحاكاة أداء هذا الدماغ.

- الكثير من المسائل التي تعتبر ممكنًا حقيقًّا لنجاعة الذكاء الاصطناعي - وعلوم الحاسوب على وجه العموم - تبلغ من العسر والتعقيد مبلغًا يصعب معه التعامل حاسوبيًّا مع أية نماذج رياضية لمكوناتها وآليات عملها، وتتجاوز الصعوبة ذلك أحيانًا فتتعدّر من الأصل وضع نماذج رياضية قابلة للحوسبة لتلك المكونات والآليات، بل تصاعد الصعوبة أحيانًا أخرى فلا تُعرف عندئذٍ مكونات ولا آليات للمسألة^(١).

ويمكن التمثيل لتلك المسائل بمسألة أساسية تمثل في «التعرف على مفردات الكلام المنطوق Speech recognition»؛ فإذا تبعنا مكونات وآليات عمل هذه الوظيفة الإدراكية لصادفنا ابتداءً الجهاز الذي يستقبل ذبذبات الصوت وهو الأذن بأجزائها الكثيرة المركبة (الصوان، القناة السمعية، المطرقة، السنдан، طبلة الأذن، القنوات الاهلاكية، القوقة، ... إلخ)، ثم العصب السمعي، ثم عمل المخ على الذبذبات التي يوصلها إليه العصب السمعي. والحقيقة أن محاولة وضع نموذج رياضي لبعض الأجزاء المنفردة من الأذن مهمٌّة عسيرة، أما وضعها معاً في نموذج واحد فهو أمر في عداد الأحلام. فإذا انتقلنا إلى عمل المخ على تفسير الذبذبات التي تصل إليه من العصب السمعي فإن آلية غير معروفةٍ من الأساس! وإذا نظرنا في وظيفة التعرف البصري على الأشكال - كمثال آخر - فإنها لا تقل صعوبة إن لم تزد عن السمع.

ويبقى المورد الوحيد الذي يمكن توفيره بيسير لوصف مثل تلك المسائل هو نماذج غزيرةً من مدخلاتها بالتوابع مع مخرجاتها، ومن ثمَّ نعمت الحاجة إلى ابتكار أساليب لمعالجة مثل تلك المسائل بالارتكاز على هذا المورد فقط، وبالفعل تمثل الاستجابة فيها بعد في ميلاد وتطور عددٍ من الأساليب المتنوعة التي تلبي هذه الحاجة والتي تكون إجمالاً ما يُعرف بمدرسة «التعلم الحاسوبي machine learning»^(٢).

١- ويلطلق أحيانًا على مثل تلك المسائل اسم «المسائل غير المهيكلة/المهيكلة unstructured problems».

٢- ستكون لنا وقفةً مع هذه المدرسة في مقاربة المسائل الإدراكية والذهنية - وفي بورتها نمذجة ومعالجة اللغات الطبيعية - في القسم الثاني وكذلك الثالث من هذا الفصل.

يَقِيَ أن نشير في ختام هذا القسم الفرعى إلى مشكلة أخرى أثارها الباحثون في نقدم لهم لحقبة «السنوات الذهبية» للذكاء الاصطناعي، ألا وهي محدودية قدرات الحواسيب آنذاك قياساً بما يتطلبه حَسْبَ تقدِيرِهِم حل بعض المسائل ذات الأهمية من موارد حاسوبية بنَسَبٍ تراوح بين مئات المرات وألاف المرات؛ فعلى سبيل المثال في عام ١٩٧٦ م قدر الباحث «هانز بيتر مورافيك» - الذي سبق ذكره وترجمته في هذا القسم الفرعى - أن محاكاة أداء شبکية العين البشرية في اكتشاف وتتبع الأجسام المتحركة يحتاج ما بين ثمانين إلى ثمانية آلاف ضِعْف القدرة الحاسوبية لأقوى «حاسوب فائق» (سوبر كمبيوتر) في ذلك الوقت. ومن الطريف أنْ نعرف أنَّه مع نهاية عِقد التسعينيات من القرن العشرين الميلادي كانت تلك القدرة الحاسوبية متوفِّرة بسهولة في العديد من الحواسيب القوية التي كانت متاحةً في الأسواق وقتذاك بأسعار لا تتجاوز خمسة آلاف دولار أمريكي. ويعني ذلك للوهلة الأولى أن مشكلة محدودية القدرات الحاسوبية كانت وقتيةً، وأنَّ الزَّمْنَ كان كفياً بحلها حيث إنَّه كما ذكرنا في القسم الفرعى ١,٤ من هذا الفصل فإنَّ هناك نمط تزايد مُرَكَّبٌ مطْرِدٌ لتنامي القدرات الحاسوبية مع الزَّمْنَ يشتهر باسم «قانون مور» ومؤداه أن تلك القدرات (من حيث عدد التعليمات المنفذة في الثانية الواحدة، وسعة الذاكرة) تتضاعف تقريباً كل ١٨ شهراً^(١)؛ فمثلاً في عام ٢٠١٥م وبعد تسعه وثلاثين عاماً من عام ١٩٧٦م تضاعفت القدرة الحاسوبية بمقدار اثنين مرفوعةً لأَسْ ستة وعشرين^(٢) أي حوالي سبعة وستين مليون مرَّةً !

ورغم أن هذا التزايد فائق التسارع مع الزَّمْنَ للقدرات الحاسوبية يبدو بِرَأْقَا داعياً لتفاؤلٍ كبيرٍ، فإنَّ الحقيقةَ مراوغةٌ بعض الشيء حيث إنَّ القدرات الحاسوبية ليست فقط هي التي تتزايد باطْرَادٍ ولكن تزايد معها أيضاً التحديات التي تمثلُ المسائل المطلوب حلُّها نوغاً وحجماً.

١- يجب الانتباه إلى أن نمط التزايد الأُسْيَ للقدرات الحاسوبية لا يمكن أن يطَرِدَ مع الزَّمْنَ إلى الأبد، وذلك لأنَّ هذا التزايد الذي استمر عبر العقود الستة الماضية يرتبط بالتصغير المطْرِد للوحدات الإلكترونية (الترانزistorات) وهو ما قارب على الاصطدام بحدودٍ فيزيائية لا يمكن تخطيها.

٢- لاحظ أن $26 = (2 \div 3) \times 3^9$.

٨، ١. الانبعاث وإعادة النهوض

كانت «النُّظم الخبيرة expert systems» أحدَ المساراتِ الأقلِ لعاناً التي كان يعمل عليها بهدوءٍ بعضُ باحثي الذكاء الاصطناعي منذ منتصف ستينيات القرن العشرين الميلادي؛ حيث يتكون «النظام الخبير» من مُكوّنَيْن أساسين؛ أولُهما هو قاعدة واسعة قدرَ الإمكان من الحقائق حول موضوعٍ محدَّدٍ ضيقٍ النّطاق (مثل: ريا الحضراوات بالتنقيط في الصّوبات الزراعية) المستقاة من خبراءٍ في هذا المجال والمصوّغة في قوالب منمَّطةٍ تركيبياً واصطلاحياً، وثانيهما هو آلية للاستدلال والاستنتاج عبر التحليل المنطقي الرياضي^(١) تعمل على المكوّن الأول: «قاعدة حقائق الخبرة». وعند تأقِيَّه سؤالاً في نطاق الموضوع الذي يغطيه، فإنَّ النَّظام الخبير عبر التحليل المنطقي الرياضي لـ«قاعدة حقائق الخبرة» المتوفرة لديه قد يُوجَّه بدوره إلى المستخدم الذي وجه إليه السؤال الأساسي بعضَ الأسئلة التكميلية ذات الأجرة البسيطة («نعم أم لا»، أو «اختيار من متعدد») ويعتمد توليد كل سؤال تكميلي في هذا الحوار على إجابات سابِقِيه، ويستمر ذلك إلى أن تتوصل آلية الاستدلال والاستنتاج إلى إجابةٍ للسؤال الأصلي^(٢).

ومن النجاحات المبكرة التي أظهرت فعاليةً وعملانيةً النظم الخبيرة نظامُ سُميَ «دندِرال Dendral» الذي بدأ تطويره عام ١٩٦٥ م بواسطة «إدوارد فِينيباوم»^(٣) وطلابه حيث حاكيَ النَّظام دورَ الخبير الكيميائي بتحديد المركبات الكيميائية عبر دراسة «تحليلها الطيفي spectrophotograph»، وكذلك نظامُ آخر سُميَ «مايسين MYCIN» اكتمل تطويره عام ١٩٧٢ م واستطاع عبر دراسة نتائج تحليل الدم تشخيص الأمراض المعدية عن طريق نقله من شخص إلى آخر.

أما النجاح الفارق للأنظمة الخبيرة الذي أعاد طرح «الذكاء الاصطناعي» في أواسط الجهات الحكومية والصناعية وفي الإعلام واسع الانتشار كحقلٍ علميٍّ وبحثيٍّ جادٌ بعد «سنواته العِجاف» - انظر القسم الفرعِي السابق - فقد أتى عام ١٩٨٠ م عِقبَ

١- والذي تناولناه في القسمين الفرعِين السابقين ٦ و ٧ و ١.

٢- أو إلى أن يستنفذ التحليل المنطقي فيها الحد الأقصى للموارد الحاسوبية المتاحة له وعندئذ يُعلن عدم العثور على حلٍّ.

٣- وكان أستاذاً لعلوم الحاسوب وباحثاً في الذكاء الاصطناعي بجامعة ستانفورد بالولايات المتحدة الأمريكية، ويعُطُّل عليه لقب «الأب المؤسس» للأنظمة الخبيرة.

اكتمال تطوير نظام أطلق عليه اسم «إكس كون XCON» في جامعة «كارنيجي ميلون» الأمريكية لصالح شركة أنظمة الحواسيب الرقمية الكبرى وقتها «مؤسسة المعدات الرقمية» (Digital Equipment Corporation (DEC).

فقد سجّل هذا النّظام الخبرُ أوَّل نجاح صناعيٌّ واقتصاديٌّ ذي باٍ للذكاء الاصطناعي؛ حيث أوردت الشركة في تقاريرها الرسمية عام ١٩٨٦ م أن هذا النظام يوفر للشركة سنويًّا أربعين مليون دولار أمريكي - وتكافئ قيمتها الشرائية عام ٢٠١٩ م على أقل تقدير خمسةٌ وتسعين مليون دولار أمريكي. وينبع هذا الوفر من استغناء كل موظف مبيعاً لدى المؤسسة بهذا النظام عن الحاجة إلى خليةٍ من الفنيين في هندسة الحواسيب تحدد مكونات «العتاد hardware» و «البرمجيات software» الصحيحة والمتوائمة معًا التي تلبي بالضبط طلب كل عميل دون نقاشٍ أو هدرٍ^(١).

وما يلفت النظر أن الأنظمة الخبرية مثلت اتجاهًا عكس التيار السائد وقتها في الذكاء الاصطناعي الذي كان ينحو إلى الاعتماد كأقصى ما يمكن في معالجة المسائل على التحليل الرياضي والخوارزميات، فيما ينحو أيضًا قدر المستطاع إلى اختزال البيانات والمعلومات المدخلة إلى عمليات المعالجة تلك. وربما كان ذلك الاتجاه من الخطوات الأولى على الطريق الذي أوصل الذكاء الاصطناعي في السنوات الأخيرة إلى نصوح مقاربة العديد من المسائل الهامة عبر «التعلم الحاسوبي» وعبر معالجة «البيانات العملاقة big data».

ويجري بناءً «قاعدة حقائق الخبرة base facts experience» لأي «نظام خبير»^(٢) باستقائها من العاملين ذوي الخبرة الراسخة في نطاق عمل هذا النظام عبر مقابلاتٍ مطولةٍ معهم يجريها مطورو النظام الذين يطرحون عليهم الأسئلة بأساليب مصممةٍ

١- لم تكن الأنظمة ذات القدرات الحاسوبية العالية وقتذاك التي تبعها المؤسسة تأتي كباقي واحده معيارية يختارها العميل من بين عدة نماذج محدودة سابقة التجهيز والتجميع تطرحها الشركة في الأسواق - كما هي العادة الآن - بل كان العمالء يصفون احتياجاتهم الحاسوبية لموظفي المبيعات وبناء على هذا الوصف تتحمل الشركة عبء انتقاء العشرات - وربما المئات - من مكونات العتاد والبرمجيات اللازمة لتلبية هذه الاحتياجات ومن ثم تقوم بتجميعها على هيئة نظام متكامل مفصل خصيصاً لكل عميل. ومع ملاحظة أن تكلفة كل عنصرٍ من عناصر العتاد أو البرمجيات كانت باهظة جدًا بالمقارنة بأسعار وقتنا الراهن - بسبب إنتاجها بكميات محدودة على نطاق ضيق - فإن إضافة عناصر لا يحتاجها العميل كانت تمثل خسارةً ماليةً مؤلمة.

لتقطية مختلف السيناريوهات المحتملة للحوار وكذلك لكشف أي تناقض في إجاباتها، كما يضعون إجاباتهم في قوالب مصممة للسماح بتحليلها تحليلًا منطقيًا رياضيًّا. وكي لا تقف معضلة «الضخامة الهائلة للمعرفة الفطرية» - التي تناولناها في القسم الفرعي السابق - سدًا منيعًا يحول دون بناء «قاعدة حقائق الخبرة» فإن كل «نظام خبير» يقييد نفسه بنطاق معرفيٍّ ضيقٍ للغاية على غرار المثالين اللذين ذكرناهما آنفًا.^(١)

وعلى صعيد آخر أثناء سنوات الابتعاث ذاتها - وهي السنوات السبع الأولى من ثمانينيات القرن العشرين الميلادي - وقعت تطوراتٌ هامةٌ أعادت إحياء المحاكاة الحاسوبية للشبكات العصبية بعد اضمحلال العمل عليها بنهاية ستينيات القرن نفسه - كما جاء في القسم الفرعي السابق - فقد ابتكر الفيزيائي الأمريكي «جون هوبفينلد» عام ١٩٨٢ م^(٢) نموذجًا جديداً من الشبكات العصبية الحاسوبية وبرهن رياضيًّا أنها تستطيع لعب دور «الذاكرة الارتباطية associative memory» التي تمثل ملمحًا مميزًا للذاكرة البشرية مقابل «ذاكرة الاستدعاء المباشر random access memory» المميزة للحواسيب الرقمية المعاصرة^(٣). وفي الوقت نفسه تقريرًا نجح «جيفري هاتتون»^(٤) مع «ديفيند رو ميلهارت»^(٥) في البناء على الأسس الرياضية التي أرساها «سيبو لينينما»^(٦)

١- ولُغة الرياضيات التي تصف معضلات بناء وتمثيل ومعالجة «المعرفة الفطرية» فإن «قاعدة حقائق الخبرة» التي يرتكز عليها أيُّ «نظام خبير» هي شريحة أونتولوجية ضئيلة الحجم للغاية - لكنها مفصلة التمثيل - من الأونتولوجيا الكلية للعالم الواقعي.

٢- وكان أستاذًا في «معهد التقنية بجامعة كاليفورنيا (كالتك) Caltech» وهو كذلك أحد مؤسسي برنامج الدكتوراة في «الحوسبة والشبكات العصبية» في المعهد ذاته.

٣- من أجل تخزينها في «ذاكرة الاستدعاء المباشر» فلا بد أن تُعطى كل معلومة «عنوانًا مفهرسًا index» ومن أجل استرجاع تلك المعلومة فلا بد كذلك من معرفة هذا «العنوان المفهرس» وهذا هو النمط الذي تعمل وفقًا له ذاكرة الحاسوب الرقمي المعاصر، أما الذهن البشري فلربما تداعي إلى ذهنه فكرةً/ معلومةً واحدة أو أكثر (مثل: «المزروعات»، «الأخضر»، «النماء»، «الفيضان»، ...) عندما تُذكَر له فكرةً/ معلومةً أخرى مرتبطةً بها (مثل: «سقوط الأمطار») ويسمى هذا النوع من الذاكرة التي تُسْتَدْعَى فيها المعلومة/ المعلومات بمعرفة معلومة أخرى مرتبطةً بها «الذاكرة الارتباطية».

٤- وهو أستاذ كنديٌّ إنجليزيٌّ بارزٌ في «علم النفس الإدراكي» و «علوم الحاسوب» في «جامعة تورونتو» في كندا ثم باحث رئيسيٌّ في «شركة غووغل» العملاقة الشهيرة للبرمجيات.

٥- وهو أستاذ أمريكي في «علم النفس الإدراكي» في «جامعة ستانفورد» ثم في «جامعة كاليفورنيا - سان دييغو» الأمريكية.

٦- وهو عالم الرياضيات الفنلندي والباحث في علوم الحاسوب في «جامعة هيلستكي» في فنلندا، ثم الأستاذ الزائر في «جامعة ماريلاند» بالولايات المتحدة الأمريكية.

في بداية عِقد السبعينيات ومن ثَمَّ إِحْكَامِ صِياغَةِ خُوازيْمِ «الانتشار الارتدادي back propagation» لتدريب «الشبَّكات العصبية متعددة الطبقات multi-layer neural nets» - وهو النموذج الأصلي للشبَّكات العصبية المحوسبة على مدخلات ومحَّلات ظاهِرةً ما من أجل محاكاة سلوكيَّها^(١)؛ وهو التطور الأهم في مسيرة هذه الشبَّكات التي سوف تترسخ مع متصرف عِقد السبعينيات على مُنْحَنِي تصاعديٍّ وصولاً إلى وقتنا الراهن كأحد الأسلحة الماضية في ترسانة «التعلم الحاسوبي machine learning» وهو ما سيلعب معنا دوراً هاماً فيما تبقى من هذا الفصل.

بعثت هذه النجاحاتُ الآمالَ بِقُوَّةٍ بعد خُفوتها في الذكاء الاصطناعي من بَابِ عُرِفَ باسم «هندسة المعرفة Knowledge Engineering» فدفعت حالة النجاح الاقتصادي الذي حققه «النظم الخبرية» - مثل حالة «إِكس كون XCON» التي أسلفنا تفصيلَها - الكثيرَ من الشركات الكبيرة حول العالم في مجال الصناعة والخدمات إلى الإقبال على إِدماج «النظم الخبرية» في عملياتها بل إلى الاستثمار السخي في إنشاء أقسام داخلها لأبحاث الذكاء الاصطناعي لتطوير «نظم خبيرة» خاصة بها، ومن أجل تلبية طلبات هذه الشركات وأقسام الذكاء الاصطناعي فيها نشأ كذلك قطاعٌ صناعيٌّ يشمل شركاتٍ للعتاد الحاسوبي الرقمي (مثل Symbolics و Lisp Machines) وشركاتٍ للبرمجيات (مثل AIon و IntelliCorp).

أما على صعيد الكيانات الأكبر فلم يكن انباع الدعم لأنشطة أبحاث وتطوير الذكاء الاصطناعي بأقلَّ من ذلك؛ ففي عام ١٩٨١ خصصت وزارة الصناعة والتجارة الدولية اليابانية ثمانمئةً وخمسين مليون دولار أمريكيّ لما أطلقت عليه وقتها «مشروع الحواسيب من الجيل الخامس» الذي تمثلت أهدافه في تطوير برمجيات وصناعة حواسيب تستطيع «إجراء حوارات باللغة الطبيعية» و «الترجمة بين اللغات الطبيعية» و «تفسير الصُّور» و «القيام بالتحليل المنطقي» بمستوى الأداء البشري، وقد اختيرت لغة «برولوغ Prolog»^(٢) لُغَةً أساسيةً لكتابَة البرمجيات في هذا المشروع. وكذلك

١- أي حساب قيمة «أيٌّ من محَّلات الظاهِرة» المُناظرة لـ «أيٌّ من المُدخلات إليها» من خارج عينة بيانات التدريب.

٢- وهي كما أسلفنا في القسم الفرعِي ١.٦ من هذا الفصل لغةً لبرمجة الحواسيب الرقمية تُيسِّر لتطوير البرمجيات استدعاء وتوظيف أدوات المنطق الرياضي التعرِيفية والتحليلية في برمجيَّاتهم بسلاسةٍ.

أنفقت الحكومة البريطانية ثلاثة وخمسين مليون جنيه إسترليني بين عامي ١٩٨٣ م و ١٩٨٧ م على مشروع مشابه تحت اسم «ألفاي Alvey»، كما تشكل تكتل من شركات أمريكية كبيرة أطلق عليه «تحالف الإلكترونيات الدقيقة والتقنيات الحاسوبية MCC» لتمويل المشروعات الضخمة في حقل الذكاء الاصطناعي وتقنيات المعلومات. أما آخر مثال نورده في هذا الصدد فهو «المبادرة الاستراتيجية للحوسبة» التي أطلقتها «وكالة مشروعات الأبحاث الدفاعية المتقدمة (داربا) DARPA» في الولايات المتحدة الأمريكية بين عامي ١٩٨٤ م و ١٩٨٨ م^(١) وعودتها - بعد انقطاع - للاستثمار في أبحاث الذكاء الاصطناعي مع مضاعفة استثمارها فيه ثلاثة مرات.

لعل القارئ بعد هذا الانبعاث يتربّع على إعلان النهاية السعيدة لهذا التاريخ الطويل باستقرار مسيرة الذكاء الاصطناعي وازدهارها المطرد حتى وقتنا الراهن، ولكن الواقع هو أن موجةً أخرى من الحفاف والسنوات العجاف كانت على وشك أن تزحف على هذا الحقل والباحثين فيه!

في عامي ١٩٨٦ م و ١٩٨٧ م تعاظم إعجاب المؤسسات التجارية والصناعية والهيئات الحكومية في العالم الغربي بالنظم الخبرية والإقبال على استخدامها والاستثمار فيها بتسارع مطرد، إلا أنه على الرغم من نجاح تلك النظم في «التعلم الأولي» من الخبراء في تخصص معين عبر استخلاص وحوسبة معرفتهم لاستخدامها عبر التحليل بأدوات المنطق الرياضي في استنتاج إجابات لسائل عملية في النطاق «الضيق» لهذه المعرفة، فقد تبدّلت مع طول الاستخدام نقاط ضعف هذه النظم؛ وأهمها الصعوبة والتكلفة العالية لتعديل «قاعدة حقائق الخبرة» مما يعني ضعف القدرة على تحقيق «التعلم المتواصل» سواء لتحسين الأداء أو للتكييف مع التغيرات في الظاهرة التي تعطيها هذه القاعدة المعرفية، وكذلك هشاشة أداء هذه النظم حيث تكاد تكون إجاباتها خطط عشوائية مع تصديها لأسئلة تخرج ولو خروجاً طفيفاً عن النطاق الضيق لقاعدتها المعرفية ما قد يؤدي إلى كوارث إذا اعتمدت العمليات الصناعية والخدمية الحيوية على تلك «النظم الخبرية».

١- وذلك لخدمة المبادرة العسكرية الكبرى التي كان يروج لها بقعة الرئيس الأمريكي «رونالد ريجان» وسماها «حرب النجوم» في إطار جهود إدارته لخسم الصراع العالمي الطويل مع «الاتحاد السوفييتي» عبر التفوق التقني الساحق عليه، وبالفعل سقط «الاتحاد السوفييتي» وتفكك سياسياً مع ختام عام ١٩٩١ م بعد استنزاف اقتصادي وعسكري كانت هذه المبادرة آخر مراحله.

وبختام عام ١٩٨٧ م وقع انقلابٌ في عالم الحواسيب الرقمية حين بزغ عصر «الحواسيب الشخصية Personal Computers» من شركتي آبل Apple و آي بي إم IBM التي تجاوَتْ قدراتِها الحاسوبية قدراتِ الحواسيب المتخصصة في الذكاء الاصطناعي (مثل Lisp machines من صناعة شركة Symbolics) في حين لم يبلغ سعر هذه الحواسيب الشخصية إلا نسبةً بسيطةً من سعر تلك الحواسيب المتخصصة، فتبخرت بين عشيةٍ وضُحاهَا خمسةٌ ملِيون دولار أمريكي مع انهيار صناعة تلك الحواسيب المتخصصة وتطوير مستلزماتها البرامجية (من أنظمة تشغيل وخلافه).

وفوق ذلك في عام ١٩٨٩ م ارتأت إدارةً جديدةً لـ «وكالة مشروعات الأبحاث الدفاعية المتقدمة (داربا) DARPA» الأمريكية أن «الذكاء الاصطناعي» ليس هو الموجة التقنية الجديدة الوعادة التي يجب أن تدعمها الوكالة وأن عليها أن تعيد توجيه ميزانيتها إلى مشروعات أخرى ذات مردودات أكثر واقعية، فكان بالفعل أن قطعت الوكالة دعمها بقسوةٍ عن مشروعات الذكاء الاصطناعي مرةً أخرى وأوقفت «مبادرةها الاستراتيجية للحوسبة».

ولم يكن الوضع بأفضل حالاً في اليابان؛ فمع الوصول إلى عام ١٩٩١ م لم تتحقق قائمة الأهداف مفرطة الطموح لـ «مشروع حواسيب الجيل الخامس» بعد عشر سنوات من انطلاقه.

ولما كان الإعلام وأوساط الأعمال وبعض الشخصيات الأكاديمية قد بالغت مَرَّةً أخرى في سنوات الثمانينيات من القرن العشرين في النفح في حدود قُدراتِ وآفاق «الذكاء الاصطناعي» بعامَّةٍ وـ «النظم الخبرية» على وجه الخصوص، فإن تلك المبالغات قد أدت مع مُضيِّ الوقت إلى خلق فُقاعةٍ اقتصادية وتقنيَّة كبيرة انفجرت بالفعل بنهاية ذلك العُقدِ نَفْسِيَّ جَرَاءَ ما أسلفنا التمثيل له لتوَّنا من نقاط قصور وتطورات تقنية وتمويلية؛ فبالوصول إلى عام ١٩٩٣ م كان عدد ما اخْتَفَى من شركاتٍ تخصصت في الذكاء الاصطناعي - أو اعتمد نشاطها عليه - ثلاثةٌ شرَّكةٌ أَنْشَئَتْ كُلُّها في أَعوام الانبعاث السبع بين ١٩٨١ م و ١٩٨٧ م.

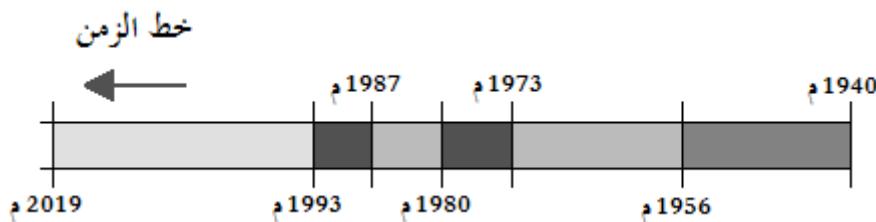
ولأنَّا المَرَّةُ الثانيةُ التي يتكرر فيها انزلاقيُّ مسار «الذكاء الاصطناعي» إلى أزمةٍ عامَّةٍ فلا بُدَّ لنا من تقريرٍ وإبرازِ الدرس المستفاد التالي من تلك الخبرة التاريخية القيمة:

قاعدة رقم ١

قد تؤدي المبالغة في قدرات أي فرع مستحدث من العلوم و/أو التقنية، والتسرب في توسيع آفاقها التطبيقية بلا قيود، ووضع مستهدفاتٍ مفرطة في التفاؤل؛ قد يؤدي كل ذلك إلى مكاسب إعلامية وتمويلية قصيرة المدى، لكنه من المؤكد أنها جمِيعاً تؤدي على المدىين المتوسط والطويل إلى أثرٍ عكسيٍّ حتميٍّ يتمثل في فقدان الثقة في هذا المسار من العلوم والتقنية وتعطيله لفترة طويلة.

وبطبيعة الحال فلا بدًّ كذلك من الاستفادة من هذا الدرس القيم وتطبيقه على حوسبة اللغات الطبيعية ومن بينها حوسبة اللغة العربية التي هي بؤرة هذا الكتاب.^(١)

و قبل متابعة هذه المسيرة في مرحلتها الأخيرة المتداة إلى وقتنا الراهن، فمن المفيد أن نلخص جميع مراحلها على خارطة زمنية يعرضها فيما يلي (الشكل ٥)؛ حيث يفصل كل خط رأسي بين مراحلتين، وتناسب درجة سطوع التظليل في كل مرحلة مع درجة النشاط والوضوح فيها.



الشكل ٥: خارطةٌ زمنيةٌ معاصرةٌ مبَسَّطةٌ لصعود وهبوط النشاط البُحثي في حقل الذكاء الاصطناعي

١- وعلى الرغم من شدة وضوحها فإن مخالفة هذه القاعدة الذهبية لا يزال يتكرر مرّةً بعد أخرى عند تفعيل مبادراتٍ طموحة لحسبة اللغة العربية في العالم العربي مسبّبةً ليس فقط فشلها بل أيضاً تشويه سمعة حقل حوسبة اللغة من حيث الجدّية والفعالية مما ينبع أثراً نواياً مستقبليةً لتكرار المحاولة، وندعو القارئ كتمرين عمليًّا إلى البحث عن أمثلةٍ معاصرةٍ مثل تلك الحالات.

في الترحال على هذه الخارطة الزمنية نصادفُ:

أولاً: المرحلة التمهيدية التي تمت من أربعينيات القرن العشرين الميلادي مع تبلور «علوم الحاسب» وبناء «الحواسيب الرقمية» الأولى حتى مؤخر «دارتقوث» الذي يعد بمثابة «الميلاد الأكاديمي» لحقل «الذكاء الاصطناعي» عام ١٩٥٦ م.

ثانياً: «السنوات الذهبية» التي امتدت إلى غاية عام ١٩٧٣ م.

ثالثاً: الموجة الأولى من «السنوات العجاف» - أو كما يطلق عليها بعض المؤرخين «الشتاء الأول للذكاء الاصطناعي» - التي امتدت لغاية عام ١٩٨٠ م.

رابعاً: «سنوات الابتعاث وإعادة النهوض» التي استمرت لغاية عام ١٩٨٧ م.

خامسًا: الموجة الثانية من «السنوات العجاف» - أو كما يطلق عليها بعض المؤرخين «الشتاء الثاني للذكاء الاصطناعي» - التي امتدت لغاية عام ١٩٩٣ م.

سادسًا: «الحقبة المعاصرة» الممتدة من ١٩٩٣ م حتى وقتنا الراهن، والتي يطرُدُ فيها التطور البحثي والإنجاز التطبيقي بعد استقرار عدد من المفاهيم والمقاربات الرياضية الحاسوبية في قناعة جهور الباحثين في هذا الحقل - كما ستبينُ الورقاتُ التالية - مما يجعلها جديرةً أن تُسمَّى أيضًا «حقبة البيانات العملاقة والتعلم الحاسوبي».

و قبل الولوج إلى «الحقبة المعاصرة» لا بدّ لنا من تقرير وإبراز قاعدة ذهبية أخرى^(١) مستفادٍ من هذه الخارطة الزمنية:

١- وفي الحقيقة فإن عدم التقيد بهذه القاعدة الذهبية يشكل للأسف مشكلةً كبيرةً في الواقع الراهن لحوسبة اللغة العربية؛ فكثيرٌ من السادة اللسانين العرب مشبعون بقناعاتٍ «منسوخة outdated» ترجع إلى «مراحل باكرة» من تطور الذكاء الاصطناعي ومعالجته لميكنة اللغات الطبيعية سابقةٍ على «حقبة المعاصرة». والسبب في ذلك هو الوقت اللازم الذي تنتقل المفاهيم والكتشفات والمنجزات التجريبية الجارية في حقل الحوسبة إلى أدبيات اللسانيات الغربية إضافةً إلى الوقت اللازم بعد ذلك الذي تُرجم إلى أدبيات اللسانيات العربية فضلاً عن الوقت اللازم ليهضمها السادة اللسانيون العرب، ويشير إلى ذلك بوجود «فارق في الطور shift phase» بين اللسانى العربي النابه الذي يعمل (في العقد الثاني من الألفية الميلادية الثانية على سبيل المثال) بقناعاتٍ قديمةً منسوخةً من حقل الذكاء الاصطناعي والحوسبة (نشأت في ثمانينيات القرن الميلادي الفائت على سبيل المثال).

فعلى سبيل المثال لا يزال عدد من اللسانين العرب النابهين الآن يجاهدون منذ سنين طويلة دون طائل في بناء نظمٍ لمعالجة اللغة العربية آلياً (كالترجمة الآلية من / إلى اللغة العربية على سبيل المثال) بمقاربات «النحو التحويلي» و «النحو التوليدية» ... وما إلى ذلك من أساليب تحليلية أهللت عنها منذ سنواتٍ طويلةً أتجه نظم حوصلة اللغات الطبيعية المعاصرة، والأوفق من أجل تجاوز مشكلة «فارق الطور» تلك هو أن يعمل اللسانيون العرب النابهون في فرق تضم إليهم باحثين نابهين في علوم الحاسوب والذكاء الاصطناعي ولا سيما «التعلم الحاسوبي machine learning» كما سنين لاحقاً في هذا الفصل.

قاعدة رقم ٢

على الباحث أو المطور أو المخطط أو مدير المشروع في حقل حoscience اللغات الطبيعية - ومنه «حoscience اللغة العربية» التي هي بؤرة اهتمام هذا الكتاب - الحذر والتبصر الشديدين عند الرجوع إلى تراث وأدبيات الذكاء الاصطناعي قبل «الحقبة المعاصرة» حيث كان التقلب في الرؤى والقناعات شديداً في مراحله الباكرة، فإذا ما ابتعى أفضل الممارسات التي استقرت في هذا الحقل فعليه بالرجوع إلى المصادر الأحدث في الحقبة المعاصرة.

وعودةً إلى ما بعد انحسار الموجة الثانية من سنواته العجاف التي امتدت قرابةً الست سنوات، فإنه يُنظر إلى مسيرة الذكاء الاصطناعي منذ عام ١٩٩٣ م حتى وقتنا الراهن كمرحلة متصلةٍ واحدةٍ؛ وفيها يلي السمات العامة التي تميز هذه المرحلة:

- في المراحل السابقة كان هناك ميلٌ واضحٌ لدى باحثي الذكاء الاصطناعي إلى الاقتصار في الأدوات التي يستخدمونها على تلك التي طورت تحت مظلة «الذكاء الاصطناعي»، وكذلك ميلٌ إلى تمييز عملهم وحقولهم عن سواه من حقول البحث والتطوير حتى تلك المتجاورة معه المرتبطة به ارتباطاً «عائلياً» وثيقاً كالفروع الأخرى لعلوم الحاسوب والرياضيات، و«علم الاتصالات Telecommunications»، و«معالجة الإشارات الرقمية Digital Signal Processing (DSP)»، و«بحوث العمليات Operations Research»، و«نظرية التحكم Control theory» ... إلخ. وفي المقابل تمتاز الحقبة المعاصرة بانفتاح بواباتها للاستقاء من روافد كل تلك الحقول^(١) حتى إنه ليصعب الآن تعين حدودٍ فاصلةٍ بين حقل الذكاء الاصطناعي وتلك الحقول الأخرى التي ترافقه.
- غَلَبَ على الأساليب التي ارتكزت عليها معاجلَاتُ الذكاء الاصطناعي للمسائل في المراحل السابقة أنها أساليب «قطعية/ تحديدية deterministic» وقد أسلفنا

١- فمثلاً: قدمت «معالجة الإشارات الرقمية» حلولاً عمليةً «للتعرف على الكلام المنطوق» و«التعرف على الأشكال المرئية».

أن ذلك نقطة ضعف كبيرة^(١) حيث إن البيانات المدخلة إلى تلك المسائل من البيئة المحيطة ذات مركبات عشوائية واضحة فضلاً عن تأصل العشوائية والتعميد في العمليات والظواهر الطبيعية. أما في هذه المرحلة فقد التحق بترسانة أساليب معالجة الذكاء الاصطناعي للمسائل مَدْ حافل من أسلحة الرياضيات الماضية من «التحليل العددي Numerical Analysis» و«نظرية الاحتمالات Probability Theory» و«التحليل الإحصائي Statistical Analysis» و«العمليات الالإحصائية Statistical Processes» و«نظرية الشوائب Chaos Theory» ... إلخ مما غير المشهد من المثالية إلى الواقعية التي حققت نجاحاتٍ تطبيقية كبيرةً كثيرةً هامةً، الواقع أن هذه الأدوات الجديدة هي التي تتصدر مشهد الذكاء الحاسوبي الآن في مرحلته المعاصرة.

- بفضل التنامي المتسارع أُسِّيَاً - كما يعبر عنه «قانون مُوور»^(٢) - في القدرات الحاسوبية التي توافر عليها الحواسب الرقمية، وبفضل بنى «الحوسبة على التوازي Parallel Processing» وبنى «الحوسبة الموزعة Distributed Processing» علاؤةً على ذلك، فقد صار من اليسير في المرحلة المعاصرة من مسيرة «الذكاء الاصطناعي» تنفيذ خوارزميات تقوم بمعالجاتٍ حسابية باللغة التعميد على أحجامٍ ضخمةٍ من البيانات - مما كان متعدّراً في مراحله السابقة - بل صارت المنصات الحاسوبية الفائقة في السنوات الأخيرة قادرةً على التهام وهضم «أحجام عملاقةٍ من البيانات».^(٣) ولذلك ترجُح الآن بقوّةٍ كَفَّةً «الذكاء القائم على المعرفة Knowledge-oriented Intelligence» - فيما تمثل المعرفة عن الظاهرة محل الدراسة بكثافةٍ عاليةٍ من دفق البيانات عن تلك الظاهرة - على كَفَّةً «الذكاء المقتصر على براعة التحليل والتركيب» كما كان الحال في المراحل السابقة.

١- راجع القسم الفرعي السابق ١،٧ من هذا الفصل؛ وبالتحديد الإطار الوارد بعد الشكل رقم ٤.

٢- الذي سبقت الإشارة إليه في نهاية القسم الفرعي ١،٤ ثم بتفصيل أكبر في نهاية القسم الفرعي ١،٧ من هذا الفصل.

٣- يلغى معامل تفاوت القدرات في هذا الصدد بين عقد ستينيات القرن العشرين الميلادي وعام ٢٠١٦م - على سبيل المثال - ما بين «واحدٍ أمامه أحد عشر صفرًا» إلى «واحدٍ أمامه اثنا عشر صفرًا».

• وفرت السُّهُّاتُ الثلاث السابقة بيئةً الذكاء الاصطناعي المطلوبة لتفعيل «التعلم الحاسوبي machine learning» وتحويله إلى النجم الألع في العُقدِين السابقيْن على مسرح الذكاء الحاسوبي مما مَثَّلَ انعطافَةً تاريخيَّةً ليس فقط في مسيرة تطوير الأنظمة الذكية بل في مسيرة الحوسبة بِكاملها؛ ففيما جرى حلُّ الكثيُّر من - أو ربما مُعظَّم - ما يمكن مقاربَتُه تحليلياً من «مسائل مهيكلةٍ structured problems» ذاتِ باٍل في المراحل السابقة لمسيرة الذكاء الاصطناعي - وربما كذلك في السنوات الأولى من مرحلته المعاصرة - فإنَّ مُعظَّم المسائل الهامة الذي بقي إيجاد حلول حاسوبيَّة ناجعة لها يُشكُّل تحديًّا كيًّراً هي «مسائل غير مهيكلة unstructured problems» لم يُيقِّن من سلاٍحٍ ماضٍ للتَّعامل معها سويًّا «التعلم الحاسوبي».^(١)

ولأنَّ معظم المساحات الهامة في حقل المعالجة الحاسوبيَّة الآلية لِلُّغَاتِ الطبيعية - ومن ضمنها «اللُّغَةُ العَرَبِيَّةُ» بطبيعة الحال - تتَّسِّي إلى طائفةٍ «المسائل غير المهيكلة»، فسوف تكون لنا وقفةُ أخرى مع «التعلم الحاسوبي» في القسم الثاني من هذا الفصل؛ حيث نبسط الفلسفَةُ التي يُنْبِيُّ عليها، وتحت أية شروط يكون هو الخيار الأمثل، وفي أية ظروفٍ ينبغي تجنبه.

• في المراحل السابقة كانت النافذة التي يطلُّ من خلاها الذكاء الاصطناعي على مستخدميه ويطلُّون منها عليه هي برمجياتٍ يكاد يقتصر كُلُّ منها على تنفيذ خُوارِزمٍ - أو أكثر - من خوارزمات الذكاء الاصطناعي بينما تبقى العناصر البراجيَّةُ الأخرى عناصر مساعدةً على الهاشم، بينما تتميز الحقبة المعاصرة أنَّ تجليات «الذكاء الاصطناعي» تكون من خلال توظيف خوارزماته كعنصرٍ مُدمَّجٍ مع عناصر أخرى عديدة - مثل: «واجهات الاستخدام التفاعليَّة»، و«الرسوميَّات»، و«تداول البيانات عبر الشبكات الرقميَّة»، و«إدارة قواعد البيانات» و«التواصل بين المستخدمين» ... إلخ - تتكامل جيًّعاً في تطبيقاتٍ رقميَّة - لا يُشترط أن ترفع لافتة «الذكاء الاصطناعي» - من أجل تحقيق ذكاءٍ

١- راجع الإطار الأخير في القسم الفرعي السابق ١,٧ لتعريف «المسائل المهيكلة» و«المسائل غير المهيكلة» و«التعلم الحاسوبي».

حاسوبيًّا يحس به ويستفيد منه مستخدم تلك التطبيقات. ونشرير كالعادة إلى أن برمجيات معالجة اللغات الطبيعية الحديثة ليست استثناءً من ذلك.

من حيث الصورة الذهنية في الوعي العام أثناء الحقبة المعاصرة؛ فقد خفَّت كثيرةً ذِكْرُ مصطلح «الذكاء الاصطناعي» وبهُتَّ أَلْقُهُ سواءً في الإعلام واسع الانتشار أو حتى الأوساط الهندسية والأكاديمية، ومقابل ذلك شاع ذِكْرُ الأساليب الناجحة - أو الوعادة - لمقاربة المسائل التي تمثل تحدياً حاسوبيًّا كبيراً؛ مثل: «نماذج ماركوف المُخْفَيَّة» (Hidden Markov Models (HMMs)، «الشبكات العصبية المحوسبة (ANNs)، «تعلم بايِّس الاحتمالي Bayesian Learning»، ... إلخ. استمر ذلك الحال حتى سنوات قليلة خلت حين أخذ الإعلام الإخباريُّ والاقتصاديُّ وحتى الدراميُّ واسع الانتشار يعيد نَصْبَ رأيات «الذكاء الاصطناعي» عالياً على منصاته ورَفَعَ الصوت بالترويج للمسابقات في مفتوح آفاقه وخارق قدراته ... إلخ مثلاً شهدت المراحل السابقة من هذه المسيرة، ولذلك يجب أن تخضع أية إخبار إعلامية حديثة عن الذكاء الحاسوبي لمزيدٍ من التدقيق والفحص قبل اعتماد صحتها وصدقيتها.

وفي ظل هذه السهات السنت العامة المميزة لحفل الذكاء الاصطناعي في مرحلته الأخيرة المعاصرة فلا بدًّ من إبرازٍ قاعدة ذهبية ثالثة يجب على من يتصدى في زمننا هذا لمشروعات أبحاث حosome اللغات الطبيعية - ومن بينها «العربية» - أو تطوير تقنيات معالجتها أن يضعها نصب عينيه:

قاعدة رقم ٣

إن أي مشروع بحثي أو تطويري تقني جاد في حقل «المعالجة الآلية للغات الطبيعية NLP» هو عملٌ (١) طبيعته هندسيةٌ ويجب أن يدار بالآليات «الإدارة الاحترافية للمشروعات» خصوصاً مع ضخامتها (٢) وأدواته هي خوارزمياتٌ مبنيةٌ على الرياضيات المتقدمة (٣) و موضوعه ومادته هي اللغة الطبيعية ومن ضمنها «العربية». وأي عملٌ ذو بالٍ في هذا الحقل يختل فيه واحد أو أكثر من هذه الأركان الثلاثة - من حيث التخطيط أو التجهيز أو الأفراد أو التنفيذ - فإن مصيره المحتَمِ إلى الإخفاق.

ماذا تحقق بالفعل إذن حتى الآن في خضم الذكاء الاصطناعي - أو الذكاء الحاسوبي؟ وما الذي يصعب - أو ربما يستحيل - أن يتحقق في المستقبل المنظور؟ هذا سؤال جدّ كبير، سنحاول قبل الانتقال إلى القسم الثاني من هذا الفصل أن نرسم ملامح عامةً لإجابتة ثم نعود في القسم الخامس من هذا الفصل للإجابة عنه مرةً أخرى بتفصيل أكبر في سياق المعالجة الآلية للغات الطبيعية.

كان النجاح الأكبر الذي تحقق حتى الآن على مستوى معالجة كميات كبيرة - وأحياناً هائلة - من الرموز من أجل:

- اكتشاف «الأنماط patterns» وأنساق ارتباطاتها؛ وهو ما يُعرف باسم «التنقيب في البيانات» (Data Mining).

- التصنيف الآلي لهذه الرموز في فئاتٍ معرفةٍ سلفاً - وهو ما يُعرف باسم «التنسيب أو مكتشفةٍ تلقائياً من سلسلة البيانات» (classification) - وهو ما يُعرف باسم «التصنيف categorization».

- تعميم الربط بين فئتين أو أكثر من الرموز بعضها يمثل مدخلاتٍ ظاهرةٍ أو عمليةٍ ما وبعضها الآخر يمثل مخرجاتٍ لها بحيث يمكن بعد ذلك التعميم لتوليد المخرجات المناظرة لأية مدخلاتٍ وهو ما يُعرف باسم «التعلم الحاسوبي» (Machine Learning).

- تعميم الربط بين فئتين أو أكثر من الرموز بعضها يمثل مدخلاتٍ ظاهرةٍ أو عمليةٍ ما وبعضها الآخر يمثل مخرجاتٍ لها بحيث يمكن بعد ذلك التعميم لتوليد/ اختيار أرجح المخرجات احتمالاً المناظرة لأية مدخلاتٍ وهو يمكن أن نطلق عليه تسمية «التعلم الاحتمالي الشرطي Conditional Probabilistic Learning» (Conditional Probabilistic Learning). وتلعب هذه العمليات الحداثية^(١) معالجة الرموز دور البطولة في توفير عنصر الذكاء الحاسوبي في صناعة البرمجيات المعاصرة.

١- إلى جانب عمليات الذكاء الاصطناعي الكلاسيكية لمعالجة الرموز مثل: التحليل عبر المنطق الرياضي، و مختلف أساليب البحث الشجري والشبكي، ... وغيرها مما أسلفنا الإشارة إليه في القسمين الفرعين ١.٦ و ١.٧ من هذا الفصل.

أما على مستوى «ربط هذه الرموز بالعالم الواقعي»؛ أو بعبارة أخرى «إكساب الرموز مدلولاتٍ في الواقع الحقيقي»، فما زال تحقيقه الآن - وربما لعقود قادمة - بعيد المنال بسبب تعذر بناء أونطولوجيا كاملة للعالم الواقعي^(١)، وبسبب تعذر التمثيل الرياضي وحوسبة «المعرفة الفطرية commonsense knowledge» وتعذر حوسبة «التحليل الفطري commonsense reasoning» تبعاً لذلك، وهي المعضلات التي تناولناها في القسم الفرعي ١,٧ من هذا الفصل^(٢). ويُستثنى من ذلك التعذر التطبيقات التي تقتصر على عالمٍ بيئة محدودةٍ للغائية؛ مثل «الألعاب الحاسوبية»، و«أنظمة المحاكاة»، وبرمجيات التحكم في الروبوتات المتخصصة ... ففي تلك الحالات نضجت الآن بالفعل الخبرات العملية لبناء الأونطولوجيا التي تُكسي الرموز مدلولاتٍ واقعيةٍ في عالمها المحدود.

ويبقى تحقيق ذروة الحلم بإكساب البرمجيات والآلات التي تنفذها «الذاتية» و«الوعي» - وهو ما يحمل للإعلام الشعبي ودراما الخيال العلمي التعبير عنها بكاروس الآلات المستقلة المتمردة التي تحكم العالم وتستبعد البشر - أقرب إلى المستحيل استحالةً مبدئيةً بسبب اصطدامها بقفز الرياضيات المتمثل في «مبرهنتي عدم الاكتفاء الرياضيات Incompleteness theorems» اللتين سبقت الإشارة إليهما في نهاية القسم الفرعي ١,٣ وفي بداية القسم الفرعي ١,٦ من هذا الفصل. ويرجع السبب في ذلك بأوجز وأبسط ما يمكن إلى أن بذرة «عدم الاكتفاء» الذي يحدد سقف المعرفة الرياضية تكمن في «استدعاء أية عملية رياضيةٍ لذاتها» أو «إحالتها على نفسها» وهو ما يصطلاح على تسميته بـ«العَوْدِيَّة» أو «الإحالَة الذاتية self reference» وهو عين المبدأ الضروري للذاتية والوعي حيث إن منشأها هو معرفة الكائن لنفسه والإشارة إلى ذاته. ولذلك فإن أية محاولة ولو أولية لغرس ذاتيةً أو وعيٍ في الآلات سوف تضمنها على مسارٍ يؤدي إلى الخوض في مساحاتٍ تعجز أدواتنا الرياضية - وبالتالي نهادجنا الحاسوبية - عن استكشافها ناهيك عن السيطرة عليها. وسوف نعود لتلك المسألة في القسم الخامس من هذا الفصل حين نحاول أن نتلمس سقف المعالجة الآلية لِللغات الطبيعية إن كان يحدُّها سقف.

١ - وفيها يختص بالمعالجة الآلية للغات الطبيعية فإن بناء أونطولوجيا كاملة للعالم تكسب الوحدات الرمزية للغة مدلولاتها تتمثل تحدياً أكبر من ذلك بكثير؛ فالأونطولوجيا المنشودة عندئذ ليس عليها فقط استيعاب العالم الحقيقي المحسوس

هائل الضخامة بل إن عليها أيضاً أيضاً استيعاب عالم المجاز والخيال وهي فضاءات غير محدودة بطبيعة تعريفها!

٢ - راجع مناقشة «مفارقة مورافيك» قبل الشكل رقم ٤ مباشراً، وكذلك المامش المحقق بهذا الشكل ذاته.

٢٠. مَدْرَسَةُ سُنَّةٍ فِي معالِجَةِ الْمَسَائِلِ فِي إِطَارِ الْذَّكَاءِ الْأَصْطَنَاعِيِّ.

في الفضاء الشاسع للمسائل - «غير التافهة non-trivial» - المطلوب التصدّي
لمعالجتها حاسوبياً يمكننا تصنیف الفئات التالية من حالاتها:

(١) «مسائل مهيكلة structured problems»: وهي مسائل نعرف عناصرها ونستطيع أن نفكّرها إلى مكوناتٍ أبسط، وكذلك نعرف آلياتِ تفاعلٍ عناصر المسألة مع بعضها البعض، وبالتالي نستطيع أن نضع لها «نموذجًا رياضيًّا صريحاً explicit mathematical model». وتنقسم «المسائل المهيكلة» بدورها إلى فئتين فرعيتين:

(١←١) «مسائل مهيكلة تُوجَد - من حيث المبدأ - حلول لنهاز جها الرياضية

». وهذه أيضاً تتشعب إلى structured problems with solvable models

شعتن:

(١) \leftarrow (١) مسائلٌ مهيكلة توجد حلول لنهاذجها الرياضية الصريحة

ويتمكن تنفيذ بعض هذه الحلول حاسوبياً بواسطة «خوارزمات مفيدة»^(٢).

(١) ← (٢) مسائل مهيكلة توجد حلول لها ذجها الرياضية الصرحة

وَلَا تُعْرِفُ خُوَارِزْمَاتٌ مَفِيدَةٌ حَاسُوبِيًّا لِتَنْفِيذِهَا عَمَلِيًّا.

(١←٢) «مسائل مهيكلة ذات نهادج رياضية عصبية على التعامل معها وحلّها

۲- مسائل معمولة با مدل‌های ناپذیر (structured problems with unsolvable models)

(٢) «مسائل غير مهيكلة unstructured problems»: وهي مسائل يتعدد تفكيرها

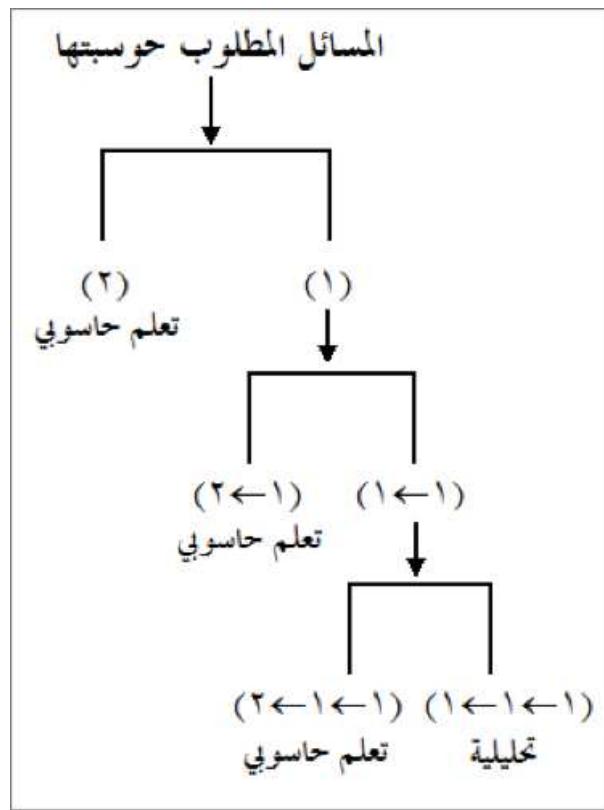
إلى مكوّنات أبسط، ولا تُعرّفُ آلّيات عملها، وأفضل ما يمكن توفيره لوصف مثل تلك

المسائل هو عَيْنَاتٌ غَزِيرَةٌ مِنْ مُدْخَلَاتِهَا بِالْتَوَازِيِّ مَعَ الْمُخَرَّجَاتِ الْمَنَاظِرَةِ لَهَا.^(٣)

١- النموذج الرياضي «الصريح» لبيان ما هو وصف - تقريري مسٌط له في العادة - له عبر نسق متراٍط من التعريفات والعمليات والعلاقات الرياضية المباشرة؛ الجبرية و/أو التفاضلية/التكاملية و/أو الشّرطية و/أو المنطقية و/أو مناولة المصففات ... إلخ، بحيث يمكن لهذا الوصف الرياضي إذا أُعطي قيم مدخلاتٍ ما أن يُتّبع المخرجات المناظرة لها باتفاقٍ مقبولٍ مع سلوك النظام الحقيقى. وسوف نتناول مجدداً في القسم التالى من هذا الفصل العلاقة بين الظاهرة ونموذجها الرياضي، والفارق بينهما.

٢- راجع تعريف «الخوارزم المفيد» الذي عرضناه مع تعريف «الذكاء الاصطناعي» قرب نهاية القسم الفرعوني ١,٥ من هذا الفصل.

٣- راجع الإطار الأخير في القسم الفرعى ١,٧ من هذا الفصل.



الشكل ٦: مقارنة فئات المسائل المختلفة حسب تصنيف تعقيدها الرياضي.

ويقترن بهذا التصنيف السابق مقارباتان للتعامل مع المسائل المطلوب حوسبتها:

الأولى: هي «المقاربة التحليلية analytic approach» للتعامل مع مسألة ما؛ وهي تنفيذ حلًّ لنموذج رياضي صريح تفديًّا حاسوبيًّا. و«النموذج الرياضي الصربي» لنظامٍ ما هو ببساطة وصفٌ للنظام عبر نسقٍ متراوطيٍ من التعاريفات والعمليات وال العلاقات الرياضية المباشرة، بحيث يمكن لهذا الوصف الرياضي إذاً أُعطيَ قيمَ مُدخلاتٍ ما وأن يحسبَ المخرجاتِ المناظرة لها باتفاقٍ مقبولٍ مع سلوكِ النظام الحقيقي.

الثانية: هي مقاربة «تُعرَفُ باسم «التعلُّم التلقائي» أو «التعلُّم الحاسوبي-machine learning» وهي تنظر إلى النظام المطلوب حوسبيه كصندوق مصمَّت لا يُعرَفُ ما بداخله؛ حيث كل ما يُسمَحُ به هو تغذية النظام بمُدخلاتٍ والحصول منه على

المُخْرِجاتِ المُناظِرَةِ لِكُلِّ مِنْهَا. وَرَكِيزةُ هَذِهِ المُقارَبَةِ هِيَ تَحْصِيلُ عِيْنَةٍ كَبِيرَةٍ (بِهَا يَكْفِي) مِنَ الْمُدْخَلَاتِ وَالْمُخْرِجَاتِ الْمُوازِيَّةِ لَهَا، ثُمَّ تَشْغِيلُ أَحَدِ الْآلَيَّاتِ الْرِياضِيَّةِ عَلَى هَذِهِ الْعِيْنَةِ الْكَبِيرَةِ مِنْ أَجْلِ اسْتِبْنَاطِ السُّلُوكِ الْعَامِ لِلنِّظَامِ بَيْنِ مُدْخَلَاتِهِ وَمُخْرِجَاتِهِ.

وَيَصُورُ (الشَّكْلُ ٦) عَالِيَّةَ المُقارَبَةِ الْمُمْكِنَةَ لِكُلِّ فِتَّةٍ مِنْ فِئَاتِ الْمَسَائِلِ حَسْبَ التَّصْنِيفِ السَّابِقِ وَفِقْ طَبِيعَةِ مَعْرِفَتِنَا بِيُنْتِهَا الْرِياضِيَّةِ، وَفِيهَا يَلِي شُرُوطَ صَلَاحِيَّةِ وَظَرُوفَ تَطْبِيقِ كُلِّ مِنَ الْمَقَارَبَتَيْنِ:

- يُمْكِنُ حُوسْبَةُ مَسَأَلَةٍ مَا عَبَرَ «مُقارَبَةٍ تَحْلِيلِيَّةً» فَقَطْ عِنْدَمَا تَوَافَرَ فِيهَا مَا جَمِيعَ الشُّرُوطِ التَّالِيَّةِ:

- مَعْرِفَةُ عِنَاصِرِ الْمَسَأَلَةِ وَإِمْكَانِيَّةِ تَفْكِيْكِهَا إِلَى مَكْوُنَاتٍ أَبْسَطِ،
- مَعْرِفَةُ آلَيَّاتِ تَفَاعُلِ عِنَاصِرِ الْمَسَأَلَةِ مَعَ بَعْضِهَا الْبَعْضِ،
- إِمْكَانِيَّةُ وَضَعِيفَ نَمْوَذْجِ رِياضِيٍّ «صَرِيحٍ» لَهَذِهِ الْمَسَأَلَةِ،
- وَجُودُ حَلُولٍ نَظَرِيَّةٍ لَهَذَا النَّمْوَذْجِ الْرِياضِيِّ الصَّرِيحِ،
- إِمْكَانِيَّةُ تَنْفِيْذِ أَحَدِ هَذِهِ الْحَلُولِ النَّظَرِيَّةِ حَاسُوبِيًّا بِوَاسْطَةِ «حُواِرِزِ مَفِيدٍ».

- إِذَا افْتَقَدَ وَاحِدٌ أَوْ أَكْثَرٌ مِنْ شُرُوطِ إِجْرَاءِ المُقارَبَةِ التَّحْلِيلِيَّةِ - السَّابِقِ ذَكْرُهَا -
- فَلَا يَتَبَقَّى خِيَارٌ عَنْدَئِذٍ إِلَّا مُقارَبَةُ الْمَسَأَلَةِ عَبَرَ «الْتَّعْلُمِ الْحَاسُوبِيِّ».

وَيُنْتَرِكُ عَادَةً إِلَى «الْمُقارَبَةِ التَّحْلِيلِيَّةِ» عَلَى أَنَّهَا الْأَسْلُوبُ التَّقْلِيدِيُّ لِحُوسْبَةِ الْمَسَائِلِ، وَإِلَى مُقارَبَةِ «الْتَّعْلُمِ الْحَاسُوبِيِّ / التَّلْقَائِيِّ» عَلَى أَنَّهَا هِيَ الْمَهَارَسَةُ «الْأَكْثَرِ حَدَائِيَّةً» لِأَنَّ الْمُقارَبَةَ الْأُخْرَيَّةَ تَحْتَاجُ بِالْفَعْلِ إِلَى رِياضِيَّاتٍ أَكْثَرَ تَقْدُّمًا وَتَعْقِيْدًا وَكَذَلِكَ إِلَى قَدْرَاتٍ حَاسُوبِيَّةٍ لِتَنْفِيْذِهَا أَعْلَى مِنَ الْأُولَى وَلَذِلِكَ تَأْخِرُ تَطْوِيرِهَا زَمِنِيًّا. وَلَكِنَّ الْمُفَارَقَةَ الْلَّا لَفْتَةَ هِيَ أَنَّ الْمُقارَبَةَ الْفِطْرِيَّةَ فِي الْطَبِيعَةِ وَكَذَلِكَ فِي سُلُوكِ الْبَشَرِ هِيَ «الْتَّعْلُمِ الْحَاسُوبِيِّ / التَّلْقَائِيِّ»؛ فَالطَّفَلُ الصَّغِيرُ مثلاً يَتَعَلَّمُ الْلُّغَةَ بِالْمَعَايِشَةِ وَالْمَحَاكَاهَ دُونَ أَنْ يَعْرُفَ أَوْ يَدْرِسَ أَوْ لَا نَظَامُهَا الصَّوْقَيَّ وَلَا الصَّرِيفَيَّ وَلَا النَّحْوِيَّ ... إِلَخُ، وَالصَّبِيُّ الَّذِي يَتَعَلَّمُ حِرْفَةً كَالْحِدَادَةِ وَالسَّبِّكِ» يَتَعَلَّمُهَا أَيْضًا مِنْ مُعَلِّمِهِ (أَوْ مُعَلِّمِهِ) بِالْمَعَايِشَةِ وَالْمَحَاكَاهَ دُونَ أَنْ يَعْرُفَ أَوْ يَدْرِسَ أَوْ لَا عَلَمَ الْمَعَادِنَ وَلَا الْمَعَاجِلَاتِ الْحَارِرِيَّةِ وَلَا فِيْزِيَاءَ الْمَوَادِ ... إِلَخُ.

وكمثالٍ على المقاربة التحليلية نذكر المسألة (الجزئية البسيطة) الخاصة بالتحقق من «كلمة السر password» الرقمية لحساب المستخدم على أية منصة براجحية حاسوبية؛ فمن الواضح أن عملية التتحقق يمكن إجراؤها عن طريق اشتراط المطابقة الكاملة بين تسلسل الرموز المدخلة وبين تسلسل الرموز في كلمة سر المستخدم المحفوظة (بعد فك تشفيرها)، وهو ما يتضمن تعريفاً رياضياً (لرموز المسموح بها وغير المسموح بها في كلمات السر) وعمليةً رياضيةً (اختبار تساوي أو عدم تساوي رموزين) وعلاقةً رياضيةً (علاقة التساوي)، وهي كلها صريحة مباشرة.

كما نذكر كمثالٍ على مقاربة التعلم الحاسوبي مسألة «التعرف على مفردات الكلام المقطوقي^(١) Speech Recognition» من متعددٍ واحدٍ التي سبق عرضها قبيل ختام القسم الفرعي ١,٧ من هذا الفصل. وأفضل ما يُعرفُ من أساليب رياضيةٍ - دون الولوج في تفاصيلها - لإنجاز التعلم الحاسوبي مع هذه «المعضلة» هي «نماذج ماركوف المخفية Hidden Markov Models» بالتزامن مع «النماذج الاحتمالية لتركيب الفونيات».

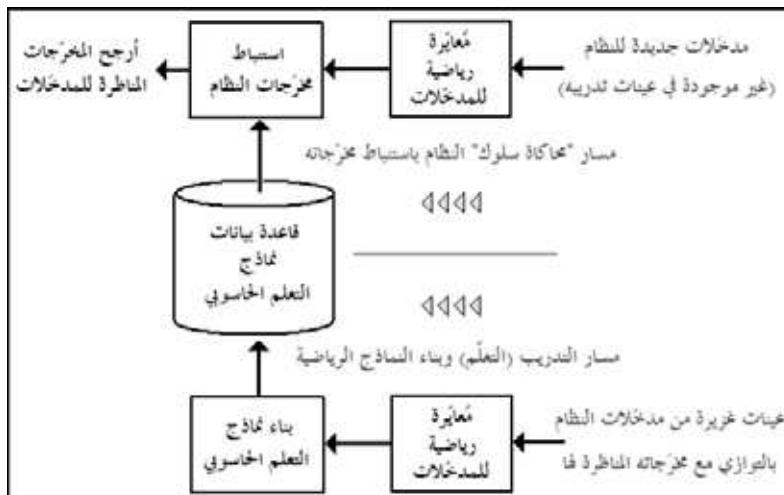
ويبين (الشكل ٧) أدناه المكونات العامة لمقاربة التعلم الحاسوبي؛ حيث تَظَهُر ارتباطُ هذه المكونات ومساراتُ تدفق البيانات فيها بينها. ومن أهم ما ينبغي ملاحظته في هذا الشكل وجود مسارين؛ أولهما تُهَضَّمُ فيه عيناتٌ وفيرةٌ من مدخلات النظام/الظاهرة قَيْدَ الحوسية والمخرجات المُناظرة لها - بعد تحويل صورتها الخام إلى صورةٍ رياضيةٍ نمطيةٍ/عياريةٍ - عبر واحدٍ من أساليب التعلم الحاسوبي حتى تُحسب منها معاملاتٌ نماذج التعلم الحاسوبي وتُخَرَّنُ في قاعدة بياناتٍ، ويسمى هذا المسار «طُورُ التعلم learning phase» أو «طُورُ التدريب training phase». أما ثانى المسارين فُيسمى «طُورُ الاستنباط» أو «طُورُ التشغيل» وهو الغرض من هذه المقاربة كلها؛ وفيه تُستتبَطُّ مُخرجاتُ النظام/الظاهرة المقابلة لأية مدخلاتٍ له ولو لم تكن في عينات طور التدريب. ومع اختلاف التفاصيل الرياضية فإن كلا الطَّورَيْن رُكْنَانُ أساسَيَّان في مختلفُ أساليب التعلم الحاسوبي المعروفة مثل «الشبكات العصبية الحوسية Artificial Neural Networks ANN»، و «آلية المتجه الداعم Support Vector Machine»،

١- بافتراض أن «حصيلة المفردات vocabulary» المطلوب التعرف عليها محدودة العدد مغلقة.

والتعلم وفق «مبدأ حالة اللا يقين القصوى»، و Maximum Entropy Principle، و «تعلم بايسيس الاحتمالي» Bayesian Learning، ... إلخ.

وينبغي التنبية على أنه إذا توافرت شروط «المقاربة التحليلية» المذكورة سلفاً في معالجة مسألة ما فلا ينبغي اللجوء إلى مقاربة «التعلم الحاسوبي» في معالجتها؛ حيث إن الأولى هي الأوضح والأضيق تعبيراً عن طبيعة المسألة وهي في الغالب الأكثر اقتصاداً في الإنجاز المعمومي الحاسوبي، أي أنها كالماء لوضوء الصلاة إذا حضر بطل التيم.

يقي قبل الانتقال إلى القسم الثالث من هذا الفصل أن نشير إلى أن بعض المسائل المركبة قد تصلح لمعالجة جزء منها «المقاربة التحليلية» بينما لا يصلح لمعالجة بقيتها إلا مقاربة «التعلم الحاسوبي» وعندئذ تسمى «المقاربة الهجينية Hybrid approach».



الشكل 7: مكونات مقاربة التعلم الحاسوبي، وارتباطها، ومسارات تدفق البيانات بينها. ⁽¹⁾

٣. تحديات نَمَذْجَةِ الْلُّغَاتِ الطَّبِيعِيَّةِ وَمَعَالِجَتِهَا آلِيَّاً.

سوف نُسقِطُ في هذا القسم بعض الاعتبارات الهامة للمقاربتين «التحليلية» و «التعلم الحاسوبي» على قضية «المعالجة الآلية اللغات الطبيعية Natural Language Processing (NLP)»، ولكن قبل ذلك نحتاج إلى التوقف مع

١- تشتهر «معاملات نماذج التعلم الحاسوبي» أي «القيم العددية المحددة حالات هذه النماذج» بين الباحثين والمطوريين في هذا المختصر («نماذج models» التعلم الحاسوبي)، لذا لمزيد التنبية على عدم الخلط بينها.

اصطلاح مركزيٌّ تكرر معنا مراراً منذ بداية هذا الفصل لتأمل معناه وعواقبه ألا وهو «النموذج الرياضي Mathematical Model» و «النَّمْذَجَة Modeling». فمن أجل دراسة ظاهرة/ مسألة طبيعية ما على نحو علميٍّ موضوعيٍّ يسمح بالقياس والمقارنة والتنبؤ والتصميم والتخطيط والتنفيذ فإن أفضل ما لدينا من أدواتٍ هي ترسانة الرياضيات بما فيها من حساب ومنطق وتحليل ... إلخ، ومن أجل إعمال هذه الأسلحة والخيل الرياضية في المسألة قيد الدراسة فلا بدًّ أولاً من صياغة وصفٍ رياضيٍّ متماسكٍ لهذه المسألة؛ فيسُمَّى هذا الوصف «النَّمْذَجَةُ الرياضيَّةُ» للمسألة كما تُسمَّى عملية بناءه «النَّمْذَجَةُ الرياضيَّةُ» لها.

وهنا يبرز سؤالٌ مُحْوَرٌ: هل النموذج الرياضي مُكافٍ للمسألة الطبيعية النَّمْذَجَةِ؟ والإجابة قولاً واحداً هي أنه يستحيل^(١) أن يكافيَ نموذجٌ رياضيٌّ المسألة التي يندرجها، وأن أفضل نموذج رياضيٌّ يمكن بناؤه للمسألة هو نموذجٌ تقريريٌّ مبسطٌ لها. ويبَرَزُ (الجدول ٢) أدناه أهم الملامح التي تختلف فيها الظواهر/ المسائل الطبيعية عن أفضل ما يمكن بناؤه من نماذج رياضية لها.

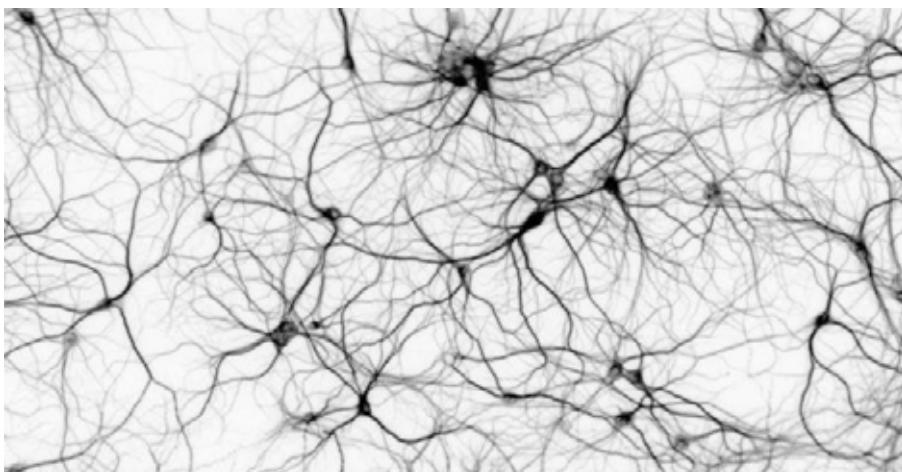
صفات النموذج الرياضي للظاهرة الطبيعية	صفات الظاهرة الطبيعية
محدودية عدد العناصر والمكونات	الكثرة المائلة للعناصر والمكونات
تنظيم وترتيب عناصر النموذج في أساق بسيطة	التعقيد المائل (شبه الفوضوي) لترتيب العناصر وعلاقتها
تقليل الاتصال والاعتماد المتبادل بين العناصر إلى أدنى حد	الاتصال والاعتماد المتبادل بين مختلف العناصر
وجود حدودٍ حاسوبيةٍ مقيّدةٍ لتنفيذ النموذج	لا حدودٍ حاسوبيةٍ تقييد تطور تفاعلات الظاهرة
وجود قيود على الدقة الحسابية في تنفيذ النموذج	لا قيودٍ حاسوبيةٍ على دقة تفاعلات الظاهرة
بعض العناصر والآليات التفاعل مجهلة	كل العناصر والتفاعلات مضمونة في الظاهرة
بعض العناصر والتفاعلات مشاركة في النموذج	كل العناصر والتفاعلات مشاركة في الظاهرة
النموذج قائمٌ بذاته معزولٌ عما سواه	الاتصال والاعتماد المتبادل بين الظاهرة وبيتها الخارجية

الجدول ٢: أهم الفوارق بين الظاهرة الطبيعية وأفضل نموذج رياضي يمكن بناؤه لها.

١- وهي استحالةٌ مبدئيةٌ لأن تلك الفوارق تظل قائمةً في عمق الطبيعة وظواهرها وفي عمق الرياضيات وعلوم الحاسوب، راجع القسمين الفرعيين ٣، ٤ و ١ من هذا الفصل، فلربما تتضاءل هذه الفوارق بعض الشيء مع تحسُّن تقنياتنا لكنها لن تتلاشى.

وكمثالٍ تطبيقيٌّ على هذه الفوارق نشير إلى «الشبكات العصبية» كظاهرة حقيقةٍ مقابل «الشبكات العصبية المحوسبة» كنموذج رياضيٍّ لها^(١)، ونترك لفطنة القارئ تعين الفوارق التي ذكرناها لتونا على هذا المثال بالاستعانة بالأشكال الثلاثة التالية (أرقام ٨ و ٩ و ١٠) التي تصور على الترتيب «الشبكة العصبية الحقيقة»، ثم «التبسيط الأوليٌّ لها»، ثم «الشبكات العصبية المحوسبة متعددة الطبقات» وهي الآن النموذج الرياضيُّ المعياريُّ للظاهرة.

وبسبب تلك الفوارق بين الظاهرة الحقيقة ونموذجها الرياضي، فلا مفرٌّ من بقاء هامشٍ خطأً غير صفرٍ (ولنسمه ط) عند تنفيذ النموذج، ومن الهام للغاية الانتباه عند المفاضلة بين نموذجين رياضيين لنفس الظاهرة/ المسألة أن المقارنة لا تكون بين دقة الأداء (ولنسمهما ق) لكلا النموذجين - وهي مغالطةٌ شهيرةٌ يلجم إليها كثيراً بعض المدلسين - بل بين هامشي الخطأ فيها؛ وبعبارة أخرى فإن: معامل المفاضلة بين نموذج ١ ونموذج ٢ يساوي: $ط \div ط = (1 - ق) \div (1 - ق)$ ، بينما المغالطة هي أن يُحسب هذا المعامل بالصيغة $ق \div ق$.^(٢)



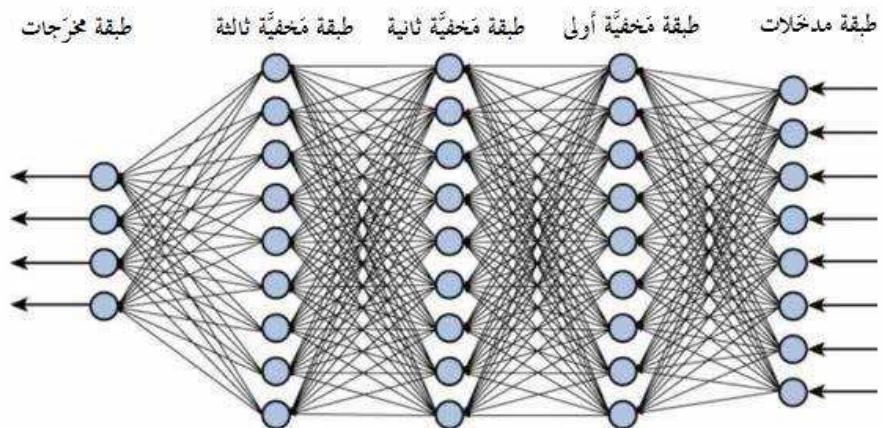
الشكل ٨: لقطةٌ مُهريةٌ من شبكةٍ عصبيةٍ حقيقة.

١- وقد تعرضنا للشبكات العصبية وحوسيتها في الإطار قبل الأخير من القسم الفرعي ٧، ثم في القسم الفرعي ٨ من هذا الفصل.

٢- مثال: مع نموذج ١ ذي دقة أداء ٩٥٪ مقابل نموذج ٢ ذي دقة أداء ٩٥٪، فإن معامل أفضلية الثاني على الأول يساوي $(1 - ٩٥\%) \div (1 - ٩٥\%) = ٢$ ولا يساوي أبداً $٩٥\% \div ٩٥\% = ١$ ، والفارق بين التقديرتين كبير جدًا.



الشكل ٩: تبسيط أولٌ للشبكات العصبية يحافظ على القنوات الاتصالية بين خلاياها.



الشكل ١٠: التموج المعياري للشبكات العصبية المحوسبة متعددة الطبقات.

ورجوعاً إلى قضية اللغات الطبيعية - وهي الظاهرة الطبيعية شديدة التعقيد - فإن التحديات التي تواجه نماذجها حاسوبياً بكلٍّيتها كبيرةً للغاية نعى إجمالاً أهمّها فيما يلي:

أولاً: تحدي الربط بين اللغة من حيث هي رموز (أصوات، مفردات، تراكيب ... إلخ) وبين دلالاتها في العالم الواقعي؛ حيث إن «المعرفة الفطرية commonsense knowledge»^(١) بالعالم الحقيقي - وهي التي يشترك فيها كل (أو معظم) البشر - هائلة الصخامة لدرجةٍ يجعل من تمثيلها رياضياً في قاعدة معرفية (تُعرف باسم «أونطولوجياً») أمراً بعيد المنال، ناهيك عن المعارف المتخصصة لفئة معينة من الناس (أونطولوجياً عَمَال المناجم، أونطولوجياً أطباء الأسنان ... إلخ).

ثانياً: لأن الخيال جزءٌ أساسيٌ في التفكير البشري فإن التفكير فيه والتعبير عنه ينعكس كذلك في اللغة الطبيعية، ولذلك فإن التحدي المُعْجِز لا يقتصر فقط على بناء أونطولوجياً للعالم الحقيقي بل يمتد إلى الحاجة إلى بناء أونطولوجياً للخيال (وهو مجال مفتوح غير محکوم)!

ثالثاً: لا يُصْمِنُ «مُتَّجِّحُ اللغة» (المتحدثُ أو الكاتبُ) عند استخدامه اللغة أداءً اتصاليةً أو قناعةً لنقل محتوى معرفيٍ على كُلِّ المعلومات التي يريد توصيلها للمتلقّي (السامع أو القارئ) في رسالته، فلو فعل لصارت العملية ثقيلةً العبء لدرجة تقارب الاستحالة، ولكنه يفترض أن السامع يشترك معه في معرفة سياق رسالته^(٢) ولذلك فإنه يُصْمِنُ في رسالته فقط المعلومات المضافة على سياقها، فتبقي المادة اللغوية (نصًا أو كلامًا منطوقًا) خُلُوًّا من السياق الذي تجري فيه الرسالة وتكون بذلك المعلومات المتاحة للألة التي تحاول معالجة اللغة الطبيعية ناقصةً نقصًا فادحًا.

رابعاً: يفترض كذلك مُتَّجِّحُ اللغة في متنقّلها أنه قادرٌ على استكمالِ كثيرٍ من المحدودات في رسالته ليس فقط لامتلاكه «المعرفة الفطرية»، بل كذلك لامتلاكه القدرة على «التحليل الفطري commonsense reasoning»^(٣) وهذا تحدي آخرٌ مُعْجِزٌ.

١- سبقت الإشارة إلى تلك المسألة الأونطولوجية في هذا الفصل في القسم الفرعي ٦ - بعد الشكل رقم ٣ - وكذلك قرب ختام القسم الفرعي ١، ٨.

٢- متكلّم بتذمّر: «غريبٌ أمر هذا الطقس!». السياق الغائب هو: أن الكلام في وقت الصيف في مكان يكون عادةً ما يكون جافاً صيفاً بينما عاصفةً ماطرةً تمر بالمكان.

٣- راجع التعليق المأمثي المترن بالشكل رقم ٤ في القسم الفرعي ٧ من هذا الفصل حيث فصّلنا مُعِضلة التحليل الفطري.

خامسًا: لا يخدم «المجاز metaphor» في اللغة فقط أغراضها الشعرية والتصويري الأدبي، بل إنه في الحقيقة يتغلغل تغلغلًا لا ينفصم عن صلب أي غرضٍ من أغراض استخدامها^(١)، والتحدي الذي يمثله المجاز هو أنه يزرع في الرسالة اللغوية عاملًا من «الالتباس المقصود» لا سبيل لإزالته آليًا حتى ولو توفرت لدينا - بمعجزة - أو نطولاً وجيا للخيال البشري.

سادسًا: كل ما سبق من تحدياتٍ ليست «استاتيكيةً» (ثابتةً في الزمن) بل «динاميكيةً» (متحرّكةً) دائمًا فيه، وتعتمد هذه الحركة في الزمن على تفاعلاتٍ اجتماعيةٍ ثقافيةٍ غايةً في التعقيد والتشابك، ولذلك فحتى لو تغلبنا - من باب الافتراض - على ما سبق من تحديات فإن علينا تعديلها باستمرار لتساير هذه التغيرات الدائبة.

سابعًا: ونأتي إلى مسأك ختام هذه التحديات وأعسرها جميًعاً ألا وهو تحقيق عاملٍ «الوعي» و «الذاتية» في اللغة والذي (يكاد) يستحيل مبدئيًّا تحقيقُها لاصطدامها بسقف الرياضيات؛ ويرجى الرجوع إلى ختام القسم الفرعي ١,٨ من هذا الفصل حيث سبق أن بيَّنا فيه سبب تلك الاستحالة. وتعني تلك الاستحالة أن أية آلة/ برجمية حاسوبية لمعالجة اللغات ستظل مجرَّد آلةٍ جامدةٍ لا تعي وجودها، ولا ذاتية لها، تنفذ فقط ما يُمْلَى عليها وما تُمْلَأُ به من تعليماتٍ منها كانت.

النتيجة الواضحة إذن هي أن أية نماذج رياضيةٍ صريحةٍ تحاول استيعاب «كامل» الظاهرة اللغوية الطبيعية عليها أن تتغلب على كل الأهوال السابق ذكرُها وهي في مقام المستحيلات أو تكاد. ولذلك فإن نمذجة الظاهرة اللغوية الطبيعية بأكملها عبر «المقاربة التحليلية» لربما كانت أكثر مهام النمذجة التحليلية عُسْرًا على الإطلاق، ولم لا والإنسان هو أَعْقَد الكائنات على الإطلاق، والظاهرة المطلوب استيعابها هي لغته التي يعكس فيها جوهر فكره ويتعدد فيها صدّى وجوده.

١- «... وأضاف أنه يتطلع إلى تحطيم الأرقام القياسية للأرباح التي حققتها الشركة في العام المالي المنصرم ...» هذه جُذَادَةً من تقريرٍ ماليٍ جادًّا عن أعمال إحدى المؤسسات التجارية، وتحتوي الجذادة عدّة استخدامات مجازية؛ منها مثلاً «يتطلع» حيث جسَّد المتكلم «تحطيم الأرقام القياسية» وكأنه شيءٌ ماديٌ يُرَى، وكذلك «تحطيم» حيث جسَّد المتكلّم أيضًا «الأرقام القياسية» وكأنها شيءٌ ماديٌ قابلٌ للتحطيم، ... إلخ.

لا مَفَرَّ إِذن من مقاربة «التعلم الحاسوبي» إن كنا نريد أن ننجز أي تقدم على صعيد المعاجلة الآلية «الكاملة» للغات الطبيعية، أليس كذلك؟ ولكن ماذا لو كنا فقط نطمئن إلى معاجلة آلية «جزئية» للغات الطبيعية؛ مثل «الصرف morphology» أو «النحو syntax»؟ هل تُجْدِي عندئذٍ نفعاً «المقاربة التحليلية» وَحْدَها؟

يمكن بالطبع بناء نموذج رياضي تحليلي لبنية الكلمة وصفاتها في أية لغة طبيعية في كلا الاتجاهين البنائي والتلفيكي ومن ثم تففيذه حاسوبياً، ولكن المشكلة التي سنكتشفها سريعاً هي تعدد الحلول التي يتتجها مثل ذلك النموذج؛ وهو ما يعرف بـ «الالتباس ambiguity» والذي لا يمكن إزالته - أي اختيار الحل الصحيح من بينها - إلا بالتغلب على التحديات التي أسلفنا ذكرها.^(١) وكذلك يمكن (بصعوبة) بناء نموذج رياضي تحليلي يستوعب التراكيب النحوية المسموح بها في أية لغة طبيعية في كلا الاتجاهين البنائي والتلفيكي ومن ثم تففيذه حاسوبياً، ولكن القاعدة الأصولية الشهيرة لدى النحاة «الإعراب فرع المعنى» تكفينا مؤونة شرح ما سيحدث حينئذ من التباس عنيف^(٢) في غياب معاجلة التحديات السابق ذكرها آنفًا.

وفي الحقيقة فإنه يندر العثور على مسائل معاجلة اللغات الطبيعية -ولو جزئياً - يمكن نمذجتها دون التباس بمقاربة تحليلية صرفة، فهل التزام مقاربة «التعلم الحاسوبي» وإهمال «المقاربة التحليلية» هو القَدْر الذي لا فكاك منه دائمًا؟ الشق الأول من الإجابة هو: نعم، مقاربة «التعلم الحاسوبي التلقائي machine learning» ضرورية

١- عندما صادفتك عزيزي القارئ هذه الفقرة «يمكن بالطبع بناء نموذج رياضي تحليلي لبنية الكلمة وصفاتها في أية لغة...» فلا بد أنك قد فَسَرْت أن كلمة «لينية» هي (حرف جر لـ) + اسم مجرور على وزن « فعلة» هو بُنْية من مادة (بـ نـ يـ) لأنك تفهم النظام الصرفي العربي ومنظومته النحوية وتعرف سياق الكلام ودلائله وتستنتاج معنى كاتبه ... إلخ. ولكن أي نموذج تحليلي يعطي النظام الصرفي العربي فقط سوف ينتهي إلى حلول متعددة: («البنية» = اسم على وزن « فعل» هو «أَبَن» من مادة (لـ بـ نـ) + ياء النسب + تاء التأنيث) أو («البنية»: حرف جر (لـ) + اسم مجرور على وزن « فعل» المحرّر بسبب الإعلال إلى «فعي» هو «بَنِي» من مادة (بـ نـ وـ) + تاء التأنيث) أو («البنية»: حرف جر (لـ) + اسم مجرور على وزن « فعل» مضئف العين واللام هو «بُنِي» + ياء النسب + تاء التأنيث) ... إلخ.

٢- الواقع أن الالتباس الحادث حينئذ ذو رُتبة أعلى كثيرة من الالتباس الصرفي الذي سبقت الإشارة إليه والتمثيل له؛ حيث إن البنية النحوية هي بُنْية شجرية في جوهرها (لأن كل مُكوّنٌ من مكوّنات التركيب النحوي قد يكون هو نفسه تركيباً نحوياً) ولذلك ينفجر عدد المسارات أَسْيَّاً أثناء إجراء التحليل - راجع مناقشة مشاكل البحث الشجري في بدايات القسم الفرعي ١، ٧ من هذا الفصل - وبالتالي يتعدّر توليد كل الحلول النحوية الممكنة من أجل الاختيار بينها، بل يتحتم تقليل المسارات المولدة أثناء إجراء التحليل أولاً بأول من أجل كبح جماح هذا الانفجار الأَسْيَ.

للتعامل الحاسوبي «الخالي من الالتباس» مع الغالبية العظمى من مسائل «المعاجلة الآلية لللغات الطبيعية». أما الشق الثاني من الإجابة فهو أنه لا يمكن إهمال «المقاربة التحليلية» analytic approach فهي أيضاً ضرورية إلى جانب مقاربة «التعلم الحاسوبي»! ومَرَد ذلك إلى سَبَّيْن؛ أو هُمَا وَأَهْمُهُمَا هو أن تصميم وبناء المُدخلات إلى آلية التعلم الحاسوبي في طَوْرِهَا التدريجي والتشغيلي - راجع (الشكل ٧) أعلاه - تتضمن بناء نماذج تحليلية للوحدات/ التراكيب اللسانية الأساسية في الظاهرة اللغوية الطبيعية المطلوب تعلمُها حاسوبياً. أما ثاني هذين السَّبَّيْن فهو أنه في بعض الأحيان يَحْسُنُ - لأسبابٍ تتصل باقتصاديات صناعة البرمجيات وبكفاءة أدائها - أن يجري حل مسألة المعاجلة اللغوية الطبيعية حلاً أولياً جزئياً عبر «المقاربة التحليلية» ثم يجري إزالة الالتباس الناتج من هذه المقاربة عبر مقاربة «التعلم الحاسوبي» وَتُعرَفُ تلك الممارسة عندئذ باسم «المقاربة الهجينة hybrid approach».^(١)

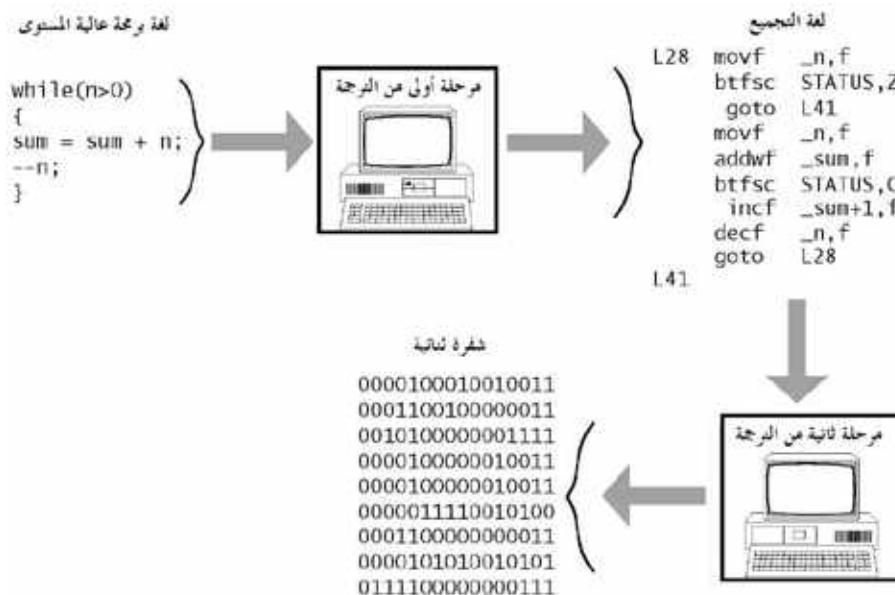
لَزِمَ قبل الانتقال إلى القسم الرابع من هذا الفصل تحرير مسألة سببٍ وما زالت تسبّبُ كثيراً من الجدل والخلط، وأدت بكثيرٍ من الباحثين الساعين بإخلاصٍ في طريق «المعاجلة الآلية للغة العربية» إلى التعرُّف في مسيرتهم البحثية والتطويرية؛ حيث لا يزال يُشَيَّع بينهم أن الطريق الوارد لتحقيق بعض الإنجازات الكبرى كالترجمة الآلية من وإلى العربية، والتحليل النحواني الآلي للغة العربية، ... إلخ يقع في دُرُوبٍ ما يشتهر بمدرسة «تشوسمكى»^(٢) من «نحو تحويلي» و«نحو توليدى» ... وما إلى ذلك من «مقارباتٍ تحليلية» أُسْسَتْ على نظرية «النحو الرياضي Formal Grammars» و«اللغات المحددة رياضياً Formal Languages». فما هي حقيقة ذلك وما القصة وراءه؟

الحقيقة هي أن هذه المدرسة بأساليبها قد حققت نصراً عملياً هائلاً بدأ من نهاية خمسينيات القرن العشرين الميلادي ويمتد أثره العميق حتى وقتنا الراهن؛ ألا وهو

١- من الأمثلة التي تصلح معها تلك «المقاربة الهجينة» التعامل مع مسألة «التحليل الصفي العربي» التي أشرنا إليها سابقاً؛ فالمقاربة التحليلية عندئذ تستخرج كل الحلول الممكنة على القياس في نظام الصرف العربي، ثم تتكثّف مقاربة «التعلم الحاسوبي» باختيار أرجح تلك الحلول بمعلومية اقتراناتها الإحصائية بما حولها من مفرداتٍ وتراتيب.

٢- «تَعُوم تشوسمكى» هو العالم الأميركي البارز في «العلوم الإدراكية»، و«علوم الحاسوب» وهو الأب المؤسس للسُّلُسانيات المعاصرة، وهو كذلك «فِيلِسُوفٌ تَحْلِيلِيٌّ»، ثم مؤرخٌ وناشطٌ سياسياً نافذٌ، وهو أستاذ في «معهد ماساتشوسيتس للتقنية MIT» ثم في «جامعة أريزونا» بالولايات المتحدة الأمريكية، ويعتبره الكثيرون من المؤرخين أحد أهم الشخصيات الفكرية المئة التي أثرت في المسيرة الحضارية للقرن العشرين الميلادي.

تمكنها من تصميم وتنفيذ مختلف «لغات الحواسيب عالية المستوى High Level Languages» وهي التي نقلت صناعة البرمجيات والمعلوماتية الحاسوبية إلى مستوى الصناعات الثقيلة الأساسية في الاقتصاد المعاصر. فقبل لغات البرمجة عالية المستوى تلك كان على من يود «إعطاء تعليمات» - أي «كتابه بـ برنامج» - لـ حاسوب رقميٍّ أن يكتبها بـ «الشفرة الثنائية machine code»Binary machine code التي لا تفهم الوحدات الإلكترونية الحاسوبية والمنطقية سواها، وهو أمر صعبٌ مكلفٌ وخصوصاً مع المهام البرمجية الكبيرة فضلاً عن ارتفاع احتمالية وقوع الأخطاء أثناء عملية البرمجة. أما لغات البرمجة عالية المستوى فهي تشبهُ لغة صياغة الحُوازِمات بنكهة اللغة الإنجليزية وهو ما يتقنه المتخصصون في الرياضيات وعلوم الحاسوب، ومهندسو المعلوماتية والبرمجيات.



الشكل 11: الترجمة من «لغة برمجية عالية المستوى» إلى «شفرة ثنائية».

و هنا يأتي الدور «السحري» لتلك المدرسة في القيام بالترجمة الآلية من اللغة البرمجية عالية المستوى إلى الشفرة الثنائية؛ وهو ما يصور (الشكل 11) عاليه مثلاً عليه. أغري هذا النجاح العظيم الكثيرين من الباحثين إلى محاولة تكراره في ملعب اللغات الطبيعية واعتماد تلك النماذج التي نجحت مع «اللغات عالية المستوى لـ برمجة الحواسيب

الرقمية» في حوسية اللغات الطبيعية توطةً لمعالجتها آليًّا. وبعد أكثر من ثلاثة عقود أثمرت المحاولات الحثيثة في ذلك المضمار عن إجراءاتٍ وأساليبٍ شتى لتحليل اللغات الطبيعية أفادت كثيراً اللسانين (البشريين) في تطبيقاتٍ كالنَّقد، وتحليل الخطاب، وعلم اللغة المقارن، وتحليل الإعاقات اللغوية ... إلخ، لكنها لم تفلح حتى الآن في تقديم حلولٍ عمليةٍ خاليةٍ من الالتباس لـ«المعالجة الآلية (الحاسوبية) للغات الطبيعية».

والسبب الجلُّي في ذلك الإخفاق هو أن «لغات البرمجة عالية المستوى» قد صُمِّمت ببالغ العناية لتفادي التحديات السبع المضليات التي ذكرناها آنفًا أمام المقاربة التحليلية «للغات الطبيعية»، ولذلك تُعرَفُ وتُعرَفُ لغات البرمجة بأنها «لغات مهذبة» التعريف «Well Defined Languages» ويمكن وفقٍ منهاجيةٍ واضحةٍ صياغةٍ خوارزماتٍ تستطيع تحليلها تحليلًا كاملاً دون أي التباس.

٤. إلى أين وصلت المعالجة الآلية لـ«لغات الطبيعية»؟

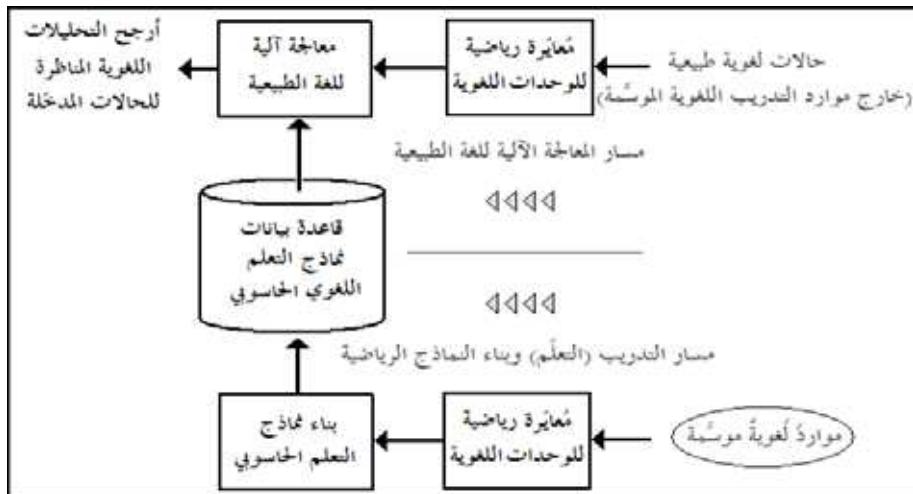
يحاول (الجدول ٣) أدناه رسم صورةٍ تقريريةٍ موجزةٍ لما عليه حالة الأبحاث والصناعة في نهاية العقد الثاني من الألفية الميلادية الثانية في حقل «المعالجة الآلية لـ«لغات الطبيعية»؛ حيث نُسَمِّي عند كُلّ مستوىً من مستويات اللغة بعضَ أهم مساحات تطبيقاتها، ثم نقوم برصد درجة أداءٍ (١) لأفضل الأعمال (٢) التي أُنجزَت في كل مساحة تطبيقية، ونذكر معه نوع المقاربة التي اتبَعها. ولربما كان أهم ملمحَين يسترعيان الانتباه في هذه الصورة هما: أن هناك علاقة عكسية بين كُلّ من «درجة الأداء performance» و «الطبقة اللسانية (أو المستوى اللغوي) language layer» فإذا زادت إحداهما انخفضت الأخرى، وأن مقاربة «التعلم الحاسوبي» حاضرةٌ دومًا إما بمفردها أو بالتهجين مع «المقاربة التحليلية».

١- وذلك على مقياسٍ تنازليٍّ من خمس درجات: {صناعيٌّ جيد، صناعيٌّ مقبول، بحثيٌّ متقدم، بحثيٌّ أوليٌّ، بعيد المنال}.

٢- وذلك بصرف النظر عن اللغة (أو اللغات) الطبيعية التي يخدمها هذا العمل - وإن كانت اللغة الإنجليزية بحكم الأمر الواقع في وقتنا هذا هي التي تحتل مركز الصدارة وبؤرة الاهتمام.

المستوى اللغوي	المساحة التطبيقية	درجة الأداء	المقارنة
المنطق	التعرف على النص المطبوع	صناعيٌّ جيد	التعلم الحاسوبي
	التعرف على النص	صناعيٌّ جيد	التعلم الحاسوبي
	التعرف على المتكلّم	صناعيٌّ جيد	التعلم الحاسوبي
	التعرف على الكلام المنطوق	صناعيٌّ جيد	التعلم الحاسوبي
	الإملاء الآلي	صناعيٌّ مقبول	التعلم الحاسوبي
	التحقق من النطق الصحيح	صناعيٌّ جيد	التعلم الحاسوبي
	توليد الكلام المنطوق من المكتوب	صناعيٌّ جيد	هجينة
	التحليل الصرفي	صناعيٌّ جيد	هجينة
	الضبط الصوتي للمفردات (التشكيل)	صناعيٌّ جيد	هجينة
	استخلاص الصفات التحويية للمفردات	صناعيٌّ جيد	هجينة
الصرفي	البحث النصيّ	صناعيٌّ جيد	هجينة
	التنقيب في النصوص	صناعيٌّ جيد	التعلم الحاسوبي
	اختصار/ تلخيص الوثائق النصية	صناعيٌّ مقبول	التعلم الحاسوبي
	مقارنة الوثائق النصية	بحسيٌّ متقدّم	التعلم الحاسوبي
	التحليل الدلالي المعجمي	بحسيٌّ متقدّم	هجينة
	صناعة المُعجمات، وإتاحتها للاستخدام	صناعيٌّ جيد	هجينة
	أنظمة تعليمية لغوية تفاعلية	صناعيٌّ جيد	هجينة
	التحليل النحوي	بحسيٌّ أولي	هجينة
	التحليل الأونطولوجي	بحسيٌّ أولي	هجينة
	الترجمة بمساعدة الحاسوب	صناعيٌّ جيد	التعلم الحاسوبي
الدّلالي	الترجمة الآلية	صناعيٌّ مقبول	التعلم الحاسوبي
	استخلاص توجهات الرأي العام	بحسيٌّ متقدّم	التعلم الحاسوبي
	إدارة الحوار الآلي	بحسيٌّ أولي	هجينة
	الفهم الآلي للنص	بعيد المنال	التعلم الحاسوبي
	التعبير الإنساني الآلي	بعيد المنال	هجينة
التدابري	الوعي بالعالم الحقيقي	بعيد المنال	هجينة
	امتلاك الذاتية	بعيد المنال	هجينة
الجدول ٣: صورةٌ تقريبيةٌ موجزةٌ لحالة الأبحاث والصناعة المعاصرة في حقل المعالجة الآلية للغات الطبيعية.			

ولأن «التعلم الحاسوبي» حاضر دوماً في هذه الصورة المعاصرة للمعالجة الآلية للغات الطبيعية، فمن المفيد إسقاط إطاره العام المبين في (الشكل ٧) سابقاً على «المعالجة الآلية للغات الطبيعية» وهذا ما يبينه (الشكل ١٢) أدناه؛ حيث نلاحظ أن أهم نقاط هذا الإسقاط هو أن التسمية العامة في (الشكل ٧) «العينات الغزيرة من مدخلات النظام بالتوالي مع خرجاته المعاصرة لها» تحول تسميتها إلى «الموارد اللغوية الموسّمة annotated language resources» في (الشكل ١٢)، وسوف يكون لـ «الموارد اللغوية annotation» و «التوسيم Language Resources» حضور هامٌ فيها تبقى من هذا الفصل.



الشكل ١٢: إسقاط الإطار العام لمقاربة «التعلم الحاسوبي» على «المعالجة الآلية للغات الطبيعية».

يشكّل «توسيم الموارد اللغوية Language Resources Annotation» المكون الأكبر في صناعة «المعالجة الآلية للغات الطبيعية» المعاصرة؛ فحيث إن هذه الصناعة ترتكز أساساً على مقاربة «التعلم الحاسوبي» فيها أساليب وأاليات «التعلم الحاسوبي» تنظر للظاهرة التي تعمل عليها كصناديق أسود مغلقة لا يعنيها منه سوى مدخلاته وخرجاته، فإن المادة المتغيرة التي ينبغي بناؤها خصيصاً لكل نظام تطبيقي هي «الموارد اللغوية الموسّمة» اللازمة لتدريسيه.

٥. هل هناك سقفٌ لمعالجة اللغات الطبيعية آليًّا؟

نعم، هناك سقفٌ لمعالجة الآلية للغات الطبيعية، وذلك لوجود سقفٍ للذكاء الاصطناعي^(١) وهو الإطار العام التي تقع داخله «المعاجلة الآلية للغات الطبيعية» وذلك بدوره لوجود سقفٍ للرياضيات^(٢) التي هي الإطار الجامع لذلك كله. فالسقف الذي يحدّد سقفَ الرياضيات وسقفَ الذكاء الاصطناعي وتقع تحته آفاق «المعاجلة الآلية للغات الطبيعية» يتمثل في المساحات اللغوية التي يتبدى فيها «وعيُّ» الكاتب أو المتحدث و «ذاتيَّة»؛ فهاتان مساحتان يتذرعان مبدئيًّا على كل أدواتنا معاجلتها آليًّا، ومن الحكمة عدم إهدار المجهود أو المال على البحث والتطوير فيها أو فيها يعتمد عليهما بقية من المساحات الأخرى في هذا الحقل^(٣).

وفيما تحت هذا السقف فلا يوجد من حيث المبدأ ما يقيّد «المعاجلة الآلية للغات الطبيعية»، ولكن من حيث القدرة العملية فإن ربط وحدات اللغة الطبيعية بمختلف مستوياتها بالعالم الحقيقي^(٤) تستلزم بناء «أونطولوجيا رقمية» لهذا العالم وهو ما يتطلب بدوره صياغةً ورقمنة «المعرفة الفطرية commonsense knowledge»، وكذلك «التحليل الفطري commonsense reasoning»، وهي جيًّاً مهامٌ هائلةً الضخامة^(٥) لا زلنا نجهل كيف نشرع فيها ويصعبُ تصورُ إنجازها في العقود القريبة القادمة، وهو ما يُقيّم سقفاً عمليًّا لالمعاجلة الآلية للغات الطبيعية تحت سقفها المبدئي المشار إليه في الفقرة السابقة.

تحت هذين السقفين إذن تعمل «المعاجلة الآلية للغات الطبيعية»، وبما أن أنظمة تطبيقاتها العملية/ الواقعية المعاصرة لا تستغني أبداً عن مقاربة «التعلم الحاسوبي» - كما بيَّنَ القسم السابق من هذا الفصل - فمن المهم أن نتعرف على كيفية أداء هذه المقاربة

١- وقد سبقت مناقشة ذلك في ختام القسم الفرعي ١، ٨ من هذا الفصل.

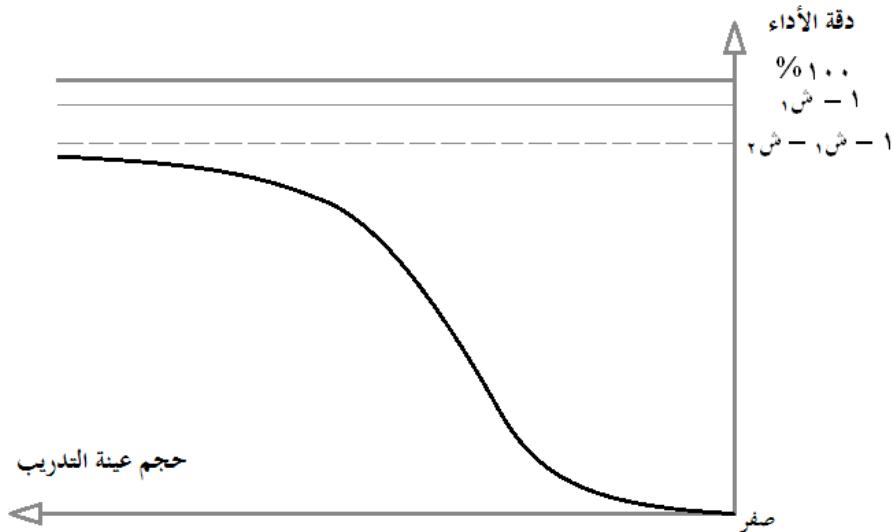
٢- وقد أشرنا لذلك في القسم الفرعي ١، ٣ ثم في القسم الفرعي ٦ من هذا الفصل.

٣- وذلك مثل التعبير عن «الأشاعر sentiments».

٤- والذي من دونه تبقى «المعاجلة الحاسوبية للغات الطبيعية Natural Language Processing (NLP)» أقرب لمعالجة الرموز أو «المعاجلة الرقمية للإشارات Digital Signal Processing (DSP)».

٥- وقد سبق التعرض لهذه المعضلات في القسم الفرعي ١، ٧ من هذا الفصل في المامش المتصل بالشكل رقم ٤، ثم أشرنا إليها مرةً أخرى مع ختام القسم الفرعي ١، ٨، ثم مرةً ثالثةً في القسم رقم ٣ بعْدَ الشكل رقم ١٠.

تحتَّهمَا. فكما يَبَّنَ (الشكل ١٢)؛ يَجِدُ تدْرِيبُ «نظام المعالجة الآلية للغة الطبيعية» على موارد لغوية موَسَّمةً تَمثِّلُ عِينَاتٍ كَبِيرَةً مَتَّنِعَةً بِمَا يَكْفِيُ مِنَ الظَّاهِرَةِ الْلُّغُوِيَّةِ المَطْلُوبَ تَعْلُمُهَا، فَتَتَشَكَّلُ نَتْيَاجَةً لِذَلِكَ (مَعَالِمَاتُ) «نَمَاذِجُ التَّعْلُمِ الْحَاسُوبِيِّ»، الَّتِي تَسْتَخَدُهَا آلِيَّةُ «الْمَعَالَجَةِ الْآلَيَّةِ لِلْلُّغَةِ الطَّبَاعِيَّةِ» فِي «وقْتِ التَّشْغِيلِ» مِنْ أَجْلِ اسْتِبَاطِ أَرْجُحِ الْحَلُولِ لِكُلِّ حَالَةٍ لغُوِيَّةٍ تُعَرَّضُ عَلَيْهِ فِي إِطَارِ الظَّاهِرَةِ الْلُّغُوِيَّةِ الَّتِي جَرَى تَعْلُمُهَا وَذَلِكَ بِدَقَّةِ أَدَاءٍ قَ.



الشكل ١٣: «مَنْحُنِيُ التَّشْبِيعِ» لِوَصْفِ السُّلُوكِ الْعَامِ لِعَمَلِيَّةِ التَّدْرِيبِ فِي مَقَارِبَةِ «الْتَّعْلُمُ الْحَاسُوبِيِّ»

وَمَعَ زِيَادَةِ حَجمِ «الموَارِدِ الْلُّغُوِيَّةِ المَوَسَّمَةِ»^(١) أَثْنَاءِ عَمَلِيَّةِ التَّدْرِيبِ، تَتَحَسَّنُ دَقَّةُ أَدَاءِ الْمَعَالَجَةِ الْآلَيَّةِ أَثْنَاءَ وَقْتِ التَّشْغِيلِ، لَكِنَّ اطْرَادَهُ هَذَا التَّحْسِنُ فِي دَقَّةِ الْأَدَاءِ يَتَبَاطَأُ بَعْدَ حَجمٍ مُعَيَّنٍ لِلْمَوَارِدِ الْلُّغُوِيَّةِ المَوَسَّمَةِ^(٢) وَيُدْخِلُ فِيهَا يَسْمَى بِمَرْحَلَةِ «الْتَّشْبِيعِ» saturation ثمَّ يَتَقَارَبُ بَعْدَ ذَلِكَ مِنْ قِيمَةٍ ثَابِتَةٍ لَا يَتَخَطَّاهَا مَهْمَا زَادَ حَجمُ «الْمَوَارِدِ الْلُّغُوِيَّةِ» الْمُسْتَخَدَّةِ لِلْتَّدْرِيبِ، وَلَا تَسَاوِي هَذِهِ القيمةُ الْوَاحِدُ الصَّحِيحُ (أَيْ لَا تَسَاوِي

١- مع اشتراطِ المحافظةِ عَلَى تَنْوِعِهَا.

٢- يَعْتَمِدُ هَذَا «الْحَجمُ الْمُعَيَّنُ» الْمُعْرُوفُ بِاسْمِ «كَمِيَّةِ التَّشْبِيعِ» عَلَى طَبِيعَةِ الظَّاهِرَةِ الْلُّغُوِيَّةِ المَطْلُوبَ تَعْلُمُهَا آلِيَّاً؛ فَيُزِيدُ ذَلِكُ الْحَجمُ كُلَّمَا زَادَ تَعْقِيدُ الظَّاهِرَةِ فِي دِيَنَامِيَّاتِهَا وَتَشَابُكِهَا فِي عِلَاقَاتِهَا، وَعَادَةً مَا يَقِعُ ذَلِكُ كُلَّمَا عَلِتَ «الْطَّبِيعَةُ الْلُّسُوَانِيَّةُ» الَّتِي تَنْتَمِي إِلَيْهَا الظَّاهِرَةُ. وَجَدِيرٌ بِالذِّكْرِ أَنَّ أَحَدَ أَسْبَابِ اتِّبَاعِ «مَقَارِبَةِ هَجْبَيَّةِ» فِي بَعْضِ نَظَمِ «الْمَعَالَجَةِ الْآلَيَّةِ لِلْلُّغَةِ الطَّبَاعِيَّةِ» هِيَ أَنَّ الْمَعَالَجَةَ الْجَزِئِيَّةَ عَبَرَ «الْمَقَارِبَةِ التَّحْلِيلِيَّةِ» قَدْ تَقْلِلُ كَثِيرًا «كَمِيَّةِ التَّشْبِيعِ» مِنْ «الْمَوَارِدِ الْلُّغُوِيَّةِ المَوَسَّمَةِ» الْمُطْلُوبَةِ لِلْتَّدْرِيبِ مَا يَخْسِنُ مِنْ اقْتِصَادِيَّاتِ بَنَاءِ النَّظَامِ كُلُّهُ.

مئةً في المئة) بل تصغره بمقدار «هامش الخطأ البشري شـ»، مضافاً إليه «هامش خطأ التعلم الحاسوبي شـ»، وهو ما يلخصه (الشكل ١٣) أعلاه.

فأهل أية لغة من البشر يخطئون ولا بدّ فيها^(١) ولو بهامش بسيط (= شـ)، ينعكس في توسيم أي مورد لغوي لتدريب أنظمة تعلمها حاسوبياً ولا يمكن أن يتجاوز أداءً أية آلية للتعلم الحاسوبي دقةً مواردتها التعليمية. كما أن الغياب الإلزامي لتمثيل أركانٍ أساسيةٍ مثل «تمثيل المعرفة الفطرية» و «الارتباط بالعالم الحقيقي» و «التحليل الفطري» و «غياب السياقات» عن المادة اللغوية قيد التعلم والتحليل، و «تجاهل الطبقات اللسانية الأعلى عند معالجة مسألة لغوية من طبقة أدنى» ... إلخ ينعكس حتى على طورِي التدريب والتشغيل كليهما ويظهر على هيئة هامش خطٍ (= شـ)، تعتمد قيمته على أثر كل هذه الأركان الغائبة في المسألة اللغوية المعينة المعينة.

٦. هل للغة العربية خصوصية مع الذكاء الاصطناعي؟

تشير في أدبيات معالجة اللغة العربية وحوسبتها مقولاتٌ تقطع بخصوصية اللغة العربية وتفردُها عن بقية لغات العالم الطبيعية الكبرى بدرجةٍ لا نظير لها من الصعوبة على جميع مستوياتها اللسانية؛ صوتيًّا وصرفياً ونحوياً ودلاليًّا ... إلخ لدرجةٍ باتت معها تلك المقولات في مقام المأثورات التي توارثها الأجيال وكاد تصدير الأبحاث والدراسات بها - تعليلاً للتقصير السابق في معالجتها و/أو تسويغاً للعمل الذي يتصدى له المؤلف - أن يكون في مقام العرف السائد الذي تُستهجنُ مخالفته. ولكن توسيع النظر في واقع اللغات الطبيعية الكبرى حول العالم يكشف الكثير من التزدد في مثل تلك المقولات؛ فائئن كان نظامُ بناء الجملة (النحو) العربي - على سبيل المثال - شديداً المرونة يسمح بالتقديم والتأخير والحدف والتقدير ... إلخ مع ما يطرحه ذلك من صعوبات والتباسات، فإن الصينية لغةٌ غير هجائية^(٢) كما أن المفردات في الصينية لا تتمايز فقط برسمها وأصواتها بل من اللازم كذلك مراعاة التنغير والنبر للتمييز بينها، ونظام بنية الكلمة (الصرف) في الألمانية يسمح ببناء كلماتٍ مرتكبةٍ شديدة التعقيد،

١- المقصود هنا أبناء اللغة الذين يتحدثونها بطلاقة الخبراء بها، والمقصود بخطائهم السهو والخلط اللذان لا مفرّ من وقوفهم ولو بنسبة بسيطةٍ لا تساوي صفرًا.

٢- المقصود هو لغة «المائدارين» النسخة الرسمية السائدة من اللغة الصينية التي لا تتكون كلها من حروفٍ أبجديةٍ/ألفبائيةٍ.

والإليابانية لا تترك فوائل أو فراغات بين كلماتها، والإنجليزية لغة هجينة إلى حدٍ بعيدٍ جداً على مستوى مفرداتها كما تتميز بمعدل نموٍ معجّميًّا مرتفع للغاية، ولغة الموسما في غرب أفريقيا (مثل نيجيريا) بها أصواتٌ غير مألوفةٌ في غيرهاً كالأصوات الشفطية، وأغلب المخطوط في الفرنسية لا يُنطقُ وبعض ما لا يُنطقُ يظهرُ نطقه تحت شروطٍ خاصةٌ ... إلخ. والخلاصة المقصودة من ذلك هي أن الصعوبات اللغوية الظاهرة لا تقتصر على العربية، فكل لغة من لغات العالم الحية الكبرى لها صفاتها المخصوصة التي تمثل صعوباتٍ في عيون غير الناطقين بها.

وبالرغم من الاختلافات البينية بين اللغات الطبيعية على مستوى الطبقات اللسانية الدينية والوسطى (الصوتية، والخطاطية، والصرافية، والمعجمية، والنحوية) التي تُعتبرُ اختلافاتٍ في الترميز والصياغة، فإن اللغات الطبيعية على مستوى الطبقات اللسانية العليا (الدلالي، والمجازي، والتدابري) تعكس في جوهرها وفحواها المشتركة الإنساني من (معرفةٍ فطريةٍ بالعالم الحقيقى، وتحليلٍ فطريٍّ، ...).

ولذلك إذا اتّبعت «المقاربة التحليلية» سبيلاً في «تحليل لغة طبيعية» فإن منبع الصعوبة الحقيقية هو تلك الطبقات اللسانية العليا (الغائية غالباً في عملية التحليل) وما يتصل بها من تعقيد الفكُر والنفس الإنسانية^(١) وليس منبعها الطبقات اللسانية الأدنى المعنية بالترميز والصياغة؛ أما ما يبدو من صعوباتٍ ظاهريّة في خصوصيات كل لغة طبيعية على تلك المستويات الترميزية والصياغية^(٢) فمقدور أهل تلك اللغة في نهاية المطاف الإحاطة بها والسيطرةُ عليها تستوي في ذلك العربية مع أخواتها من ألسنة البشر.

أما إذا ما اتّبعت مقاربة «التعلم الحاسوبي» في معالجة اللغة العربية فإن تلك المقاربة - كما أوضحنا سابقاً في نهاية القسم رقم ٤ ثم في القسم رقم ٥ من هذا الفصل - لا تفرق

١- راجع التحديات السبع المعضلات أمام نبذة الظاهرة اللغوية الطبيعية بكلّيّها المفصلة في القسم رقم ٣ من هذا الفصل.

٢- في الواقع أن العربية إن كانت لها خصوصية في هذا الصدد بين ألسنة البشر الحية فهي أنها أكثرها نظاماً، خذ مثلاً الاتساق الكبير بين نطق النص ورسمه، ومثلاً آخر غلبة انتظام الاشتراق الصري في العربي من عدد صغير نسبياً من المواد وعدده أصغر من الصيغة الصرافية، ومثلاً ثالثاً اتساقها الدلالي المعجمي إلى حدٍ بعيد، ومن ثم قارئه مع ضعف أو غياب مثل ذلك الانتظام في كثير من مظاهر اللغات الأخرى - راجع المعجم الإنجليزي المعاصر على سبيل المثال - وهي أطوطها اتصالاً في التاريخ (عبر حوالى ألفي عام) وهي أكثرها توثيقاً وتقعيداً عبر تاريخها الطويل المتصل. وإن كان هناك من أثر لتلك الخصوصيات في مظاهر اللغة العربية فلا بدّ أن يكون أثراً إيجابياً للغاية عند اتباع أية مقاربة تحليلية لها على نقىض ما تزعم المقولات الشائعة التي بدأنا القول بها.

بين اللغة العربية وأية لغةٍ طبيعيةٍ أخرى؛ حيث لا تعنيها التفاصيلُ الداخليةُ للظواهر (اللغوية) التي تقوم بتعلُّمها فهي تنظرُ إليها جميًعاً كصناديقٍ أسودَ مغلقٌ لا يميُّزها عن بعضها البعض سوى الموارد اللغوية الموسَّمة الالازمة لتدريب آلية «التعلم الحاسوبي». ^(١)

قاعدة رقم ٥

إن تعليل التقصير والتأخر في حقل حوسبة اللغة العربية وتطبيقات المعالجة الآلية لها بدعوى اتصف اللغة العربية بخصوصياتٍ تجعل من حوسبتها ومعالجتها الآلية أمراً بالغ الصعوبة مقارنةً مع غيرها من اللغات الطبيعية لكرّ محضٍ وهمٍ ينبغي تَنْحِيَتُه؛ فدرجة الصعوبة في حوسبة اللغة العربية ومعالجتها آلياً لا تختلف عن صعوبة حوسبة أية لغة طبيعية أخرى ومعالجتها آلياً، سواء اتبعت في تلك المعالجة أساليب تحليلية أو أساليب للتعلم الحاسوبي أو أساليب هجينةٍ بينهما.

٧. ما هي آفاق مستقبل المعالجة الآلية لِلُّغة العربية عبر الذكاء الاصطناعي؟

مثلاً مثل سواها من اللغات الطبيعية الكبرى في العالم، فإن آفاق «المعالجة الآلية للغة العربية» تقع تحت السقفين اللذين جرى تعرِيفُهُما في القسم رقم ٥، ويمكن لها خلال ما لا يجاوز عقدهاً من الآن (أي بنهائيات العِقد الثالث من الألفية الميلادية الثانية) أن تدارك ما فاتها كي تلحق بالصورة - التي رصدها سابقاً الجدول رقم ٣ في القسم الرابع من هذا الفصل - حالة الأبحاث والصناعة المعاصرة في حقل المعالجة الآلية لِلُّغات الطبيعية، فتتحقق بذلك مكاسب اقتصادية كبيرة جراء المنافسة في هذا القطاع العظيم من اقتصاد المعلومات والمعرفة، وتتحقق آثار ثقافية إيجابية عميقه سوف تمثل دفعه قوية للقوة

١- ولذلك تحرص كبريات شركات التقنية المعلوماتية في عصرنا الراهن التي تتجه نظراً لمعالجة اللغات الطبيعية تغطي مختلفاً (أو عديداً من) اللغات الطبيعية الحية على اتباع مقاربة «تعلم حاسوبي» صرفة وتحاول تفادي أية «مقاربات تحليلية» ولو كانت جزئيةً؛ حيث يكفلها ذلك إنتاج نظام معالجة واحدٍ مع إنتاج موارد لغوية متعددة بتنوع اللغات المختلفة التي يغطيها، وذلك بدلأً من إنتاج أنظمة معالجة متعددة بتنوع تلك اللغات المختلفة. خذ مثلاً على ذلك «خدمة غووغل الشهيرة للترجمة» التي تغطي عدداً = ن من اللغات الطبيعية، فمع مقاربة التعلم الحاسوبي - وهي المتّعة فعلاً - فإن نظام الترجمة واحدٌ وما يتعدد فهو الموارد اللغوية الالازمة لتدريبه على الترجمة بين كل زوج من أزواج اللغات المشمولة البالغ عددها ن×(ن-١) من هذه الأزواج، فإذا فكرت «شركة غووغل Google» أن تتبع مقاربة تحليلية لاحتاجت إلى تطوير ن×(ن-١) من الأنظمة المختلفة كي تغطي الترجمة بين كل أزواج اللغات المشمولة.

الناعمة العربية لهذه الأمة، وتحقق أيضًا نجاحاتٌ علميةٌ وتقنيةٌ تسهم في النهوض بالبحث والتطوير في بلادنا في حقل الحوسبة والمعلوماتية على وجه العموم.^(١)

قاعدة رقم ٦

من أجل أن تتحقق «المعالجة الآلية للغة العربية» خلال نحو عقدٍ من الزمان بأفضل ما وصلت إليه أبحاث وصناعة «المعالجة الآلية للغات الطبيعية»، يلزم إجراءً ما يلي: أولاً: إعداد جيلٍ من الأساتذة النابهين في «هندسة وعلوم الحاسب» الخبراء في الذكاء الاصطناعي بمدرسته المعاصرة لجعلهم على إمامٍ ودراءٍ جيدةٍ بأسس علوم اللغة واللسانيات العربية وتحدياتها.

ثانياً: إعداد جيلٍ من الأساتذة النابهين في «علوم اللغة واللسانيات العربية» لجعلهم على إمامٍ ودراءٍ جيدةٍ بمبادئ «علوم الحاسب» ونُظُمه، ومبادئ «الذكاء الاصطناعي» وقضايا حosomeة اللغة الطبيعية ومعالجتها آلياً.

ثالثاً: أن يؤهلَ الصنف الأول من الأساتذة قاعدةً أوسع من طلبة هندسة وعلوم الحاسب النابهين بنفس الكيفية التي جرى إعداد هؤلاء الأساتذة بها.

رابعاً: أن يؤهلَ الصنف الثاني من الأساتذة قاعدةً أوسع من طلبة علوم اللغة واللسانيات العربية النابهين بنفس الكيفية التي جرى إعداد هؤلاء الأساتذة بها.

خامسًا: تقوم على هذه القاعدة الحاسوبية اللسانية التي جرى بناؤها في الخطوات السابقة صناعة «الموارد اللغوية العربية وتوسيمها» كصناعةٍ معلوماتيةٍ ثقيلةٍ، كثيفة العمالقة.

سادسًا: لا بدّ أن يضمّ كُلّ مشروع لصناعة «الموارد العربية اللغوية وتوسيمها»، و/أو توظيفها في مقاربة «تعلم حاسوبيّ» أو «مقاربةٍ هجينةٍ» لصناعة تطبيقٍ على «المعالجة الآلية للغة العربية» الأصناف الثلاثة الآتية من العاملين بالأوزان المناسبة لحاجة المشروع: المتخصصين في علوم اللغة واللسانيات العربية، والمتخصصين في هندسة وعلوم الحاسب ولا سيما الذكاء الاصطناعي، و«محترفي إدارة المشروعات Project Management Professionals (PMPs)»^(١).

ولعل هذا الفصل وهذا الكتاب أن يكونا خطوةً على هذا الطريق.

١- ويقوم هؤلاء بتطبيق آليات التخطيط والتنفيذ والمتابعة التي تساعد هذه المشروعات «كثيفة العمالقة» و«متشاركة المهام» و«ذات الحساسية العالية للدقة» على الوصول لبر الأمان؛ أي وفق مواصفات الجودة، ودون تجاوز الميزانية ولا الإطار الزمني.

٨. الخاتمة والتاتج

ارتحلنا عبر هذا الفصل في تاريخ حلم محاكاة الآلات للذكاء البشري منذ التاريخ القديم بالتدريج مع تطور الأدوات الممكنة له خلال القرون القليلة الماضية، ثم فصلنا في ظروف تأسُّس حقل «الذكاء الاصطناعي» في العصر الحديث ولا سيما مع بدء بناء الحواسيب الرقمية واستعراضنا بالتفصيل المراحل التي مر بها صعوداً وهبوطاً إلى أن استقر على مسار تصاعديٍّ سريع منذ حوالي ربع القرن.

قدم هذا الفصل بعد ذلك المقارنة بين الأساسيات حل المسائل المعترضة في إطار الذكاء الاصطناعي، والتحديات الأساسية أمام قضية «المعاجلة الآلية لِللغات الطبيعية» وهي إحدى المباحث الكبرى للذكاء الاصطناعي، ومضي الفصل ليبيّن حال هذه القضية مع كل مقاربةٍ منها، ولماذا مالت الكفة نحو مقاربة التعلم الحاسوبي «مقابل المقاربة التحليلية»، ثم انتقلنا بعد ذلك لرسم صورة الحالة الواقعية الراهنة لتطور الأبحاث والصناعة في حقل «المعاجلة الآلية لِللغات الطبيعية»، وبعدها أجبنا على سؤال السقف الأعلى لما يمكن أن تبلغه الأبحاث والصناعة في هذا الحقل - على الأقل في المستقبل المنظور.

وأخيراً انتقل التركيز على إسقاط ما سبق على «المعاجلة الآلية لِلغة العربية» وهل هناك حقاً خصوصيات لها تجعلها أعصى على المعاجلة الآلية من اللغات الأخرى؟ وما هي الأفق المستقبلية لحوسبة العربية ومعاجلتها آلياً؟

وفي الختام نؤكّد أن هذا الفصل - رغم كل ما يحشّد به تفاصيلٍ تاريخيةٍ وعلميةٍ - ليس في مقام تقديم شرح أو تحليلٍ تقيّيٍّ مستوفٍ لأية آلية رياضية أو حاسوبيةٍ من تلك التي يعرضها، فتلك الشروح والتحليلات وفيّرةٌ مستقرّةٌ في بطون الكثير والكثير من الأسفار المتخصصة الموجّهة إلى المتخصصين في علوم الحاسوب والرياضيات المتقدمة. ولكن رسالة هذا الفصل توجه أساساً إلى العاملين النابحين في علوم اللغة العربية وما يرْفُدُها من حقولٍ^(١) من يتطلعون و/ أو يتصدون بالفعل لتوظيف الحوسبة وتقنياتها

١- وتنوّجه أيضاً هذه الرسالة بطبيعة الحال إلى كل مثقّفٍ عربيٍّ مهتمًّا بالإسلام بيهودية الذكاء الاصطناعيٍّ وتاريخه وواقع ما يوفره من آلياتٍ وتقنيّاتٍ لمعالجة اللغات الطبيعية؛ وخصوصاً العربية، وحدودٍ إمكانياتٍ هذه الآليات والتقنيات وآفاقها.

في تطوير معالجة اللغة العربية لبناء تطبيقات ومشروعات ذات مردودٍ تربويٍّ و/أو ثقافيٍّ و/أو معلوماتيٍّ و/أو اقتصاديٍّ، وذلك بُعْدَة تعريفهم بوضوح بجذور وواقع هذا المجال في إنجازٍ وتبسيطٍ غير مُخلِّينٍ مما يساعدهم على وضع أهدافٍ واقعيةٍ لتطبيقاتهم ومشروعاتهم واتخاذ خياراتٍ صحيحةٍ تؤدي إلى إنجاز هذه الأهداف بكفاءةٍ وجودةٍ عاليةٍ.

بليوجرافيا مرجعية.

1. Aggarwal, C.C. Springer (2018) *Machine Learning for Text*, Springer.
2. Buchanan, B.G. (2006: Winter 2005) *A (Very) Brief History of Artificial Intelligence*, AI Magazine, AI Magazine: Volume 26 – Number 4, pp. 53–60, USA.
3. Domingos, P. (2015) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York - USA.
4. Goodfellow, I., Bengio, Y., Courville, A., (2016) *Deep Learning*, MIT Press, Massachusetts – USA.
5. Hofstadter, D. (1979) *Gödel, Escher, Bach: an Eternal Golden Braid*, Basic Books, New York USA.
6. Indurkhy, I. (Editor), Damerau, F.J. (Editor), (2010) *Handbook of Natural Language Processing* (2nd edition), Chapman & Hall Book/ CRC Press, Florida – USA.
7. Jurafsky, D., Martin, J. H., (2008) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Processing*, (2nd edition), Prentice Hall, USA.
8. Lane, H., Howard, C., Hannes, M.H., (2019) *Natural Language Processing in Action*, Manning Publication, USA.
9. Mitkov, R. (editor), (2005) *The Oxford Handbook of Computational Linguistics*, Oxford University Press – UK.
10. Nilsson, N.J. (2015) *The Quest for Artificial Intelligence*, Cambridge University Press, UK.
11. Russell, S., Peter Norvig, P., (2009) *Artificial Intelligence: A Modern Approach*, (3rd edition), Prentice Hall, USA.
12. Schutze, H., Manning, C.D., (2000) *Foundations of Statistical Natural Language Processing*, MIT Press, USA.

الفصل الثاني
المُعَالَجَةُ الْآلَيَّةُ لِلْغَةِ الْعَرَبِيَّةِ الْمَكْتُوبَةِ
مُقْدِمَةٌ فِي ذَكَاءِ الْآلَةِ

د. المُعْتَزِّ بِاللَّهِ السَّعِيد

ملخص

يُقدم هذا الفصل مدخلاً إلى معالجة اللغة العربية المكتوبة عبر بوابة الذكاء الاصطناعي، في محاولة للإجابة عن السؤال البحثي: كيف يمكن للألة أن تُحاكي ذكاء الإنسان في فهم اللغة العربية ومعالجتها في بنيتها المكتوبة؟. ويعرض الباحث لذلك في مقدمة وأربعة محاور أساسية تُمثل مُستويات معالجة اللغة المكتوبة، وهي (مستوى المعالجة الكتابية، مستوى المعالجة الصرفية، مستوى المعالجة التركيبية، ومستوى المعالجة الدلالية). ويعرض الباحث في كل مستوى لوحدة المعالجة (التحليل)، والتوجيه الآلي للمعالجة السطحية والعميقة لها، وأبرز تطبيقاتها في اللغة العربية.

ويستهدف الفصل القارئ العربي ذا الخلفية اللغوية الذي ينشد التعرُّف على طائق المعالجة الآلية للغة العربية المكتوبة و منهاجها، على النحو الذي يفتح أمامه الطريق إلى فهم أكثر عمقاً. لهذا، يسعى الباحث إلى عرض أفكاره انطلاقاً من منطق اللغة العربية [القائم على الكيف، عبر السِّياع والقياس]، مع موائمة مُبسطة بمنطق الآلة [القائم على الكم، عبر المحاكاة الرياضية]. ويلتزم الباحث منهاجاً وصفياً تحليلياً، يسعى من خلاله إلى وصف واقع اللغة العربية المكتوبة، وتحليلها في مستوىها المختلفة، بعية الوقف على أساليب توجيه الآلة إلى فهمها ومعالجتها على الوجه الموافق لطبيعتها.

الكلمات المفتاحية:

معالجة اللغات الطبيعية Natural Language Processing (NLP)، اللغة العربية المكتوبة Written Arabic ، الذكاء الاصطناعي Artificial Intelligence (AI)، تقنيات اللغات الإنسانية Human Language Technology (HLT)

١. الذكاء الاصطناعي واللغة المكتوبة.

اللغة المنطقية أسبق إلى الوجود من اللغة المكتوبة؛ إذ لم يبتكر الإنسان أبجديات الكتابة إلا في حقيقة زمنية متأخرة نسبياً عن حقيقة وجوده. ومع ظهور الحواسيب وتطورها في منتصف القرن العشرين، ظهرت محاولات وتجارب أولية تهدف إلى إخضاع الآلة لمحاكاة ذكاء الإنسان في فهم اللغات الطبيعية وتحليلها؛ إلا أن هذه المحاولات والتجارب قد انطلقت من اللغة المكتوبة، لأسباب عديدة، أهمها: الطبيعة القياسية النسبية لها، وإمكانية التفاعل مع بنيتها الملموسة، وسهولة حفظها واسترجاعها، وثبتت صورتها الخارجية بصرف النظر عن مستخدميها أو المتفاعلين معها.

يتناول هذا الفصل محاولة للإجابة عن السؤال المحوري بشأن تعاطي الآلة مع اللغة الطبيعية في صورتها المكتوبة، أي: كيف يمكن للآلة أن تحاكي ذكاء الإنسان في فهم اللغة الطبيعية المكتوبة؟ ولا ننسد الحديث عن عمق الفكرة وإجراءاتها التي قد تبدو معتقدة إلى حد ما؛ لا سيما في أبعادها الرياضية والهندسية المتشعبة؛ لكننا نلتمس تقديم طرح تمهدى، يمكن من خلاله بناء قاعدة مستقبلية، تكون مطلقاً لتناول أكثر عمقاً المعالجة اللغات الطبيعية (NLP). وسيكون حديثنا عن اللغة الطبيعية المكتوبة بصورة عامة، مع التركيز على اللغة العربية.

حين نسعى إلى إخضاع الآلة لفهم اللغة الطبيعية، فعلينا أن نقدر طبيعة المعطيات اللغوية الموجهة للآلة ومدى قابلية هذه المعطيات للتحليل الكمي. ذلك لأننا نخلق بيئات تحاكي قدرة الإنسان على استخدام اللغة الطبيعية؛ وهذه القدرة لم تصنعها قوى الطبيعة؛ لكنها جاءت نتيجة عمليات متدرجة لاكتساب المهارات اللغوية (الاستماع، والتحدث، والكتابة، والقراءة). فالطفل الناشئ في مجتمع يستخدم لغة معينة يستمع إلى أصوات اللغة في محيط مجتمعه، وينطق في مراحل النشأة الأولى ببعضها من هذه الأصوات، ثم يؤلف بينها في كلمات، ويصنع بحلاً بسيطة ومركبة، ثم يتعلم كتابة المحارف وقراءتها، ثم الكلمات والجمل، وهكذا، حتى يصير قادراً على استخدام اللغة في بيئه تفاعلية. وترتبط هذه القدرة بالمعطيات اللغوية التي توافرت له؛ فإذا نشأ مثلاً في مجتمع ذي ثقافة لغوية واسعة، فمن البدهي أن يتأثر بهذه الثقافة في تكوينه، والحال كذلك إذا نشأ

في مجتمع محدود الثقافة. والآلية قريبة من ذلك؛ حيث تنشأ علاقة طردية بين قدرتها على فهم اللغات الطبيعية من ناحية، واتساع مدى المعرفة اللغوية الموجهة لها من ناحية أخرى. وهذا يفسّر وجود تباين بين محركات أدوات معالجة اللغات الطبيعية، سواء في اللغة الواحدة، أم بين اللغات المُتعدّدة.

يكمن الإشكال الحقيقي في وجود فجوة بين الذكاء الطبيعي [ذكاء الإنسان] والذكاء الاصطناعي [ذكاء الآلة]. ومَرَد ذلك إلى أن محاكاة الطبيعة لا يمكن أن تحدث بصورة كُلّية؛ حيث تظل بعض سمات الشيء الطبيعي [الأصل] خاصة به، وغير قابلة للاستحداث في الشيء المصنوع [التقليد]. والأمر يتصل بها يمكن إخضاعه للقياس في الشيء الطبيعي. وفي النموذج الذي نحن بصدده، بشأن محاكاة ذكاء الإنسان في الآلة، يمكن مثلاً أن نحاكي القدرة على ممارسة نشاط خارجي، كالقدرة على القراءة أو الكتابة؛ لكننا لا نستطيع محاكاة المشاعر الداخلية، كالحب والآلام والغيرة. وأقصى ما نستطيع أن نصل إليه في هذا الشأن أن نُخضع الآلة إلى اكتشاف وجود بعض المشاعر، دون أن تتولّ هذه المشاعر فيها. ولا يتم ذلك إلا عبر قرائن [لغوية وغير لغوية] قابلة للقياس.

والواقع أن إخضاع الآلة لفهم اللغة الطبيعية يتجاوز حدود اللغة ذاتها إلى عوامل خارجية تتصل بطبيعة الإنسان النفسية ومحيطه الاجتماعي. ويُوضّح (الجدول ١) مثلاً على ذلك.

السياق الاجتماعي	السياق اللغوي	الجملة	م
مُتغير	ثابت	ان فعل الأب على ابنه، لأنَّه مُشاغب	١
		ان فعل الأب على ابنه، لأنَّه مُغاضب	٢

الجدول ١: نموذج توضيحي لاختلاف السياق اللغوي عن السياق الاجتماعي.

يحتوي الجدول على جملتين متفقتين تماماً في السياق اللغوي؛ وهو السياق الذي يمكن التحكم فيه؛ لكنهما تختلفان في السياق الاجتماعي؛ وهو السياق الذي يفهمه الإنسان ويتفاعل معه بصورة ديناميكية. ويستدعي هذا الفهم إدراك العلاقة بين الأب والابن، لتحديد من يتَّخذ سمة (المُشاغبة) ومن يتَّخذ سمة (المُغاضبة). وسنصل بذلك إلى أنَّ

- الجُملة ١ = الابن مُشاغب.

- الجملة ٢ = الأئمّة.

وَمِثْلُ ذَلِكَ يُفْسِرُ الْحَاجَةَ الدَّائِمَةَ إِلَى التَّدْخُلِ الْبَشَرِيِّ فِي مُرَاجِعَةِ مُخْرَجَاتِ الْآلَةِ، مِهْمَا بَلَغَ مُسْتَوْى ذَكَائِهَا الْاِصْطَناعِيِّ. وَفِي إِطَارِ سَعْيِنَا إِلَى اسْتِكْشافِ تَدْرُجِ الْآلَةِ مِنْ مُحَاكَةِ الْلُّغَةِ الطَّبِيعِيَّةِ الْمُكْتَوِيَّةِ فِي أَبْسِطِ أَشْكَالِهَا إِلَى مُحَاكَاتِهَا فِي أَعْقَمِ مَدَاهَا، لَا سِيَّما فِي الْلُّغَةِ الْعَرَبِيَّةِ، فَسَنُعْرُضُ فِي الصَّفَحَاتِ التَّالِيَّةِ أَرْبَعَةَ مَحَاوِرَ أَسَاسِيَّةً، تُمَثِّلُ مُسْتَوَيَّاتِ مُعَالَجَةِ الْلُّغَةِ الْعَرَبِيَّةِ الْمُكْتَوِيَّةِ، بَدِئًا بِمُسْتَوْى مُعَالَجَةِ مَحَارِفِ الْعَرَبِيَّةِ، وَمُرْوُرًا بِمُسْتَوْى مُعَالَجَةِ الْبُنِيَّةِ وَمُسْتَوْى مُعَالَجَةِ التَّرْكِيبِ، وَانتَهَاءً بِمُسْتَوْى مُعَالَجَةِ الْمَعَانِي.

٢. المعالجة الآلية الكتابية [المحرفة / الحبر افيمة].

١٢. وَحدَةُ الْمُعَالَةِ الْكِتَاسَةِ.

ب ت ث

الشكل ١: صورة الجرافيم في اللغة العربية

وقد يأخذُ الجرافيم شكلًا كتابيًّا واحدًا أو أكثرَ من شكل، بحسب النّظام الكتابي للّغة الطّبيعية، وبحسب موضعه في الكلمة [أو السّياق التّعبيريّ]. ويُعرَفُ الشّكلُ الواحدُ للجرافيم بـ (اللوجراف Allograph). وعلى سبيل المثال، يأخذُ جرافيم (A) في الإنجليزية اللوجرافين فقط، هما (a, A)، بينما يأخذُ جرافيم (ء) [الهمزة] في العربية خمسةً اللوجرافات، هي (ء، أ، إ، ئ، ء).



الشكل ٢: صورة اللوجراف في اللغة العربية

٢. التّوجيه الآلي لمعالجة الوحدات الكتابيّة في اللغة العربية.
كُلُّ حرفٍ محرَفٌ، وليس كُلُّ محرَفٍ حرفًا؛ لأنَّ المحارفَ تشملُ الحروفَ والرموزَ الكتابيَّة الأخرى في اللغات الطّبيعية. ويتمُّ توجيهُ الآلة لمعالجة المحارف باعتبارِها وحدات الكتابة على إحدى صورَتَين:

١) صورة مُفردة: ونعني صورة الحروف والرموز.

٢) صورة مُركَبة: ونعني صورة الكلمات أو الماجمِيع الكتابيَّة.

وسواءً أ جاءَت هذه المحارفُ [الوحدات الكتابيَّة] مُفردةً أم مُركَبة، يُمكِّنُ إخضاعُها للمُعالَجة الآليَّة عبرَ ما يُعرَفُ بـ (خوارزمات التّعلُّم الآليَّ Machine Learning) التي تمثُّلُ أساليبَ منهجيَّةً لتوجيهِ الآلة إلى فهم طبيعة اللّغة الطّبيعية Algorithms

على أساس رياضيٍّ. ويعني هذا أنَّ علينا أن نقدم لـالآلة معطياتٍ ثابتة (محصاة) لـتستند إلىها في محاكاة ذكاء الإنسان في فهم اللُّغة. ومع وجود عددٍ كبيرٍ من مناهج التَّعلُّم الآلي، فإنَّا نُركِّز على منهجين رئيسيين في معالجة اللُّغة العربيَّة، هما:

١) التَّعلُّم على أساس القواعد Rule-based Machine Learning

٢) التَّعلُّم على أساس المُدَوَّنات [الذَّخائر اللُّغويَّة] Corpus-based Machine Learning

وعبر هذين المنهجتين، يُمكِّن تقسيمُ مستوى المعالجة ذاتِه إلى قسمين، على النَّحو الآتي:

١، ٢. المعالجة السَّطحية للوحدات الكتابية.

إذا تناولنا الوحدات الكتابية العربيَّة في صورتها المُفردة، فسنجدُها مجموعةً من المحارف المعلومة والمُمحصاة، وإن تفاوتَ إحصاؤها وفقاً لـأسلوب الحصر ذاتِه. وإذا تناولناها في صورتها المُرَكَّبة، فسنجدُها كلهات معلومةً في أذهان أبناء اللُّغة وفي المعاجم العربيَّة المكتوبة. وأيًّا كانَ الأمرُ، فالمعالجة السَّطحية للوحدات العربيَّة المكتوبة تقوم أساساً على مطابقة المادَّة الخاضعة للمُعالجة بواقع اللُّغة.

وحتَّى نستبين ذلك نقول:

إذا اعتمدنا أسلوب التَّعلُّم الآلي على أساس القواعد، فالمعالجة السَّطحية تستدعي مُطابقة الوحدات الكتابية الخاضعة للمُعالجة بقواعد الكتابة العربيَّة التي تُوجَّهُ الآلة إليها عبر قواعد بياناتٍ حاويةٍ لها. ومن هذه القواعد مثلاً: أشكالُ حُرُوف العربيَّة في أول الكلمة وأوسيطها وأخرها، وحُرُوف الزيادة العربيَّة، ومواطن هذه الحُرُوف حين تأتي في صورة السَّوابق، ومواطنها حين تأتي في صورة اللَّواحق، وغير ذلك. وتبدو هذه القواعد بسيطةً على حدِّ بعيدٍ؛ إذ تكون معلومةً لأبناء اللُّغة، ولا يستدعي استكشافها معالجاتٍ سابقةً في أكثر الأحيان. وبيَّنَ (الجدول ٢) بعض النَّماذج لهذه القواعد.

القاعدة	المحارف العربية	م
لا تتشابك مع حروفٍ لاحقة تأتي مستقلةً في أول الكلمة دائمًا	ء، ؤ، د، ذ، ر، ز، و	١
تأتي مستقلةً في وسط الكلمة أو آخرها، إذا سبقت بالمحارف: (ء، ؤ، د، ذ، ر، ز، و)	١	٢
تأتي غير مستقلةً في وسط الكلمة أو آخرها، إذا سبقت بالحروف الأخرى		
تأتي في نهاية الكلمة، دون أو لها أو أوسطها	ة	٣
تأتي سابقةً لإفاده التعريف	ال	٤
تأتي مستقلةً أو مسبوقةً بحروفٍ معينةٍ (مثل: و، ف، ك)		
تأتي لاحقةً للدلالة على ضمير الغائبة	ها	٥
تقبل السوابق	الضمائر المعنفصة	٦
لا تقبل اللواحق	(أنا، نحن، أنتم، ...)	

الجدول ٢: من قواعد الكتابة العربية المستخدمة في المعالجة السطحية للوحدات الكتابية

أًما إذا اعتمدنا أسلوب التعلم الآلي على أساس المدونات، فالمعالجة السطحية تستدعي مطابقة الوحدات الكتابية الخاضعة للمعالجة بمقابلاتٍ لها. وتبدو هذه المقابلات في صورة كلماتٍ مستمدَّةٍ من مدوناتٍ لغويةٍ Linguistic Corpora تمثلُ لواقع اللغة؛ حيث تُعاد هيكلة نصوص المدونات لظهور أشباه بمعجم لمفردات اللغة. وحال المطابقة، تُوجه الآلة إلى التعرُّف على الكلمات التي تكون لها م مقابلاتٍ في هذا المعجم، وتُوجه كذلك إلى تجاهل ما عدتها.

٢، ٢. المعالجة العميقية للوحدات الكتابية.

السمة الأساسية للمعالجة العميقية للوحدات الكتابية في العربية أنها تعتمد على معطياتٍ غير معلومةٍ مسبقاً؛ بمعنى أنَّ موجَّهات الآلة تتجاوزُ المعطيات السطحية لأشكال المحارف وسماتها الظاهرة إلى معطياتٍ أخرى ترتكزُ على ظواهر لغويةٍ غير ظاهرةٍ للمحارف، أو لِنُقل إلَيْها غير معلومةٍ بالضرورة لأبناء اللغة. ومن ثُمَّ، تستدعي

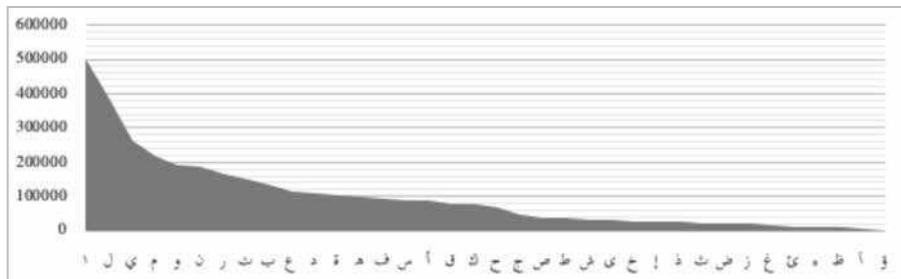
الالمعالجة العميقهُ إجراءاتٍ أوليةٍ تسبق المعالجة الآلية للمحارف، بهدف استنباط قواعد تبني على معطياتٍ جديدة، ثم تدريب الآلة على استيعاب هذه المعطيات. وفي إطار معالجة المحارف العربية، سواءً أكانت مفردةً أم مركبةً، يمكن التمثيل على المعالجة العميقه بظاهرتين، هما: ائتلاف الحروف، ودوران الحروف. وذلك على النحو الآتي:

١) ائتلاف الحروف [المحارف] العربية.

يمكن توجيه الآلة إلى التعرّف على أشكال المحارف العربية ومواضع ورودها في الكلمة، ويمكن توجيهها كذلك إلى الأقسام الصرفية (السوابق، والجذوع، واللّواحق) لأجل تحليلها كتابياً. لكن هناك قواعد أخرى لائتا الحروف العربية فيما بينها. ومنها مثلاً: أنَّ حرف العين لا يتألفُ مع الغين، وأنَّ الحاء لا يتألفُ مع الهاء، ومثل ذلك في حروف أخرى. ويساعد استخلاص قواعد ائتلاف الحروف على توجيه الآلة إلى معالجة أخطاء الكتابة العربية.

٢) دوران الحروف [المحارف] العربية.

يشتمل النّظام الكتابي في اللغة العربية على ثمانية وعشرين حرفاً [جرافيماً] أساسياً؛ وإذا أردنا حصر تنوعات هذه الحروف، فسنجدُها ستةً وثلاثين تنوعاً، تشمل: سبعةً تنوعات للهمزة (ا، آ، إ، ئ، ء، ئ)، وتنوعين لللّاء (ت، ة)، وسبعةً وعشرين تنوعاً للأحرف السبعة والعشرين الأخرى. لكن معالجة عميقه هذه التنوعات تجعلنا نكتشف أنَّ حروف العربية وتنوعاتها ليست في ذات المستوى من الدوران في النصوص العربية؛ حيث تردد بعضها بكثرة، مثل (ا، ل، ي، م، ن)، وتردد بعضها بصورة متوسطة، مثل (ق، ك، ح، ج)، ويندر ورود بعضها الآخر مقارنةً بغيره، مثل (ء، ظ، آ، ئ)، على النحو الوارد في (الشكل ٣). ويُستفادُ من هذه المعطيات في توجيه الآلة إلى الاحتمالات الغالبة للمحارف العربية حال ورودها في النصوص، الأمر الذي يُساعدُ على تعين مواطن الخطأ، واقتراح البديل المناسب.



الشكل ٣: دوران المحارف [الأوغرافات الحروف] في النصوص العربية

٣، ٢. من تطبيقات معالجة اللغة العربية المكتوبة في المستوى الجراميميّ.

تحظى المعالجة الآلية لمحارف اللغات الإنسانية بعناية العاملين في ميادين حوسية اللغة؛ نظراً للحاجة إليها في رقمنة التراث الإنساني المكتوب، وإيجاد بيئه تفاعلية لذوي الاحتياجات الخاصة من الراغبين في قراءة الكتب، وتطوير عمل الهاتف الذكي، ودوعاً أخرى. ولا يختلف منهج معالجة الجراميمات كثيراً بين اللغات الإنسانية؛ حيث يعول أساساً على أشكال المحارف ومواضع الاختلاف بينها. ويمكن التمثيل على تطبيقات معالجة اللغة العربية المكتوبة في المستوى الجراميمي بما يأتي.

١، ٢، ٣. التَّعْرُفُ الْآيَيِّ عَلَى الْحُرُوفِ الْمُطَبَّوِعَةِ Typewritten OCR

من تطبيقات التَّعْرُفُ الْآيَيِّ [الصَّوْئيِّ] على المحارف Optical Character Recognition (OCR) النصوص المطبوعة من صيغتها الورقية (مثل الكتب والوثائق ونحوها) إلى صيغة رقمية [محوسبة]، يمكن التَّحْكُمُ في محارفها وكلماتها بالإضافة أو الحذف أو التعديل؛ أو بعبارة أخرى: تحويل الصورة إلى نص مكتوب. ويستدعي هذا التطبيق مُراعة التَّبَاعِينِ الواقع في أشكال المحارف المطبوعة؛ من حيث الوانها وأنواعها وأحجامها؛ كما يستدعي مُراعة قواعد ائلاف المحارف، ومواطن وجود النقاط على المنقوط منها، ومواطن وجود علامات الفَبْط ونحوها.

٢. التَّعْرِفُ الْآلَىٰ عَلَى الْحُرُوفِ الْمُخْطُوْتَةِ .Handwriting OCR

هو أيضاً من تطبيقات التَّعْرُفُ الْآلَى على المحارف OCR؛ ويُستفادُ منهُ في تحويل النُّصُوص المخطوطة [المكتوبة بخطّ اليد] إلى صيغة رقميَّة. ويُساعدُ هذا التطبيق في مجالاتٍ عديدة، مثل: معالجة بصمة الكتابة، وتحقيق النُّصُوص، وتحرير الوثائق المخطوطة. ويستدعي هذا التطبيق مُراعاة التَّبَاعُون الواقع في أشكال المحارف المخطوطة؛ وهو - بطبيعة الحال - أكثرُ تعرِيقاً من سابقه، نظراً لاختلاف أنماط الكتابة عبرَ الزَّمان والمكان والأفراد، بالإضافة إلى تأثير الضَّواعِيَّات البصريَّة المُحيطة بالوثائق المخطوطة.

٢. قراءة الكُتُب .Book Reader

يهدفُ هذا التَّطبيقُ إلى توجيهِ الآلةِ إلى القيامِ بدورِ القارئِ الآليِ للنصوصِ في الكُتُبِ والوثائقِ المكتوبة؛ حيثُ يُساعدُ ذوي الاحتياجاتِ الخاصةَ من فاقدِي القدرةِ علىِ الإبصارِ علىِ القراءةِ. ويَمْرُّ بمرحلَتَينِ رئيسَتَينِ؛ حيثُ تُعنى المرحلةُ الأولىِ بالتعرفِ علىِ المخارفِ [الجغرافياتِ]، لتحولَ المادَّةَ بذلكَ من (نصٌّ مُصَوَّرٌ جامدٌ) إلىِ (نصٌّ مكتوبٌ قابلٌ للمُعالجةِ). أمَّا المرحلةُ الأخرىِ، فتُعنى بتحويلِ النَّصِّ المكتوبِ إلىِ نَصٌّ منطوقٌ، عبرَ تطبيقٍ آخرَ لتحويلِ النَّصِّ إلىِ كلامٍ (Text To Speech (TTS)).

٣. المعالجة الآلية البنوية [الصَّرْفَ فِيهَا].

١٣. وَحدَةِ الْمُعَالَةِ الصَّفَّةِ.

تعنى المعاجلة البنوية [الصّرفيّة] بمجموعةٍ من العناصر المكوّنة لبنيّة الكلمة في اللّغات الطّبيعيّة، سواءً بتحليلها إلى مكوّناتٍ صُغرى أم بتوسيعها من مكوّناتٍ صُغرى. ووحدة المعاجلة البنوية هي (المورفيم Morpheme)؛ ويمكن تعريفه بأنه أصغر وحدةٍ لعوّيّةٍ محرّدةٍ لها معنى. ويكونُ هذا المعنى مُعجميًّا، على نحو ما نجدُ في مورفيمات (إنسان، حبّ، سماء) أو وظيفيًّا، على نحو ما نجدُ في مورفيمات (ال) الدّالّة عال، التّعّبف، و (س) الدّالّة عال التّسويف، و (ة) الدّالّة عال التّأنيث؛ ونحو ذلك.

وَتَدَالُّ الْمُوْرَفِيَّاتُ فِي الْكَلْمَةِ الْوَاحِدَةِ لِتُتَبَّعَ عِنَاصِرَ بَنِيَّتِهَا الْمُمْتَمِّلَةُ فِيهَا يَأْتِي:

- الجذر Root: يقصد به مجموعة الحروف الأصلية المكونة الكلمة المشتقة أو القابلة للاستقاق [مثل: الكلمات العربية والمغربية]. ويمكن أن يطلق عليه (الأصل) إذا كانَ الكلمة غير مشتقة [مثل: الكلمات الدخيلة والضمائر].
- الجذع Stem: يقصد به صورة الكلمة المعينة بعد تحريرها من زوائدِها [السابق والواحد].
- الفرع Lemma: هو المُقابلُ الصَّرِيفُ المُجرَدُ للوحدة المُعجميَّة [Lexeme] في الكلمة المعينة. ويأتي في العربية على صورة الفعل الماضي المُسند إلى الغائب [ما لم يلزم صورة أخرى]، أو الاسم المفرد في صيغة التَّذكير [ما لم يلزم صورة أخرى]، أو الأداة في شكلها الرَّئيس.
- الزَّائدة Affix: وهي مُورفِيْمُ لاصقُ، يسبِّبُ جذع الكلمة في بدايته، فيكونُ (سابقة Prefix) أو يلحقُ بها في نهايتها، فيكونُ (لاحقة Suffix).

ويمكن التَّمثيلُ على هذه العناصر في الكلمة (سيعاودون) على النحو الوارد في الجدول (٣).

التمثيل	العنصر	م
ع و د	الجذر	١
يُعاوِد	الجذع	٢
عاوَدَ	الفرع	٣
س-	السابقة	الزَّائدة
ون	اللاحقة	

الجدول (٣): العناصر المُكونة لبني الكلمة العربية - نموذج (سيعاودون)

٢، ٣. التَّوجيه الآلي لمعالجة الوحدات الصَّرفية في اللغة العربية.

يُعَدُ النَّظامُ الصَّرِيفُ العربيُّ مثالِيًّا في حَوْسَبِهِ، نظرًا لقياسيَّةِ جُزِءٍ كبيِّرٍ من قواعدهِ، الأمرُ الَّذِي يسمِحُ بتطويع الآلة للتعاطي مع هذه القواعد، بعد ترجمتها من اللغة الطَّبَيعيَّةِ إلى لغات الآلة. ومع هذا، فإنَّ هناك جانبًا سُمِّيَّاً ليس هُنَيَا في الصَّرف العربيِّ، وسُنُحاوِلُ الإِبانَةَ - فيما يأْتِي - عن التَّوجيه الآليِّ لمعالجة الوحدات الصَّرفية العربية، عبرَ المُسْتَوَيَّينِ: السَّطْحِيِّ والعميق.

١، ٢، ٣. المعالجة السطحية للوحدات الصّرفية.

يُعني الصّرف العربي بالأفعال والأسّماء دون الأدوات التي تلزم حالة الجمود. ويُصنّف الفعل والاسم إلى عدة أبوابٍ صرفية، وفقاً لاعتباراتٍ مختلفة. ويُوضّح (الجدول ٤) بعضًا من هذه الأقسام.

م	قسم الكلام	مِعيار التصنيف	الباب الصّرفي	التمثيل
ال فعل	١	باعتبار الزَّمن	الماضي	كتَبَ
			المضارع	يَكْتُبُ
			الأمر / الطلب	اَكْتُبُ
	٢	باعتبار الصّحة والاعتلال	الصّحيح	بَعَثَ
			المُعْتَلُ	هَوَى
			المُجَرَّد	دَرَسَ
الاسم	٣	باعتبار التجُّرد والزيادة	المُزِيد	دَارَسَ
			الجَامِد	الرَّجُل
			الْمُسْتَخْدِمُ	الْمُسْتَخْدِمُ
	٤	باعتبار الجمود والاشتقاق	الصّحيح	الكتاب
			المنقوص	القاضي
			المقصور	الفتى
٦	٥	باعتبار الصّحة والاعتلال	المُعْتَلُ	السَّمْدُود
			المُجَرَّد	السَّهْم
			المُزِيد	الاستقدام

الجدول ٤: من أبواب الصّرف العربي

وُتُّصنّفُ هذه الأبواب بدورها إلى أبوابٍ فرعيةٍ أخرى. فعلى مستوى الأفعال - مثلاً - يأتي الفعل المُجَرَّد ثلَاثِيًّا ورباعيًّا؛ ويأتي الثلاثيُّ منه في ثلاثة أبواب (سِماعيَّة) بصيغة الماضي، هي (فَعَلَ، فَعِلَّ، فَعُلَّ)، ويأتي كذلك في ستة أبواب (سِماعيَّة) باعتبار صيغتي الماضي والمضارع معاً، هي (فَعَلَ يَفْعُلُ «مثلاً: نَصَرَ يَنْصُرُ»، فَعَلَ يَفْعُلُ «مثلاً: جَلَسَ يَجْلِسُ»، فَعَلَ يَفْعُلُ «مثلاً: ذَهَبَ يَذْهَبُ»، فَعِلَّ يَفْعُلُ «مثلاً: عَلِمَ يَعْلَمُ»، فَعُلَّ يَفْعُلُ «مثلاً: حَسُنَ يَحْسُنُ»، فَعَلَ يَفْعُلُ «مثلاً: حَسِبَ يَحْسِبُ»)، وهكذا. وعلى مستوى

الأسماء، يأتي الاسم المستقى من الفعل الثلاثي في صورة اسم الفاعل على زنة (فاعِل)، وفي صورة اسم المفعول على زنة (مفعول)، ويكون قياسياً في أكثر الأحيان، وسياعتَّ في مواضع قليلة، كأنْ يُستَغنَّى عن اسم الفاعل مثلاً مع الفعل (شَجَعَ) لتحلَّ محلَّه الصفة المُشَبَّهة (شُجاعٌ)، وأنْ يأتي الاسم المستقى في صورة اسم المفعول على زنة (فَعِيل)، مثل (فَتَيْلٌ). وهكذا.

وإذا أخذنا في الاعتبار:

- أنَّ التَّمَيِّزَ بَيْنَ الْأَبْوَابِ الْصَّرْفِيَّةِ يَسْتَدِعِي ضَبْطَهَا بِالشَّكْلِ، لِلتَّمَيِّزِ مُثْلًا بَيْنَ اسْمِ الْفَاعِلِ (مُعَلَّمٌ) وَاسْمِ الْمَفْعُولِ (مُعَلَّمٌ) الْمُشَتَّقَيْنِ مِنَ الْفَعْلِ الْرُّبَاعِيِّ (عَلَمٌ)،
- وَأَنَّ الْمُفَرَّدَاتِ فِي أَبْوَابِ الْصَّرْفِ الْعَرَبِيِّ تَقْبِلُ دُخُولَ الرَّوَابِدِ عَلَيْهَا، مِنَ السَّوَابِقِ وَاللَّوَاحِقِ،
- وَأَنَّ أَبْوَابَ الْصَّرْفِ الْعَرَبِيِّ الرَّئِيْسَةَ تَتَفَرَّعُ إِلَى أَبْوَابٍ أُخْرَى، تُتَجْزَعُ قَوَاعِدَ سِيَاسِيَّةً وَقِيَاسِيَّةً.

فسيكونُ علينا أن نقفَ على الإشكالات المُحتمَلة حَالَ توجيهِ الآلة إلى المعالجة الصَّرْفِيَّةِ.

وعلى سبيل المثال، إذا أردنا إجراء معالجة صرفيَّة لكلمة (فلَك) بِصُورَتِها المُجَرَّدة من علامات الضَّبْط، فستكونُ التَّيَّجَةُ عَدَّة احتمالاتٍ، يُصَنَّفُ بعُضُّها الكلمة فعلاً، ويُصَنَّفُ بعُضُّها الكلمة اسمًا، ويُصَنَّفُ بعُضُّها الكلمة حرفاً أو أداة. ويُبَيَّنُ (الجدول ٥) بعضاً من هذه الاحتمالات.

م	الكلمة	السَّوَابِقِ	الجذع	اللَّوَاحِقِ	قِسْمِ الْكَلَامِ
١	فَلَكَ	-	فَلَكَ	-	فعل
٢	فَلَكَ	-	فَلَ	ك	فعل
٣	فُلَكٌ	-	فُلَكٌ	-	اسم
٤	فَلَكٌ	-	فَلَكٌ	-	اسم
٥	فَلَكَ	ف	ل (لِلْجَرْ)	ك	حرفاً / أداة

الجدول ٥: مِن احتمالات المُعَالَجَةِ الصَّرْفِيَّةِ لِكَلْمَةِ (فَلَكَ) بِصُورَتِها المُجَرَّدة

لأجل معالجة مثل هذه الإشكالات في اللغة العربية، فإننا نقوم بتزويد الآلة بمجموعةٍ من قواعد البيانات المشتملة على معطياتٍ لغوية تتبع منظومة التصريف العربي (بجانبها: القياسي، والسماعي)؛ وتمثل هذه المعطيات أساساً لتوجيه الآلة إلى معالجة الوحدات الصّرفية، عبر خوارزمياتٍ معيّنةٍ موافقةٍ لطبيعة العربية وبنيتها الصّرفية. ويُوضّح (الجدول ٦) نماذج من هذه المعطيات وأمثلةً عليها.

التمثيل	النموذج	المعطيات	م
(ساعدَ، تَائِي، استَخدَم)	الأفعال الرباعية والخماسية والسداسية	قوائم الأفعال والأسماء ذات الطبيعة القياسية، وقواعدها التصريفية والاشتقاقية	١
اسم الفاعل (مُسْعِين) من الفعل السُّداسي (استعان)	الأسماء المشتقة من الأفعال الرباعية والخماسية والسداسية		
قرأً - يقرأً كتَبَ - يكتُب	الأفعال الثلاثية المُجردة (بصيغتي الماضي والمُضارع)	قوائم الأفعال والأسماء ذات الطبيعة السماعية	٢
(الحُبُّ، العَيْنُ، الإِنْسَانُ)	الأسماء الجامدة		
(إنجليزية، إنكليزية، إنجليزية)	تنوعات الكلمات الأعجمية	قوائم التنوعات الكتابية	٣
(إن، لم، «لا» النَّاهِيَة)	أدوات جزْم المُضارع	قوائم الأدوات	٤
التَّاءُ والأَلِفُ في (تَقَارِبٍ)	س، ء، ل، ت، م، و، ن، ي، هـ، أـ	قوائم حُرُوف الزِّيادة على الأفعال	٥
(و، فـ)	سوابق العَطف	قوائم السَّوابق	٦
(سُون، يِن، سَات، تَـة)	لواحق الجُمُع والتَّائِيَت	قوائم اللَّواحق	٧

الجدول ٦: من معطيات معالجة السطحية للوحدات الصّرفية في اللغة العربية

تجري معالجة الوحدات الصّرفية استناداً إلى هذه المعطيات؛ ويتحدد شكل المخرجات بناءً على النّصوص المستهدفة؛ حيث تختلف مخرجات الكلمات

المشكولة كلياً عن الكلمات المشكولة جزئياً، وتحتفل محرجات الكلمات المشكولة عن محرجات الكلمات المجردة من الضبط. وبطبيعة الحال، فإن احتمالات المعالجة الصرفية تقل إلى الحد الأدنى إذا كانت الكلمة تامة التشكيل، وتزداد تدريجياً كلما قل الضبط، حتى تصل إلى الحد الأقصى في الكلمات المجردة من الضبط.

٢، ٣. المعالجة العميقـة للوحدات الصـرفـية.

تتسم اللـغـة العـربـيـة بـوـجـود ظـاهـرـة الإـعـرـابـ. وـمـع هـذـا، فـوـاقـع الكـتـابـة العـربـيـة أـن تـخـلـو الـكـلـمـات من عـلـامـات الضـبـطـ، إـلـا في مـوـاضـع الـالـتـبـاسـ أو بـعـض آـنـاط النـصـوصـ المـقـدـسـةـ أو التـرـاثـيـةـ أو نـحـوـ ذـلـكـ. وـمـن ثـمـ، فـالـتـيـجـةـ الطـبـيـعـيـةـ لـمـعـالـجـةـ الـوـحدـاتـ الصـرـفـيـةـ أـن تـعـدـدـ اـحـتـمـالـاتـ إـلـىـ الـحـدـ الـأـقـصـىـ، لـأـنـ الـآـلـةـ تـقـدـمـ كـلـ اـحـتـمـالـ مـكـنـ لـلـكـلـمـةـ الـمـجـرـدـةـ منـ الضـبـطـ حـالـ ضـبـطـهـاـ بـمـخـلـفـ الـأـشـكـالـ الـمـمـكـنـةـ. وـعـلـىـ سـبـيلـ المـثالـ، سـتـعـدـدـ اـحـتـمـالـاتـ مـعـالـجـةـ كـلـمـةـ (ـحـمـلـ) بـصـورـتـهاـ الـمـجـرـدـةـ بـيـنـ الـأـسـمـاءـ وـالـأـفـعـالـ؛ وـإـذـاـ حـوـلـنـاـ إـلـىـ صـورـةـ مـضـبـوـطـةـ جـزـئـيـاـ بـوـضـعـ السـكـونـ فـيـ وـسـطـ الـكـلـمـةـ عـلـىـ هـيـةـ (ـحـمـلـ)، فـهـذـاـ يـعـنـيـ اـسـتـشـاءـ الـأـفـعـالـ؛ وـإـذـاـ حـوـلـنـاـ إـلـىـ صـورـةـ مـضـبـوـطـةـ كـلـيـاـ [ـعـلـىـ مـسـتـوـيـ الـبـيـنـةـ]ـ عـلـىـ هـيـةـ (ـحـمـلـ)، فـهـذـاـ يـعـنـيـ الـاـقـتـصـارـ عـلـىـ نـتـيـجـةـ وـاحـدـةـ [ـأـوـ نـتـائـجـ مـحـدـودـةـ الـاحـتـمـالـاتـ]ـ؛ وـهـكـذـاـ.

وـثـمـةـ نـقـطـةـ مـهـمـةـ أـخـرـىـ تـسـتـرـعـيـ الـاـنـتـبـاهـ بـشـأنـ مـوـافـقـةـ مـحرـجـاتـ الـمـعـالـجـةـ الـصـرـفـيـةـ لـلـغـةـ العـربـيـةـ الـمـسـتـخـدـمـةـ. فـالـوـاقـعـ أـنـ عـدـدـاـ لـيـسـ هـيـنـاـ مـنـ الـمـحرـجـاتـ تـقـعـ فـيـ دـائـرـةـ الصـحـيـحـ؛ لـكـنـهـاـ لـاـ تـوـجـدـ فـيـ الـوـاقـعـ؛ وـإـذـاـ وـجـدـتـ فـهـيـ نـادـرـةـ لـلـغـاـيـةـ. وـيـمـكـنـ التـمـثـيلـ عـلـىـ ذـلـكـ بـالـوـحدـاتـ الـصـرـفـيـةـ فـيـ جـمـلـةـ (ـهـذـاـ هـوـ أـخـيـ)ـ. فـاستـخـدـامـ الـكـلـمـاتـ الـثـلـاثـةـ يـشـيـعـ فـيـ الـعـربـيـةـ عـلـىـ صـورـةـ الضـبـطـ (ـهـذـاـ هـوـ أـخـيـ)ـ بـدـلـالـةـ (ـهـذـاـ)ـ عـلـىـ اـسـمـ الإـشـارـةـ، وـ (ـهـوـ)ـ عـلـىـ ضـمـيرـ الغـائـبـ، وـ (ـأـخـيـ)ـ عـلـىـ اـلـاسـمـ الـمـسـنـدـ إـلـىـ يـاءـ الـمـتـكـلـلـ؛ إـلـاـ أـنـ مـحرـجـاتـ الـمـعـالـجـةـ الـصـرـفـيـةـ سـتـشـيـرـ إـلـىـ اـحـتـمـالـ أـنـ تـكـوـنـ (ـهـذـاـ)ـ فـعـلـاـ مـاضـيـاـ مـسـنـداـ إـلـىـ أـلـفـ الـاثـيـنـ (ـهـذـاـ)، وـأـنـ تـكـوـنـ (ـهـوـ)ـ فـعـلـاـ لـلـطـلـبـ (ـهـوـ)، وـأـنـ تـكـوـنـ (ـأـخـيـ)ـ كـلـمـةـ مـصـغـرـةـ (ـأـخـيـ)ـ؛ وـنـحـوـ ذـلـكـ. وـمـثـلـ هـذـهـ الـاحـتـمـالـاتـ صـحـيـحـ مـنـ النـاحـيـةـ الـمـعـيـارـيـةـ؛ لـكـنـهـ نـادـرـ الـوـجـودـ جـدـاـ فـيـ وـاقـعـ الـلـغـةـ.

زُبَدَةُ القَوْلُ أَنَّا نَشُدُّ فِي الْمُعَالَجَةِ الْعُمِيقَةِ لِلْوَحْدَاتِ الْصَّرْفِيَّةِ أَمْرَيْنِ، هُمَا:

١) توجيه الآلة إلى دلالات الكلمات المُجرَّدة من الضَّبط.

٢) ترتيب احتمالات معالجة الوحدات الصرفيَّة بحسب دورها في واقع اللغة.

ويستدعي الأمران تدريب مُدوَّنة لُغويَّة على المعالجة العميقَة؛ حيث تُعدُّ هذه المُدوَّنة مورداً لمعطيات القرائن المساعدة على ترجيح الدلالات الأقرب إلى الصَّواب؛ مثل قرينة أن تكون الكلمة (من) اسمًا موصولاً إذا لحقتها كلماتٌ تبدأ بحروف (ت، ن، ي)؛ حيث تُرجح إشارة هذه الكلمات إلى أفعالٍ مضارعة، أو نحو ذلك. وبطبيعة الحال فإنَّ دوران المفردات في المُدوَّنة يُعدُّ موجَّهاً للآلة لترتيب احتمالات وتقسيتها من مُخرَجاتٍ لا تُعبِّرُ عن واقع اللغة، وإن كانت خاضعةً للقياس.

٣.٣. من تطبيقات معالجة اللغة العربيَّة المكتوبة في المستوى الصرفي.

ما دُمنا نتعامل مع مفردات النُصوص [الأبنية]، فالحاجة ماسَةٌ ومستمرةٌ إلى حُوَسَّةِ الصرف العربي في مختلف ميادين حُوَسَّةِ اللغة، بطريقٍ مُباشرٍ أو غير مُباشر. وتبُرُّ الحاجة إلى المعالجة الآلية لقواعد الصرف العربي في العديد من التطبيقات عبر عمليَّتين أساسيتَين، هما (التوَلِيد) الذي ينطلق من الجذر [الأصل] اللُّغويّ بهدف تعين الوحدات [المورفيات] وتشكّلاتها الصرفيَّة، و(التحليل/ التَّفْكِيك) الذي ينطلق من الكلمة [المجموع الكَتابِيّ] بهدف تعين المكوِّنات الصرفيَّة وأصولها التي تكونَت عنها. ويُمْكِن التَّمثيل على تطبيقات معالجة العربيَّة المكتوبة في المستوى الصرفي بما يأْتِي.

١.٣.٣. الفهرسة الآلية Automatic indexing

تُعنى الفهرسة الآلية ببناء قوائم الكلمات وترتيبها من مجموعات النُصوص، وترتيبها على نَسَقٍ مُعَيَّنٍ، كأن تُرتب وفقاً لألفبائية اللغة، أو بحسب دوران الكلمات في النَّصّ، أو نحو ذلك. وتتأثَّر مُخرَجات أدوات الفهرسة الآلية باختلاف النَّظام الصرفي للغة الطبيعية؛ حيث تبدو في صورة وحداتٍ مُتسقة مع البنية المُعجمية في اللغات الإلَاصَقَيَّة (مثلاً: الإنجليزية، والفرنسية)، وتبدو أقلَّ اتساقاً في اللغات الاشتراقَيَّة (مثلاً: العربيَّة، والعربيَّة). والإشكالُ بالنسبة لغة العربيَّة أنَّ أدوات الفهرسة الألفبائية

تأثير بزوراً الكلمة وتنوعها الاشتراكية، فتقوم بتوزيعها على مداخل عديدة. وسعياً إلى حل هذا الإشكال، فإننا نلجأ إلى بناء ما يُعرف بـ (أدوات الفهرسة الجذرية) التي تقوم على معالجة صرفيّة للكلمات بهدف تجاوز الزوائد وتحميم التنوعات الاشتراكية.

٣،٣. التحليل الصّرفي Morphological Analysis

تُؤدي أدوات التحليل الصّرفي دوراً مهماً في معالجة النص العربي؛ ويعول عليها بصورة كبيرة في حوسبة المُعجم العربي، لا سيما فيما يتصل بمبانيه [الوحدات والمداخل المُعجمية]. وتفاوت مُحرّجات التحليل الصّرفي بحسب وسيلة المعالجة؛ إذ يمكن الاقتصر على تعين الوحدات الصّرفيّة وتصنيفاتها؛ ويمكن تجاوزها إلى عمليات أخرى، تشمل: (التجذير Rooting) الذي يعني باستخلاص جذور الكلمات أو أصولها اللغويّة، و (التّجذيع Stemming) الذي يعني باستخلاص جذوع الكلمات بعد تحريرها من الزوائد [السّوابق واللّواحق]، و (التفريغ Lemmatization) الذي يعني بتعيين ما يُعرف بـ «الفرّوع»؛ وهي في العربية صورة الوحدات المُعجمية المُنثقة عن الوحدات الصّرفيّة. ومن ناحية أخرى، تعدد التنوعات الشّكليّة للكلمات بحسب الضّبط بالشكل؛ ويعتمد في إدارتها على خوارزماتٍ مُعينة.

٤. المعالجة الآلية التّركيبية.

١،٤. وحدة المعالجة التّركيبية.

إذا كانت المعالجة البنوية معنية بالmorphemes ذاتها، من حيث بنيتها التصريفية، فإن المعالجة التّركيبية تعنى بموقع هذه المُفرّدات في التركيب النّحوّي، وتعنى كذلك بالعلاقات الشّكليّة بين هذه المُفرّدات. ويمكن القول - بعبارة أخرى - إنَّ المعالجة التّركيبية ترتكز على (أقسام الكلام Parts of Speech) حين تقع في كيانٍ تركيبيٍّ مُكتمل الأركان؛ سواءً أكان هذا الكيان بسيطاً أم مركباً.

في ضوء ذلك، لا نستطيع الوقوف على وحدة ثابتة للمعالجة التّركيبية، نظرًا لتباعد الكيانات التي ترد فيها أقسام الكلام في اللغات الطبيعيّة. فقد تكون هذه الوحدة (عبارة Phrase) أو (جملة Clause)، أو (فقرة Sentence)، وفقاً لطبيعة التركيب المُتضمن

لأقسام الكلام وموضعه في السياق. وتوضيحاً لذلك، يمكن التَّمثيل على الوحدات التَّركيبية في العربية على النحو الوارد في (الجدول ٧).

الكيان التَّركيبية	م	نماذج الوحدة التَّركيبية
العبارة	١	الأُمُّ الْمُتَّحِدَة
الجملة	٢	الأُمُّ الْمُتَّحِدَةُ مُنْظَمَةٌ دُولَيَّةٌ فَعَالَةٌ
الفقرة	٣	يُمْكِنُ اللُّجُوءُ إِلَى الأُمُّ الْمُتَّحِدَةِ؛ حِيثُ تُعَدُّ مُنْظَمَةً دُولَيَّةً فَعَالَةً

الجدول ٧: أنماط الوحدات التَّركيبية في اللغة العربية [العبارة، الجملة، الفقرة]

سنلاحظُ من خلال هذه النماذج ما يأقِي:

- النَّمط (١) - العبارة: تدلُّ العبارة (الأُمُّ الْمُتَّحِدَة) على كيانٍ تركيبٍ، حِيزٌ أَقْلُ من الجملة؛ لكنَّ هذا الكيانَ يُمثِّلُ وحدةً تركيبيةً كاملةً حينَ يرُدُّ في سياق عُناوِينٍ أو لافتَةٍ أو نحو ذلك.
- النَّمط (٢) - الجملة: تدلُّ الجملة (الأُمُّ الْمُتَّحِدَةُ مُنْظَمَةٌ دُولَيَّةٌ فَعَالَةٌ) على كيانٍ تركيبٍ، حِيزٌ الجملة الكاملة؛ حيثُ يُعبِّرُ عن جملةً اسميةً رُكناها المُبتدأ والخبر. وَيُمثِّلُ وحدةً تركيبيةً حينَ يرُدُّ في سياق لُغويٍّ مُكتمل.
- النَّمط (٣) - الفقرة: تدلُّ الفقرة (يُمْكِنُ اللُّجُوءُ إِلَى الأُمُّ الْمُتَّحِدَةِ؛ حِيثُ تُعَدُّ مُنْظَمَةً دُولَيَّةً فَعَالَةً) على كيانٍ تركيبٍ، حِيزٌ أكبر من الجملة [في هذه الفقرة جُملتان]. وَيُمثِّلُ هذا الكيانُ وحدةً تركيبيةً حينَ يرُدُّ في سياق مجموعَةٍ من الجمل المُترابطةِ التي تدورُ حول فكرةٍ معينةٍ، وَيُعوَّلُ على إِدراها في فهم تركيب غيرها.

٢، ٤. التَّوجيه الآلي لمعالجة الوحدات التَّركيبية في اللغة العربية.

حينَ نتعاطى مع الجملة باعتبارها وحدة المعالجة التَّركيبية، فإنَّا نُمِيزُ في العربية بينَ نوعَيْها: الفعلية والاسمية. ويكونُ النوعُ الأوَّل (الجملة الفعلية) من عُنصرَين رئيسيَّين - في حالة الْلُّزُوم - هما: الفعل والفاعل (أو الفعل ونائب الفاعل)، ومن ثلاثة عناصرٍ - في حالة التَّعدي، هي: الفعل والفاعل والمفعول. أمَّا الجملةُ الاسمية، فلها

عنصران رئيسيان، هما: المبتدأ والخبر. وثمة عناصر أخرى ثانوية، تؤدي وظائف الإضافة والمعطف والوصف والتوكيد وغير ذلك.

وحيث نوجّه الآلة إلى فهم طبيعة هذه العناصر، فعلينا أولاً أن نحدد أقسام الكلام التي تنتهي إليها عناصر الجملة؛ وعلينا تاليًا أن نوجّد قرائن [أو علامات] مميزة لكل عنصر على حدة؛ بحيث تكون هذه العلامات مقياساً يُستندُ إليه في تمييز العنصر الواحد عن العناصر الأخرى. وبيدو الأمر بسيطاً للوهلة الأولى، إذ يفترض أن لكل قسم من أقسام الكلام (الاسم، والفعل، والحرف/الأداة) علاماتٍ مميزة وضّعها النّحّاة عند التّقعيد للغة العربية؛ كالتعريف والتّنّوين لتمييز الاسم، وتناء الفاعل ونون التّوكيد لتمييز الفعل، والتّجرّد من سمات الاسم والفعل لتمييز الحرف.

لكنَّ الأمر لا يخلو من إشكالاتٍ غير قليلة حين نوجّه الآلة للمعالجة التركيبية؛ وعلى سبيل المثال: إذا وَجَّهنا الآلة إلى اعتبار الكلمات المسبوقة بـ(الـ) أسماء، فالنتيجة لن تكون صحيحةً في مثل الأفعال (التحمّل، التّزم، التّقى، التّمّس، ...); وإذا وَجَّهنا الآلة إلى اعتبار الكلمات التي تلحّن بها الثناء (تـ) أفعلاً، فالنتيجة لن تكون صحيحةً في مثل الأسماء (بَيْت، سَبْت، نَبْت، نَعْت، ...); وقس على ذلك مختلف موجّهات القواعد التركيبية للغة العربية. ومن ثم، فالقواعد المفهومة، مما يستطيع العقل البشري تمييزه، قد لا تكون مفهوماً للآلة على الوجه الأمثل. وسنحاول أن نستبيان ذلك فيما يأتي:

١٢٤. المعالجة السطحية للوحدات التركيبية.

في هذا المستوى من المعالجة، يكون الهدف تعين أقسام الكلام. والواقع أنَّ اللغة العربية تصنّيفاتٍ متعددةً لأقسام الكلام، منها التّصنيف الثلاثي التقليدي (الاسم، والفعل، والحرف)، والتصنيف الرباعي (الاسم، والفعل، والحرف، والضمير)، والخماسي (الاسم، والفعل، والحرف، والضمير، والظرف)، والسباعي (الاسم، والفعل، والحرف، والضمير، والظرف، والصفة، والخالفة). وأيًّا كان التّصنيف الذي نعتمده في توجيه الآلة، فالمنهج واحد؛ إذ يقوم على تدريب الآلة على مدوناتٍ لغوية مُوسَّمة Annotated Corpora PoS بعلاماتٍ معيّنةٍ يُستدلُّ بها على أقسام الكلام.

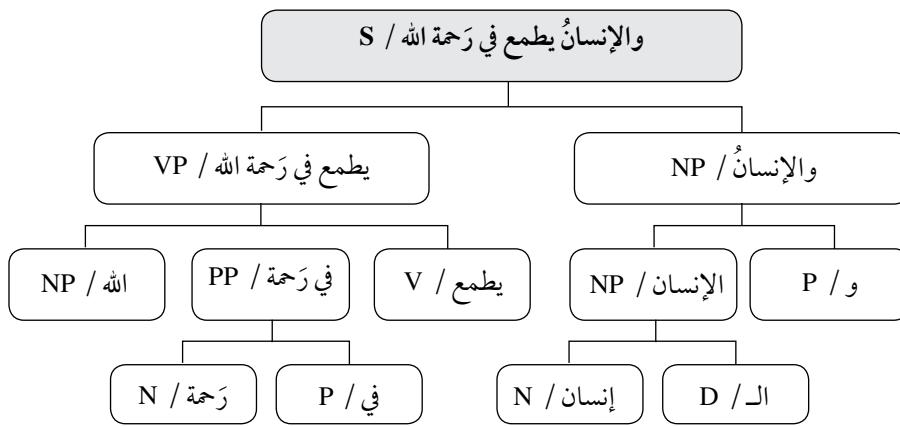
ولتوضيح ذلك يمكن القول إن المعاجلة السطحية للوحدات التركيبية تتم عبر عملية، تُعرف بـ «العنونة التركيبية Syntactic Annotation». ويترافق عن هذه العملية إجراءان، على النحو الآتي:

١) **التوسيم Tagging**: يتم على مدونة لغوية كبيرة نسبياً؛ ويقصد به إلحاق علامات معيينة، تُعرف بـ «اللوسوم التركيبية Syntactic Tags» بالكلمات المضمنة في هذه المدونة، لتحول بذلك من مدونة خام Raw Corpus إلى مدونة موسّمة Annotated Corpus. ولم يُست للوسوم التركيبية صورة ثابتة، سواءً من حيث الكلم أم الكيف؛ إذ تتحكم فيها أهداف المعاجلة. ويمكن التّمثيل عليها بالنموذج الوارد في (الشكل ٤).

النص قبل التوسيم
الآن .. كيف يمكن أن تُحيي تجوم السماء ؟
النص بعد التوسيم
الآن[AD]/كيف[QU]/يمكن[VI]/أن[PO]/تحيي[CN]/تجوم[CN]/السماء[CN]؟
PoS Tags وسوم أقسام الكلام
الظرف[AD]، الاستفهام[QU]، الفعل المضارع[VI]، الأداة[PO]، الاسم الشائع[CN]

الشكل ٤: نموذج لمادة نصية مُستمدّة من مدونة لغوية موسّمة تركيبياً

٢) **التجزئة Parsing**: ويتأصل هذا الإجراء بالوحدة التركيبية في صورتها المُكتملة؛ حيث تقوم من خلاله بتقسيم هذه الوحدة على مكونات أصغر، لنتمكن من التعامل مع كل منها على حدة. ويمكن التّمثيل على هذا الإجراء بالنموذج الوارد في (الشكل ٥).



الشكل ٥: نموذج لتجزئة الوحدة التركيبية في اللغة العربية

يُعوَّل على مُحرَّجات العنونة التركيبية - التي تتمُّ في مرحلة ما قبل المعالجة - في توفير المورد الرئيسي للمعالجة التركيبية؛ ونعني (المُدوَّنة اللُّغويَّة المُوسَّمة تركيبياً Syntactic Annotated Corpus). ونقوم - عبر هذا المورد - بتدريب النُّصوص، لنخلُص إلى قاعدة بياناتٍ لمُفرَّدات اللُّغة ووُسومها المُمثَّلة لأقسام الكلام. وعبر عمليَّات المُطابقة، يُمكِّن للالة أن تُحاكي مُحرَّجات قاعدة البيانات، حال توجيهها إلى المعالجة التركيبية السطحية بتعيين أقسام الكلام.

٤، ٢. المعالجة العميقَة للوحدات التركيبية.

لا تكفي القاعدة التركيبية الصُّرفة لتوجيه الآلة إلى معالجةٍ سليمة؛ إذ تُحاطُّ القاعدةُ في كثيرٍ من الأحيان ببعض الاستثناءات. وقد يتخلَّلها التباسٌ ناتجٌ عن تشابه بعض أقسام الكلام أو تجرُّدها من علامات الضَّبط [التشكيل] المُميَّزة لها. ومثال ذلك ما نجده في الكلمة (بل) التي يصعب تحديده قسمها الكلامي في صورتها المُجرَّدة؛ حيث تتحمل أن تكونَ اسمًا (بل) أو فعلًا (بل) أو حرفاً (بل)؛ ومثل ذلك كثيرٌ في العربية. لهذا، تستدعي معالجة العميقَة للوحدات التركيبية تجاوزَ البنية السطحية للجملة إلى تكوين هيكلٍ من القواعد المُنظمة التي تخضع لمقاييسٍ كميَّة، يمكن الاستنادُ إليه في توجيه الآلة إلى معالجةٍ مُوافقةٍ لطبيعة اللغة. وبطبيعة الحال، فإنَّ هذه القواعد تتوافقُ في جزءٍ منها مع منطق اللُّغة؛ لكنَّها تخرجُ كذلك في أحيانٍ كثيرةٍ عن هذا المنطق [غير المُطرَّد] إلى منطق الآلة [المُطرَّد].

إنَّا نعتمدُ - في المعالجة العميقَة للوحدات التركيبية - على ثلاثة عناصر أساسية، هي:

١) القواعد النحوية Grammar: تمثل المُعطيات اللغوية؛ ومصدرها في العربية قوانين النحو العربي [في الجانب التركيب منه]; لا سيما تلك التي تتصل بأقسام الكلام وال العلاقات بينها، وأنماط الإعراب والبناء [الرفع، والنصب، والخض، والجزم] لأقسام الكلام المختلفة.

٢) المُدوّنات اللغوية المشكولة والمُوسمة تركيبياً: مصدرها مجموعات من النصوص الممثلة لواقع اللغة؛ ويفترض أن تكون هذه المُدوّنات مقياساً للاستعمال اللغوي. ووجود هذا العنصر في المعالجة التركيبية العميق يعني أن إنجازها على نحو سليم يستدعي أن تسبق بمعالجَة سطحية. ونشدُّ في هذه المُدوّنات أن تُضبطَ باستخدام علامات الضبط العربية Arabic Diacritics؛ حيث يُعَوَّل عليها في تعين الموضع الإعرابي لأقسام الكلام [بأنماطها المعرفة والمبنية]؛ ويعوَّل عليها كذلك في الاستدلال على الأنماط الإعرابية للكلمات المُتضمنة في الوحدة التركيبية المُعينة، سواءً أكانت هذه الكلمات ملزمةً حالة الإعراب أم البناء.

٣) خوارزمية التحليل التركيب Syntactic Analysis Algorithm: وهي المُسلسلة الرياضية التي تُعبِّر عن مراحل المعالجة التركيبية. ويعوَّل فيها على الطرق الإحصائية التي تُساعدُ على حصر أنماط الجملة وتصويفها على نحوٍ موافقٍ لغة المستخدمة فعليًّا.

تتألَّف هذه العناصر فيما بينها لتكوين منظومة المعالجة التركيبية في اللغة العربية. فالقواعد النحوية هي الأساس الذي يتحدد من خلاله شكل الوحدة التركيبية؛ إذ إنها الموجِّه الرئيسي لمعرفة أركان الجملة [الأساسية والثانوية]، وقواعد ضبطها على نسقٍ سليم. والمُدوّنة المُوسمة هي المورد الذي نستعينُ من خلاله أنماط الجملة العربية ونعرفُ حُدوَّتها وأساليب صوغيها في اللغة الحية. وخوارزمية التحليل التركيب هي الوسيلة القياسية التي تقدُّمنا إلى أنماط التراكيب الشائعة وغير الشائعة، وتقودنا كذلك إلى تعين احتمالات المعالجة الأقرب إلى واقع اللغة حال وجود التباس في النصوص.

ولمزيد من التوضيح، نعرض في (الجدول ٨) نماذج موجّهات العناصر الثلاثة، مع التّمثيل.

العنصر	الموجّه	الشّمثيل	م
القواعدة النّحوية	القاعدة ١: «لن» أداة مبنية، تسبق الفعل المضارع، وتنصّبه.	لن [أداة] تكون [فعل مضارع منصوب].	١
المُدوّنات اللّغوية	القاعدة ٢: كُلُّ ما بدأ بـ«يس» يُوسّم بـ«نه» (فعل مضارع).	يسّح - يَسْتَمِر	٢
خوارزمية التّحليل التّركيبيّ	في القاعدة ١: القاعدة الصحيحة، مع وجود استثناءاتٍ يُظهّرُها الضّبط.	لن [فعل طلب] لاخيك	٣
	في القاعدة ٢: القاعدة الصحيحة، مع وجود استثناءاتٍ يُظهّرُها النّصّ.	يسّوّع استثناء مُؤكّدة: يَسِير استثناء مُحتَمل:	

الجدول ٨: نماذج من موجّهات عناصر المُعالجة التّركيبيّة العميقّة في اللّغة العربيّة

سنلاحظ أنَّ القاعدة النّحوية تُعبّر عن أحد القوانين التّركيبيّة في العربيّة؛ لكنَّ توجيه الآلة إلى هذا القانون يفرض تزويدّها بمعطياتٍ استثنائيّة، لتعاطي مع الاحتمالات الأخرى. وسنلاحظ كذلك أنَّ القاعدة المستخلصة من المُدوّنة اللّغوية تُعبّر عن أحد القوانين الوصفيّة التي يفرضها النّصّ. ونقوم في هذه الحالة أيضًا بتزويد الآلة بمعطياتٍ استثنائيّة [مؤكّدة ومحتملة]. وفي القاعدتين [النّحوية والنّصيّة]، يُعوّل على خوارزمية التّحليل التّركيبيّ في تتبع الاستثناءات وتقدير احتماليّة ورودها.

٣، ٤. من تطبيقات مُعالجة اللّغة العربيّة المكتوبة في المستوى التّركيبيّ.

يُمثّلُ مستوى المُعالجة التّركيبيّة تحدّيًّا للعاملين في حوسّبة اللّغة، نظرًا لطبيعة الجملة العربيّة وتعدد أنماطِها وثراء العربيّة بالظواهر التّركيبيّة ذات الطّبيعة الخاصة [كالأسماء الخمسة، والثّثنية، والجُمُوع السّالمة] ووجود ظاهرة الإعراب [الّتي تتحكّم في شكل الكلمات في حالَيِّ البناء والإعراب] ونحو ذلك. ويُمكّن التّمثيل على تطبيقات مُعالجة العربيّة المكتوبة في المستوى التّركيبيّ بما يأتي.

٤، ٣، ٤. التَّدْقِيقُ الْإِمْلَائِيُّ Spell and Grammar Checking

يُعنِي التَّدْقِيقُ الْإِمْلَائِيُّ باستكشاف أخطاء الكتابة؛ سواءً أكانت في قواعد إملاء الكلمات أم في القواعد التَّرْكِيَّة (النَّحْوَيَّة)؛ ويَعْقُبُ استكشاف الأخطاء تقديم اقتراح لمجموعة من احتمالات الصَّواب. وَيُعَدُّ هذا التَّطْبِيقُ أَحَدَ أَبْرَزِ تَطْبِيقَاتِ الْمُعَالَجَةِ التَّرْكِيَّةِ في الْعَرَبِيَّةِ، نَظَرًا لِحَاجَةِ إِلَيْهِ مِنْ قَبْلِ مُسْتَخْدِمِيِّ الْعَرَبِيَّةِ فِي التَّفَاعُلِ مَعَ الْآلَةِ بِأَشْكالِهَا الْمُخْتَلِفَةِ [الْحُوَاسِيبُ الْلُّوْحِيَّةُ، وَالْهَوَافِنُ الْذَّكِيَّةُ، ...].

ويَسْتَمِدُّ هَذَا التَّطْبِيقُ مُعْطِيَاتِهِ مِنْ مُدوَّنَةِ لُغَوَيَّةٍ مُتَعَدِّدَةِ الْأَهْمَاطِ التَّرْكِيَّةِ، عَلَى أَنْ تَكُونَ نُصُوصُهَا مَكْتُوبَةً بِصُورَةٍ سَلِيمَةٍ [إِمْلَائِيًّا وَنَحْوِيًّا]؛ حِيثُ يَتَمُّ تَدْرِيبُ الْمُدوَّنَةِ عَلَى مُطَابَقَةِ الْمُفَرَّدَاتِ وَالْتَّرَاكِيبِ الْمُسْتَهْدَفَةِ بِالْمُفَرَّدَاتِ وَالْتَّرَاكِيبِ الْمُتَضَمِّنَةِ فِي قواعدِ يَبَانَاتِ أَدْوَاتِ التَّدْقِيقِ، ثُمَّ اقتراح احتمالاتِ الصَّوابِ عَنْدَ اكتشافِ وُجُودِ خطَا مَا، اسْتَنادًا إِلَى مُعَجَّمٍ [أَوْ مَسَرَّدٍ] يَحْوِي قَدْرًا كَبِيرًا مِنَ الْمُفَرَّدَاتِ وَالْمُتَلَازِمَاتِ؛ حِيثُ تُرَجَّحُ الْكَلِمَاتُ الصَّحِيَّةُ الَّتِي تَشَرَّكُ مَعَ الْكَلِمةِ مَوْضِعُ الْخَطَا فِي عَدْدٍ مِنَ الْمَحَارِفِ أَوْ الَّتِي يَكْثُرُ تَلَازُمُهَا مَعَ كَلِمَاتٍ سَابِقَةٍ أَوْ لَاحِقَةٍ بِالْكَلِمةِ الْمَعْنَيَّةِ.

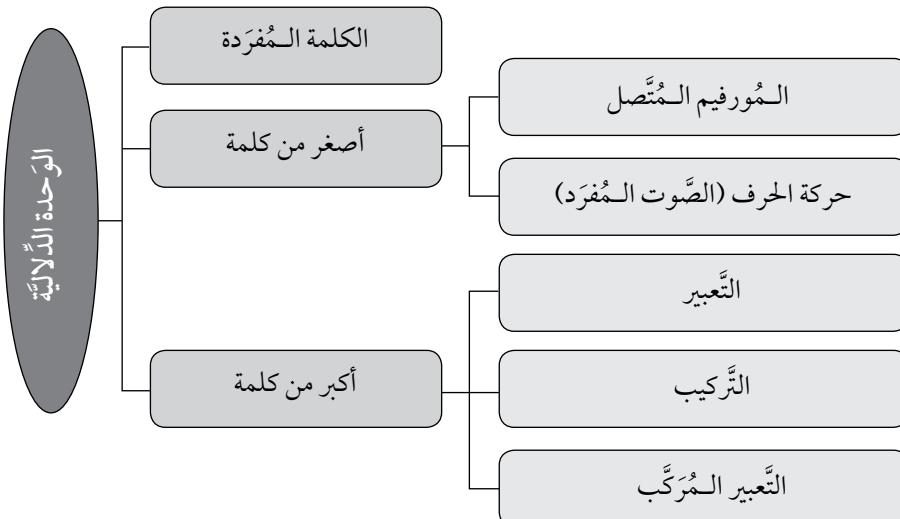
٤، ٣، ٤. التَّشْكِيلُ الْأَلْيَ Automatic Diacritics

يَفْرُضُ النَّظَامُ التَّرْكِيُّ لِلْغُوَيَّةِ وُجُودَ مِثْلَ هَذَا التَّطْبِيقِ الَّذِي يُعنِي بِضَبْطِ شَكْلِ الْكَتَابَةِ بِاسْتِخْدَامِ (عَلَامَاتِ الضَّبْطِ). وَيَعْمَلُ هَذَا التَّطْبِيقُ فِي الْعَرَبِيَّةِ عَلَى مُسْتَوَيَّيْنِ؛ يُعنِي الْأَوَّلُ مِنْهُمَا بِضَبْطِ الْأَبْنِيَّةِ [الْكَلِمَاتُ الْمُجَرَّدَةُ]، وَيُعنِي الْآخَرُ بِضَبْطِ الْإِعْرَابِ [أَوْ الْكَلِمَاتُ]. وَيَسْتَمِدُّ هَذَا التَّطْبِيقُ مُعْطِيَاتِهِ مِنْ مُدوَّنَةِ لُغَوَيَّةٍ مُوَسَّمَةٍ وَمَضْبُوَطَةٍ بِالشَّكْلِ ضَبِطًا تَامًا؛ وَيَعْتَمِدُ فِي ضَبْطِ الْأَبْنِيَّةِ بِصُورَةٍ كَبِيرَةٍ عَلَى دَوَانِ شَكْلِ الْكَلِمَةِ فِي الْمُدوَّنَةِ الْمُوجَّهَةِ لِلتَّدْرِيبِ عَلَى الضَّبْطِ؛ إِذْ تَرَدُّدُ احْتِمَالَيُّ الضَّبْطِ بِشَكْلِ مُعَيَّنٍ بازديادِ دورانِ الْكَلِمَةِ بِهَذَا الشَّكْلِ. أَمَّا فِي مُسْتَوَى الْإِعْرَابِ، فَيَعْتَمِدُ عَلَى قِرَائِنَ لُغَوَيَّةٍ مُعَيَّنَةٍ تُصَاحِبُ الْكَلِمَاتِ الْعَرَبِيَّةِ (مِثْلُ: أَدْوَاتُ نَصْبِ الْأَفْعَالِ، وَحُرُوفِ جَرِ الْأَسْمَاءِ، وَنَحْوِهَا)، بِالإِضَافَةِ إِلَى اعْتِمَادِهِ عَلَى مُخْرَجَاتِ إِحْصَاءِ تَدْرِيبِ أَهْمَاطِ الْجُمَلِ [الْبِسِيَّةُ وَالْمُرَكَّبَةُ] فِي الْمُدوَّنَةِ الْلُّغَوَيَّةِ.

٥. المعالجة الآلية الدلالية.

١، ٥. وحدة المعالجة الدلالية.

وحدة المعالجة الدلالية هي (السيميم Sememe)، وهو أصغر وحدة دلالية لها معنى. ونظرًا لتوزُّع مفردات اللغة بين (الكلمات التامة Full Words) التي يعني بها المعجم، كالأسماء والأفعال والصفات، و (الكلمات الصورية Form Words) التي يعني بها علم التركيب [النحو]، كالضمائر والأدوات؛ فإن الوحدة الدلالية تأخذ أحد ثلاثة صور، على نحو الموضح في (الشكل ٦). ويمكن التمثيل على صور الوحدات الدلالية بالنماذج الموضحة في (الجدول ٩).



الشكل ٦: أنماط الوحدة الدلالية في اللغات الطبيعية

م	أنماط الوحدات	نماذج الوحدة الدلالية
١	المُورفيم المُنْتَصل (الكلمة المُفردة)	إنسان
٢	حركة الحرف (المُورفيم المُنْتَصل)	(س) يعمل = السُّويف كتَبَ (تَ) = خطاب المُفَرَد المُذَكَّر
٣	التعَبِير	صَرَبَ أَخْمَاسًا في أَسْدَاسٍ = تَحِيرٌ
	الترَكِيب	تَابَطَ شَرًّا = الْعِلْمِيَّة
	التعَبِير المُرْكَب	الحافلة النَّهْرَيَّة

الجدول ٩: أنماط الوحدات الدلالية في اللغة العربية

٢، ٥. التَّوْجِيهُ الْآلَى لِمُعَالَجَةِ الْوَحْدَاتِ الدَّلَالِيَّةِ فِي الْلُّغَةِ الْعَرَبِيَّةِ.

إذا جاءَت الْوَحْدَةُ الدَّلَالِيَّةُ عَلَى نَمَطٍ أَصْغَرَ مِنَ الْكَلْمَةِ الْوَاحِدَةِ، فَإِنَّهَا تُشَيرُ إِلَى مَعْنَى وَظِيفِيِّ Functional Meaning؛ وَتُوَجَّهُ آلِيًّا حِينَئِذٍ لِلِّمُعَالَجَةِ الْصَّرْفِيَّةِ أَوِ التَّرَكِيَّةِ؛ وَإِذَا جَاءَتْ عَلَى نَمَطٍ أَكْبَرَ مِنَ الْكَلْمَةِ، فَإِنَّهَا تُشَيرُ إِلَى مَعْنَى مُصْطَلِحِيِّ Terminological Meaning، يُهْتَدَى إِلَيْهِ عَبْرَ قَوَاعِدِ بَيَانِاتٍ خَاصَّةٍ، تَحْوِي قَوَائِمَ الْمُصْطَلِحَاتِ وَالْتَّعَابِيرِ وَالْتَّرَاكِيبِ الْمُتَلَازِمَةِ وَمَعْنَاهَا؛ وَتَكُونُ هَذِهِ الْقَوَاعِدُ أَشَبَّهَةً بِالْمَعَاجِمِ الْمُصْطَلِحِيَّةِ. أَمَّا إِذَا جَاءَتْ الْوَحْدَةُ الدَّلَالِيَّةُ عَلَى نَمَطِ الْكَلْمَةِ الْمُفَرَّدَةِ، فَإِنَّهَا تُشَيرُ إِلَى مَعْنَى مُعَجمِيِّ Lexical Meaning؛ وَهُوَ الَّذِي تَعَاطِي مَعْنَى الْآلَةِ عَنْدَ تَوْجِيهِهِ إِلَى الِّمُعَالَجَةِ الدَّلَالِيَّةِ. وَيَتِمُّ هَذَا التَّوْجِيهُ عَبْرَ الْمُسْتَوَيَّينِ: السَّطْحِيِّ وَالْعَمِيقِ عَلَى النَّحوِ الْآتِيِّ.

١، ٢، ٥. الِّمُعَالَجَةُ السَّطْحِيَّةُ لِلْوَحْدَاتِ الدَّلَالِيَّةِ.

إِذَا بَحَثَنَا فِي (الْمُعَجمِ الْوَسِيْطِ) عَنْ كَلْمَةِ (الْبَدَاوِيَّة)، فَسَنَجِدُ الْمَعْنَى الْآتِيَّ:

(الْبَدَاوِيَّةُ الْحَيَاةُ فِي الْبَادِيَّةِ؛ وَيَعْلُبُ عَلَيْهَا التَّنَقُّلُ وَالْتَّرَحالُ).

وَإِذَا بَحَثَنَا فِي الْمُعَجمِ ذَاتِهِ عَنْ كَلْمَةِ (الْعَيْنِ)، فَسَنَجِدُ الْمَعْنَى الْآتِيَّ:

(الْعَيْنُ) عُضُوُّ الْإِبْصَارِ لِلْإِنْسَانِ وَغَيْرِهِ مِنَ الْحَيَّانِ،

وَ: يَنْبُوْعُ الْمَاءِ يَنْبُوْعُ مِنَ الْأَرْضِ وَيَجْرِيُ،

وَ: أَهْلُ الْبَلَدِ،

وَ: أَهْلُ الدَّارِ،

وَ: الْجَاسُوسُ،

وَ: رَئِيسُ الْجَيْشِ،

وَ: ...

فَالْوَحْدَةُ الدَّلَالِيَّةُ تَحْتَمِلُ أَنْ تُشَيرَ إِلَى مَعْنَى وَاحِدٍ، وَتَحْمُلُ كَذَلِكَ أَنْ تُشَيرَ إِلَى أَكْثَرِ مَعْنَى. وَفِي الْحَالَةِ الْأُولَى (عِنْدَ الدَّلَالَةِ عَلَى مَعْنَى وَاحِدٍ)، يُمْكِنُ تَوْجِيهُ الْآلَةِ إِلَى هَذَا الْمَعْنَى عَبْرَ (الْمُعَجَّمِ) Dictionary باعتِبَارِهِ مُورِّدًا رَئِيْسًا لِلْمَعْنَى الْمُعَجمِيَّةِ؛ مَعْ مُرَاعَاةِ أَنَّ

بعض الكلمات تُشير إلى معانٍ مجازية ليست مُضمنة بالضرورة في المعجم. أما في الحالة الأخرى (عند الدلالة على أكثر من معنى)، فعلينا حينئذ أن نُوجه الآلة إلى هذه المعاني جميعاً عبر معاجلة سطحية تستند إلى قواعد بيانات معجمية؛ وسيكون الإشكال في هذه الحالة في توجيه الآلة إلى ترجيح المعنى المقصود من جملة المعاني المقترنة.

والواقع أنَّ المعاني المعجمية لا تكفي وحدتها لمعالجة الوحدات الدلالية؛ إذ يُعَوَّل كذلك على العلاقات القائمة بين الوحدات الدلالية. ويُبيَّن (الجدول ١٠) بعضًا من هذه العلاقات.

التمثيل	المفهوم	العلاقة	م	
الخوف = الحشيشة	إذا كان (أ) يُعبر عن (ب)، و (ب) يُعبر عن (أ)، فإن (أ) و (ب) مترافقان.	الترافق	أفقية	١
الأخذ × العطاء	إذا كان (أ) ضد (ب)، و (ب) ضد (أ)، فإن (أ) و (ب) مُتضادان.	المُتضاد		٢
الرَّجُل ≠ المرأة (إنسان)	إذا كان (أ) يشتمل على (ب) و (ج)، وكان (ب) ≠ (ج)، فإن (ب) و (ج) مُتباينان.	المُتباين		٣
الثَّيْن - الفاكهة	إذا كان (أ) نوعاً من (ب)، فالعلاقة بين (أ) و (ب) نوعية.	النوعية		٤
الرَّأْس - الجسد	إذا كان (أ) جزءاً من (ب)، فالعلاقة بين (أ) و (ب) جزئية.	الجزئية		٥
اللُّغَة / الكلام	إذا كان (أ) يؤدّي إلى (ب)، وكان (ب) لا يتم إلاؤه عن طريق (أ)، فإن (أ) و (ب) مُتلازمان.	المُتلازمانية		٦

الجدول ١٠: من العلاقات الدلالية في اللغة العربية

تقوم المعاجلة السطحية للوحدات الدلالية على توظيف موارد لغوية في توجيه الآلة إلى معانٍ هذه الوحدات وال العلاقات بينها. وبالإضافة إلى (المعجم)، فَمَمَّا موارد أخرى، منها على سبيل المثال:

- الشبكة الدلالية Semantic Net: وَتُعَدُّ هيكلًا معرفياً يربطُ بين الوحدات [الكلمات] باستخدام العلاقات الدلالية في مستوياتٍ أفقيةٍ ورأسيَّةٍ متعددة.

- شبكة الكلمات WordNet: وتعُد قاعدة بيانات مُعجمية دلالية؛ حيث تربطُ بين الوحدات والمفاهيم من ناحية، وبين الوحدات وال العلاقات الدلالية من ناحية أخرى.

- الأنطولوجيا Ontology: وهي شبكة معرفية ضخمة، تجمع في كيانها بين الشبكات الدلالية وشبكات الكلمات؛ وتجاور الأنطولوجيا المعاني المُعجمية إلى مختلف أشكال المعرفة؛ كما تتجاوز العلاقات الدلالية المباشرة إلى العلاقات غير المباشرة بين الوحدات.

٢٢. المعالجة العميقية للوحدات الدلالية.

إذا كانت المعالجة السطحية للوحدات الدلالية تعنى بتوجيه الآلة إلى المعنى الواحد أو المعاني المتعددة، فالمعالجة العميقية تعنى بالبحث في احتمالية وجود معانٍ أخرى للوحدات ذات المعنى الواحد، وتعنى كذلك بالترجيح بين احتمالات المعاني في الوحدات التي تشير إلى معانٍ متعددة. وتمت المعالجة العميقية عبر ما يُعرف بـ «فك الالتباس الدلالي» Word Sense Disambiguation؛ وهو إجراء متطور يهدف إلى تعين المعنى المقصود، اعتماداً على قرائن لغوية تُستخلص من السياق الذي ترد فيه الوحدات الدلالية. وتقوم فكرة هذا الإجراء على البحث عن المعنى المقصود في مجاورات الوحدة المعينة؛ حيث يُغلب أن تتجاوز الوحدة الدلالية الذالة على معنى معين مع أنها مُعینة من مفردات اللغة؛ سواءً أكان هذا التجاوز مُباشراً (قبل الوحدة أو بعدها) أم غير مُباشراً.

في اللغة العربية، يمكن التمثيل على التجاوز المباشر بكلمة (ابن) حين تأتي مجاورةً [سابقة] لوحدة دلالية معينة. فإذا رمزاً مثلاً للكلمة (ابن) بالرمز (S) وللوحدة المُعجمية المعنية [التالية] بالرمز (N)، فإن ورودهما مجاورتين يعني ترجيح احتمالية أن تشير الوحدة (N) إلى العلمية. ومن ناحية أخرى، يمكن التمثيل على التجاوز غير المباشر بكلمة (يُغمض) حين تأتي في سياق الوحدة الدلالية (عين) في الجملة (حين يُغمض الإنسان عينه يشعر بشيء من السكينة)؛ حيث أفادت الكلمة (يُغمض) التي جاءت في صورة الفعل المضارع دلالة الوحدة (عين) على (عضو الإبصار)؛ وهكذا.

ونظراً لطبيعة اللغة العربية في أبنيتها وتراكيبيها، فإنَّ فكُ الالتباس الدلالي لوحاتها يستدعي تداخلاً بين الأنظمة الثلاثة (البنيوي، والتركيبي، والدلالي) - إلى حدٍ كبير؛ حيث يُعوَّل على النَّظام البنوي في تعين الوَحدة الدلاليَّة ذاتها، وتجريدها من زوائدَها (السوابق واللَّواحق)، ويُعوَّل على النَّظام التَّركيبي في تعين قِسم الكلام الذي تنتهي إليه الوَحدة، واستكشاف مُجاوراتها، وتعين حدود السياق اللُّغوي؛ ويُعوَّل على النَّظام الدلالي في تعين معاني الوَحدة، ثمَّ ترجيح المعنى الأقرب إليها.

يُقَيَّ أن نُشير إلى وجود العديد من خوارزميات فك الالتباس الدلالي WSD. ومنها مثلاً: (خوارزم Lesk، خوارزم بايز Naïve Bayesian Algorithms، خوارزم Yarowsky). الواقع أنَّ طرائق تطبيق هذه الخوارزمات مُبَيَّنة إلى حدٍ ما، وإن كانت جميعاً تنطلق من قرينة السياق اللُّغوي. وفي العموم، يستدعي فك الالتباس الدلالي في العربية تلاحمًا بين مجموعة من الموارد التي تشمل: (المعاجم والموسوعات، وشبكات الكلمات، والأنطولوجيات، والمُدَوَّنات اللُّغوية، والمُحَلَّلات التَّركيبية).

٣،٥. من تطبيقات معالجة اللغة العربية المكتوبة في المستوى الدلالي.

يُعوَّل على التَّحليل الدلالي في معالجة اللغة العربية المكتوبة في العديد من التطبيقات التي تتَّصل بالمعنى؛ مع التَّأكيد على التَّهَاجُر الاحتمي بين مستوى التَّحليل الدلالي ومستويات التَّحليل اللُّغوي الأخرى، وفقاً للغاية من المعالجة. ويمكن التَّمثيل على هذه التطبيقات بما يأقي:

١،٣،٥. استرجاع المعلومات النَّصيَّة :Text Information Retrieval

يهدف هذا التطبيق إلى البحث أو الاستدلال عن معلوماتٍ مُعَيَّنةٍ في مُحتوى كبيرٍ نسبياً من النصوص غير المُنظمَة. ويُستفادُ منهُ في بناء محركات البحث Search Engines وهيكلة مُسَوِّدَعات البيانات Data-Warehouses. وتُبرِّزُ أهميَّة المعالجة الدلاليَّة في هذا التطبيق عندما تتحمل الكلمة البحث المستهدفة دلالاتٍ مُتعددة، ويرادُ التركيزُ على نتائج البحث عن إحدى هذه الدلالات، على نحو ما نجد مثلاً عند البحث عن كلمة (الأسد) حالَ قَصْد توجيه البحث إلى العَلميَّة.

٢، ٣، ٥. تحليل المشاعر (التأثر على الآراء) : Opinion Mining

من تطبيقات التّنقيب في النّصوص Text Mining؛ ويهدف إلى الكشف عن المشاعر أو الآراء التي تعبّر عن وجهات نظر الأفراد والجماعات [تلميحاً أو تصريحاً] في مجموعة من النّصوص. ويُستفادُ من هذا التّطبيق في مجالاتٍ عديدة، مثل: البحث الجنائي، والاستكشاف المُخابراتي، والتسويق، بالإضافة إلى الدراسات المسحية المعنية بتحليل الآراء. ولهذا التّطبيق أهميّة كبيرة في تحليل البيانات النّصيّة في وسائل التّواصل الاجتماعي Social Media؛ نظراً لثراء مادة هذه الوسائل بالنّصوص المكتوبة من ناحية، وتعدد توجّهات مستخدميها وأراءهم من ناحية أخرى.

٣، ٤، ٥. التّأثر على أسماء الكائنات : Named-entity Recognition

يُعدُّ كذلكَ أحدَ تطبيقات التّنقيب في النّصوص؛ ويهدفُ هذا التّطبيق إلى تمييز الأسماء الدالّة على العلميّة [الكائنات أو الكيانات]، كأسماء الأعلام والمعالم والمؤسّسات ونحوها؛ وهي أسماء يكثرُ وجودُها في البيانات النّصيّة المُمتنعة [أو شبه المُمتنعة]، كالمُدوّنات اللّغويّة والمعاجم والموسوعات ونحوها. وفي اللّغة العربيّة يُستفادُ من هذا التّطبيق في جوانبٍ عديدة، مثل: التّحليل التّركيبيّ، وتحليل الخطاب، وبناء قواعد بيانات الأعلام، بالإضافة إلى جوانبٍ أخرى تتبع ميدان التّنقيب في النّصوص، مثل: التّلخيص الآلي للنّصوص Automatic Text Summarization الذي يعني بإعادة بناء النّص في صورة مختصرة، والتّقييم الآلي Automatic Text Scoring الذي يعني بقياس صحة النّص.

بليوجرافيا مرجعية.

1. Abdel-Fattah, Y. (2018). *Arabic Corpus Linguistic*. Edinburgh University Press.
2. Arulmozi, S.; Dash, N. (2018). *History, Features, and Typology of Language Corpora*. Springer.
3. Blokdyk, G. (2018). *Text Mining Complete Self-Assessment Guide*. Emereo Pty Limited.
4. Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
5. Collins, L. (2019). *Corpus Linguistics for Online Communication: A Guide for Research*. Routledge.
6. Dash, N. S., Ramamoorthy, L. (2018). *Utility and Application of Language Corpora*. Springer.
7. Deng, L.; Liu, Y. (2018). *Deep Learning in Natural Language Processing*. Springer.
8. Ertel, W. (2018). *Introduction to Artificial Intelligence*. Springer.
9. Farr, F.; Murray, L. (2016): *The Routledge Handbook of Language Learning and Technology*. Routledge.
10. Feldman, R.; Sanger, J. (2006): *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. CUP.
11. Gomez, P. C. (2013): *Statistical Methods in Language and Linguistic Research*. I^{sd}.
12. Handford, M. (2018). *Corpus Linguistics for Discourse Analysis: A Guide for Research*. Routledge.
13. Haynes, D. (2018). *Metadata for Information Management and Retrieval: Understanding metadata and its use*. Facet Publishing.
14. Jo, T. (2018). *Text Mining: Concepts, Implementation, and Big Data Challenge*. Springer.

15. Karwowski, W. (2019). *Intelligent Human Systems Integration 2019*. Springer.
16. Kulkarni, A.; Shivananda, A. (2019). *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python*. Apress.
17. MacIntyre, J.; Maglogiannis, I.; Iliadis, L.; Pimenidis, E. (2019). Artificial Intelligence Applications and Innovations: AIAI 2019 IFIP WG 12.5 International Workshops: MHDW and 5G-PINE 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings. Springer.
18. Marr, B. (2019). *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*. John Wiley & Sons.
19. McEnery, T. (2014). *Arabic Corpus Linguistics*. Edinburgh University Press.
20. McEnery, T.; Meurers, D.; Rebuschat, P. (2017). *Experimental, Corpus-based and Computational Approaches to Language Learning: Evidence and Interpretation*. Wiley.
21. Mel'čuk, I. (2015). *Semantics: From meaning to text*. John Benjamins Publishing Company.
22. Neamat El, G.; Yee, S. (2018). *Computational Linguistics, Speech and Image Processing for Arabic Language*. World Scientific.
23. Shaalan, k.; Hassanien, A.; Tolba, F. (2017). *Intelligent Natural Language Processing: Trends and Applications*. Springer.
24. Sharp, B.; Sedes, F.; Lubaszewski, W. (2017). *Cognitive Approach to Natural Language Processing*. Elsevier.
25. Sidorov, G. (2019). *Syntactic N-grams in Computational Linguistics*. Springer.

26. Soudi, A.; Bosch, A.; Neumann, G. (2007). *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Springer Science & Business Media.
27. Srinivasa-Desikan, V. (2018). *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python*, Gensim, spaCy, and Keras. Packt Publishing.
28. Tran, T. X. (2019). *Artificial Intelligence: Learning Artificial Intelligence*. Thanh Tran.
29. Walker, B.; McIntyre, D. (2019). *Corpus Stylistics: Theory and Practice*. Edinburgh University Press.
30. Zizka, J.; Darena, F.; Svoboda, A. (2019). *Text Mining with Machine Learning*. Taylor & Francis Group.

الفصل الثالث

المُعَالَجَةُ الْأَلْيَّةُ لِلْغَةِ الْعَرَبِيَّةِ الْمَنْطَوَقَةِ

د. أحمد راغب أحمد

ملخص

مع الطفرة التي أحدثتها التقنية في عالمنا المعاصر، وتزامناً مع الثورة الإلكترونية التي صبّغت كل مجالات الحياة، أصبحت تقنيات معالجة الصوت البشري مطلباً مهّماً، ليس فقط لتسهيل عمليات التواصل بين المتحدثين، وإنما لأغراض اقتصادية وتجارية وسياسية لا حصر لها، بالإضافة إلى الأبعاد الأكاديمية والاجتماعية التي تمثل في مساعدة ذوي الاحتياجات الخاصة عن طريق تحويل الزخم العلمي والثقافي المكتوب إلى مادة مسموعة يفيد منها ذوي الاحتياجات الخاصة والأسيّاء على حد سواء.

وقد حاول هذا الفصل التعرض للمبادئ التطبيقية التي يتبعها المختصون في مجال تحليل الصوت اللغوي وتوليدّه، وذلك من خلال إلقاء الضوء على خصائص مكونات الصوت اللغوي التي قُمت بتقسيمها إلى مجموعتين، تتناول المجموعة الأولى منها الصفات الأكoustية التي ترسم ملامع هذا الصوت بغض النظر عن قائله، في حين تتناول المجموعة الأخرى الصفات الذاتية الشخصية التي ترتبط ارتباطاً وثيقاً بِمُتّجِّع الصوت، والتي تتفاوت خصائصها حسب الفئات الناطقة ذكوراً وإناثاً، أطفالاً أو شباباً أو شيوخاً، بل إنَّ هذه الاختلافات تلقي بضوئها على خصائص الصوت حسب اختلاف الأفراد في الفتة الواحدة، وهو ما يترتب عليه وجود صفات أو خصائص صوتية لكل فرد تميّزه.

وقد اختتمت الفصل بشرح تطبيقي لعملية إنتاج الصوت العربي حاسوبياً من خلال عرض لتقنية تحويل النص العربي المكتوب إلى صوت منطوق (Text To Automatic Speech)، وعرضٌ آخر لتقنية التعرف على الكلام المنطوق (Speech Recognition).

الكلمات المفتاحية:

التحليل الصوتي الحاسوبي، مكونات الصوت اللغوي، الناطق الآلي، التمثيل الأدائي، التباين الأدائي.

١. الصَّوْتُ الْلُّغُوِيُّ وَعِلْمُ الْأَصْوَاتِ.

الكلام هو أكثر وسائل التواصل الطبيعية والبدنية وأكثرها تفضيلاً لدى البشر، وقد أورد ابن جنبي أنَّ حَدَّ اللغة أصوات يعبر بها كل قوم عن أغراضهم (ابن جنبي، ١٩١٣: ٣٣)، وعليه فإنَّ الكلام المنطوق ما هو إلا تعبير عن الأفكار والمشاعر من خلال إصدار ذبذبات مسموعة تنتقل عبر الجهاز الصوتي لدى الإنسان، ليتم استقبالها بواسطة الجهاز السمعي ومن ثم تتم معالجتها دلاليًّا وفق العرف اللغوي الكامن في أذهان الناطقين باللغة.

وقد عرَّف «روبن» الصوت بأنه اضطراب مادي في الهواء يتمثل في قوة أو ضعف ثم في ضعف تدريجي ينتهي إلى نقطة الزوال النهائي (العطية، ١٩٨٣: ٦).

وينشأ الصوت البشري كما ذهب إبراهيم أنيس من ذبذبات مصدرها عند الإنسان الحنجرة، فعند اندفاع النَّفَس من الرَّئَتين يمر بالحنجرة فتحدث تلك الاهتزازات التي بعد صدورها من الأنف أو الفم تنتقل خلال الهواء الخارجي على شكل موجات حتى تصل إلى الأذن (أنيس، ١٩٦١: ٧).

وبوجه عام فإنَّ الصوت هو الأثر السمعي الذي به ذبذبة مستمرة مطردة حتى ولو لم يكن مصدره الجهاز الصوتي (حسَّان، ١٩٩٠: ٥٩).

فالصوت إذن عبارة عن ذبذبات ناتجة عن قوة تنتقل عبر الهواء، وهو كذلك عرض يخرج من النفس مستطلياً متصلة حتى يعرض له في الحلق والفم والشفتين مقاطع تثنية عن امتداد واستطالة فيسمى المقطع أيّنما عرض له حرف (ابن جنبي، ١٩٥٤: ١٨٦).

ويرى غساني أنَّ الصوت ظاهرة طبيعية وشكل من أشكال الطاقة، وهو يستلزم وجود جسم في حالة اهتزاز أو تذبذب، وهذه الاهتزازات أو الذبذبات تنتقل عبر وسط معين حتى تصل إلى أذن الإنسان (جميداني، ٢٠١٠: ١٤).

أما إذا انتقلنا إلى علم الأصوات فيمكن تعريفه بأنه علم لغوي في المقام الأول يُعْنِي بدراسة أصوات اللغة حيث ينظر هذا العلم في الأصوات في حد ذاتها من حيث إخراجها، بل وحتى من حيث سماها، لكن بعض اللغويين يطلقونه ويريدون

به دراسة التغيرات والتحولات التي تحدث في أصوات اللغة نتيجة تطورها (البهنساوي، ٢٠٠٤: ١١).

٢. مصطلحات أساسية.

قبل الولوج إلى مسائل تحليل الصوت اللغوي حاسوبيًّا ينبغي أن تتم الإشارة إلى أهم المصطلحات التي تؤطر لهذا النوع من الدراسات، وأول هذه المصطلحات وأكثرها ذيوعاً مصطلح تقنيات اللغات البشرية (Human language Technology)، وهو مجال علمي وتطبيقي حديث يجمع بين دراسة اللغة من ناحية وتطبيقات التحليل الحاسوبي من ناحية أخرى، وتهدف تقنيات اللغات البشرية التي يتم الإشارة إليها اختصاراً بـ (HLT) إلى دراسة المناهج التي تتبعها البرامج الحاسوبية والأجهزة الإلكترونية لتحليل النصوص اللغوية والكلام البشري، كما تدرس طريقة إنتاج تلك النصوص والأصوات وتحليلها أو تعديلها والاستحابة لها بطريقة آلية، وتعتمد هذه التقنيات على مجالين فرعين، أوهما: معالجة اللغات الطبيعية (Natural Language Processing) أو ما يعرف اختصاراً بـ (NLP) وهو حقل فرعي لعلوم الكمبيوتر وهندسة المعلومات والذكاء الاصطناعي المهتم بالتفاعلات بين أجهزة الكمبيوتر واللغات البشرية (الطبيعية)، وخاصة كيفية برمجة أجهزة الكمبيوتر لمعالجة وتحليل كميات كبيرة من بيانات اللغة الطبيعية. وثانيهما مجال اللغويات الحاسوبية (Computational Linguistics) ويعرف اختصاراً بـ (CL). مع الاستفادة من معطيات مجالات مكملة أخرى مثل: تقنيات الكلام البشري (Speech Technology) وعلم الإحصاء التطبيقي (Applied Statistics) (Uszkoreit, ٢٠١٨).

وإلى جانب ما سبق أضحت حقول رياضية أخرى - وثيقة الصلة بهيكلة اللغة حاسوبيًّا، مثل نظرية المخططات وبالأخص علم الإحصاء (Statistics) - تحتل مكانة تزداد أهميتها يوماً بعد يوم داخل خريطة اللغويات الحاسوبية وتمثيلها الصوري وطبيعتها الديناميكية مما يؤدي إلى فهم آلي أعمق وتحليل أدق للغات الطبيعية (wataonline.net).

وإذا كانت الدراسات اللغوية الحاسوبية العامة قد قطعت شوطاً كبيراً على مستوى الإنجازات والنتائج فإنَّ الدراسات اللسانية العربية أيضاً قد طرحت عدة محاولات

تهدف إلى تطوير تقنيات الحاسوب للغة العربية بما يتوافق مع شخصيتها ومحارفها ورسومها من جهة، وبما يتواءم مع قواعد اللغة العربية وخصائصها الحاسوبية من جهة أخرى، بدءاً بترجمة الحروف والنصوص، ومروراً ببناء التقنيات الصوتية اللغوية لهدف تحسين الاتصال الآلي بين الإنسان والحاسوب (الوعر، ١٩٨٩: ٢٣).

٣. التحليل الحاسوبي لمكونات الصوت اللغوي.

إذا أردنا أن نقدم تحليلًا حاسوبيًا لمكونات الصوت اللغوي فينبعي أن نجزء هذه المكونات إلى مجموعتين، تتناول المجموعة الأولى الصفات الثابتة للصوت المنطوق، وأقصد بها تلك الصفات الأكoustية التي ترسم ملامح هذا الصوت بغض النظر عن قائله، وقد اصطلحت على تسميتها بالتماثل الأدائي، في حين تتناول المجموعة الأخرى الصفات الذاتية الشخصية التي ترتبط ارتباطاً وثيقاً بمتّج الصوت، وهو هنا المتكلم، وتتفاوت خصائصها حسب الفئات الناطقة ذكوراً وإناثاً، أطفالاً وشباباً وشيوخاً، بل إنَّ هذه الاختلافات تلقي بضوئها على خصائص الصوت حسب اختلاف الأفراد في الفئة الواحدة، وهو ما يترتب عليه وجود صفات أو خصائص صوتية لكل فرد تميزه عن غيره من الناطقين الذين يتمون إلى نفس المجموعة الصوتية، وهذا الأمر يؤدي بدوره إلى وجود بصمة صوتية تميز كل متحدث عن الآخر، وقد اصطلحت على تسمية هذه المجموعة بالتبابين الأدائي.

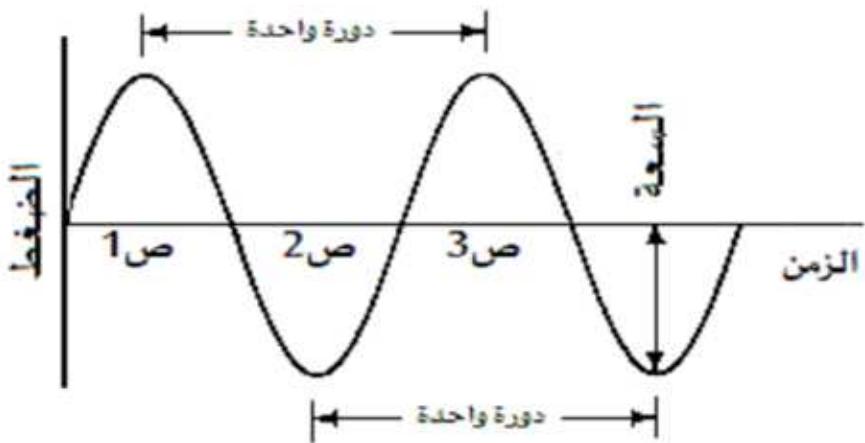
١، ٣. التماثل الأدائي:

المراد بالتماثل الأدائي تلك الخواص الصوتية الأكoustية (Acoustics features) التي تتصف بالثبات وتبيّن مواصفات الصوت اللغوي بغض النظر عن قائله، وبالتالي فلا نجد اختلافاً ملحوظاً في هذه المكونات مرتبطة بشخصية الناطق أو الصوت باعتبار نوعه أو جنسه أو عمره، وعليه فهي أقرب ما تكون إلى الخواص الإستاتيكية، حيث إنها أكثر عمومية وأقل خصوصية، وتحدد ملامح هذا التماثل الصوتي في الموجة الصوتية (Sound wave)، وسرعة الصوت (Speed of sound)، وشدة الصوت (Sound intensity)، والرسم التَّذبذبِيُّ (شكل الموجة)، والصورة الطيفية (Spectrogram).

١، ٣. الموجة الصوتية (Sound wave):

الموجة الصوت هي نمط الاضطراب الناتج عن حركة الطاقة التي تنتقل عبر وسيط (وسط) مادي - مثل الهواء أو الماء أو أي مادة سائلة أو صلبة أخرى - حيث تنتشر بعيداً عن المَصَوْت أو مصدر إنتاج الصوت، ونعني به في سياقنا هذا الناطق البشري أو الآلي؛ حيث يُصدر هذا المصَوْت طاقة تُسبِّب اهتزازاً ينتقل من المصدر إلى أقصى نقطة سمعية عبر الوسط المادي، ويُحِدِّث هذا الاهتزاز زعزعة لاستقرار جزيئات هذا الوسط المحيط بشكل جزئي؛ ثم تُحِدِّث هذه الجزيئات إزعاجاً مماثلاً للجزيئات التي تليها وهلم جرا إلى نهاية المدى الصوتي. ومن ثم تنشأ الدوائر الصوتية التي تشبه وتماثل الدوائر التي تحدث عن إسقاط كتلة مادية في مياه البحر، وتحمل تلك الموجات الطاقة الصوتية عبر الوسيط في كل الاتجاهات، وتقل كثافتها كلما بُعدت عن مصدرها.

وعليه فإنَّ الإشارة إلى الموجة الصوتية تعبَّر عن إحدى أشكال انتقال الطاقة الصوتية، والتي تنتقل عبر أي وسط مادي يسمح بانتقال الطاقة من المصدر إلى حيز الاستقبال بدون إزاحة جزيئات هذا الوسط بشكل دائم، بل يقتصر الانتقال على الشكل الجزئي، أي أنه لا تنتقل أي كتلة مع انتقال الموجة، ولكن جزيئات الوسط تتحرك بشكل متعمد أو مواز لاتجاه حركة الموجة حول موقع ثابت. وتتصف أي موجة بصفة الدورية، وهو ما يتربَّط عليه أن تكون تكرار النمطِ ما من الشدة في فترات زمنية متتابعة، ويسمى هذا التكرار بالتردد، وهو عبارة عن عدد الموجات المارة في مقطع ما مقسوماً على وحدة الزمن. بينما تسمى المسافة الأفقية التي تقطعها الموجة الواحدة «طول الموجة»، وهو أصغر جزء متكرر مكون للموجات، ويساوي المسافة بين قمتين أو قاعدين متتابعين، أو هو المسافة المحسورة بين ثلات قيم صفرية متتابعة للموجة، كما يوضّحه الشكل التالي.

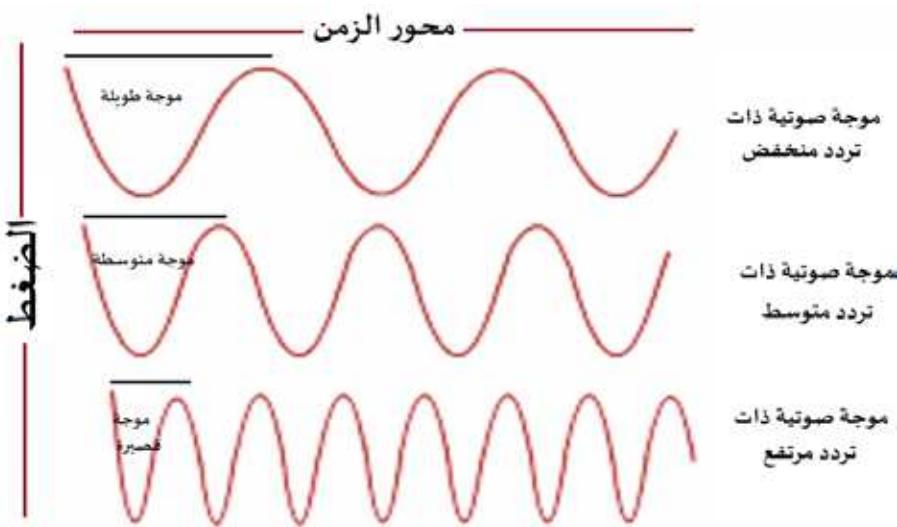


الشكل ١: رسم تخطيطي للموجة الصوتية مشتمل على طول الموجة وسعتها والقيم الصفرية، موزعة على محوري الزمن والضغط

وصفة الدورية للموجة عبارة عن التردد أو التواتر، وهو مقدار تكرر الموجة الواحدة ذات الطول الموجي المتفق عليه في كل وحدة زمن.

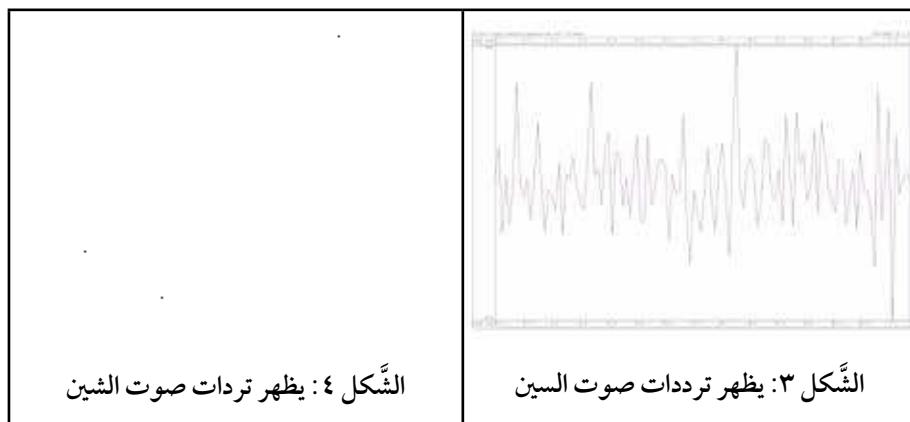
وتكون درجة الصوت أعلى كلما كانت الذبذبات أسرع، وكلما زادت سرعة هذه الذبذبات كان الصوت دقيقاً أو حاداً، ومن ثم فإن الصوت يكون سميكاً إذا قل عدد الذبذبات في الثانية الواحدة. «فالفرق بين شوكة رنانة ذات درجة صوتية عالية وأخرى ذات درجة صوتية منخفضة أن الأولى تعمل عدداً أكبر من الذبذبات في الثانية الواحدة» (عمر، ١٩٩٧: ٦٧). «ويقاس التردد بعدد الذبذبات التي تحدث في الثانية الواحدة» (الغامدي، ٢٠٠٠: ٣١).

وإذا كان تردد الصوت (Sound Frequency): هو عدد الذبذبات أو الموجات الكاملة في الثانية الواحدة (أيوب، ١٩٨٤: ٥٦)، فيمكننا رصد هذا التردد كما في (الشكل ٢) الذي يعرض موجة صوتية ذات تردد مرتفع وأخرى ذات تردد منخفض.



الشكل ٢: موجة صوتية ذات تردد مرتفع (صوت رفيع) وموجة صوتية ذات تردد متوسط وموجة صوتية ذات تردد منخفض (صوت غليظ)

ونلاحظ هنا أن التردد يناسب عكسيًا مع طول الموجة، أي كلما زاد التردد قصر طول الموجة، أي: التردد 4000 هرتز له طول موجة أقصر من التردد 500 هرتز. ويعتبر التردد من العناصر التي تؤدي إلى إدراك بعض الأصوات وتمييزها؛ حيث يؤدي اختلاف تردد موجات بعض الأصوات اللغوية إلى تغير في إدراكيها، فالصوت /س/ ذو تردد عال يفوق 4000 هرتز، فإذا انخفض تردداته ليقترب من 2000 هرتز فإن السامع يدركه /ش/. (الغامدي، ٢٠٠٠: ٧٨).



٢، ٣. سرعة الصوت (Speed of sound):

هي سرعة انتقال الطاقة الصوتية في الوسط، وهي ثابتة في الوسط الواحد بغض النظر عن نوع الصوت وتردداته، ولكنها تختلف من وسط إلى آخر فتتغير طبقاً لكتافة الوسط وإعاقته للصوت ودرجة الحرارة. وتمثل خصائص الوسط التي تحدد سرعة الصوت في الكثافة وقابلية الانضغاط. وكلما زادت الكثافة وزادت قابلية الانضغاط، قلت سرعة الصوت. وتزداد سرعة الصوت في السوائل عن الغازات وفي الأجسام الصلبة عن السوائل؛ وذلك لتقارب الجزيئات بها، فالسوائل والأجسام الصلبة بصفة عامة أكثر كثافة من الهواء، وهي أيضاً أقل من الهواء بكثير في قابلية الانضغاط. ولذلك، فإن الصوت يتقلّب بسرعة أكبر خلال السوائل والأجسام الصلبة.

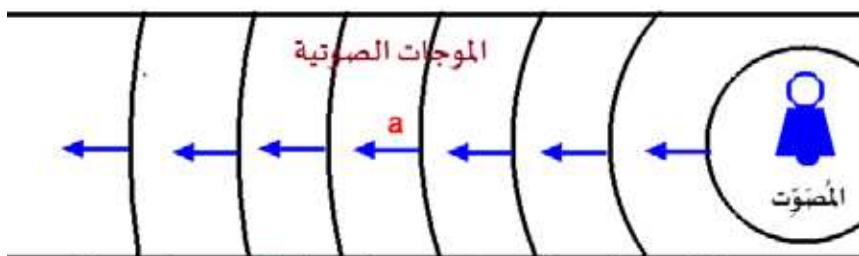
وعليه فإنَّ الصوت عبارة عن موجة طولية ينتجها ضغط المادة وتنافرها. وعموماً فإنَّه ينبغي أنَّ نضع في الحسبان عند حساب سرعة الصوت مجموعة من الخصائص التي تؤثر على تلك السرعة مثل الضغط والكتافة والكتلة الجزيئية؛ فكلما قلَّت كثافة الوسيط، زادت سرعة الصوت، وكلما زاد الضغط، كان الصوت أبطأ.

سرعة الصوت هي المسافة المقطوعة لكل وحدة زمنية بواسطة موجة صوتية أثناء انتشارها عبر وسيط مرن، عند ٢٠ درجة مئوية (ما يعادل ٦٨ درجة فهرنهايت)، وتبلغ سرعة الصوت في الهواء حوالي ٣٤٣ مترًا في الثانية (١٢٣٤,٨ كم / ساعة)؛ أو ١,١٢٥ قدمًا في ثانية؛ أو ٧٦٧ ميلًا في الساعة. وتعتمد سرعة الصوت على درجة الحرارة، وعلى نسبة تشبع الهواء بالغازات. ويتم حساب سرعة الصوت في الهواء بواسطة الصيغة الآتية:

$$\text{سرعة الصوت (م / ث)} = ٥٣١ + ٣٣١,٥ \cdot \text{ـ (درجة مئوية)}$$

وإذا كان الوسيط غير هوائي فإن تلك السرعة تكون أكثر تباعيًّا؛ حيث يتقلّب الصوت بصورة أبطأ في الغازات؛ وبصورة أسرع في السوائل وفي المواد الصلبة. وعلى سبيل المثال إذا كانت سرعة الصوت في الهواء ٣٤٣ مترًا في الثانية فإنها تنضاعف في السوائل لتصل إلى ١٤٨٠ مترًا في الثانية، وتبلغ ٥١٢٠ مترًا في الثانية عبر الوسط الصلب مثل الحديد، بل وتبلغ في المواد شديدة الصلابة مثل الماس حوالي ١٢٠٠٠ مترًا في الثانية، أي حوالي ٣٥ ضعفًا من سرعته في الهواء (Bannon, ٢٠١٥).

سرعة الصوت



Speed of sound (a) depends on the type of medium and the temperature of the medium.

For calorically perfect air: $\gamma_{perf} = \text{ratio of specific heats} = \text{constant} = 1.4$

$$a^2 = R T \gamma_{perf}$$

For calorically imperfect air: $\gamma = \text{ratio of specific heats} = \gamma (T)$

$$a^2 = R T \left\{ 1 + \frac{(\gamma_{perf} - 1)}{1 + (\gamma_{perf} - 1) \left[\left(\frac{\Theta}{T} \right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} \right]} \right\}$$

$\Theta = 3056^{\circ}\text{K}$

$R = \text{gas constant } (286 \text{ m}^2/\text{s}^2/\text{K}^{\circ} \text{ for air}) \quad T = \text{absolute temperature } ^{\circ}\text{K}$

الشكل ٥: سرعة الصوت، المصدر (www.grc.nasa.gov).

١.٣. شدة الصوت (Sound intensity)

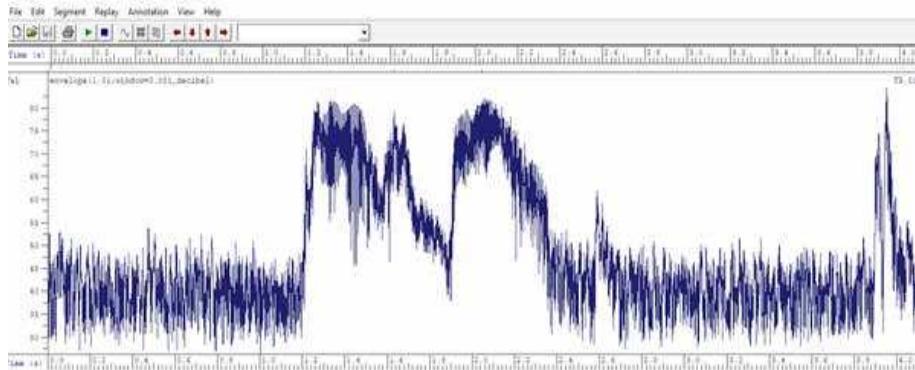
يُعرَّفُ مستوى شدة الصوت والذي يشار إليه بالمصطلح (Sound intensity) أو (Acoustic intensity) بأنه القدرة أو الطاقة، التي تحملها الموجات الصوتية لكل وحدة مساحية في اتجاه عمودي لتلك الوحدة، والسياق المعتمد لقياس شدة الصوت هو الهواء. ويتم قياس شدة الصوت بوحدة الواط لكل متر مربع أو لكل سنتيمتر مربع (واط/م² أو واط / سم²)، وغالباً يتم قياس شدة الصوت وفقاً لمعيار كثافة السمع القياسي الذي يشار إليه بالديسيبل (decibel) الذي تعبّر عنه هذه المعادلة:

$$I_0 = 10^{-12} \text{ watts/m}^2 = 10^{-16} \text{ watts/cm}^2$$

والديسيبل هو وحدة شدة الصوت، وهو وحدة ليس لها تمييز؛ لأنها نسبة بين كميتين من الطاقة، وهو الأساس لقياس كل الأصوات في الطبيعة مثل الكلام، والضوضاء، وأصوات الحيوانات وغيرها، ويمثل الديسيبل «مقيماً عملياً» لقوة الصوت وضغطه،

والديسيبل ١٠/١ بيل، والبيل هو ضغط أضعف صوت تدركه الأذن الإنسانية» (أيوب، ١٩٨٤: ١١٦).

ومن ثم فإن مصطلح شدة الصوت (Sound intensity) لا يرادف مصطلح ضغط الصوت (Sound Pressure)، وإن كان هناك ترابط بينهما مرجعه أن الصوت المسموع يتكون من موجات ضغط، وملعون أن إحدى الطرق المتتبعة في تقدير شدة الصوت هي تحديد مقدار تباين الضغط الصوتي مقارنة بالضغط الجوي. ويعتبر الداين هو وحدة قياس الضغط. ومقداره $1/1000$ من الجرام، «وضغطة موجة ما هو عبارة عن عدد الداينات التي تضغط على مساحة سنتيمتر مربع واحد (أيوب، ١٩٨٤: ١١٧). ويوضح الشكل الآتي شدة الصوت.



الشكل ٦: منظور توضيح لشدة الأصوات المكونة لكلمة (صوت) بنطق المؤلف

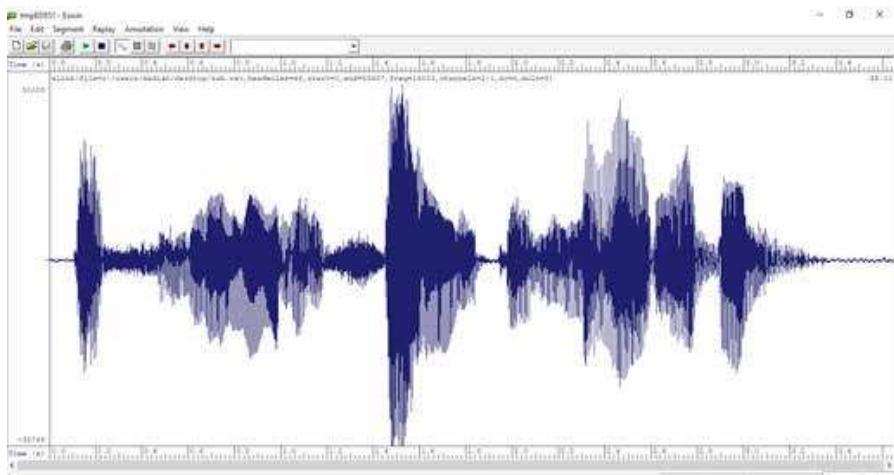
٤، ٣. الرسم التَّدَبْدِيُّ (شكل الموجة) (Waveform):

الصوت عبارة عن: «اَصْطَرَابٌ مَادِيٌّ فِي الْهَوَاءِ يَتَمَثَّلُ فِي قُوَّةٍ أَوْ ضَعْفٍ سَرِيعَيْنِ لِلضَّغْطِ الْمَتَحَرِّكِ مِنَ الْمُصْدَرِ فِي اِتِّجَاهِ الْخَارِجِ، ثُمَّ فِي ضَعْفٍ تَدْرِيجِيٍّ إِلَى نُقْطَةِ الزَّوَالِ النَّهَائِيِّ». (الموسوى، ١٩٩٨: ٩٨) و (هلال، ١٩٨٨: ١٢٤). وعليه فإنَّ التوصيف الفيزيائي للصوت يتمثل في كونه سلسلة لتغيرات الضغط في الوسط بين المرسل والمستقبل. فعندما يبدأ المرسل عملية إصدار الصوت عن طريق إنتاج مجموعة من الذبذبات التي تضغط على الكتلة الهوائية في شكل موجات متلاحقة تصطدم بالمستقبل - ميكروفون أو أذن مثلاً - وهنا تحدث عملية تخلخل مصاحبة لهذا الضغط،

وأكثر الأمور شيوعاً لتمثيل معالم الصوت الحسية ورصدها هو الرسم التذبذبي، الذي يطلق عليه الشكل الموجي. ويمثل المحور الأفقي فيه محور الزّمن وتم قراءته من اليسار إلى اليمين بينما يمثل المنحنى الرأسي قوة هذا الصوت في لحظة معينة، في الوقت الذي تظهر فيه عدد الترددات الذبذبية للحظة الواحدة على هيئة تتابع وتلاحق هذه الخطوط الأساسية لهذا المنحنى.

ويمكن الإشارة إلى الشكل الموجي (Waveform) على أنه مقابل منظور للصوت المنطوق، وعليه فإنه - أي الشكل الموجي - أقرب ما يكون إلى رسم بياني لمحور الجهد موزعاً على محور الزّمن. وقد سبقت الإشارة إلى أنَّ الصوت ينشأ عن أشكال موجية تتعاقب بصورة دورية. وبالتالي فإنَّ الأدوات الإلكترونية والحاوسبة يمكنها رسم هذه الأشكال الموجية الدورية وعرضها على هيئة رسم بياني، يظهرها على شكل أنواع من الخطوط المترجة. وتم قراءة هذا الشكل الموجي من اليسار إلى اليمين عبر محور الزّمن الأفقي، في الوقت الذي يعرض فيه المحور الرأسي - من أعلى إلى أسفل - إشارات السعة الصوتية أو ما يمكن تسميته بالجهد النسبي (Relative Voltage).

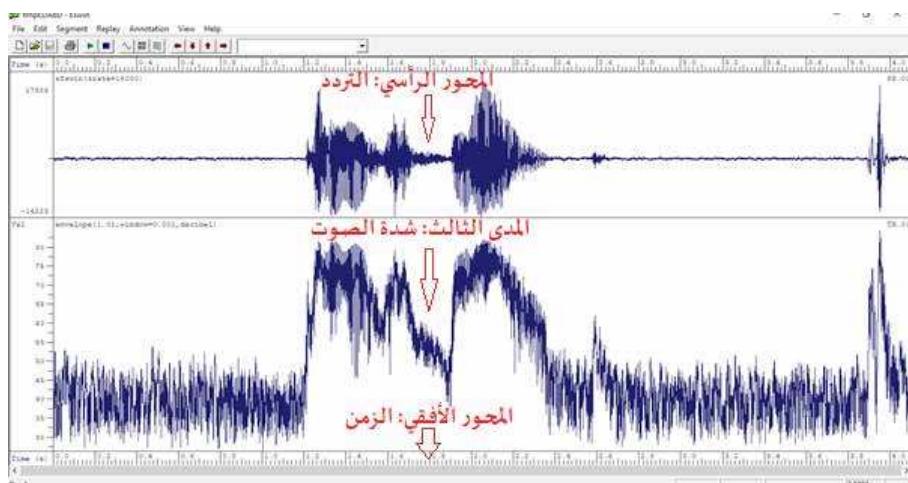
فعندما ننطق أية كلمة (دقة صوتية واحدة) في البداية سيتم التنفس من الرئتين ثم يخرج الهواء إلى الحنجرة الذي قد يسبب عند مروره بها اهتزاز الحبلين الصوتين ويستمر في مسيره حتى يصل إلى القناة الصوتية ويتنقل منها إلى التجويف الفموي، وفي الأصوات الأنفية يضطر الهواء للمرور بالتجويف الأنفي (nasal cavity) وبعدها تخرج الكلمة وتصبح مسموعة وتحول إلى شكل الموجة. ويظهر (الشكل ٧) صورة للرسم التذبذبي من خلال برنامج (SFS).



الشكل ٧: الرسم التَّذبذبِيُّ لجملة (دراسة الصوت اللغوي) كما يعرضه برنامج (SFS)

١٥. الصورة الطيفية (Spectrogram):

الصورة الطيفية للصوت عبارة عن تمثيل مرئي ثلاثي الأبعاد لتردد الإشارات الصوتية وتَغَيِّرِها عبر محور الزمن وشدها، ويمثل المحور الأفقي في الصورة الطيفية منحني الزَّمن باليللي ثانية، ويمثل التردد المحور الرأسي بالميرتز، أما بعد الثالث وهو مدى التردد أو شدة الصوت فِيمَثُله قرب أو بعد لون الصورة عن لون الخلفية. وقد تعتبر الصورة الطيفية عدداً من السلاسل المتوازية إذا نظرنا إليها من أعلى (مسقط أفقي)؛ حيث تُمثَّل الزيادات أو القمم الرأسية في الصور الطيفية بالبقع المظلمة في الصورة الطيفية، وكلتاها تعبر عن شدة الصوت المنطوق.



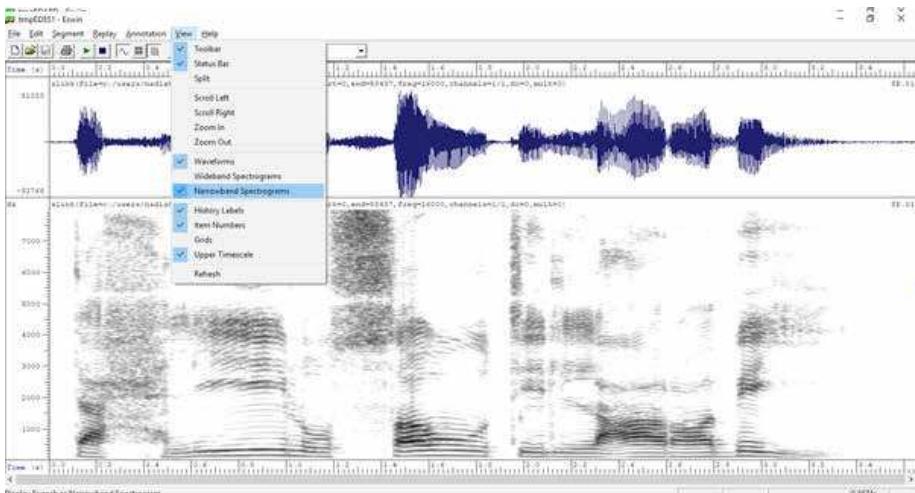
الشكل ٨: مكونات الصورة الطيفية: التردد والشدة والزمن

وأحياناً يطلق على الصورة الطيفية مسميات مماثلة أو مقاربة مثل مسمى المطبوعات الصوتية (voiceprints) أو المخطط الصوتي (voicegrams) أو الصورة الصوتية (sonographs) وهي عبارة عن أداة تحليلية للصوت ينتج عنها رسوم تمثّل الصوت حسب الترددات المكونة له. وكان استخدام الصورة الطيفية للصوت مخصوصاً بادئ الأمر في مجالات الموسيقى، والسونار، والرادار، وقياس الزلازل، وما لبث أنَّ تطور استخدامها في مجال معالجة الكلام (JL Flanagan, 1972: 23)، ثم ذاع استخدامه في مجال إنتاج الكلام وتوليد وتحليله بصورة آلية.

وتسمى الآلة المنتجة للصور الطيفية المطياف البصري (Optical Spectrometer)، ويعتبر المطياف واحداً من أهم الأجهزة التي ساعدت في رصد الخواص الأקוסطية للصوت اللغوي وتحليله؛ حيث يقوم بعرض صورة مرئية للصوت المنطوق كاشفة عن خصائصه، وقد عَوَّل عليه علماء الصوتيات في تحديد صفات الأصوات اللغوية وترجح أوجه الخلاف في وصف بعض الأصوات التي دار حولها الجدل قديماً وحديثاً، «ونظراً لأنَّ موجات الصوت اللغوية من النوع المركب فإنَّ عرضها باستخدام جهاز عارض الذبذبات لا يقدم كل التفاصيل عن الموجة الصوتية. بينما يقدم المطياف ثلاثة أبعاد للموجة المرسومة وهي: التردد، والشدة، والزمن. وهذا يعين الباحث على معرفة زمن الصوت، والتردد الأساسي، والنطاق الرئيسي وشدةه» (العامدي، ٢٠٠٠: ١١٤).

وقد ظلت معظم دراسات أكoustique الصوت اللغوية تعتمد على المطياف بشكل أو بآخر لمدة أربعة عقود حتى وجدت البديل الأمثل في البرامج والأدوات الحاسوبية.

فالبرامج الحاسوبية الآن يمكنها عرض تلك الصور الطيفية بطريقة أكثر دقة وسهولة وفاعلية، بل وتحويل بيانات تلك الصور إلى قيم رقمية ثابتة ومحددة، ومن أمثلة هذه البرامج (Speech filing) (Gold wife) (Cool edit pro)، وبرنامج (System) (البهنساوي، ٢٠٠٤: ١١).



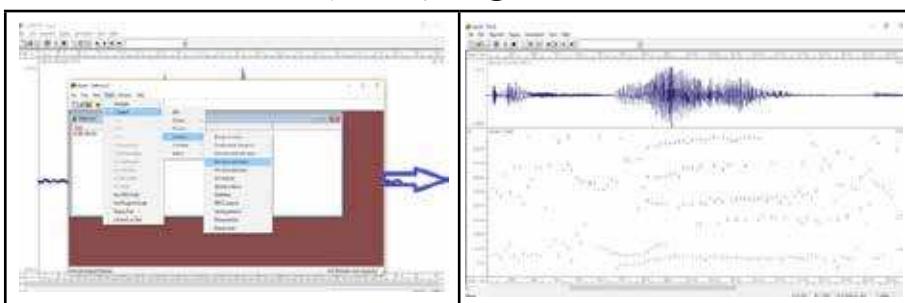
الشكل ٩: كيفية عرض الصورة الطيفية لجملة [تحليل الصوت العربي] من خلال برنامج (SFS)

ونستطيع أن نقرأ مجموعة من الخصائص الأكoustique للأصوات التي تم نطقها ومن ثم عرضها في هذه الصورة الطيفية، ويمكننا ملاحظة اختلاف طبيعة الأصوات المنطوقة بناءً على معطيات الصورة الطيفية، فالصوت الأول في الجملة هو صوت التاء، وهو صوت مهموس انفجاري، ويظهر أمامنا كغيره من الأصوات الانفجارية المهموسة التي تبدأ بفترة صامتة يحتبس فيها النَّفَس، وبعدها بحوالي ٨٢ ميلي ثانية تظهر طاقة الطَّاقة فجأة وبقوة في نطاقات التَّرَدُّد أو الحزم، على هيئة انفجار. وبعدها تظهر الفتحة القصيرة مستغرقة زمناً مقداره ٩٨ ميلي ثانية بمناطق تردد عالية ناتجة عن صفة الجهر التي تنس بها كل الحركات طويلاً وقصيرها بالإضافة إلى بعض الصوامت، ثم يبدأ حرف الحاء ابتداءً من الخط الزمني ١٨٠ ، ٠ ثانية وحتى الخط الزمني ٣٧٠ ، ٠ ثانية،

وهو صوت احتكاكٍ مهموس تكون الطاقة عند النطق به مركزة بكثرة في نطاق التردد أو الحزم، وفوضويٍ إلى حدٍ ما (شبيه بالضوضاء) في مظهره.

أما الأصوات المجهورة مثل اللام والياء والواو والعين مثل فتبدو أكثر تنظيماً، وتبدو أعلى السلسلة (البعض المظلمة) في الواقع على هيئة خطوط في وضع أفقى الشكل عبر الصورة الطيفية. هذه الحزم الصوتية (المعلم formants) هي الترددات أو مجموعة الترددات (Groups of frequencies) التي تشكل نوع الصوت (Tamper) وتميّزه عن الأصوات الأخرى ذات الأنواع المختلفة. وتمثل هذه الحزم الذبذبات (عمر، ١٩٩٧: ١١٦)؛ حيث يعطي شكل الفم رنين الصوت، ويتم ترقيمها من الأسفل إلى الأعلى على هذا النحو: F_1 أو F_2 ، F_3 إلخ؛ وتخالف مواضع هذه الحزم باختلاف نوع الصوت المنطوق، ومن الممكن تحديدها بشكل تقريري مع كل فونيم.

ومن خلال (الشكل ١٠) نرى توزيع المعلم أو الحزم الصوتية لكلمة مفردة.



الشكل ١٠: يوضح طريقة استخراج الحزم الصوتية من خلال برنامج (SFS)؛ بينما يوضح الشكل ١١: أبعاد الحزم الصوتية لكلمة [صوتيات]، التي سُجّلت بصوت المؤلف

٢، ٣. التبادل الأدائي:

تبادل مستويات الكلام البشري بين مختلف اللغات واللهجات التي يمارسها المجتمع الإنساني تعبيراً عن أفكارهم وأغراضهم، وتصف اللغة بشكل عام بخصائصها الديناميكية التي تؤدي إلى تنامي مكوناتها واتساعها بشكل مضطرب لا سيما على مستوى المفردات التي تعد واحدة من أهم لبنات اللغة، وإذا انتقلنا إلى المستوى الصوتي الذي يعد محور حديثنا فسنجد أن ذلك التبادل يصبح أشدّ تعقيداً وأكثر فردية، فلكل فرد في المجتمع سمات صوتية فسيولوجية وأدائية تميّزه عن غيره من ناطقي اللغة، وهذا ما

يجعل البحث في قضايا التحليل الصوتي الحاسوبي أكثر اتساعاً عن غيره من مستويات التحليل اللغوي، لاعتماده على التباين اللغوي الواضح بين الناطقين باللغة، حيث تعتبر دراسة أثر اختلاف الجنس واللكلة والعمر من القضايا المهمة في التعرف على الكلام المنطوق (8: Huang, 2001).

وتتحدد ملامح هذا التباين الصوتي في المدى الزمني (Duration) والنغمة الصوتية (Pitch) والتردد الأساسي (Fundamental Frequency) والأداء الشخصي أو الذاتي (Speaker Variability).

١، ٢، ٣. المدى الزمني (Duration):

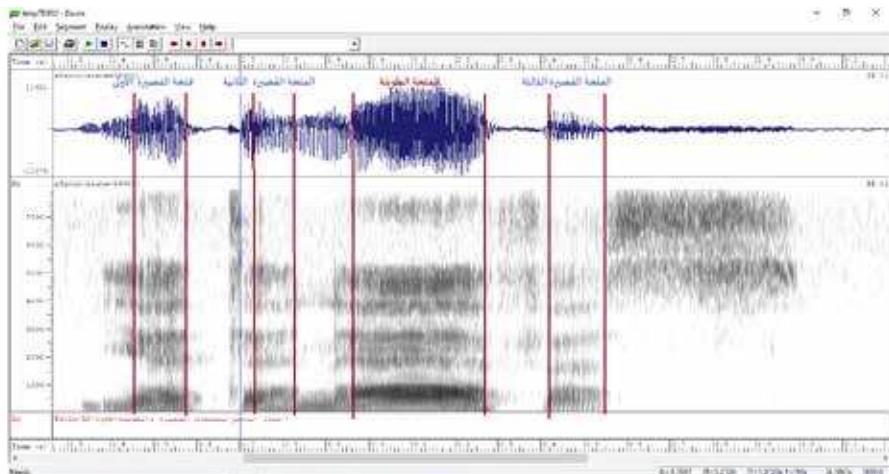
يمثل الطول أو المدى الزمني للفونيم أحد أهم السمات التي تلعب دوراً جوهرياً في التفرقة بين صوت وآخر، ويختلف هذا المدى الزمني طولاً وقصراً وفقاً للسمات الصوتية (Khattab, 2014: 235)، فهناك فارق ملحوظ بين زمن النطق بالأصوات المهموسة عن زمن النطق بالأصوات المجهورة، وكذلك يظهر فارق واضح بين زمن النطق بالأصوات الرخوة عن زمن نطق الأصوات الشديدة، فلكل مجموعة صوتية حيز متقارب من المدى الزمني، وتعد الحركات الطويلة والصوات المضعفة أو (المشدة) والتي يطلق عليها (geminates) من أطول الأصوات في اللغة العربية (Blust, 2013: 17). وملعون أن الحرف المضعف أو المشدد في اللغة العربية عبارة عن مزج بين حرفين متباينين أوهما ساكن والآخر متحرك، ويرمز لهذا التضعيف برأس سين توضع أعلى الحرف المشدد، ومثاها حرف الدال في الكلمة (شَدَّ). وظاهرة التشديد ليست ظاهرة مقصورة على اللغة العربية، بل هي سمة لغوية في لغات عدّة مثل: اللغة العربية والبربرية والمالطية والدنماركية والإستونية والفنلندية والهندية والهنغارية والإيطالية واليابانية واللاتينية والتاميلية. وهاتان الظاهرتان (الحركات الطويلة والتشديد) متراّبطتان في بعض اللغات مثل: الإيطالية والإستونية والنرويجية؛ إلا أنها غير متراّبطتين في لغات مثل العربية واليابانية والسويدية.

وقد أبدى علماء العربية اهتماماً بالغاً بالحركات طويلة وقصيرةها. ذلك أن لها تأثيراً على الدلالة، فقسموها باعتبار زمنها إلى حركات قصيرة وهي الفتحة والضمة والكسرة، وحركات طويلة وهي ما أطلقوا عليه (حروف المد) بأنواعها الثلاثة:

الأَلْف السَّاكِنَة المَفْتُوحُ مَا قَبْلَهَا دَائِمًا (ـا)، وَالوَاوُ السَّاكِنَة المَضْمُومُ مَا قَبْلَهَا (وُو)، وَالْيَاء السَّاكِنَة المَكْسُورُ مَا قَبْلَهَا (يـ) (أحمد، ٢٠٠٤: ٢٨٦)، كما تضم اللغة العربية مجموعة من أنصاف العلل متمثلة في آلياء والواو المفتوح ما قبلهما، بالإضافة إلى أصوات العلة وهي الياء والواو المتحركتان.

وقد ردّ علماء التجويد العرب سبب الاختلاف بين الحركة الطويلة من ناحية وبين الحركة القصيرة والصوامت المشددة من ناحية أخرى، مستندين إلى عاملين أحدهما اتساع المدى الزمني والآخر الامتداد النطقي، «إذا استوفى حرف المد نصيبيه من المدى انتقل بذلك من الحركة إلى الحرف، وهذه الخاصية ثابتة لحرروف المد دون غيرها من الأصوات الجامدة «لا سِيَّا الشَّدِيدَة (أي الانفجارية) فإنها آنية الحدوث، وكذلك الرخوة (الاحتكاكية) فإنها وإن كانت زمانية يمتد بها الصوت مدة، لكن ذلك الامتداد لا يبلغ مقدار ألف، أي مقدار نطق حرف المد» (الحمد، ٢٠٠٧: ٥٣٦).

ويُظهر الشكل التالي الفارق في المدى الزمني بين الفتحة القصيرة والفتحة الطويلة.

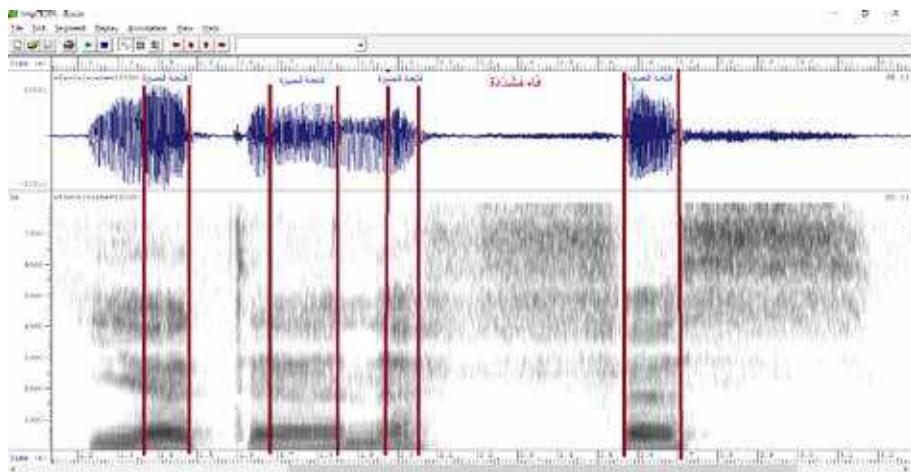


الشكل ١٢: الفارق في المدى الزمني بين الفتحات القصيرة والفتحة الطويلة (ألف المد) في كلمة (يتنافس) بصوت المؤلف

وبالنظر إلى (الشكل ١٢) نجد أن الفتحة الأولى التي أعقبت صوت الياء بدأت من الثانية ٤٧١، وحتى الثانية ٥٧٢، مستغرقة زماناً قدره ١٠١٠ ميللي ثانية. في حين أنَّ الفتحة الثانية التي أعقبت صوت التاء بدأت من الثانية ١٧٠٧، وحتى الثانية

١,٨٠٩ مستغرقة زمناً قدره ١٠١ ، ٠ ميللي ثانية. بينما جاءت الفتحة الطويلة المتمثلة في ألف المد مستغرقة زمناً قدره ٣٥٢ ، ٠ ميللي ثانية، حيث بدأت من الثانية ١,٩٥٢ وحتى الثانية ٢,٢٧٨ . وأخيراً جاءت الفتحة الثالثة التي أعقبت صوت الفاء مبتدئة من الثانية ٤١٣ ، ٢ وحتى الثانية ٥٣٧ ، ٢ مستغرقة زمناً قدره ١٢٣ ، ٠ ميللي ثانية.

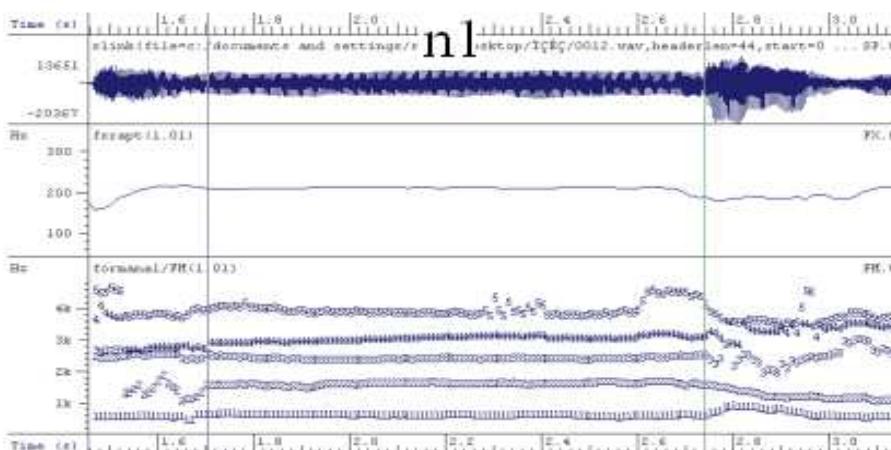
وإذا ما أردنا أن نقارن بين زمن الحركات من ناحية وبين المدى الزمني للحرف المشدد فإننا سنجد فارقاً ملحوظاً على النحو الذي يوضحه الشكل الآتي:



الشكل ١٣: الفارق في المدى الزمني بين الفتحات القصيرة والحرف المشدد في الكلمة (يتنفس) بصوت المؤلف وبالنظر إلى (الشكل ١٣) نجد أن الفتحة الأولى التي أعقبت صوت الياء بدأت من الثانية ١,٣١٤ وحتى الثانية ١,٤٧٧ مستغرقة زمناً قدره ١٦٣ ، ٠ ميللي ثانية. في حين أنَّ الفتحة الثانية التي أعقبت صوت التاء بدأت من الثانية ٦٩٢ ، ١ وحتى الثانية ١,٨٤٤ مستغرقة زمناً قدره ١٥١ ، ٠ ميللي ثانية. ثم بدأت الفتحة الثالثة التي أعقبت حرف النون في الثانية ١,٩٤٧ ، ١ وانتهت عند الخط الزمني ٢,٠٦٨ مستغرقة زمناً قدره ١٢٠ ، ٠ . بينما جاء صوت الفاء المشدد مستغرقة زمناً قدره ٥٢١ ، ٠ ميللي ثانية أي بزيادة ١٦٩ ، ٠ وهو ما يمثل نسبة مئوية تقارب ١٤٨٪، حيث بدأت من الثانية ٢,٠٤٢ وحتى الثانية ٥٦٣ ، ٢ . وأخيراً جاءت الفتحة الثالثة التي أعقبت صوت الفاء مبتدئة من الثانية ٥٦٣ ، ٢ وحتى الثانية ٦٩٢ ، ٢ مستغرقة زمناً قدره ١٣١ ، ٠ ميللي ثانية.

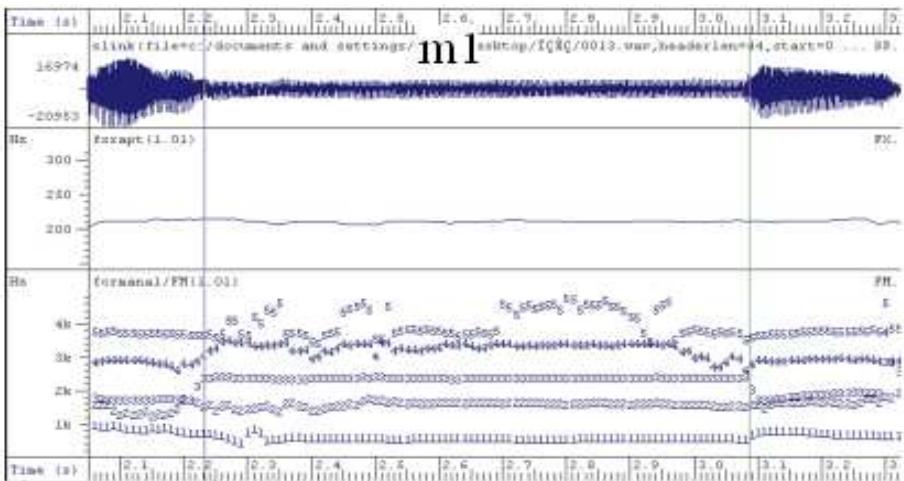
وقد لفت نظري أثناء تحليلي للأمداد النسبيّة للأصوات العربية اتساع المدى الزمني مع الصوامت المشددة بشكل عام، ومع صوتي النون والميم المشددين في الأداء القرآني بشكل خاص، وللتتأكد من ذلك قمت بجمع قاعدة بيانات احتوت على ٢١٥ جملة صوتية متنوعة من القرآن الكريم، اشتملت على سورة الفاتحة، وما تيسر من سورة البقرة والكهف وبعض سور جزء عم، وقد اشتملت قاعدة البيانات هذه على ١٠٥٢٧ فونيهً، ويمكننا حصر نتائجها فيما يتعلّق بالنون والميم المشددين على هذا النحو:

أولاً: النون المشددة



الشكل ١٤: صورة طيفية للنون المشددة، وتظهر فيها ثلاثة مستويات من مستويات التحليل الطيفي، يبدأ من الوسط (wave form) ويتوسّطه منحنى التغيم (Fundamental Frequency) (f1,f2,f3..)، وينتهي بمنحنى الفورمان (f0)

ثانيًا: الميم المشددة:



الشكل ١٥: صورة طيفية للميم المشددة، وتظهر فيها ثلاثة مستويات من مستويات التحليل الطيفي، يبدأ من الـ (wave form) ويتوسطه منحنى التغيم (Fundamental Frequency)، وينتهي بمنحنى الفورمان (f1,f2,f3..).

إذا ما تأملنا تحليلنا لصوتي الميم والنون المشددين فسنلمح أنَّ كليهما صوتٌ احتكاكٌ بيئيٌّ مجهورٌ أَعْنَ مُرَقَّقٍ، وقد وردت النون المشددة في هذه النهاذج ٦٩ مرة، وبلغ متوسط مداها الزمني حوالي ٥٤٢ ميلي ثانية، وهذا المدى الزمني أكثر من ثلاثة أضعاف زمن النون الساكنة، حيث يبلغ متوسط مداها الزمني الطبيعي ١٥٩ ميلي ثانية.

أما الميم المشددة فقد وردت في هذه النهاذج ٥٩ مرة، وبلغ متوسط مداها الزمني حوالي ٥١٠ ميلي ثانية، وهذا المدى الزمني يقترب من ثلاثة أضعاف زمن الميم الساكنة، حيث يبلغ متوسط مداها الزمني ١٦٩ ميلي ثانية.

وتلعب معرفة المدى الزمني للفونيم دورًا مهمًا في فهم الكلام، إلا أنَّ معظم أدوات التعرف على الصوت آليةً تتجاهل هذا الدور، بسبب الاعتماد على أكثر النهاذج شهرة في التعرف على الصوت المنطوق/ وهي نهاذج (Hidden Markov Models) والتي تعرف اختصارًا بـ(HMMs). وهي نهاذج إحصائيةٌ مخفيةٌ تعبر عن أبسط شبكة ديناميكية. وطورها كل من Coworkers L. E. Baum وزميله (L. E. Baum, 2017: 27) توصيف الفونيم أكثر من اعتمادها على الجانب الكمي.

٢، ٣. النغمة الصوتية (Pitch):

التنغيم ظاهرة صوتية عامة، وكثيراً ما تؤدي هذه الظاهرة إلى توجيه دلالة الخطاب دون الحاجة إلى تغيير مفردات الجملة، «ويؤدي التنغيم دوراً فاعلاً في التقرير والتوكيد والتعجب والاستفهام والنفي والإنكار والتهكم والزجر، وغيرها من أنواع الفعل الإنساني كالغضب واليأس والفرح والحزن بوساطة التلوين في الدرجات التنغيمية بمستوياتها العليا والمتوسطة والهابطة، ولذلك عدّها علماء اللغة من الفوئيات غير التركيبية التي من شأنها أن تعرّفنا على مواقف المتكلمين من خلال تنوع ظهورها» (محمد، ٢٠٠٩: ٩).

وقد دأب علماء الأصوات على تناول التنغيم في إطار وصفه بأنه فوئيم غير تركيبي، ويقصدون: الظواهر الصوتية التي لا تتشكل ضمن حيز زمني مستقل، «فالتنغيم ليس «جزءاً من التركيب اللغوي في الجملة، بل هو حدثٌ طارئٌ على التركيب يصاحب، ويتغير نتيجة تغيره في السياق اللغوي الجاري فيه، إذ يربطُ التنغيم عناصر التركيب ببعضها ببعض» (عطية، ٢٠١٥: ٢٣).

فالتنغيم الكلامي بمثابة مقامات موسيقية يتبعها المتحدث لإضافة لمسة فنية على خطابه، فحين ينطق يتكلم الإنسان «لا يتبعُ درجةً صوتيةً واحدةً في النطق بجميع الأصوات، فالأصواتُ التي يتكونُ منها المقطعُ الواحدُ قد تختلفُ في درجة الصوت وكذلك الكلماتُ قد تختلفُ فيها، ومن اللenguاتِ ما يجعلُ لاختلافِ درجة الصوت أهميةً كبرى، إذ تختلفُ فيها معانٍ الكلمات تبعاً لاختلاف درجة الصوت حين النطق بها، ويمكن أن نسمى نظام تواли درجات الصوت بالنغمة الموسيقية» (أنيس، ١٩٦١: ١٧٥). وقد دأب الباحثون على تكسير مستويات التنغيم إلى ثلاثة أضرب:

١. النغمة الصاعدة: وتعني وجود درجةٍ منخفضةٍ في مقطعٍ أو أكثر تليها درجةً أكثر علوّاً منها.

٢. النغمة الهابطة: وتعني وجود درجةٍ عاليةٍ في مقطعٍ أو أكثر تليها درجةً أكثر انخفاضاً.

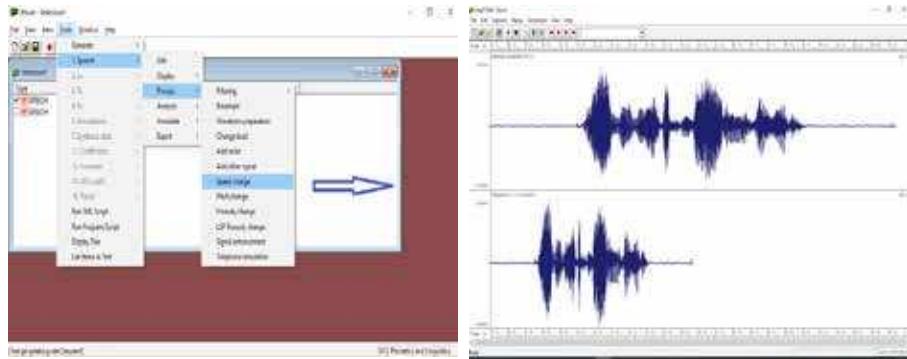
٣. النغمة المستوية: وتعني وجود عددٍ من المقطاعات تكون درجاتها متحدة.

إنَّ الصوت البشري أداة سحرية. فعن طريقه يمكننا تمييز من نعرفهم؛ وتسمح لنا الإمكانيات الصوتية بخلق نغمات غنائية صوتية تؤثر الأسماع؛ كما أنَّ تنوع النغمات الصوتية تمكننا من التعرف على المشاعر الإنسانية المصاحبة للتواصل اللفظي؛ بالإضافة أنَّ تلك النغمات تمثل السمة الأولى لتكوين البصمة الصوتية التي تميز كل فرد عن بقية الأفراد الناطقة بنفس اللغة. حيث يتم إنشاء الصوت البشري الصوت من خلال عدد لا يحصى من حركات العضلات، الأمر الذي يترتب عليه أن يتالف هذا الصوت من العديد من المكونات المختلفة والمترفة (different components)، وهي النغمة (pitch) والتنغيم (tone) والسرعة (rate).

وتمثل النغمة الصوتية أحد أهم مكونات الصوت الإنساني، فهي جزء لا يتجزأ من الصوت، ويتم تعريفها على أنها معدل اهتزازات التدفقات الصوتية، وكلما زاد معدل هذه الذبذبات كلما ارتفعت درجة الصوت، وبالتالي ينشأ عنها صوت حاد، وبالتالي فعندما تكون معدلات هذا الاهتزاز بطيئة فإنه ينشأ عنها أصوات أكثر عمقاً، وهو أمر مرتب بشكل وثيق بمدى سُمك وطول الرقيقين الصوتيتين (vocal cords) من ناحية، وبمدى شد أو جذب العضلات المحيطة بهما من ناحية أخرى (Stevens, ٢٠٠٠: ١٨). وتلعب الرقيقان الصوتيان هنا الدور الأهم في تشكيل الذبذبات الصوتية، وتُعرَفُ الرقيقان الصوتيان أيضاً بالحبال الصوتية، وهي عبارة عن فتيلتين متماثلتين من الأنسجة المكونة من الغشاء المخاطي، وتمتد أفقياً من الخلف إلى الأمام عبر الحنجرة، وتقوم بعملية تعديل تدفق الهواء القادم من الرئتين، في الوقت الذي تهتز فيه هاتان الرقيقان أثناء مرور هذا الهواء المنبعث من الرئة، والتي تلعب دوراً أساسياً في عملية تكوين الأصوات أثناء النطق. و يؤثر حجم الحبال الصوتية على طبيعة الصوت.

وهذا ما يفسر سبب ارتفاع أصوات النساء عموماً عن أصوات الرجال؛ ذلك نظراً لطبيعة تكوين الرقيقين الصوتيتين لديهن والتي تميز بالقصر مقارنة بمشيلاتها عن الرجال، ومع ذلك فليست هذه هي العوامل فقط هي التي تؤثر على النغمة الصوتية؛ حيث تتأثر نغمة الصوت بالعواطف والحالات المزاجية والنفسية. فعندما تتعري الإنسان حالة خوف أو إثارة تقلص العضلات حول الحنجرة (larynx) بشكل لا شعوري أو لا إرادي، ويترتب على ذلك مزيدٌ من الضغط على الرقيقين الصوتيتين، فترتفع النغمة

الصوتية. والأمر لا يقتصر على العامل اللا إرادي؛ بل قد يحدث ذلك بطريقة إرادية في حالة التمرين الصوت والإلقاء وتقليل الأصوات، أو محاكاة بعض الأصوات البشرية أو غير البشرية. والأمر لم يعد يقتصر على العاملين الفسيولوجي والنفسي فقط، بل إن الأدوات الحاسوبية أتاحت لنا تغيير درجات النغم وسرعته، وأصبح بالإمكان تحويل أصوات ذكرية إلى أصوات أنثوية عن طريق البرامج الحاسوبية.



الشكل ١٦: طريقة إدخال تعديلات على سرعة النغمة الصوتية من خلال برنامج (SFS)

٣، ٢، ٣. التردد الأساسي (Fundamental Frequency)

غالباً ما يتم تعريف التردد الأساسي بشكل عام على أنه أدنى تردد في الشكل الموجي (waveform)، وفي الدراسات الصوتية يتم تعريفه بأنه مقياس مدى ارتفاع أو انخفاض تردد صوت الفرد، غالباً ما يرتبط هذا الارتفاع أو الانخفاض بالنغم (pitch) الذي يمثل توتر الحبلين الصوتين (vocal fold vibration) نتيجة اندفاع الهواء من الرئتين مما يؤدي إلى اهتزازات صوتية. وتحتختلف قيم هذا التردد الأساسي بين الذكور والإناث، فعلى سبيل المثال تبلغ قيمه عند الذكور البالغين ما بين ٨٥ و ١٨٠ هيرتز في حين تبلغ قيمه عند الإناث البالغات ما بين ١٦٥ و ٢٥٥ هيرتز . هذا بشكل عام أثناء النطق العادي في الظروف العادية، أما عندما يصاحب هذا النطق تغييرًا خاصًا أو تجويديًا أو تغييرًا فإن هذا النطاق سيكون أكثر اتساعًا بنسبة تتوافق مع درجة النغم. ولا يقتصر التردد الأساسي في التمييز بين أصوات الذكور والإناث بل يمثل في الوقت نفسه أحد العوامل الفردية التي تميز صوت متحدث عن آخر، والذي يمكن عن طريقه إنتاج أصوات يختلف الواحد منها عن الآخر. وحين تسمع أي صوتين يمكن أن تقارنها من

هذه الجوانب المختلفة، مثل شوكه رنانة وأرغن، فهـما يصدران صوتين مختلفين نتيجة عامل أو أكثر» (عمر، ١٩٩٧: ١٧٩).

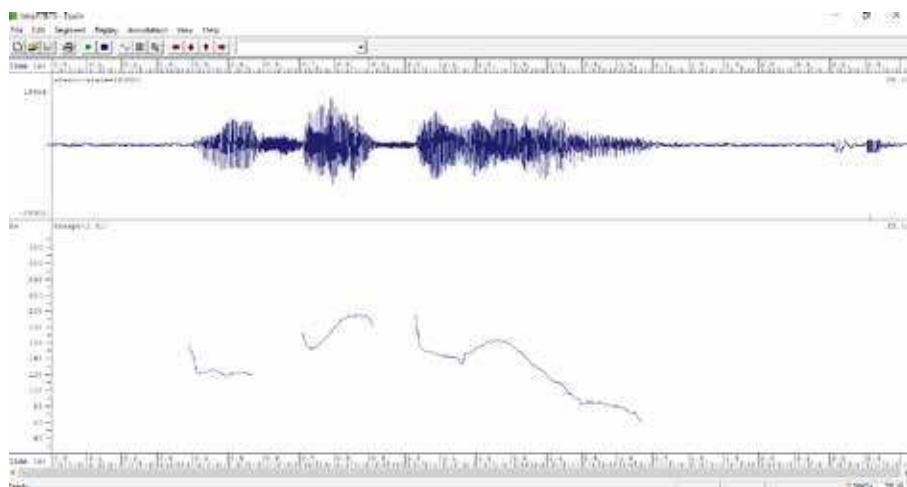
ويعبر شكل الموجة عن قيم شكل الموجة أو التردد الأساسي لهذا الصوت، فالصوت ينظر إليه عادةً على أنه عملية فيزيائية مادية تتكون من جزأين أساسين:

- منتج حقيقي للصوت؛ ويمثله الحبلان الصوتيان.
- منتج أو مشكل فرعي للصوت؛ وهو عبارة عن منتج مساعد، يقتصر دوره على إتمام عملية التوجيه واكتساب الصفات التمييزية، وتمثله باقي أعضاء النطق مثل: اللسان والشفاه والأسنان... إلخ.

وتحاول عملية تحليل التغيم رصد هذا التردد الأساسي للصوت المنطوق. وهناك عدّة قواعد لمعرفة التردد الأساسي تخلص إلى أنه العنصر الأقوى الذي يجعلنا ندرك علاقة التلازم بين الصوت وصاحبـه، فبمجرد أن نسمع جملة نعرف قائلـها بسبـب معرفـنا لـتنـغيـمه. «ويمـكن أن نـميـز صـوتـاً عن صـوتـ آخر بـشـكلـ المـوجـةـ التـيـ تـتـبـعـ كـلـاً مـنـهـاـ، وـلـيـسـ شـكـلـ المـوجـةـ مـقـيـاسـاـ بـالـعـنـيـدـ الـدـقـيقـ، وـلـكـنـهـ مـجـرـدـ وـسـيـلـةـ لـتـمـيـزـ الـأـصـوـاتـ بـتـحـلـيـلـهـاـ إـلـىـ الـمـوجـاتـ التـيـ تـتـكـوـنـ مـنـهـاـ» (أـيـوبـ، ١٩٨٤: ١٨٠).

وتشـهـرـ مـعـالـمـ التـرـددـ الأسـاسـيـ عـلـىـ هـيـةـ منـحـنـيـاتـ أـفـقـيـةـ مـواـزـيـةـ لـخـطـ الزـمـنـ، وـتـكـوـنـ أـكـثـرـ وـضـوـحـاـ مـعـ الـأـصـوـاتـ الـمـجـهـورـةـ؛ حـيـثـ يـهـنـزـ الـحـبـلـانـ الصـوـتـيـانـ، وـتـوـاجـدـ الـقـيـمـ لـهـذـهـ الـحـزـمـ بـيـنـ ٨٠ـ٢٠٠ـ هـرـتـزـ بـالـنـسـبـةـ لـلـمـتـحـدـثـيـنـ الـذـكـورـ، وـتـزـاـيدـ هـذـهـ الـقـيـمـ مـعـ الـإـنـاثـ لـصـيـقـ الـخـنـجـرـةـ عـنـ حـنـجـرـةـ الـذـكـورـ، وـقـلـةـ سـمـكـ الـأـحـبـالـ الصـوـتـيـةـ؛ فـتـبـلـغـ قـيـمـةـ هـذـهـ الـحـزـمـ مـعـ الـأـنـثـىـ الـبـالـغـةـ ٣٥٠ـ١٥٠ـ هـرـتـزـ.

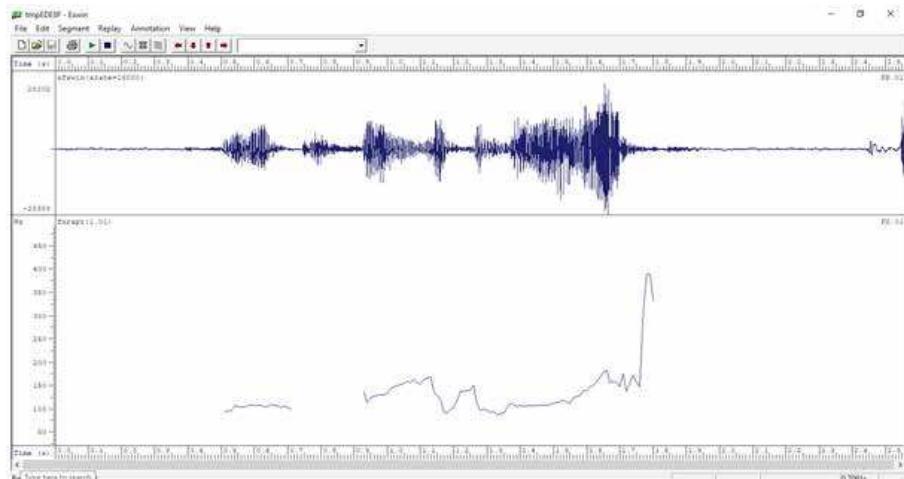
ويـوضـحـ (الـشـكـلـ ١٧ـ) منـحـنـيـاتـ التـرـددـ الأسـاسـيـ لـكـلـمـةـ مـفـرـدةـ.



الشكل ١٧: منحنى التردد الأساسي لكلمة مفردة

في هذه الصورة تتوارد قيم التردد الأساسي بين ١٠٠ و ١٥٠ هرتز، والصورة عبارة عن رسم طيفي لكلمة «مسافرون» بصوت ذكر المؤلف، ونلاحظ أن التردد الأساسي في المقطع الثاني أعلى منه في بقية المقاطع.

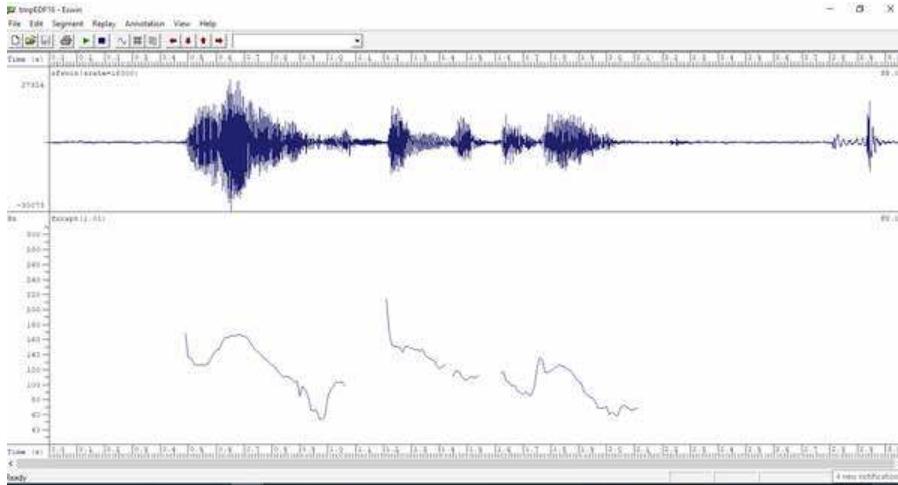
ويعرض الشكل الآتي منحنى التردد الأساسي لجملة استفهامية.



الشكل ١٨: منحنى التردد الأساسي لجملة استفهامية

يعرض (الشكل ١٨) مخططًا للتردد الأساسي لجملة استفهامية تمثل النغمة الصاعدة، حيث نلمح فيها أنها لم تختوِّ على ذلك الانخفاض الذي رأيناه في نهاية منحنى التردد الأساسي في الشكل السابق، والجملة (هل تصدق ذلك؟) بصوت المؤلف.

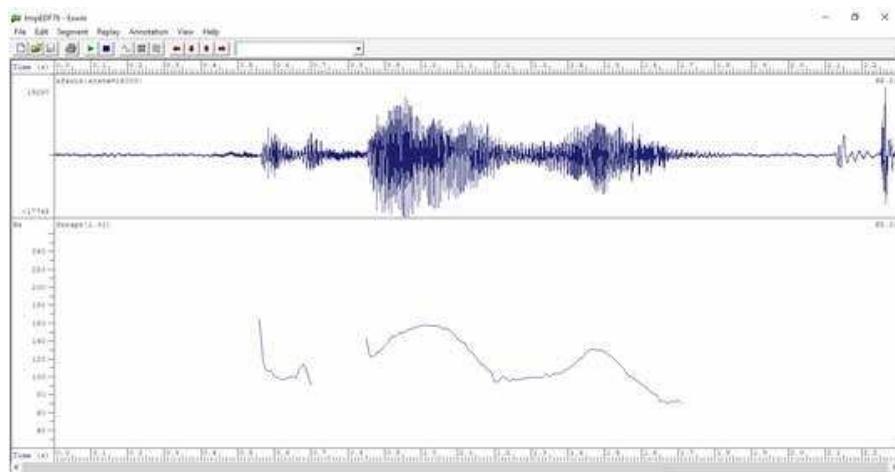
بينما يظهر (الشكل ١٩) منحنى التردد الأساسي لجملة خبرية.



الشكل ١٩: منحنى التردد الأساسي لجملة خبرية

ويعرض الشكل مخططًا للتردد الأساسي لجملة خبرية تمثل النغمة المهاطنة؛ حيث يبدأ فيها التردد الأساسي منخفضًا ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي (نعم أصدق ذلك) وتم تسجيلها بصوت المؤلف.

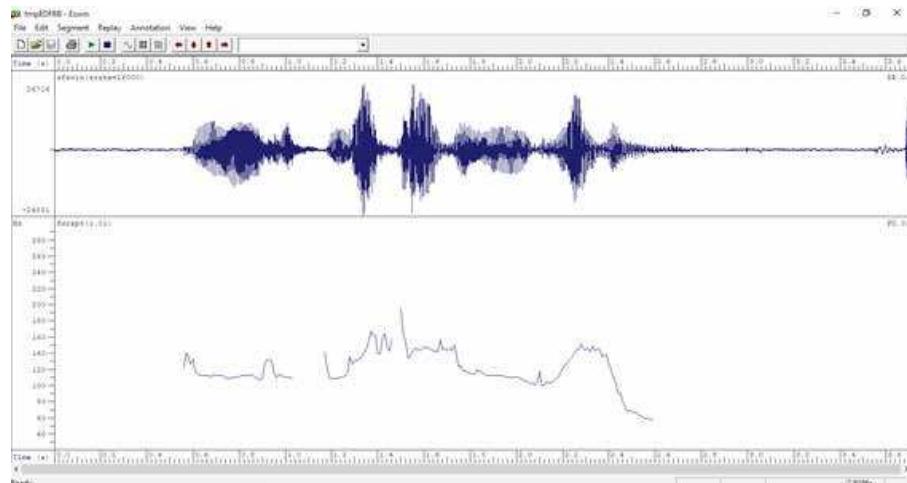
وقد أداَب بعض اللغويين المعاصرين على الزعم بأنَّ هناك شكلاً ثالثاً لمنحنى التنغيم السابقين - الاستفهامي والخبري - وهو المنحنى التنغيمي الخاص بالجملة التعجيبة، إلا أنني قد قمت بجمع قاعدة بيانات صوتية لتسعين جملة تعجيبة منطوقة من ثلاثة أشخاص، فلم أهتدِ إلى أي خلاف ظاهر بين طبيعة نطق هذه الجمل وبين الجمل الخبرية.



الشكل ٢٠: منحنى التردد الأساسي لجملة تعجبية

ويعرض (الشكل ٢٠) خططاً لتردد الأساسي لجملة تعجبية، وقد ظهرت بنغمة هابطة تمايل النغمة الهاابطة الملازمة للجمل الخبرية؛ حيث يبدأ فيها التردد الأساسي منخفضاً ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي (سبحان الله!).

ويعرض الشكل الآتي نموذجاً آخر من الجملة التعجبية المعتمدة على أداة تعجب.



الشكل ٢١: منحنى التردد الأساسي لجملة تعجبية معتمدة على أداة تعجب

يعرض هذا الشكل خططاً للتردد الأساسي لجملة تعجبية معتمدة على أداة تعجب، ولم يظهر فيها أي خلاف عن الجمل الخبيرة؛ حيث يبدأ فيها التردد الأساسي منخفضاً ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي (ما أصدقه من رجل!).

٤، ٢، ٣. الأداء الشخصي (Speaker Variability):

يتمثل الأداء الشخصي (Speaker Variability) في مجموعة من الفوارق الصوتية المرتبطة بأداء المتكلم، مثل: الجنس واللكلة والعمر وسرعة الأداء الكلامي من ناحية، وبوسيلة انتقال هذا الصوت سواء بطريقة مباشرة أو عبر الهاتف العادي أو المحمولة أو عبر الأثير الإذاعي من ناحية أخرى. إنَّ كل عنصر من هذه العناصر يضفي بظاهره على طبيعة هذا الصوت و يؤثر على عملية إدراكه و تمييزه. و تعد هذه النقطة من الأهمية بمكان في عمليات التعرف الآلي على الصوت المنطوق. والتي عادة ما تتطلب الاعتماد على عمليتين متوازيتين من التحليل الإحصائي للمتغيرات المتعددة للصوت، و هما تحليل المكون الرئيسي (PCA) و تحليل المكونات المستقلة (ICA).

و تمثل عملية التباين هذه إحدى الصعوبات الأساسية في تمييز الإشارات الصوتية، حيث إنها شديدة الارتباط ببعضها البعض، و هناك صعوبة أخرى في التعرف على الكلام، تتمثل في صعوبة التعرف على الإشارات الصوتية عند نقلها عبر الهاتف المحمولة أو عبر الأثير الإذاعي، حيث يتوقع حدوث التباس كبير بين بعض الأصوات المتشابهة مثل صوتي اليسن والشين. ولذلك يتم تحليل هذه الإشارات الصوتية ينبغي الاعتماد على مجموعة من الأدوات الإحصائية المتاحة في هذا الصدد، مثل تحليل المكونات الرئيسية (principal component analysis) والتي تعرف اختصاراً بـ (PCA) و تحليل العناصر المستقلة (independent component analysis) والتي تعرف اختصاراً بـ (ICA) (H. Hotellings. 2005: 423) و (Hyvarinen, 2000: 418).

٤ . من التقنيات الصوتية الحاسوبية .

سبقت الإشارة إلى تعريف ابن جني للصوت اللغوي باعتباره «أصواتاً يعبر بها كل قوم عن أغراضهم». فالكلام هو الوسيلة الطبيعية للتواصل بين الجنس البشري. فلا يجد الناس أدنى مشقة تذكر في التواصل الصوتي حيث إنه لا يتطلب أي تدريب خاص. ومع انتشار الحاسوب واعتباره أحد متطلبات الحياة اليومية باتت الحاجة ملحة إلى التعامل معه كما يتعامل بعضاً مع البعض، وأصبح الناس يتطلعون إلى استخدام الكلام كوسيلة اتصال للفيسبوك مع أجهزة الكمبيوتر، بدلاً من استخدام لوحات المفاتيح وأجهزة التأثير، وهو ما أدى إلى تسارع العمل في تقنيات إنتاج الكلام وتحليله والتعرف عليه (R. Arun, 2004: 14).

لقد أدى التطور التكنولوجي والحسوبي إلى إحداث طفرة كبيرة في تقنيات الصوت اللغوي، التي باتت تلعب دوراً أكبر في المجالات الأكاديمية والاقتصادية والسياسية على حد سواء، ومن أبرز هذه التقنيات تقنية تحويل النص العربي المكتوب إلى صوت منطوق (Text To Speech) وتقنية التعرف الآلي على الصوت المنطوق (Speech Recognition) وتقنية البحث الصوتي (Audio Indexer) وتقنيات الترجمة الصوتية (Audio Translation) التي تعتمد على مزج تقنية تحويل النص المكتوب إلى صوت منطوق مع تقنية التعرف الآلي على الصوت المنطوق. وسوف نحاول الوقوف على تقنيتين من هذه التقنيات بمزيد من الإيضاح، هما: تقنية تحويل النص إلى صوت، وتقنية التعرف على الكلام المنطوق.

١ ، ٤ . تقنية تحويل النص العربي المكتوب إلى صوت منطوق (TTS) .

مع الطفرة التي أحدثتها التقنية في عالمنا المعاصر، وتزامناً مع الثورة الإلكترونية التي صبعت كل مجالات الحياة أصبحت تقنيات معالجة الصوت البشري، وعلى رأسها تقنية تحويل النص العربي المكتوب إلى صوت منطوق text-to-speech مطلباً منها ليس فقط لتسهيل عمليات التواصل بين المتحدثين وإنما لأغراض اقتصادية وتجارية بل وتنعداد إلى جوانب سياسية وأيدلولوجية، بالإضافة إلى الأبعاد الأكاديمية والاجتماعية التي تمثل في مساعدة ذوي الاحتياجات الخاصة عن طريق تحويل الكتب والدوريات والجرائد والنصوص المكتوبة بشكل عام إلى منطوقه ليستفيد منها المكفوفون والأسوياء على حد سواء (الغامدي وآخرون، د.ت: ٧).

ويمكن القول: إن تقنية تحويل النص العربي المكتوب إلى صوت منطوق عبارة عن تقنية حاسوبية تهدف إلى قراءة أي نص بصورة آلية، سواء تم تقديمها على شكل نص مكتوب بامتداد (*.pdf) أو (*.doc) أو (*.text) ... إلخ أو تم تقديمها بعد مسحه ضوئياً بامتداد (*.BMP) أو (*.JPEG) أو (*.GIF) إلخ.. عن طريق الاستعانة بنظام التعرف الضوئي على الأحرف (OCR).

وبالتالي فهي تقنية مختلفة تماماً عن أنظمة عرض الكلام المسجل عبر أجهزة الكاسيت أو الفيديو أو الكمبيوتر، حيث إن المدف هنا تخلق الكلام آلياً وليس تخزين كلمات أو جمل محددة ثم عرضها وقت الحاجة.

وتجدر الإشارة هنا أن هذه التقنية تختلف كلياً عن تطبيقات الاستجابة الصوتية التي يتم استخدامها على شكل واسع في محطات القطار أو خدمة العملاء لدى شركات الاتصالات العالمية؛ حيث إن ذلك النوع من آلات التحدث (المحددة سلفاً) التي يطلق عليها أنظمة الاستجابة الصوتية تنتج خطاباً اصطناعياً وليس تخليقياً، وذلك عن طريق سلسة من الكلمات المعزلة (isolated words) أو أجزاء الجمل (parts of sentences). وغالباً ما يتم الاعتماد على هذا النوع من الأنظمة عندما تكون المفردات المقصودة قليلة نسبياً، وغالباً لا تزيد عن ١٠٠ كلمة في سياق محدد سلفاً، أما في سياق تقنيات TTS فإنه من المستحيل تسجيل أو تخزين جميع كلمات اللغة، وبالتالي فمن الأنساب والأيسر تعريف «تقنية تحويل النص إلى كلام» على أنه تخلق أو إنتاج الكلام بواسطة الآلات، عن طريق تحويل هذا النص إلى صوت التلقائي (Thierry, 1997: 13).

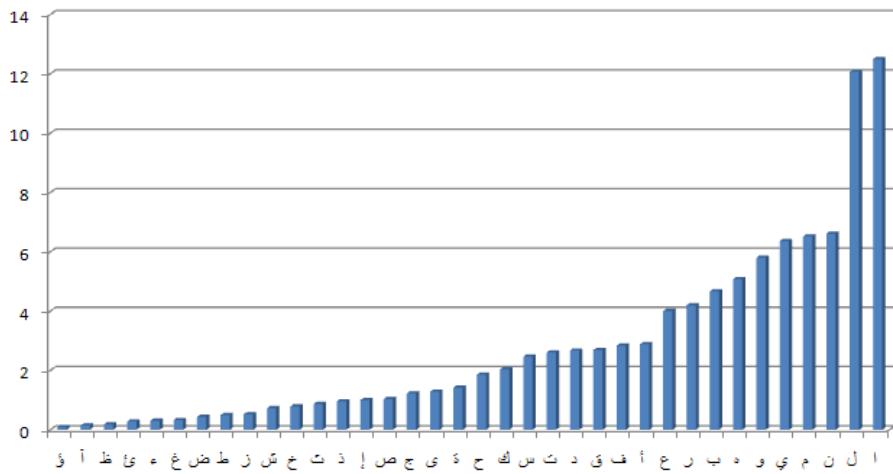
٤، ١. مراحل معالجة النص العربي وموائمه مع صوت منطوق.

أولاً: جمع المادة النصية.

يعد جمع المادة النصية المكتوبة أولى مراحل معالجة النص العربي في مشروع تحويل النص العربي المكتوب إلى صوت منطوق (TTS)، ولا بد أن تكون المادة النصية متوافقة مع الغرض من الناطق الآلي، بمعنى أنه إذا كان المدف من إعداد الناطق الآلي قراءة الأخبار السياسية أو الرياضية أو الدينية فينبغي أن تكون كل قاعدة البيانات النصية مركزة على تلك الأخبار السياسية أو الرياضية أو الدينية، وإذا كان الغرض عاماً بمعنى

أنه ناطق آلي لكل أنواع النصوص فينبعي أن تكون قاعدة البيانات النصية على نفس
القدر من العمومية.

ولا يقتصر الأمر على نوع النصوص فقط، بل يجب مراعاة نسب توزيع الحروف والحركات العربية الواردة في النص؛ لتغطي كافة التراكيب الصوتية، بمعنى أنه يجب أن تتوارد كل صوامت اللغة وحركاتاتها، وفي الوقت نفسه يجب مراعاة أن تكون نسب تردد هذه الصوامت والحركات موزعة توزيعاً عادلاً، وهذا التوزيع العادل لا يقتضي التساوي، فلا يعقل أن يتكرر صوت كثير التردد في النظام الصوتي العربي مثل صوت اللام أو الميم أو النون بنفس درجة تكرار صوت آخر قليل التردد في النظام الصوت العربي مثل صوت الضاد والغين والظاء. والشكل الآتي يوضح نسب توزيع الأصوات العربية في النص القرآني.



الشكل ٢٢: توزيع تردد إحصاءات الحروف العربية الواردة في النص القرآني، مرتبة على حسب تردد الصوت يوضح الشكل السابق نسب ورود الأصوات العربية في القرآن الكريم؛ حيث ورد صوت اللام في القرآن ٦١٨٤٦٥ مرة بنسبة مئوية (١٢,٠٪)، وورد صوت النون ٣٣٨٦٤٦ مرة بنسبة مئوية (٦١,٦٪)، وجاء صوت الميم ٣٣٣٨٥٨ مرة بنسبة مئوية (٥٢,٦٪)، بينما لم يرد صوت الضاد إلا ٢٢٦٤٠ مرة بنسبة مئوية (٤٤,٠٪)، وصوت الغين ١٦٦٩٣ مرة بنسبة مئوية (٣٣,٠٪)، وصوت الظاء ٨٩٥٩ مرة فقط بنسبة مئوية (١٨,٠٪).

ثانيًا: المطابقة بين النص المكتوب والصوت المتوقع نطقه.

ويتم ذلك بتطبيق قاعدة كل ما يقرأ يكتب وما لا يقرأ لا يكتب، وذلك لسد الفجوة بين النظم الكتابي في اللغة العربية وبين النظام الصوتي، وقد أكد منصور الغامدي أن الباحثين العاملين في مجال النطق الآلي والتعرف الآلي على الكلام العربي يواجهون صعوبة إلى حد ما في التعامل مع النص العربي وذلك من حيث تحويل رموزه المكتوبة إلى رموز صوتية. ورغم أن العربية تكاد تكون أقدم لغة حددت أصواتها ودون نظامها وقوانيتها الفونولوجية قبل أكثر من ١٢ قرناً إلا أنه لم توضع هذه الخصائص والقوانين الفونولوجية بشكل يمكن للعاملين في مجال الحوسبة الاستفادة منها بطريقة مباشرة (الغامدي وأخرون، د.ت: ٧)؛ حيث يُلزِمُ منا نظام الكتابة العربي بإدراج حروف لا يتم نطقها وحذف حروف يجب نطقها، ومن أمثلة النوع الأول ما يلي:

- اللام الشسمية أيّنا وجدت، وهي لام التعريف التي تسبق الاسم النكرة في صير معرفة، وتأتي مع ١٤ حرفاً، مجموعة في البيت التالي، وتمثل أوائل الكلمات:

طب ثم صل رحـا تـفـزـ زـرـ ذـانـعـم --- دـعـ سـوـءـ ظـنـ زـرـ شـرـيفـاـ لـلـكـرـمـ

فأيّما كلمة بدأت باللام الشسمية تقوم بحذف هذه اللام من الكتابة الصوتية.

- الألف الفارقة، وهي الألف التي تلحق واو الجماعة في نحو: قاموا ، جلسوا ॥ ذهباوا، لأنَّ هذه الألفات لا تنطق، وإنما وضعت لعلل صرفية.

• الواو في كلمة عمرو لأنها لا تنطق وصلاً أو وقاً

• الواو الأولى في نحو: أولئك، وأولو، وأولى.

ومن أمثلة النوع الثاني ما يلي:

• إضافة واو ثانية في كلمة داود. فتكتب: داود.

• إضافة ألف في بعض أسماء الإشارة نحو: ذلك، وهذا، وهذاذان، وهذين، وهذه، وهؤلاء. فتكتب ذلك، وهذا، وهذاذان، وهاذين، وهاده، وهؤلاء.

• إضافة ألف في الأعلام مثل: الله، واللهم، وإله، والرحمن. فتكتب: الـلـاهـ، والـلـاهـمـ، وإـلـاهـ، والـرـحـمـنـ.

- إضافة همزة متحركة مع ألفات الوصل عند البدء بها، لأن هذا البدء يرتب عليه نطق هذه الهمزة مثل همزة القطع، فمتلاً كلمة اسمع، تكتب هكذا: إِسْمَعْ.

ثالثاً: معالجة الاختصارات الكتابية.

المختصرات الكتابية في اللغة الإنجليزية عبارة عن «صورة ختصرة لكلمة word أو لاسم مركب compound noun أو لعبارة phrase ، تنشأ عن طريق ترك بعض حروف الكلمة أو استخدام أحرف معينة من كل الكلمة، وعلى سبيل المثال فإن ”m“ هي اختصار لكلمة ”meter“ في أحد التغيرات عن الطول، ويقابلها في اللغة العربية الحرف ميم (م) الذي يدل على المعنى نفسه اختصاراً الكلمة (متر)، كأن يقال مثلاً: طول الجدار ٣٠ British Broadcasting Corporation، كذلك فإن المختصر (BBC) هي اختصار للعبارة Corporation، وقريب منها في العربية قولك (ش م م) اختصاراً القولوك شركة مساهمة مصرية، أو (س ح م) اختصاراً لعبارة سكك حديد مصر، وعادة ما يشيع في اللغات الحية استخدام بعض الاختصارات بصورة أكثر من الصورة الكاملة لهذا المختصر، ويتفاوت ذلك بين لغة وأخرى كثرة أو قلة (أحمد، ٢٠١٣: ٣٢٦).

ومع انتشار وسائل الإعلام المسموعة المرئية والمكتوبة، وتطور سبل الاتصال عبر الهواتف المحمولة والإنتernet أصبح استعمال المختصرات الكتابية ظاهرة تحتاج من القائمين على مشروعات تحويل النص العربي إلى صوت منطوق أن يولوا هذه الظاهرة الاهتمام اللائق بها؛ ومن ثم جمعها ثم إدراجها في أنظمتهم الناطقة.

ولا يخفى على كل متخصص أننا نستعمل في كتاباتنا بعض الرموز والمختصرات للدلالة على أمر ما يكوف معروفاً عند من نخاطبهم، مثل: (ص.ب) التي نستعملها للدلالة على صندوق بريد، و (هـ) التي نستخدمها للتعبير عن التاريخ المجري، أو (م) التي نستخدمها للتعبير عن التاريخ الميلادي.

رابعاً: تشكيل النص آلياً.

إنَّ وضع علامات التشكيل يؤدي إلى فك الالتباس الدلالي ووضوح المعنى، فعلى سبيل المثال إذا قمنا بتكوين كلمة ثلاثة من الجذر (ك ت ب)، فمن الممكن أن تكون كَتَبَ أو كُتِبَ أو كُتُبْ، أو كَتْبَ أو كَتَبْ ... إلخ، وقد تقوم السليقة اللغوية بفك هذا

الالتباس تلقائياً عن طريق المعرفة اللغوية الثابتة في العقل الفردي، وإن كانت تلك المقدرة تختلف من قارئ إلى آخر، أما الحاسب الآلي فإنه لا بد من برمجته ليحمل المعرفة اللسانية التي عند المتحدث العربي ليصل إلى التسليمة نفسها، وهي عملية في غاية التعقيد. وقد بذلك مراكز علمية وبحثية عديدة جهداً مضنياً لتحقيق هذه الغاية إلا أن الهدف لم يتحقق بشكل مرض إلى الآن. ومن ثم كانت هناك محاولات لما يعرف بالتشكيلالجزئي diacritization partial؛ وهو تشكيل بعض الحروف وترك غيرها، كتشكيل الحروف المتصلة ببنية الكلمة وترك الحروف في نهايتها التي غالباً ما يكون لها علاقة بالنظام النحوي العربي أو تشكيل الكلمات الأكثر شيوعاً وترك النادر منها (الغامدي وآخرون، د.ت: ٨).

وبدون تشكيل الحرف العربي لا يمكن لنظام نطق آلي معرفة التضعيف من عدمه والتنوين والصوات التي تلي الصوامت، مما يجعل عملية النطق الآلي للنص الغير مشكل عملية مستحيلة (الغامدي وآخرون، د.ت: ٨).

خامسًا: قراءة النص بصوت واحد أو أكثر من القراء أو المذيعين الماهرين.

ونقصد بها مرحلة التسجيل وإعداد قاعدة البيانات الصوتية، وهي أهم مرحلة، حيث يرتبط مدى نجاح المشروع بجودة التسجيل وموافقتها للمواصفات الفنية المعتمدة، ويجب أن نراعي النقاط الآتية في التسجيل:

- لا بد أن تكون النصوص المختارة ذات تنوع فوني يتواءم مع نسب توزيع الأصوات في اللغة الفعلية.
- يتم الإحصاء على مستوى الفوئيم الثلاثي وليس الثنائي أو المفرد.
- جودة تسجيل الصوت يجب ألا تقل عن ٣٢ * ١٦ بت.
- التسجيل بصيغة استريو.
- تتزامن عملية تسجيل الصوت مع تسجيل الذبذبات الحنجرية (EGG).
- يكتفى بتسجيل قاعدة البيانات بصوت مذيع واحد.
- لا تقل المدة الصوتية الخالصة - بعد حذف الفترات الصامتة قبل قراءة كل جمل وبعدها - عن ثلث ساعات.

- تم عملية التسجيل داخل استوديو محكم، وباستخدام جهاز لاقط صوت (مايك) على درجة عالية من الدقة، ومجهز بتقنية استبعاد الضوضاء المصاحبة للثوت، ويتم استخدام بطاقة صوتية معتمدة (Sound card).

سادساً: تشكيل النص صوتيًّا، وفقاً للقراءة الفعلية.

في هذه المرحلة نقوم بإعادة النظر في تشكيل النص بعد قراءته من لدن المذيع أو القارئ الذي ستستخدم صوته في النظام، وهنا ينبغي عدم الاعتماد على القواعد النحوية المعيارية، بمعنى أنه إذا لحن المذيع أو القارئ الذي سبني النظام على صوته في جملة مثل: شرح المدرسين الدرس، فإنَّ هذا الخطأ لن يترتب عليه أية مشكلة في إعداد قاعدة البيانات، لأن الاعتماد هنا على الوحدات الصوتية وليس على القواعد النحوية.

سابعاً: تخليل الكتابة الصوتية (Transcription).

ونقصد بها تحويل النص المكتوب باللغة العربية وفقاً لقواعد الأبجدية الصوتية الدولية (IPA) International Phonetic Alphabet أو وفقاً لمنهج التقييم الفونيمي للأبجدية الصوتية (SAMPA)، وقد تم ابتكار النظام الأول (IPA) منذ أواخر القرن التاسع عشر، واستطاع الباحثون من خلاله ابتكار ألفبائية جديدة للتعبير عن أصوات اللغات المختلفة. وتحاول هذه الألفبائية الحفاظ على قرب العلاقة بين الرموز الكتابية والصوت المنطوق. وكانت المراجعة الأخيرة لها في عام ١٩٨٩ م. ويعمل نظام IPA على إيجاد رمز لكل فونيم في كل لغة بشرية، ووضع علامات تشكيلية للتغييرات التي تعتري أية لغة من اللغات (MacMahon ١٩٩٦: ٨٣٢).

وقد قسمت هذه الجمعية اللغوية الأصوات الإنسانية إلى ستة أقسام، هي:

- رموز لأصوات صامتة رئوية.
- رموز لأصوات صامتة غير رئوية.
- رموز لأصوات صائمة (الحركات).
- رموز لأصوات نغمية.
- رموز لعلامات تمييزية.
- رموز الفوقطعيات.

أما النظام الآخر وهو نظام (SAMPA) فيعد أكثر الأنظمة استخداماً في برامج التحليل الصوتي الحاسوبي، وهو برنامج تم تطويره أواخر ثمانينيات القرن الماضي ليغطي ستة لغات أوروبية ضمن برنامج EEC ESPRIT للبحث والتطوير في مجال تكنولوجيا المعلومات. وفيه تم الاعتماد على معظم الرموز الصوتية التي ابتكرتها IPA؛ ومن ثم إضافة علامات أخرى بدلاً لبعض الأصوات لتكون أكثر ملائمة مع الحروف الحاسوبية، فمثلاً تم استبدال الرمز [ə] بالرمز [ء] الذي تم استخدامه من لدن IPA، ليعبر عن إحدى صوائر اللغة (L. DeMiller, 2000: 125).

وقد زاد من أهمية نظام (سامبا) ما تتمتع به من التوافق وسهولة التعامل مع الحاسوب الآلي؛ بحيث يجد كل رمز صوتي مكاناً له على لوحة المفاتيح بالجهاز، مما سهل استخدامه لدى الباحثين الحاسوبيين. وقد تم إدراج الرموز الصوتية العربية في هذا النظام على النحو المبين في الجدول الآتي:

المجموعة الأولى: الحركات:

الكتابة الصوتية	المثال	الرمز
D'il	ظل	I
X\al	حل	A
umr`?	عمر	U
i:d`?	عيد	:i
ma:l	مال	:a
fu:l	فول	:u

المجموعة الثانية: أنصاف الحركات:

الكتابة الصوتية	المثال	الرمز
wa:hid	واحد	W
Jawm	يوم	J

المجموعة الثالثة: الصوامت:

وتتفرع عنها صوامت انفجارية، وصوامت احتكاكية، وصوامت أنفية، وصامت تكراري، وصامتان جانبيان، على النحو المبين أدناه:

الصوامت الانفجارية:

الكتابة الصوتية	المثال	الرمز
ba:b	باب	B
ˋ?tis	تسع	T
da:r	دار	D
ˋ?t'a:bi	طابع	ˋt
d'arab	ضرب	ˋd
kabi:r	كبير	K
gami:l	جميل	G

الصوامت الاحتكاكية:

الكتابة الصوتية	الكلمة المثال	الرمز
fi:l	فيل	F
nivi:n	نفین	V
Tala:T	ثلاث	T
Dakar	ذكر	D
D`ala:m	ظلام	ˋD
sa?i:d	سعيد	S
zami:l	زميل	Z
s`aGi:r	صغرى	ˋs
Sams	شمس	S
Zami:l	جميل	Z
xit'a:b	خطاب	X
Garb	غرب	G

الكتابه الصوتيه	الكلمة المثال	الرمز
X\ilm	حلم	\X
alam`?	علم	(`)?
?:hawa	هواء	H

الصوامت الأنفية:

الكتابه الصوتيه	المثال	الرمز
ma:l	مال	M
nu:r	نور	N

الصوت التكراري:

الكتابه الصوتيه	الكلمة المثال	الرمز
rima:l	رمال	R

الصوتان الجانبيان:

الكتابه الصوتيه	الكلمة المثال	الرمز
:la	لا	L
al'l`ah?	الله	l`

ولا يعني ذلك بالضرورة أنه يجب على الباحث في المجال الصوتي الحاسوبي أن يقوم بتحويل كل حرف أو حركة إلى رموز الكتابة الصوتية السابقة بطريقة يدوية، بل هناك الكثير من الأدوات الحاسوبية التي تحول النص العربي المُشكّل إلى أحد أنظمة الكتابة الصوتية بطريقة آلية، على النحو الموضح في الشكل الآتي:



الشكل ٢٣: برنامج محول الكتابة الصوتية (RDI Transcribe V1.01 Alpha) الذي ابتكرته الشركة الهندسية لتطوير النظم الرقمية (RDI)

ثامنًا: موائمة الجمل النصية مع التدفقات الصوتية.

حيث نقوم في هذه المرحلة بإعداد مجموعة من الملفات لكل جملة منطقية، وهذه الملفات هي:

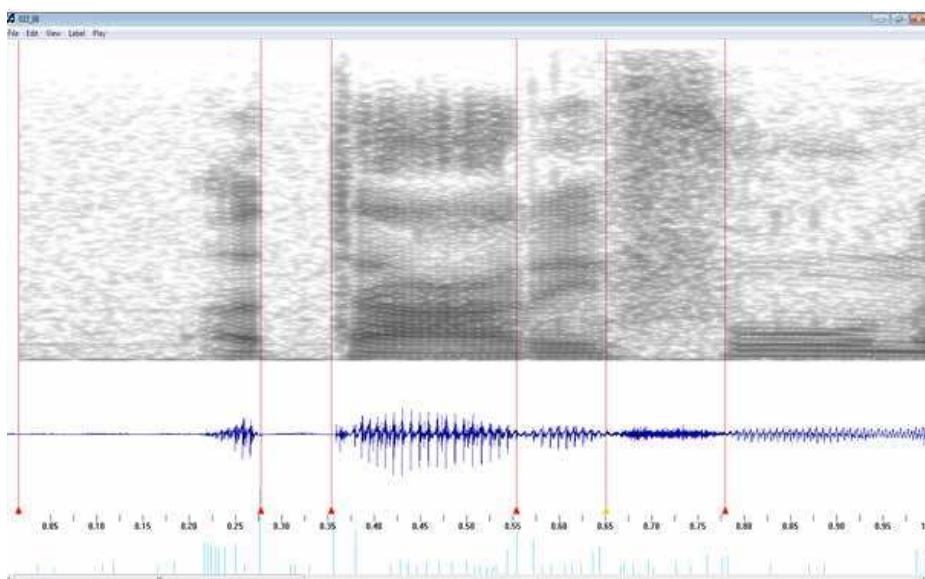
- ملفات الصوت: بامتداد (Wave)، ويمثل المادة الأساسية للمشروع.
- ملفات النصوص: بامتداد (Text) ، ونكتب فيها النص الموافقة للصوت المنطوق.
- ملفات الكتابة الصوتية: (Transcription)، متوافقة مع إحدى الأبجديات الدولية.
- ملفات الأنو: (Ano) لتحليل خصائص كل فونيم (الخصائص الكيفية).
- ملفات الlap: (Lab)، لتحديد زمن كل فونيم (الخصائص الكمية).

تاسعاً: استنباط بداية ونهايات الفونيمات آلياً.

يتم تحديد بدايات ونهايات كل فونيم في كل كلمة من كلمات قاعدة البيانات الصوتية عن طريق تمريرها على تقنية (HMMs) عن طريق تحديد الخواص الأكoustية لكل فونيم، والتي تتمثل في: الجهر أو الهمس، الشدة أو التوسط أو الرخاوة، الاستعلاء أو الاستفال، الإذلاق أو الإصمات ... إلخ.

عاشرًا: تحديد بدايات ونهايات الفونيمات.

ويتم هذا التحديد أولاً بشكل آلي عبر (HMMs) ثم يتم ضبط بدايات ونهايات كل فونيم من لدن خبراء صوتيين، وهناك الكثير من الأدوات والبرامج التي تسهم في إنجاز هذه المهمة، ومنها برنامج التحديد الزمني الآتي.



الشكل ٢٤: برنامج التحديد الزمني (Segment Lab) الذي أنتجته الشركة الهندسية لتطوير النظم الرقمية (RDI) وبذلك تكون قاعدة البيانات صالحة لدمجها في نظام تحويل النص العربي المكتوب إلى صوت منطوق.

٤. تقنية التعرف على الكلام المنطوق (ASR).

تعد تقنية التعرف على الكلام المنطوق واحدة من أهم فروع تخصصات اللسانيات الحاسوبية التي تهدف إلى تطوير المناهج والتقنيات لتكون أكثر قدرة على التعرف على اللغة المنطقية وترجمتها أو تحويلها إلى نصوص إلكترونية (Lee, Chin-Hui, 1991: 232).

كما يعرف هذا التخصص أيضًا باسم التعرف التلقائي على الكلام (automatic speech recognition) ويشار إليه اختصاراً بـ (ASR) أو تقنية تحويل الصوت المنطوق إلى نص مكتوب (speech to text) والتي يشار إليها اختصاراً بـ (STT). إنه نوع من العلوم يمكنه دمج المعرفة والبحوث اللغوية مع علوم الحاسوب داخل إطار مجال الهندسة الكهربائية.

وما يميز تلك التقنية اعتمادها بشكل أساسى على «تدريب النظام» (training) أو (enrollment) عن طريق قاعدة بيانات صوتية وعنونتها بناءً على معطيات التحليل الصوتي الحاسوبى، وغالبًا ما يتم الاعتماد على عدد كبير من المتحدثين للغة أو للهجة الواحدة، شريطة أن تختلف فئاتهم العمرية وخصائصهم الصوتية، فيقرأ كل متحدث منهم بشكل فردي نصاً لغويًا، وربما يتم الاكتفاء بقراءة مجموعة من المفردات المعزولة في السياق، وذلك وفقاً للغرض من إنشاء النظام سواء أكان الغرض طرح النظام لسرقة كبيرة من المتحدثين أو قصره على فرد واحد حيث يكتفى النظام بتحليل صوت هذا الشخص ويستخدمه من أجل التعرف على الصوت نفسه لاحقاً، ويسمى هذا النظام بالدقة الشديدة، وتسمى هذه الطريقة التابعة للمتحدث "speaker dependent" في حين تسمى الطريقة غير المعتمدة على المتحدث (Lee, Chin-Hui, 1991: 232) و (Martin, E. A. 1987: 778).

وهناك الكثير من التطبيقات المتاحة التي تقدم خدماتها اعتماداً على تقنية التعرف الآلي على الصوت المنطوق، ومنها على سبيل المثال التطبيقات المثبتة في الهواتف النقالة أو المحمولات الإلكترونية بشكيل عام، حيث تتيح هذه التطبيقات إعطاء مجموعة من الأوامر الصوتية لفتح برنامج ما على الجهاز أو الاتصال بشخص من قائمة المراسلات أو إرسال بطاقة معايدة أو تهنئة أو ما إلى ذلك.

ومنها أيضاً ما يتم تركيبه في أجهزة التحكم عن بعد، وعن طريقها يمكنك مثلاً فتح جهاز التكييف وضبط درجة حرارته عن طريق الأوامر الصوتية، ومن ذلك أيضاً التحكم في توجيه الكراسي المتحركة التي يستخدمها ذوو الاحتياجات الخاصة .(P. Nguyen 2010: 127)

كما يتم استخدامها بشكل أوسع في الكتب الإلكترونية التي يمكنك من خلالها إعطاء أوامر محددة للتطبيق فيستجيب لها كتقليل صفحات الكتاب أو قراءة فصل من فصوله، أو التحكم في حجم الخط ونوعه ولونه إلخ.

والأهم من ذلك كله استخدام تلك التقنية وإدارتها في تقنية مشابهة لها، وهي البحث الصوتي الآلي (Speech Audio Indexer) التي تتيح للمتصفح العثور على مواضع معينة من ملفات صوتية مطولة (Lori Lamel, 2008: 8).

ببليوجرافيا مرجعية.

١. ابن جني، أبو الفتح عثمان. (١٩١٣)، *الخصائص*، تحقيق محمد على التجار، دار الكتب المصرية، الطبعة الرابعة، الجزء الأول، الطبعة الأولى.
٢. ابن جني، أبو الفتح عثمان. (١٩٥٤) *سر صناعة الإعراب*، تحقيق مصطفى السقا وآخرون، دار الكتب العلمية، بيروت، لبنان، الجزء الأول، الطبعة الأولى.
٣. أحمد، أحمد راغب. (٢٠٠٢) دور المؤثرات السياقية في تحديد المدى الزمني للفونيم، *مجلة الدراسات اللغوية والأدبية*، السنة الثالثة، العدد الأول.
٤. أحمد، أحمد راغب. (٢٠٠٤)، *فونولوجيا القرآن: دراسة لأحكام التجويد في ضوء علم الأصوات الحديث*، رسالة ماجستير غير منشورة، كلية الآداب جامعة عين شمس.
٥. أحمد، أحمد راغب. (٢٠١٣) *المختصرات الكتابية في اللغة العربية* - دراسة استقرائية تحليلية. *مجلة الدراسات العربية والإسلامية*، دار العلوم جامعة المنيا، العدد ٢٧.
٦. أحمد، راغب أحمد. (٢٠٠٩) *نظام التعرف الآلي على الصوت القرآني*، حفص - دراسة توثيقية اختبارية، المؤتمر الدولي الأول للتعلم والتعليم عن بعد، الرياض.
٧. أنيس، إبراهيم. (١٩٦١). *الأصوات اللغوية*. القاهرة: دار النهضة العربية.
٨. أيوب، عبد الرحمن. (١٩٨٤): *الكلام إنتاجه وتحليله*، مطبوعات جامعة الكويت.
٩. البهنساوي، حسام. (٢٠٠٤)، *علم الأصوات*، مكتبة الثقافة الدينية، القاهرة، مصر، ط١.
١٠. حسان، تمام. (١٩٩٠)، *مناهج البحث في اللغة العربية* ، مكتبة الأنجلو المصرية.
١١. الحمد، غانم قدوري. (٢٠٠٧)، *الدراسات الصوتية عند علماء التجويد*، دار عمار للنشر والتوزيع، عمان، الطبعة الثانية.

١٢. حيداني، عيسى واضح. (٢٠١٠). *الصوت اللغوي دراسة وظيفية تشريحية*، دار غيداء للنشر والتوزيع، الطبعة الأولى.

١٣. العطية، خليل إبراهيم، (١٩٨٣)، في البحث الصوتي عند العرب، منشورات دار الجاحظ للنشر، بغداد، الطبعة الأولى.

١٤. عطية، سليمان. (٢٠١٥)، *الفنونيات فوق التركيبية في القرآن الكريم (المقطع النبر - التغيم)* الأكاديمية الحديثة لكتاب الجامعي.

١٥. عمر، أحمد مختار. (١٩٩٧)، دراسة الصوت اللغوي، القاهرة، عالم الكتب.

١٦. الغامدي، منصور بن محمد. (٢٠٠٠)، *الصوتيات العربية*، مكتبة التوبية، الطبعة الأولى.

١٧. الغامدي، منصور بن محمد. (د. ت)، *حسني المحتسب، مصطفى الشافعي. قوانين الفونولوجيا العربية*، منشورات مدينة الملك عبد العزيز للعلوم والتقنية.

١٨. محمد، سهام. (٢٠٠٩) *الفنونيات غير التركيبية (الثانوية): التغيم أنموذجاً*، بحث غير منشور، الجامعة الأردنية، كلية الآداب.

١٩. الموسوي، مناف. (١٩٩٨)، *علم الأصوات اللغوية*، عالم الكتب، بيروت، الطبعة الأولى.

٢٠. هلال، عبد الغفار حامد. (١٩٨٨)، *أصوات اللغة العربية*، مكتبة الأنجلو المصرية، ط. ٢.

٢١. الوعر، مازن. (١٩٨٩)، دراسات لسانية تطبيقية، دار طلاس، دمشق، الطبعة الأولى.

1. Bannon, Mike; Kaputa, Frank. “*The Newton–Laplace Equation and Speed of Sound*”. Thermal Jackets. Retrieved 30 May 2019.
2. Lee, Chin-Hui, Rabiner, R. Lawrence and Pieraccini, Roberto (1991) ‘*Speaker Independent Continuous Speech Recognition Using Continuous Density Hidden Markov Models*’. NATO ASI Series, Vol. F75.

3. Lori Lamel, Jean-Luc Gauvain, *Speech Processing for Audio Indexing, International Conference on Natural Language Processing*, GoTAL 2008, Advances in Natural Language Processing pp 4-15
4. Martin, E. A. (1987), 'A Two-stage Isolated-word Recognition System using Discriminant Analysis', MIT Lincoln Laboratory, Technical Report 773
5. P. Nguyen (2010). *Automatic classification of speaker characteristics*, International Conference on Communications and Electronics 2010, INSPEC Accession Number: 11698809
6. H. Hotellings, "Analysis of a complex of statistical variables into principle components", J. Educ. Psychol., 24, pp.417-441, 498-520, 1933.
7. Huang, C., Chen, T., Li, S.Z., Chang, E., & Zhou, J. (2001). *Analysis of speaker variability*. INTERSPEECH.
8. Hyvarinen and E. Oja, (2000) "Independent component analysis: algorithms and application" Neural Networks 13. pp.411~430.
9. JL Flanagan, *Speech Analysis, Synthesis and Perception*, Springer-Verlag, New York, 1972
10. Jump up to:a b Blust, Robert. (2013). *The Austronesian Languages* (Rev. ed.). Australian National University.
11. Khattab, Ghada & Al-Tamimi, Jalal. (2014). *Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure*. Laboratory Phonology, 5 (2), 231-269.
12. L. DeMiller, Anna & Rettig, James (2000). *Linguistics: A Guide to the Reference Literature* (2nd ed.). Libraries Unlimited. ISBN 1-56308-619-0.
13. L. E. Baum, A *statistical estimation procedure for probabilistic functions of Markov processes*, IDA-CRD Working Paper No. 13

14. MacMahon, Michael K. C. (1996). “*Phonetic Notation*”. In P. T. Daniels and W. Bright (eds.). *The World’s Writing Systems*. New York: Oxford University Press. pp. 821–846. ISBN 0-19-507993-0.
15. Mr. R. Arun Thilak & Mrs. R. Madharaci (2004), “*Speech Recognizer for Tamil Language*”, Tamil Internet, Singapore.
16. Stevens, K.N.(2000), *Acoustic Phonetics*, MIT Press, ISBN 0-262-69250-3, 978-0-262-69250-2
17. Thierry Dutoit, *An Introduction to Text-to-Speech Synthesis*, Springer; 1997 edition (April 30, 1997), P 13
18. Uszkoreit, Hans. “*DFKI-LT - What is Language Technology*”. Retrieved 16 November 2018.
19. Wall, Joan (1989). *International Phonetic Alphabet for Singers: A Manual for English and Foreign Language Diction*. Pst. ISBN 1-877761-50-8.
20. <http://hyperphysics.phy-astr.gsu.edu/hbase/Sound/intens.html> , Retrieved 30 May 2019.
21. <http://www.intellaren.com/articles/ar/a-study-of-arabic-letter-frequency-analysis>, Retrieved 30 May 2019.
22. http://www.wataonline.net/site/modules/newbb/viewtopic.php?topic_id=1777, Retrieved 30 May 2019.
23. <https://www.cooleditpro.com.> , Retrieved 30 May 2019.
24. <https://www.goldwave.com/goldwave.php.>, Retrieved 30 May 2019.
25. <https://www.grc.nasa.gov/WWW/BGH/sound.html>, Retrieved 30 May 2019.
26. <https://www.phon.ucl.ac.uk/resource/sfs/> , Retrieved 30 May 2019.

الفصل الرابع

الذكاء الاصطناعي وتعليم اللغة العربية نحو منصة تعليمية متكاملة

د. نعيم محمد عبد الغني

ملخص

هذا الفصل ليس تأصيلاً وتاريخاً لعلاقة الذكاء الاصناعي باللغة العربية، وإنما هو مشروعٌ طموحٌ لمنصةٍ علميةٍ تتكاملُ فيها التقنياتُ التي تعالجُ اللغاتِ الطبيعيةَ في منصةٍ واحدةٍ تناهٍ للجميع على اختلافِ مستوياتِهم وأعمرِهم. وقد سلطتُ في هذه الدراسةِ الضوءَ على أهمِ التقنياتِ ذاتِ التطبيقاتِ المتعددة، مع التركيز على البُعد التعليمي.

وقد أبنتُ عن كيفيةِ توظيفِ هذه التطبيقاتِ في تعليمِ العربيةِ، كما أبنتُ عن جوانبِها الإيجابيةِ والسلبيةِ، وكيف يمكنُ تطويرُها؛ لتحقيقَ النتيجةِ المرجوةِ من توظيفِها في بيئةِ التعليمِ.

وانطلقَ الفصلُ بتسلاسلٍ منطقيٍ يُعرِّفُ اللغةَ ويرصدُ حاجةَ الإنسانِ لها، ووسائلِ تعليمِها التقليديةِ وغير التقليديةِ، ثم يرصدُ تعليمَ اللغةِ في مستوياتها: الصوتيةِ والصرفيةِ والتركيبةِ والدلاليةِ، ثم عرضتُ نموذجاً مُصوّراً لمشروعِ منصةٍ تعليميةٍ متكاملةٍ.

الكلمات المفتاحية:

اللغة العربية، الحاسوب، التعرف على الصوت، التعرف على الصورة، تكنولوجيا التعليم.

١. اللُّغَةُ وَالْتَّعْلُمُ.

اللغةُ أكثرُ ما يميّزُ الإنسانَ عن باقي المخلوقاتِ، فالوظيفةُ الأساسيةُ المميّزةُ لعقلِ الإنسان هي مقدرتُه على إنتاجِ أنظمةٍ رَمْزِيَّةٍ والتمكنُ من استعمالِها، وعلى رأسِ هذه الأنظمة الرمزيَّةِ النَّظَامُ اللُّغُوِيُّ المستعملُ في التَّوَاصِلِ وِتَمْثِيلِ المُعْلَمَةِ وِتَخْزِينِ المُعْرِفَةِ وِتَقْلِيلِها.

إنَّ اللُّغَةَ وَعَاءُ حضارةِ الإنسانِ وَفَكْرُهُ، فمن خلاَّلِهَا يُسْتَطِعُ التَّعْبِيرُ عن احْتِياجَاتِهِ، وَمِنْ ثُمَّ فَهُوَ بِشَكْلِ تَلْقَائِيٍّ يَسْعَى لِاِكْتَسَابِ لِغَتِهِ مِنْ مُحِيطِهِ الَّذِي يَعِيشُ فِيهِ بِدَأْيَةِ مِنْ أَمْهُ وَأَسْرِهِ وَانْتِهَاءً بِمَجَمِعِهِ الَّذِي يَتَفَوَّتُ بِضِيقِهِ وَسِعَةِ حَسْبِ سُعِيِّ الإِنْسَانِ وَاحْتِكَاكِهِ بِالْقَنَافِسِ الْأُخْرَى، ذَلِكَ الْاحْتِكَاكُ الَّذِي قَدْ يُدْفِعُهُ لِتَعْلِمُ لُغَةً أُخْرَى غَيْرِ لُغَتِهِ الْأَمِّ.

إِذَا فَالْإِنْسَانُ عِنْدَهُ دَافِعٌ لِتَعْلِمُ اللُّغَةَ؛ كَيْ يُسْتَطِعَ أَنْ يَتَوَاصَلَ مَعَ مُحِيطِهِ الَّذِي يَعِيشُ فِيهِ، غَيْرُ أَنْ هَذَا التَّعْلِمُ فِي حَدَّ ذَاهِبِهِ عَمَلِيَّةٌ مُرْكَبَةٌ وَصَعِبَةٌ، تَسْتَدِعِي مِنَ الْمُعْلَمِ تَوْظِيفَ مَعَارِفَ ذَهْنِيَّةٍ وَآلَيَّاتِ نَفْسِيَّةٍ وَاجْتِمَاعِيَّةٍ. وَهَذَا الْعَبُءُ لَا يَحْسَسُ بِهِ إِلَّا الْمُعْلَمُ النَّظَامِيُّ، أَمَّا مِنْ يُلْعَمُ اللُّغَةَ بِشَكْلِ فَطْرِيِّ كَالْأَمِّ الَّذِي تَعْلُمُ أَبْنَاهَا عَنْ طَرِيقِ الْاحْتِكَاكِ الْمَبَاشِرِ بِهَا فَإِنَّهَا لَا تَضُعُ خَطَّةً مُنْظَمَةً، بَلْ يَكُونُ التَّعْلِيمُ بِشَكْلِ فَطْرِيِّ سَلْسِلٌ وَيَكُونُ مُحَدُودًا بِطَبَيْعَةِ تَفَاعُلِ الْأَمِّ وَالْأُسْرَةِ مَعَ هَذَا الطَّفْلِ، أَمَّا الْمُعْلَمُ النَّظَامِيُّ لِلُّغَةِ فَإِنَّ عَلَيْهِ عَبِيْنَا فِي تَعْلِيمِ اللُّغَةِ؛ لِأَنَّهُ مَطَالِبٌ بِتَوْظِيفِ مَعَارِفِهِ وَمَهَارَاتِهِ لِيَعْلَمَ الطَّلَابُ، بِدَأْيَةً مِنَ التَّخْطِيطِ لِلْمَنْهِجِ، ثُمَّ التَّدْرِيسِ وَالْتَّوَاصِلِ بَيْنِ طَلَابِهِ وَانْتِهَاءً بِالْأَخْتِيَارَاتِ وَالْتَّقِيمِ.

بِهَذَا الْمَدْخُلِ يَتَبَيَّنُ لَنَا أَنَّ الإِنْسَانَ يَمْرُ بِمَرْحَلَتَيْنِ فِي تَعْلِيمِ اللُّغَةِ، الْأَوَّلِيَّةُ هِيَ الْمَرْحَلةُ الْعَشَوَائِيَّةُ الْطَّبِيعِيَّةُ الَّتِي يَكْتُسُ فِيهَا الإِنْسَانُ لِغَتَهُ مِنْ مُحِيطِهِ، وَالْمَرْحَلةُ الثَّانِيَةُ تَمَثَّلُ فِي الْتَّعْلِيمِ النَّظَامِيِّ لِلُّغَةِ.

وَاللُّغَةُ كَأَدَاءٍ اِتِصَالٍ تَتَكَوَّنُ مِنْ مَسْتَوَيَاتٍ عَدَّةٍ: تَبْدِأُ بِالصُّوتِ ثُمَّ الْصَّرْفِ ثُمَّ التَّرْكِيبِ وَأَخِيرًا الدَّلَالَةِ، وَمَتَعَلِّمُ اللُّغَةِ يَنْبَغِي أَنْ يَتَعَلَّمَ هَذِهِ الْمَسْتَوَيَاتِ بِالْتَّوَازِيِّ، فَيَتَعَلَّمُ أَصْوَاتَ الْحُرُوفِ وَيُفَرِّقُ بَيْنَ الْحُرُوكَاتِ وَالصُّوامِاتِ ثُمَّ يُكَوِّنُ ثُرُوَّةً لُغُوِيَّةً مِنَ الْمَفَرَدَاتِ يُسْتَطِعُ أَنْ يُكَوِّنَ مِنْهَا جُمَلًا صَحِيحةً تَرْكِيبًا وَلَا مَعْنَى دَلَالِيَّ يَعْهُمُهُ الْمُخَاطَبُ. وَالْمَحَصَّلَةُ النَّهَايَةُ أَنْ يَكُونَ مُتَعَلِّمُ اللُّغَةِ قَادِرًا عَلَى اسْتِخْدَامِهَا تَحْدُثًا وَكَتَابَةً وَقِرَاءَةً وَسَمَاعًا، وَهَذِهِ الْعَمَلِيَّةُ هِي

الغاية التي يسعى إليها متعلم أية لغة، ومنها اللغة العربية، بعض النظر عمّا تمتاز به العربية من خصوصيات تميّزها عن باقي اللغات، لكن هذه العملية معقّدة؛ حيث تتطلّب بحثاً فيسيولوجيّاً وفيزيائياً لتوضيح طريقة إنتاج الكلام، ثم بحثاً لغوياً في دراسة خصائص هذا الكلام، فتحنّن اللغوين - في واقع الأمر - لا ندرس طريقة إنتاج الكلام، بل ندرس الكلام بعد إنتاجه، فتحلّ خصائصه الصوتية والصرفيّة والنحوية والدلاليّة. والذكاء الاصطناعيُّ أمّا إدّا تحدّد كبير، الأول: في كيفية إنتاج الكلام ومحاكاة العقل البشري في ذلك، والثاني في كيفية المحافظة على خصائص الكلام وقواعده في كلّ لغة.

٢. طبيعة إنتاج اللغة.

تتّخذ التعرّيفاتُ المختلفة للغة عدّة مُنظّلاتٍ تصلُّ غالباًها إلى نتيجةٍ واحدةٍ، هي أنّ اللغة أداةً للتواصل الإنساني، فهي أصواتٌ يُعبّرُ بها كُلُّ قومٍ عن أغراضِهم (ابن جنّي، ١/٣٣) أو هي نظامٌ من الرموز الاعتباطية يتمُّ بواسطتها التعارفُ بينَ أفرادِ المجتمع، وتختضُّ هذه الأصواتُ للوصفي من حيث المخارج أو الحركاتُ التي يقومُ بها جهازُ النطق ومن حيث الصّفاتُ والظواهرُ الصوتيةُ المُصاحبةُ لهذا النطق (عمايرة، ١٩٨٧: ١٥).

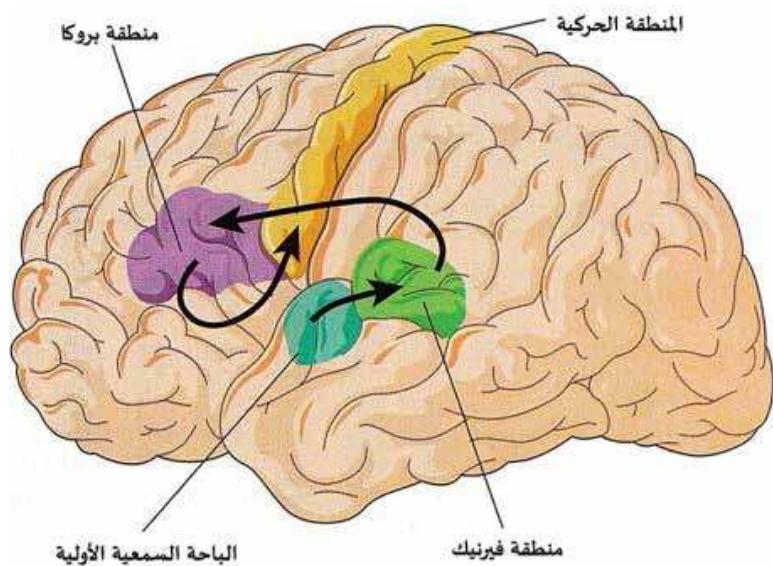
وإذا كانت اللغة أصواتاً أو إشاراتٍ متفارقةً - على حد تعبير دي سوسيير - فهذا يعني أنها رموزٌ ووحداتٌ لغويةٌ تعملُ ضمنَ نظامٍ يقومُ بوظيفةٍ اتصاليةٍ بينَ أفرادِ المجتمع. وهذه الرموز يخزنها العقلُ البشري؛ ليُتّجَّ منها أصواتاً وكلامًا له معنى ودلالةً وفق آليةٍ معقّدة، هذه الآلية تناوَلَها الدارِسُونَ على مُستوييْن، المستوى الأول هو إنتاجُ الكلام في المخ، والمستوى الثاني فيسيولوجية النطق. فأما إنتاجُ الكلام في الدّماغِ فيكونُ بتفعيلِ تَرْمِيزاتٍ لغويةٍ للتحدّث والفهم والقراءة والكتابة، فالمتحدث يبدأ في التفكير في معنى ما يريد أن يتحدث به ثم يختار له المفرداتِ المناسبة، وهذه المفرداتُ لها أصواتٌ وشكلٌ صرفيٌّ، ثم بعدها يركّبُ الجملةَ في رأسه لينطّقها في نهاية المطافِ، هذه العملية تترجمُ إلى حركة الفم والفكين واللسان والحنك والحنجرة، وتنتمي إلى الإنسانِ الطبيعيِّ بشكلٍ سريع. إنَّ آلية إنتاج الكلام واستقباله هي حالةٌ من حالاتِ الإعجازِ في خلقِ البشر، حالةٌ وقفَ علماءُ اللغةِ والطبِّ أمامها كثيراً، فعالمُ النّفسِ فيجوتسكي ذكرَ أنَّ اللغة تتم من

خلال الاتصال الخارجي مع البشر، وترتيب الأفكار في العقل، فهناك نظامان يعملان: نظام داخلي، وآخر خارجي، وكلا النظامين يستخدمان شفرة لغوية واحدة (جوديث جرین، ١٩٩٢: ١٣ - ١٥).

أما تشومسكي فإنه يرى أن إنتاج اللغة يمر من خلال الكفاءة والأداء، فكل إنسان يحتوي دماغه على معرفة لغوية يتم تغذيتها من محيط مجتمعه، وبعد إضافة المدخلات اللغوية يصبح الإنسان قادرًا على التكلم (Chomsky, ٢٠٠٠: ٧٧ - ٧٩).

إن هذه المسألة درسها باستفاضة علم اللغة العصبي (Neuro linguistique)، وله أفرع متعددة يهمنا منها الآن ما يتصل بطريقه إنتاج الكلام، ففي تشريح الدماغ - كما في (الشكل ١) هناك مناطق مسؤولة عن اللغة هي: منطقة بروكا واللُّحاء السمعي واللُّحاء الحركي الرئيسي ومنطقة فيرينك، وعند نُطُقِ آيَةٍ كلمة مقرودة يجب أن تصل المعلوماتُ أولاً إلى القشرة البصرية التي ترسل بدورها المعلومات إلى منطقة الكلام الخلفية بما فيها منطقة فيرينك، ثم تنتقل المعلومات إلى منطقة بروكا، ثم إلى اللُّحاء الرئيسي.

إن هذه المناطق مسؤولة عن السمع والُّنطِق، فهي أدوات الاتصال المباشرة، وأي خلل فيها يُحدث صمماً أو فقدان النطق أو على الأقل اضطراب فيهما (شهداء، ٢٠١٥، ٨٣ - ٨٩).



الشكل ١: مناطق إنتاج الكلام في الدماغ البشري

٣. الذكاء الاصطناعي وإنتاج اللغة.

والسؤال الآن: هل يستطيع الذكاء الاصطناعي محاكاة هذا النظام المعتقد في تركيبه وطريقة أدائه؟ وما الذي توصل إليه حتى الآن؟ وهل يمكن للألة أن تخل محل المعلم في تعليم اللغة؟

والجوابُ نعم؛ لأنَّ بين اللغة والذكاء الاصطناعي عاملًا مشتركًا يتمثلُ في الترميز، فإذا كان العقل البشري يحتاج إلى أنْ تُعرَّفَه على شكل الحروف التي هي رموز في الأساس، فبِالإمكان أنْ ندرب الحاسوب على قدر كبير من هذه الرموز، فالعلاقة بين اللغة والحاوسب تأخذ طابعًا تبادلِيًّا، والالتقاء بين اللغة والحاوسب أمرٌ حتميٌّ لأنَّ الإنسانَ محور النشاط اللُّغويِّ، والإنسانُ هو الذي منحه الله القدرة على تصميم الحاسوبِ وتشغيله، وهذا يمكِّنه من صناعة لغات البرمجة التي تخدم لغته التي يتكلم بها (علي، ١٩٨٨: ١١٣، ١١٥).

إنَّ صناعةَ تكنولوجيا المُعْلَمَاتِ تقومُ على الانطلاقِ من اكتشافِ حاجاتِ البشرية؛ لِتَنْدُرسَ إمكانيةَ تحقيقِها في صورةِ تطبيقاتٍ سهلةِ التناولِ، ومن حاجاتِ البشرية الملحَّةِ الآن ضرورةُ التواصُل اللُّغويِّ، سواءً أكانَ هذا التواصُلُ بينَ أبناءِ اللُّغةِ الواحدةِ، أمَّ عن طريقِ الترجمَةِ بينَ لُغاتٍ متعددةٍ، وفي الأمرين يتَدْخُلُ الذَّكاءُ الاصطناعيُّ الذي ما زالَ يُواجِهُ تحدياتٍ كبيرةً حتى الآن، ويبلغُ هذا التحدي ذروتهُ في الترجمَةِ الآليةِ التي مَا زالتُ حُلْمًا يُراودُ علماءَ التكنولوجيا الذين عَكَفُوا الآن على دراسةِ إمكانيةِ تحقيقِه؛ فبدأوا بدراسةِ خَصَائِصِ اللُّغاتِ في أصواتِها ومفرداتها وتراسِيَّها لِمَعْرِفَةِ المعاني التي ستكونُ مُرْتَجَمَةً بِدَقَّةٍ.

وتم هذه الدراسات عبر خبراء في هندسة التكنولوجيا وخبراء في اللغات؛ حيث شغلوا الآن في حصر الموارد المتاحة في محاولة حل بعض المشكلات التي ستواجههم في الترجمة الآلية؛ رغبة في تواصل البشر في كافة أنحاء العالم بشكل كامل مفروءة ومسموعة، فالكل يرجو أن يطّلع بسهولة على الكتب والصحف المدونةً بغير لغته التي يتكلم بها، ويريد أن يفهم محادثة فورية بجهاز بسيطٍ عندما يتكلم مع أحدٍ من غير أبناء لغته، كما يود أن يوصل ما عنده لآخرين الذين يتكلمون بلسانٍ غير لسانه، والكل

يتساءل: هل سيأتي اليوم الذي أتكلم فيه بالهاتف بلغتي ليترجم كلامي عبر لغة أخرى فيفهمها الأجنبي، ويأتي رده الذي هو بلغة أخرى مترجماً للغتي في سرعة ودقة؟!

وكل هذه الأماني ينظر إليها الباحثون باهتمام فيتطلعون لحلها، فيدفعهم أملهم للعمل، فتكون البداية في تحليل اللغات ومعرفة مشكلاتها والقدر المشترك بين اللغات الذي من الممكن أن يكون قاعدةً لهذه الترجمة؛ ليوفر الجهد والوقت. وحتى لا تكون البداية من فراغ، فقد بدأ الغرب في البحث عن خصائص اللغات ووضعوا مناهج مختلفة للبحث كان لها دورها في صناعة تكنولوجيا عربية؛ فعل سيل المثال عندما بدأ الإنجليز في محاولة صنع تطبيقاتٍ تكنولوجية على اللغة الإنجليزية انطلقو من الصوت لصنع برنامج لتعليم اللغة الإنجليزية للناطقين بغيرها، ويعتمد البرنامج على تقنية التعرّف على الصوت، فقاموا بصنع برنامج يُسمى (Tell me more) الذي يُعلم كيفية نطق اللغة الإنجليزية بصورة صحيحة، وإذا استقام النطق للمتعلم استطاع بسهولة أن يكتسب اللغة بمفرداتها وتركيبيها.

أما الترجمة الآلية فنجحت في جر القائمين في تكنولوجيا المعلومات إلى آفاق بحثية أخرى لا بد من ارتياidiها لخدمة البشرية؛ مما خلق رؤية واضحة إلى حد كبير للخارطة البحثية في عالم التكنولوجيا، عن طريق تحديد الأهداف، ثم تعين المتطلبات الضرورية؛ لتحقيق هذه الأهداف التي سستُخدم في تطبيقات مفيدة للبشرية، ومن يطالع مثلاً موقع منظمة (ELRA) - وهي منظمة تهتم بدراسة اللغات الأوروبية - يرى أنهم وضعوا خطةً بدأت من ٢٠٠٣ وحتى ٢٠١١، وحدّدوا هدفهم النهائي في ٢٠١١ الماضي والمتمثل في الترجمة الآلية للغة الهولندية، وما زالت خطّتهم تسير بنجاح حتى الآن، بعد أن حقّقوا جزءاً كبيراً من هدفهم بالنسبة للغة الهولندية.

٤. الذكاء الاصطناعي واللغة العربية.

إذاً بعد أن تعرفنا على الخارطة البحثية للذكاء الاصطناعي وتعامله مع اللغات بشكل عام، فمن الممكن استثمار هذا النجاح في اللغات الأخرى للتعامل مع اللغة العربية، وخاصةً في مجال التعليم، لكن الأمر ليس سهلاً؛ فهناك معضلات كبرى خاصةً باللغة العربية نفسها، وبطريقة التفكير في إنتاج الكلام واستقباله، فهل يمكن جعل الحاسوب محاكيً للعقل البشري في تعليم اللغة العربية بتعقيداته التي أشرنا إليها؟

إن اللغة العربية تمتاز بخصائص فريدة تساعدُها على برمجتها آلياً، وبشكل يندرُ وجوده في لغاتٍ أخرى، فالانتظام الصوتيُّ في اللغة العربية والعلاقة الوثيقة بين طريقة كتابتها ونطقها يدلُّ على قابلية اللغة العربية للمعالجة الآلية بشكل عام، وتوليد الكلام وتميزه آلياً بصورة خاصة (علي، ١٩٨٨: ١٧٤). واللغة العربية وُصفَت بأَنَّها لغة حبرية؛ لشدة انتظام كثير من خصائصها الصوتية والصرفية والإعرابية، وهذا يؤكِّد قابلية العربية لأن تختلف في البرمجة الآلية من خلال معادلات رياضية (علي، ١٩٨٧: ٧٥٣، ٧٥٤).

وفي العقود الثلاثة الماضية كانت المحاولاتُ جادةً في اللغة العربية؛ بجعلِ الحاسوب مُحاكيًّا للعقل البشريٍّ، من خلال العمل في مستويات اللغة العربية كإطارٍ منهجيٍّ، حيث كانت هناك جهود صوتيةٌ وصرفيةٌ ونحويةٌ ومعجميةٌ ودلاليةٌ، ونسبة النجاح تفاوتت حسبَ طبيعةِ كُلِّ مستوىٍ، فالنجاحُ على مستوى الصوت والصرف كان مُرضيًّا، بينما على مستوى التركيب والدلالة كانت النتائجُ غير مُرضية، وإن أمكن الاستفادة منها؛ وهذا ما سيظهرُ في الصفحاتِ التالية.

٥. الذكاء الاصطناعيٌّ وتعلم النطق.

تعلم النطق من خلال الذكاء الاصطناعيٍّ يكونُ عبر تقنية «التعرف على الصوت»، وهي تقنيةٌ لها تطبيقاتٌ كثيرةٌ، ونلحظُها في الهواتف المحمولة؛ حيثُ القراءةُ الآلية للأسماء والرسائل وغيرها، وهذه التقنيةُ يمكنُ استثمارُها في تعليمِ نطقِ أصواتِ اللغة بشكلٍ عامٍ واللغة العربية بشكلٍ خاصٍ.

من الأمثلة على ذلك تقنية «حفظ» لتعليمِ أحكام التجويد. وتوافقُ فكرةُ هذه التقنية مع بحث لدكتور صلاح حامد؛ حيث استطاع بطريقةٍ رياضيةٍ توظيف التقنيات لمعرفةِ أخطاءِ التلاوة (حامد، ٢٠٠٤)، وساركُز على دراسةِ حامد في هذا المقام؛ لأنَّي كنتُ أحدَ الذين عملوا مع الباحث في هذه التقنية التي تُعلمُ نطقَ أصواتِ العربية، من خلال مرحلتين:

المرحلة الأولى: بناء نموذجٍ معياريًّا لأصواتِ اللغة العربية، ويشملُ هذا النموذج الفوئيات الأساسيةَ من الصوامتِ والحركاتِ والفوئياتِ فوق التركيبيةِ من النبرِ والتنغيم.

أما المرحلة الثانية فإنها تشمل جمع احتمالات الخطأ عند المتعلمين في الشرحية المستهدفة، فإذا كنّا نستهدف الأطفال فإننا نحاول جمع احتمالات الخطأ عندهم وتحليل أسبابه وإعطاء النموذج الأمثل للنطق الصحيح.

وما يحدث بعد هاتين المراحلتين هو أننا ندرب الحاسوب على معيارية النطق الصحيح للفونيم فنعطيه مثلاً أن الباء تتكون من فونيمين مرققين (Ba) فمن أتي بخلاف ذلك النطق يوجهه الحاسوب إلى وجود خطأ، ويمكّنه من التّكّرار بشكل سليم حتى يخرج لنا هذا النطق الصحيح، وإذا قلنا إن الطاء حرفٌ مفخّمٌ فإننا نبني نموذجاً معيارياً صحيحاً للطاء المفخّمة؛ بحيث لا يمكن قبول أي صوتٍ يخالف هذا النموذج المعياري، فهذه مرحلةٌ ببناء النموذج، لكنَّ هذه المرحلة بها عدة مشكلات لم تتمكن التكنولوجيا حتى الآن أن تحلّها، من ذلك مثلاً:

أولاً: الفونيمات الأساسية، يمكن ضبطها في الغالب - خاصة في مجال التفخيم والترقيق - لكن الصوّات أو الحركات تختلف أطوالها حسب أداء المتكلّم، وهذا يتضح جلياً في قراءة القرآن الكريم، فمن الممكن أن تقرأ (بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ) في أزمنة مختلفة تتفاوت في أطوالها حسب أداء بطيء وسريع وكُلُّها صحيحة، والحاسوب يضبطُ أطوال المدود بعملية رياضية بحثة، فإذا افترضنا أن طول الألف في قراءة «الحمد» مثلاً يساوي (X) فإن طوله أيضاً في قراءة «التحقيق» يساوي خمسة أضعاف، وهذا أيضاً يختلفُ في الكلام العادي عند اختلاف اللهجات وتتنوعها.

ثانياً: أما الفونيمات التركيبية، وهي النبر والتنغيم فإنها أيضاً تمثل مشكلة؛ لأنها لا تعتمد على الأداء الصوقي فقط، وإنما تعتمد على تلوين الصوت الذي يختلف بشكلٍ كبيرٍ من شخص لآخر، كما تعتمد على لغة الجسد وما يعتريه من افعالات أثناء الكلام، وهذا ما يُحاول الذكاء الاصطناعي حلّه حتى الآن.

أما المرحلة الثانية فإنها تمثل جمع الأخطاء المحتملة من الشرحية المستهدفة، وهذا الجمع يتم من خلال توزيع جمل ونصوص تم قراءتها من الشرحية التي تضمّ لها التقنية، ثم بعد ذلك يتم تحليلها، فنذكر أن احتمالات الخطأ في الطاء أن تُنطق تاءً مثلاً، ويكثر هذا عند الإناث، أو أن تُنطق مثيرةً بضاد لتكون طاء شديدة، كما في بعض لهجات

الصعيد بجنوب مصر. وهكذا نتتبع اهتمالات الخطأ وأسبابه في العينة التي نُدربُ الحاسوب عليها، وكلما كانت العينة كبيرةً صارت الدقة عاليةً في اكتشاف الأخطاء.

بهذا المفهوم أنتجت شركة (RDI) تطبيقات لتعليم أصوات العربية، ومنها برنامج (حفص) المشار إليه، والذي يمكن اختصار الكلام حوله على النحو التالي:

- يستخدم هذا البرنامج -كما أسلفنا - تقنية (التعرف على الصوت)، وبدأت بتعليم أحكام التجويد من خلال جمل مختارة من القرآن الكريم، تتمثل في الكلمة وكلمتين حتى خمس كلمات أو ست كلمات.
- يبدأ النموذج المعياري في نطق الكلمة حسب القواعد المعروفة بها.
- يستمع المتعلم جيداً، ويحاول ترديد ما سمع فإن استطاع تقليل ما سمع بشكل صحيح؛ لتكون فوئيات المتدرب مطابقةً لفوئيات النموذج، حينها تخرج له رسالة تفيده بأنه على صواب، وإن أخطأ تحدّد له بشكل بصريًّا مكان الخطأ، ثم يستمع إلى رسالة تُفيد أنه أبدل الثاء سيناً مثلاً أو جعل المد سبع حركات والصواب أن يكون سِتاً وهكذا..

إن الأمثلة التي اعتمد عليها برنامج «حفص» مختارةً بعناية شديدة، وهي تطبيقات على دروس التجويد بتصنيفاتها التقليدية في الكتب، ومصحوبةً بشرح للمصطلحات والقواعد الصوتية لنطق القرآن الكريم، حسب مستوى المتعلم. وهذه التقنية كي تصل إلى مستوى يناسب متدربين من كافة المستويات احتاجت لتحليل الأخطاء عند شريحة كبيرة من المجتمع العربي على تنوع هُجَاجَاته واختلاف ثقافاته ومستوياته الاجتماعية وأعمار أفراده، فكان تجربةً رائدةً بدأت بواكيرها في مطلع القرن الحادي والعشرين، ولا يزال التطوير فيها مستمراً.

لقد وفر البرنامج بيئة تعليمية تحاكي المعلم التقليدي، لكن بطبيعة الحال لا غنى عن المعلم التقليدي، وتبقى التقنية وسيلةً مساعدةً تحقق نتائجً جيدةً، وبتجربة حفص على فضليين دراسين من الطلاب أُجريت لهما اختبار تحديد مستوى، ثم كان لفصل منهما تدريس حكم التفخيم والترقيق بالطريقة التقليدية، والفصل الثاني تعلم الحكم من خلال برنامج حفص، ثم أُجري للفضليين اختبار، وكانت النتيجة أن أداء المتعلم عن

طريق التكنولوجيا أكثر نجاعة، وربما يرجع ذلك لأن التكنولوجيا توفر وسائل بصرية وسمعية للمتعلم بخلاف الدارس بشكل تقليدي؛ لأنه يعتمد على مهارات الأستاذ وما يبتكره من وسائل تعليم - غالباً - ما تكون تقليدية.

ويمكن استئثار هذه التقنية وتوسيعها لتعليم أصوات العربية بشكل عام، وخاصة في سن الطفولة، فالتعليم الصحيح للغة يبدأ من الأصوات التي تساعد على الكتابة بشكل صحيح، كما يمكن تطبيق هذه التقنية في تعليم اللغة العربية للناطقين بغيرها، وتعليم ذوي الاحتياجات الخاصة من الذين يعانون مشكلات في النطق، فالوسائل الإثرائية المصاحبة من الأنشطة والأسئلة والمسابقات وغيرها من محفزات التعليم تكون قيمة مُضافةً لهذه التقنية الأساسية في تعليم النطق الصحيح.

إن الموجود من برامج تعليم أصوات اللغة الآن لا تعدو أن تكون حروفاً يتعرف المتعلم أشكالها منفردةً ومتصلةً في كلمات يكررها المتعلم، لكنه لا يتلقى تغذيةً راجعةً لطريقة النطق الصحيح، ومن ثم فتقنية «التعرف على الصوت» تُكمّل هذا النقص وتجعل الآلة أكثر تفاعلاً مع المتعلم، لتنشأ بذلك بيئة أقرب إلى الطبيعية.

إن الذكاء الاصطناعي كأداة مُساعدة في تعليم أصوات اللغة العربية أثبت نجاعة ونجاحاً، فهناك دراسة أُجريت بجامعة اليرموك على طلاب الصف الثاني الإعدادي، وعدهم أربعون طالباً وطالبة، قسموا إلى مجموعتين متكافتين، مجموعة تجريبية وأخرى ضابطة، واستمرت الدراسة ثمانية أسابيع بواقع حصة كل أسبوع، وتعلم الطالب في هذه الفترة سوري الصفات وص، ثم أُجري للمجموعتين اختبار يقيس مستوى الأداء في التلاوة، فتفوقت المجموعة التجريبية التي استخدمت التكنولوجيا السمعية والبصرية على المجموعة الضابطة التي درست بشكل تقليدي (عبد الله، ١٩٩٠: ٦٢، ٦٣).

وهذه النتيجة أشرنا إليها في التجربة التي حاول فيها الدكتور صلاح حامد أن يتعرف فيها على أداء تقنية حفظ في تعليم أحكام التجويد التي تُعد في صلب الدرس الصوقي للغة العربية، والتجربة مبتعدتان في الزمان والمكان والأهداف، لكنهما تثبتان شيئاً واحداً هو جدوى استخدام التكنولوجيا في تعليم النطق الصحيح.

٦. برامج إثراء الثروة اللغوية.

في العموم تكون المعاجم اللغوية هي المصدر الذي يحوي الألفاظ والشواهد التشرية والشعرية، وهذه المعاجم يُلحظ فيها أنها تعتمد على بعضها بشكل كبير، وبها ألفاظ مهجورة غير مُستَعْمَلَةٍ، أو بها ألفاظٌ تصعب على الدارسين. ولا يُمكن أن نطلب من المتعلم أن يحفظَ مُعجمًا كي يتمكن من اللغة، بل إننا نكسيه ثروةً لغويةً من خلال تطبيقِ عملي لمارسةِ اللغة تَحْدِيثًا وكتابَةً وقراءَةً، وفي هذه الحالة يُكَوِّنُ المتعلم مُعجمًا خاصًا به يكتسبُ فيه المفردات بطريقةٍ تراكميَّةً، وهذه هي الطريقة التقليدية في تنمية الثروة اللغوية، ويفيدنا الذكاء الاصطناعي بتحديد أكثر المفردات والأساليب اللغوية انتشارًا، ومن خلال هذا الحصر يُمكِّنُنا بناءً منهج تعليميٍّ تتفاوتُ مستوياته حسبَ المُتَلَقِّي؛ ومن ثَمَّ فإنَّ الذكاء الاصطناعي بإمكانه أنْ يَبْيَنَ مُعجمًا حديثًا لِتَسْتَعِيُّ الألفاظ والأساليب الأكثر استعمالًا في اللغة العربية، وكيفية استخدام هذه الألفاظ والأساليب في سياقاتها المُخْتَلِفة، فالحاسوب يُحدِّدُ جذرَ الكلمة آليًا ووسائلَ اشتراطتها والمواطنَ المُخْتَلِفة التي تُسْتَخَدَمُ فيها الكلمة.

إن هذا المنهج له جذور ترجع إلى العالم الأمريكي إدوارد ثورنديك (١٨٧٤ - ١٩٤٩) الذي سعى إلى إيجاد وسيلة لتعلم اللغة الإنجليزية فألف كتاب «قائمة المفردات للمعلمين» Teacher's Wordbook ونشر عام ١٩٢١م؛ وفيه استخلص أكثر الكلمات استخدامًا وشيوعاً بناءً على مدونة نصية مؤلفة من واحد وأربعين كتاباً أبرزها: الإنجيل والأساطير الأمريكية والصحف والكتب المدرسية. (السعيد، ٢٠١٦).

وبعد هذه المحاولة الجادة كانت هناك محاولات مهمة مثل مشروع معجم كوبيلد الإنجليزي التعليمي (Cobuild English Dictionary)، وهو معجم يعتمد في بنيته الأساسية على مدونة لغوية إلكترونية، وفي هذا المعجم حصرٌ للمفردات والتراكيب الشائعة واستعمالاتها اللغوية، واعتمد المعجم على مدونة من عشرين مليون كلمة جمعت من مصادر مختلفة، وخرج المعجم بعد هذا الجهد الكبير إلى النور عام ١٩٨٧م، ثم أعيدت طباعته في عام ٢٠٠٠م بعد زيادة حجم المدونة إلى مئتي مليون كلمة، (السعيد، ٢٠٠٧: ٦٩ وما بعدها).

وهناك دراسة فتحي يونس عام ١٩٧٤ عن الكلمات الشائعة عند تلاميذ الصفوف الأولى من المرحلة الابتدائية (يونس، ١٩٧٤)، والمفردات الشائعة في اللغة العربية لداود عبده عام ١٩٧٩ م (عبد، ١٩٧٩). ومن الدراسات أيضاً التي اعتمدت رصد المفردات الشائعة دراسة بعنوان «تصميم وتطوير مدونة لغوية للغة العربية المعاصرة a Corpus of contemporary Arabic» للباحثة لطيفة السليطي وهي دراسة ماجستير بجامعة ليدز البريطانية أُنجزت عام ٢٠٠٤ (Al-Suliti, 2004).

وأسأوف قليلاً عند دراسة السليطي التي اعتمدت فيها على مدونة من ٤١٥ مقالة تقارب ٩٠٠ ألف كلمة، وشملت مجالات السياسة والاقتصاد والرياضة والإعلام، وبعد فهرستها وتحليلها استطاعت استخلاص أكثر ألف كلمة مكررة في اللغة العربية، وكان لفظ الجلالة الأكثر ترددًا في هذه المدونة، ثم قامت بإحصاء أكثر ٥٠٠ متلازمة في اللغة العربية وتعني بالمتلازمة أن تستدعي الكلمة كلمةً أخرى بشكل عفوي، مثل مكة المكرمة، المدينة المنورة، وقد أطلق عليها الدكتور تمام حسان مصطلح التوارد وأطلق عليها غيره التصاحبات اللفظية (عبد الغني، ٢٠٠٥: ٣٥).

وبناءً على هذا المفهوم استخرجت لطيفة من مدونتها ٥٠ متلازمة، مثل: القرآن الكريم، العالم الإسلامي، الصحة العالمية... إلخ، وصنفت الباحثة متلازماتها إلى متلازمات ثنائية الكلمة أو ثلاثة أو أكثر من ذلك، فمثلاً من المتلازمات الثلاثية ثاني أكسيد الكربون، ومن المتلازمات الرباعية: حقوق الإنسان في الإسلام.

ولم تكتف السليطي بهذا بل استخرجت من مدونتها إحصاء بأكثر مئة مسكون أو تعير اصطلاحاً؛ ويقصد به التعبير اللغوي الذي يضم أكثر من من وحدة لغوية تستعمل باطراد في اللغة، وهذا المسكون تكون له دلالة ثابتة تختلف عن الدلالة المعجمية للمفردة (عمر، ١٩٨٥: ٢٢، ٢٣) مثل: (العلاقات العامة، على قلب رجل واحد، بعض النظر عن، على سبيل المثال... إلخ)، ومثل هذه المسكونات تشير معرفة الطالب وتجعله قادراً على التحدث بطلاقة وتحمل لديه رشاقة في الإنشاء وإنتاج النص الكتابي.

كما أحصت السليطي نسبَ استخدام الروابط والأدوات في المدونة، فكانت حروف الجر أعلاها ثم بعدها أسماء الإشارة، ثم الأسماء الموصولة.

إن مثل هذا العمل يمكن استثماره في إعداد منهج لإثراء الشروة اللغوية لكل مراحل التعليم، سواء لتعليم العربية لأهلها أو للناطقين بغيرها؛ لأن الاعتماد على المستعمل من الكلام يشعر الطالب بأهمية ما يدرسه، والذكاء الاصطناعي يمكن أن يساعدنا في بناء مدونة نستخرج منها المشهور من الكلام سواء في المفردات والجمل والتركيب، ويساعدنا على تنقية المناهج من الحشو الذي لا يستفيد منه الطالب، وفي هذا المقام هناك عدة دراسات قام بها الدكتور المعز بالله السعيد تمثل بشكل عملي كيفية توظيف المدونات اللغوية في تعليم العربية وتطوير مقرراتها التعليمية. ومن ذلك (مدونة اللغة العربية لأغراض تعليمية) التي تناول فيها ثلاثة مستويات مؤثرة في المهارات اللغوية الأربع (الاستماع والتحدث والقراءة والكتابة). وتمثلت المستويات الثلاثة في: (الحروف، والكلمات، والتركيب)، وخلص الباحث إلى أن الأكثر شيوعاً يكون الخطأ فيه قليلاً فمثلاً تلاميذ المرحلة الابتدائية لا يجدون مشكلة في (ا، ل، ي، و، م) لأنها مألوفة لديهم في كلمات كثيرة تأتي أمامهم، بينما يخطئون في كتابة: (ؤ، ئ، ظ)، لقلة ورودها أمام هؤلاء الطلاب (السعيد، ٢٠١٦: ٩١-٥٧).

إن توظيف مثل هذه المدونات في منصة علمية سيعطي لنا نتائج إيجابية في تعليم اللغة العربية للجميع. وإذا اخترنا من المشهور مسكونات ومتلازمات فإن هذا يمكن استثماره في تعليم اللغة العربية للناطقين بغيرها، فالمسكونات أو التعبيرات الاصطلاحية التي تختلف في دلالتها عن الموجود في المعجم يمكن أن تجعل المتعلم يقارن بين مسكونات لغته ومسكونات العربية، فتنشأ عملية احتكاك أو تفاعل إيجابي، وقد تكون هذه بداية لمعجم مسكونات لغوي يشمل مسكونات اللغة العربية وما يقابلها من مسكونات في اللغات الأخرى.

إن بناء المدونات وصناعة المعاجم بهذه الصورة تحتاج إلى جهد مؤسسي، ولا يمكن للأفراد أن ينهضوا بذلك. لقد حاولت شركة صخر إنجاز معجم حديث للغة العربية احتوى على ٣٦ ألف مدخل بشان وأربعين ألف كلمة مع عشرة آلاف تركيب مصاحب، وذلك مع سماتها النحوية والدلالية بمرادفاتها وأضدادها ومشتقاتها وسياقاتها المختلفة وترجمتها للإنجليزية وإمكانية نطقها بالعربية.

لقد عملت شركة صخر على بناء ذخيرة لغوية هائلة اعتمدت فيها على مدونة لكتابات معاصرة تبدأ من خمسينات القرن الماضي بالإضافة إلى مئة وعشرين صحيفة عربية وضمت المدونة أيضاً البرامج الوثائقية المرئية التي أتيح محتواها الكتبي على الإنترن特، ولم تغفل المدونة مصادر اللغة الأساسية من القرآن والحديث والشعر. وبعد هذا الجمع تأتي مرحلة المعالجة التي تتطلب وجود محلل صرفي يعقبه مشكل آلي ليُبني عليه بعد ذلك الناطق الآلي ثم المترجم الآلي، وهذه خطوات تسلم كل واحدة إلى الأخرى فلا يمكن البدء بالترجمة الآلية مثلاً لأنها مبنية على ما سبقها من منجزات.

لكن هذه التجربة لم تكتمل حتى الآن، فأنماط نصوص اللغة العربية كثيرة جداً، وتحتاج إلى فريق كبير من الباحثين الذين يعملون على تطويرها وتوسيع المدونة ووضع قوانين للعلاقات بين المفردات، وهذا لا يمكن أن تنهض به مؤسسة خاصة، بل يحتاج مؤسسة عامة تدعم رسمياً وشعبياً وتكون لها صفة الاستمرارية والاستدامة.

إن هذه المدونة لا بد أن تعرض على المجتمع وبيدي رأيه فيها وتكون له صلاحية في إبداء الملاحظات التي يمكن الاستفادة منها في تحرير المدونة وتصحيح المعجم، وهناك تطبيق لفت انتباهي في هذا المقام، وهو تطبيق الرديف، الذي يهتم بالبحث عن معنى الكلمة ومفردها وجمعها وما يذكر لها من شواهد شعرية إن وجدت، ويتيح للمستخدم أن يقترح كلمات تضاف إلى قاعدة بياناته، لكن هذا التطبيق بسيط ولا يعتمد على مدونة نصية، بل اعتماده على ما ذُكر في المعجم، ولذلك فهو وإن سهل البحث عن بعض الكلمات فإنه يظل محدوداً. فعدد مترادفاته تزيد قليلاً عن خمسة آلاف مترادف، وبه ما يقرب من تسعه آلاف ضد، وجمع شواهد شعرية للكلمات التي يبحث عنها بالمتراوف أو الضد ما يقرب من ثلاثة وثلاثين ألف كلمة، ورغم هذه المحدودية أجريت عليه أكثر من مليون عملية بحثية، وهذا يؤكّد لنا أن أيّة مبادرة جادة في هذا المجال ستجد لها قراءً ودارسين.



الشكل ٢: تطبيق «الرِّدِيف» لتنمية الثروة اللفظية

ومن التطبيقات المفيدة في مجال إثراء الثروة اللفظية «القاموس العربي Arabic Dictionary»؛ وفكرةه قائمة على التحليل الصرفي وتحويل النص المكتوب إلى صوت منطوق باستخدام تقنية التعرف على الصوت، بحيث إنك إذا أدخلت له نصاً، وأردت أن تعرف معنى الكلمة فيه، فإنه يعرض لك معناها وذرها، كما ينطبقها لك باستخدام تقنية تحويل النص إلى صوت، وهذا إلى حد كبير يفيدنا في مجال التعليم، لكنه ليس مبنياً على مدونات تراعي الفروق الفردية والمعايير التربوية والتعليمية؛ ليظل محسوراً في إطار صلاحية كونه مرجعاً وليس منهجاً، فالمعاجم لا تُقرّر على الطلاب، بل تكون مرجعاً لهم، ونحن بحاجة إلى منهج به نصوصٌ تعالج حاسوبياً صوتياً وصرفياً ونحوياً ودلالياً، ويكون الحاسوب له القدرةُ مثل الإنسان على تعليم مهارات الكتابة والقراءة والتحديث والاستماع.

إن حاجتنا إلى المعجم المعاصر لا تقتصر على رصد الكلمات الشائعة والمُسْتَعْمَلَة، بل تحتاج إلى تطبيقات الذكاء الاصطناعي للحصول على الكلمات من خلال تقنية تحويل النص إلى كلام منطوق، وبهذه الخطوة نلتقي في التعليم مع المرحلة السابقة التي أشرنا إليها في تقنية التعرف على الصوت، فيبدأ المتعلم بربط ما يسمعه ويُكررُه بما يقرؤه من مفردات وألفاظٍ في هذا المعجم، خاصةً وأن في العربية إشكالات عدم التطابق بين المكتوب والمنطوق، مثل (هذا) تُكتب بدون ألفٍ، لكنّها تُنطق بـالـف، ومن

ثم فإننا نحتاج إلى معلومات صوتية عند كل مدخل مُعجمي للتسهيل على الدارسين، وتشمل هذه المعلومات: كيفية النطق مثل عمرو - عمر، و الجنس الكلمة هل هي مذكر أم مؤنث، و عددها من حيث الأفراد والثنائية والجمع، وما يتعلّق بها من معلومات صرفية هل الكلمة تكون اسمًا أم فعلًا أم حرفًا وهل هي اسم فاعل أو مفعول.. إلخ، كما نحتاج في هذا المعجم التعليمي أن نعرّف المتعلم هل هذه الكلمة عربية خالصة أم مُعرَّبة؛ وهكذا نعطي المتعلم في هذا المعجم ذخيرة لغوية تتعلّق بكل ما يتصل بالكلمة بداية من نطقها وانتهاء بها يتعلّق بها من معلومات وأوصاف لغوية.

ولنا أن نتخيل أن هناك وجود منصة لمعجم مدرسي متاح لجميع الطلاب تساعدهم على قراءة الكلمة بشكل صحيح وتدرّبهم على النطق، ثم تعرّفهم بأحوال الكلمة الصرفية وطريقة كتابتها وتحتاج لهم سياقات استعمالها من النصوص من خلال عرض النتائج الأكثر شهرةً من المدونة التي يبني منها هذا المعجم. إن هذه المنصة تجعل التطوير والتحديث في المناهج الدراسية المتعلقة باللغة العربية مستمراً وبشكل منظم يبني الملكات النطقية والقدرات العقلية، وأرى أنه بات فريضة الوقت الذي نعيش فيه.

٧. الذكاء الاصطناعي والكتابة.

تعلم الكتابة يتطلب عدة مهارات، أولها معرفة شكل الحرف ورسمه بالطريقة الصحيحة، والثانية ترجمة النطق السليم إلى كتابة صحيحة، والمستوى المهاري الثالث هو كيفية التأليف بين المفردات في جمل مفيدة. والذكاء الاصطناعي يمكن أن يقدم لنا حلولاً في هذه المستويات الثلاث من خلال تقنية تقنية (OCR) التي تعني باختصار لكلمة (Optical character Recognition) وتعني التعرف الضوئي على الحروف، وهذه التقنية يمكن توظيفها في تعليم الخط والكتابة الصحيحة للحروف بنفس الفكرة المنطقية التي عرضت لتقنية التعرف على الصوت، فيمكننا بناء نموذج خط النسخ بصفته أول خط يمكن تعلمه للأطفال، ثم بناء نموذج خط الرقعة، ويطلب من الطلاب محاكاة هذا الخط على ورق عادي ثم مسح هذا الورق ضوئياً، فيُعرفه الحاسوب مواطن الخط وأبعاد الحروف وكيفية كتابتها على السطر وما إلى ذلك من إرشادات، ويعلمه طريقة كتابة الحرف منفصلاً ثم متصلًا في كلمات.

لقد حفظت تقنية (التعرف الضوئي على الحروف) تقدماً في معرفة الخطوط، لكن بها مشكلات في التعامل مع الخط العربي الذي يتميز بأنه يكتب بشكل متصل بخلاف الإنجليزية التي تكتب حروفها منفصلة، وهذا الاتصال يصعب على التقنية معرفة حروف الكلمة، لكن هذه المشكلة يمكن حلّها بناءً على اختيار أشهر الكلمات التي استخرجناها من **مدونة المعجم** الذي أشرنا إليه في مرحلة بناء معجم معاصر، ثم تدريب التقنية على صورٍ مثالٍ ومعيارٍ للحروف في هذه الكلمات، ثم جمع عينة من كتابات الطلاب واستخراج ملامح الخط في كتابة الحروف، وبشكل رياضي يمكن لهذه التقنية أن تتطور على الكلمات المختارة كبداية لمدونة صغيرة لتعليم الخط تتسع شيئاً فشيئاً كلما زاد استعمال الطلاب لتقنية (التعرف الضوئي على الحروف).

إن اتصال الكلمات ليس المشكلة الوحيدة في تقنية (التعرف الضوئي على الحروف)، لكن هناك مشكلات أخرى تمثل في نقط الحروف أو الإعجام، ونصف حروف العربية منقوط والنصف الآخر ليس منقوطاً، وهذا ليس موجوداً في الإنجليزية، ومن ثم قد تخطئ تقنية (التعرف الضوئي على الحروف) في التفرقة بين الحاء والخاء والجيم بسبب النقطة، والباء والياء والدال والذال، ومعالجة هذه المشكلة تمثل في إعطاء التقنية الكم الأكبر من صور هذه الحروف ساعة اتصالها وانفصالها.

ومن المشكلات التي تواجه تقنية (التعرف الضوئي على الحروف) أيضاً مشكلة التشكيل، فالإنجليزية ليست بها فتحة أو ضمة أو كسرة أو شدة أو سكون، أما العربية فإنها تعتمد بشكل أساسٍ على هذا التشكيل الذي تحتاجه لضبط النطق؛ والتقنية (التعرف الضوئي على الحروف) تخطئ في معرفة هذا التشكيل خاصة مع تعدد الخطوط من رقعة ونسخ وديوانٍ وفارسي.. إلخ، والحل هو الاقتصار على نوع واحد من الخط في عدد محدود من الكلمات كما أسلفنا.

إن وضع قواعد محددة لكتابية الخط ليس أمراً صعباً، وهذه القواعد تمثل في معرفة أشكال الحروف متصلة ومنفصلة وطريقة نقطها وتشكيلها، فمثلاً ندرب الحاسوب على أن الحرف لا تُرسم عليه دائرة السكون إذا جاء في بداية الكلمة، ولا يمكن أن يجتمع حرفان ساكنان في كلمة، وغير ذلك من قواعد الكتابة والتشكيل.

إن هذه التقنية رغم مشكلاتها فإنها تمثل حجر أساس يمكن بناء صرح تعليمي ذكي عليها، لتعليم الخط الحسن الذي نفتقده في كثير من المتعلمين اليوم الذين يعتمدون على الكتابة الإلكترونية، ومعها يقل استعمالهم للقلم، رغم أن الكتابة اليدوية شيء مهم ولا يمكن الاستغناء عنه منها تقدمت التكنولوجيا.

٨. الذكاء الاصطناعي وتعليم الإملاء.

تفيدنا تقنية (التعرف الصوتي على الحروف) أيضاً في تعلم الإملاء حيث تقارن التقنية بين الكلمات المكتوبة في قاعدة البيانات وما يتوجب المتعلم من كلماتٍ تُملأ عليه، ويحدد له الخطأ فيها، هذا على مستوى الخط اليدوي أو أي خط آخر أدخل بشكٍ ضوئي.

إن التدقيق الإملائي في الحاسوب خطأ خطوات متقدمة من خلال تطبيقات المدقق الإملائي، وأشهرها المدقق الإملائي لشركة مايكرو سوفت والمدقق الإملائي لشركة صخر. وفكرة المدقق الإملائي تقوم على إيجاد بدائل واحتمالات لتصحيح الكلمة الخطأ. ودقة هذين المدققين في الكلمات المفردة أعلى منها في الكلمات المركبة، ويتميز مدقق صخر عن مدقق (مايكرو سوفت) بإمكانية التصحيح التلقائي وتقليل عدد الاقتراحات ومراعاة السياق والتعامل مع المصطلحات والتشكيل.

إن محركات البحث المعنية بالتدقيق الإملائي الحالية تعاني من مشكلات كثيرة، فمثلاً هناك تطبيق يُسمى: «اكتب صح ektebsa7»، يُصحح النصوص إملائياً، وبه قائمة من المتلازمات الأسلوبية والأخطاء الشائعة، لكن عند تجربة هذا التطبيق بكتابة جملة قصيرة ظهرت به عدة مشكلات منها:

- أولاً: لا يُفرق المدقق في كثير من الأحيان بين الحروف المُعجمة وغير المُعجمة، مثل الحاء والجيم، والدال والذال، والطاء والظاء.

- ثانياً: لا يستطيع اكتشاف الأخطاء النحوية في الكتابة إذا طال الفصل، فمثلاً همزة (إن) تُكسر في سبعة مواضع، لو جاء المثال: قالت: إن أبي يدعوك، هنا لن يخطئ المدقق، لأن النموذج المدخل في قاعدة بياناته ينص على هذا الشكل، أما لو جاء المثال: قالت ابنة شعيب لموسى: (إن) أبي يدعوك، هنا طال الفصل ولن يتمكن المدقق الإملائي من التصحيح، فلو كتبنا (أن) ما اكتشفها.

إن التدقيق الإملائي يعتمد على فهم النص، وهذا يحتاج إلى خوارزمية رياضية معقدة، ومُدوّنةٌ كبيرةٌ من خلالها نستطيع أن نضع القواعد التي من خلالها يمكننا الحصول على دقة عالية، لكن إن اقتصرنا على تقنية التعرُّف الضوئي من خلال رصد النهاج الصحيحة في الكتابة، ومقارنتها بما يكتبه الطلاب، ثم بدأنا في توسيعة المدونة بناء على ما نجده من أخطاء وما نضعه من قواعد نحدث بها البرنامج، في هذه الحالة سُيُّبني مدقق إملائيٌّ عالي الدقة.

ومن الأمانة أن نذكر أن من يتاحون هذه التطبيقات لا يزعمون أن دقتها عالية، بل يعترفون بأنها لا تزال قيد التطوير وأنها تقتصر على المشهور من الأخطاء فقط.

إننا بهذه التقنيات الثلاث يمكن أن نحقق طفرة في تعليم العربية، فإذا افترضنا إمكانية تحويل الكلمات المكتوبة يدوياً أو الممسوحة ضوئياً إلى نصٌ رقمي، ثم أدخلنا هذه الكلمات على المدقق الإملائي والنحوبي، إننا بذلك نحصل على نتيجة تعليمية هائلة، مع الاعتراف بأنَّ الدقة لن تكون عاليةً في البداية، ولكن كلما كثُرت المادة المدخلة للتقنية زادت دقتها.

إن تقنية المدقق الإملائي لا تكتفي فقط بإعطاء الشكل الصحيح للكلمة إملائياً ونحوياً، بل تعطيه معلوماتٍ عن هذا الخط الذي وقع فيه، فإذا كتب مثلاً (سُيَّلَ)، فإنَّ المدقق يعطيه احتمالات: سُيَّلَ، أم سَأَلَ، ثم إذا وقف على الكلمة الأولى استخرج منها معلوماتٍ لغوية حول الكلمة منها: أنَّ الهمزة تكتب على نبرة إذا كانت مكسورة، ومنها أنَّ الفعل مبني للمجهول، وما يأتي بعده يكون نائباً للفاعل.. وهكذا في كل الكلمات.

وإذا افترضنا جدلاً أننا أعطينا المتعلم نصاً من خمسة أسطر، ودرربنا بتقنية (التعرُّف على الصوت) على نطقه وأدائه، ثم أعطيناه معاني الكلمات من خلال المعجم الذي يُبني تلقائياً باستعمال الدارسين، ثم طلبنا منه أن يقلد الخطَّ وبعدها قام الحاسوب بقراءة النصِّ أمامه وأعطاه فرصة للاستماع والإملاء، ثم يعرض ما تعلَّمه على الحاسوب، فيقارنُ الخطوط ويضع له إرشاداتٍ لكتابَة الكلمة بخطوطٍ صحيحة، ثم يُدرِّبه على الإملاء ويُعرِّفُه أخطاءه ويعطيه المعلومات الواافية حول الكلمة من حيث النطق والكتابَة والمعنى المعجمي والاستعمال السياقي، إن تقنيةً بهذه الجودة لا بد وأن تحدثَ

طَفْرَةً هائلةً في تعليم العربية، وهذه التقنية ليست أضفافاً أحلاً بل مشروعٌ قابلُ للتطبيق، ولكنه يحتاج فقط إلى التطوير المستمر، فأساسُ التقنية موجودٌ، ولكنَّ هذا الأساس يحتاج أن يُبني عليه.

٩. الذكاء الاصطناعي وتعليم النحو.

إذا وصلنا بالتعلم إلى أنه صار قادراً على نطق الكلمات وكتابتها ومعرفة معانيها، يمكن توجيهه حتىٌّ إلى تعلم النحو لتركيب جمل مفيدة. والنحو لا يزال يمثل مشكلةً عند كثير من متعلمي العربية، والسبب في ذلك - من وجهة نظري - فصل القواعد النحوية عن النصوص، فصار النحو قواعد مجردة تنسى سريعاً، ولم يكن ملكرةً أو مهارة تُقْوِّمُ اللسان.

إن الذكاء الاصطناعي لم يصل في مرحلة تعليم النحو إلى تقدُّم يمكن وصفه بالكبير، مقارنة بالذكاء الاصطناعي في مجال الصوت والصرف والمعجم؛ لأن النحو يعتمد على الفهم، ويحتاج إلى فك الالتباس الدلالي بين الكلمات لتحديد مواقعها الصحيحة في الجملة، فمثلاً: جملة (قال الرجل) تتكون من فعل وفاعل، لا يمكن معرفة دلالتها إلا بإضافة مكمل، فنقول (قال الرجل كلمة الحق) فتكون (قال) من القول والكلام، أو نقول: (قال الرجل وقت القيلولة)، أي نام في وقت القيلولة والمعنى هنا تغيير جذريّاً بسبب مكملاً الجملة.

ورغم هذه المشكلات في النحو فإننا نعيّد التأكيد على أن معالجة عينة أشهر الكلمات وبناء منصة تعليمية عليها يعد نواة للتلافي الأخطاء الصادرة عن التقنية وتكوين منصة يمكن تطويرها فيما بعد. فالمدقق النحوي سيعطي بناء على الكتابة الصحيحة والضبط السليم احتمالات أكثر دقة للمتعلم، فمثلاً إذا كتب الدارس: (رأيت المتهمن) فالتطبيق يعطيه اقتراحات لضبط الميم: (رأيت المتهمِّن) أم (رأيت المتهَمِّن)؛ ليفرق بين المثنى والجمع، أما إذا كتب (رأيت المتهمن)، فإن المدقق النحوي في كل الأحوال سيفيده أن (المتهمن) خطأ لأنها مفعول به منصوب وعلامة نصبه الياء لأنه جمع مذكر سالم أو مثنى حسب التحديد الذي سيكون أمامه في النص.

ويعتمد المدقق النحوي على السوابق واللوائح، فمثلاً التركيب (لن يأكلون) لا يمكن أن تأتي بهذه الصورة؛ لأن (لن) من أدوات النصب، واللاحق (النون) لا يمكن أن تأتي مع لن الناصبة، وهنا تعطيه التقنية معلومة أدوات نصب الفعل المضارع المتمثلة في (أن - لن - كي - لام التعليل - حتى) كما تعطيه علامات النصب الأصلية والفرعية المتمثلة في الفتحة والألف والياء وحذف النون، وهكذا في كل الأدوات والمحروف.

إن المدقق النحوي عندما يشرح كلّ خطأ من خلال معلومات سريعة مباشرة، فإنها تكون أكثر رسوخاً في ذهن المتعلم، وبكثرة التدريب ستستقر المعلومة في رأسه ويمكنه تلافيها بعد ذلك.

ولبناء تطبيق يعتمد على الذكاء الاصطناعي في تعليم النحو يمكننا أن نضع المنهج التالي:

- ١) اختيار أكثر دروس النحو استعمالاً في الكلام.
- ٢) تصفية الدرس النحوي من الخلافات والكلمات الغريبة، والأساليب غير المستعملة، مثل ما انفك وما برح وغيرها.
- ٣) بناء نموذج لحمل سهلة لم تحدث فيها عوارض التركيب مثل الحذف والإضافة، والتقديم والتأخير.
- ٤) بناء نموذج للتوارد المعجمي والدلالي للتواافق في بناء الجملة، فعندما يكتب الطالب كلمة شرب فإن الحاسوب لا بد أن يحصر خياراته في أن يكون الفاعل كائناً حياً، فإذا كتب الطالب: شرب محمد، فإن الحاسوب يحدد خيارات المفعول به في المفردات التي تصلح للشرب، فلا يأتي المفعول، الجبل أو الحائط، بل يكون شرب محمد العصير، أو شرب محمد الماء.. وهكذا
- ٥) تدريب الحاسوب على تحديد علامات الإعراب الأصلية والفرعية، وذلك اعتناداً على تقنية المحلل الصري، فمثلاً تكون هناك معادلة أركانها: أن الفعل يأتي بعده الفاعل مرفوعاً بالضمة إذا كان مفرداً أو بالألف إذا كان مثنى أو بالواو إذا كان جمع مذكر سالم.

٦) تدريب الحاسوب على المتلازمات المقيدة والمركبات التي تأتي على صورة واحدة، وهي قليلة ويمكن حصرها، مثل: عدد من المركبات الاسمية كمركب الصلة والموصول، ومركب المصدر.

٧) محاولة التفرقة بين مركبات النعت والإضافة؛ لأنها أشهر المركبات استعمالاً في الكلام.

إن عدداً غيرَ قليلاً من قواعد النحو يمكن تحويلها إلى معادلاتٍ رياضيةٍ من خلالها نستطيع أن نقدم نموذجاً معيارياً صحيحاً وندرِّب الطلابَ على الجُمل القصيرة كمرحلة أولى حتى نتقدم شيئاً فشيئاً من فكِّ اللبس الدلالي في المُدقَّقاتِ النحوية.

١٠. الذكاء الاصطناعي وإعادة بناء النصّ.

نَفْصِدُ بإعادة بناء النص هنا التلخيص، وإعادة ترتيب الكلام في الجملة بشكل صحيح. وتقنيَّة التلخيص تنقسم إلى نوعين: التلخيص الاستخراجي، ويتضمنُ اختيار عباراتٍ وجملٍ من المستند الأساسي؛ لتكوين الملاخص الجديد دون إعادة صياغة هذه العبارات والجمل. والتقنيات المعتمدة على هذا النوع تتضمنُ تقييمَ أهميَّة الجملِ والعبارات، ومن ثم اختيار تلك التي لها التقييم الأعلى على افتراض أنَّها هي الجمل والعبارات التي تحوي المعنى الذي أراد الكاتب إيصاله.

أما التلخيص التوضيحي للنص فيتضمنُ تلخيصَ النصّ بعبارات وجمل جديدة كلياً؛ حيث يتم توليد جمل جديدةٍ تحوي الرسالة الأساسية للمستند الأصلي. ويعُدُّ هذا الأسلوب أكثر تحدياً من الأسلوب السابق، ولكنه أكثر حماكاً للأسلوب البشري.

إن تقنية التلخيص قام على تطويرها مجموعة من الباحثين منهم المهندس إبراهيم صبح في رسالته للماجستير (صبح، ٢٠٠٩). ويعتمد في هذه التقنية على التلخيص الاستخراجي، وفيه محاولة لحذف الجمل التي يراها غيرَ ضرورية للمتعلم، واعتمد في ذلك على عينة عشوائية من المقالات المنشورة في الصحف، وقد شرُفتُ بمعاونته في هذا الجانب، حيث كنت أقوم بحذفِ جمل أراها غيرَ ضرورية في فهم المُحتوى، ثم تُدفع هذه المُدخلات اليدوية إلى التقنية التي يُمكِّنها بعد ذلك أن تستطيع ما يمكن حذفه في النص ولا يؤثُّ في معناه.

وأما إعادة إنتاج جمل أو بنائها فتعتمد على تقنيات التدقيق النحوي والتلخيص والتحليل الصفي، لأننا سنعطي الكلمات سمات حسب موقعها في الجملة، فمثلا الكلمات النكرة لا يمكن أن نبدأ بها الجملة، والفاعل يكون مرفوعاً، والفعل الماضي يكون مبنيا على الفتح، والمضاف إليه يكون مجروراً وهكذا، فعند إعطاء المتعلم كلمات ليطلب منه إنشاء جملة صحيحة مفيدة فإن التقنية ستراعي كل هذه المحددات سلفا. فمثلا الكلمات (المسلمون، رمضان، شهر، صام) تقوم بتقديمها للمتعلم ثم نطلب منه ترتيبها وعند الخطإ نرشده إلى موقع الكلمة في الجملة.

إن الاستفادة من هذه التقنية يمكن أن تساهم في تدريب المتعلم على رسم خارطة ذهنية لما يقرؤه ويتعلم، فهو إذا تدرّب على التلخيص الاستخراجي فإنه يكتسب مجموعة من المهارات ستعجله قادراً على الاستخراج التوضيحي ومن ثم بناء نصٍ جديد.

إن آلية التلخيص الاستخراجي تقوم على تحويل النص إلى شكل رياضي تجري عليه مقاييس تُسقّفها قبلها معالجة للنص، وتحذف في هذه المعالجة الكلمات غير المهمة، مثل الحروف والأدوات، وكذلك إعادة الكلمات إلى جذورها وقياس التشابه بين جمل النص وتحليل العلاقات الدلالية بين الكلمات والجمل من خلال مصفوفة رياضية، وهناك طرق عدّة لاستخراج الملخص من هذه المصفوفة، أبسط تلك الطرق، هو اختيار الجمل ذات الارتباط الأقوى بأهم المفاهيم. فمثلاً لو أردنا أن يحتوي الملخص على عشر جمل، فإننا نختار الجمل العشر ذات القيم الأعلى في أول عشرة صفوف من المصفوفة.

إن هذه التقنية يمكن استعمالها في تطبيقات الفهم والاستيعاب للنص، بحيث يمكن أن نذكر للمتعلم بعض المعلومات وهو يكملها، وهناك برامج منتشرة في عدد من الدول تسمى برامج كشف النص المخفي، وتستخدم لتدريب الطالب على التنبؤ بالنص اللغوي المخفي، وهذه النصوص تكون دُرُوساً في القراءة والاستيعاب أو قطعاً أدبيّاً من النثر والشعر، ويُصَمَّم البرنامج على شكل لغةٍ تربوية تحسب فيها نقاطاً عند التنبؤ بالمحذوف في الوقت المُحدد.

١١. مشكلات الذكاء الاصطناعي في التعامل مع اللغة العربية.

تمثل مُعالجة اللغة العربية تحدياً كبيراً، وهذا التحدي مختلفٌ مستوياته، ففي الجانب الصوتي التحدي أقل؛ لأنَّ المعلومات الصوتية في اللغة العربية أكثر تحدياً من المعلومات الصرفية ثم المعجمية ثم النحوية والدلالية، وفيما يأتي تحديد لبعض هذه المشكلات.

١) اللبس الدلالي.

العقبة الرئيسيةُ التي تصادِفنا هي اللبس الدلاليٌ سواء على مستوى البنية أم التركيب، فاللبس يقف عائقاً دون تحقيق دقةٍ عاليةٍ في معالجة النصوص العربية، فمثلاً كلمة الولَدَان (مثُل) كلمة (ولد) تلتبس بالولَدَان وهي جمع، والأولى تُرْفَعُ بالألف وتنصب بالياء، وأما الثانية فتكتوَنُ على حالِها رفعاً ونصباً وجراً، ولا يستطيعُ الحاسوبُ أن يفرق بين (بالولَدَان) و (بالياء)، بل يخلطُ الحاسوبُ بين فطائر جمع فطيرة، وبين فطائر مفرد طائر التي أَلصقت بها الفاء كسابق (جمعة، ١٤٣٧ : ٧٤).

وعلينا أن نعترف أن الدراسة البحثية في معالجة الدلالة لا تزال تحتاجُ الكثير من الجهود المُكَلَّفةِ مادياً، فلا يمكنُ للقطاع الخاص أن يتحملها، ومن ثمَّ لا بد من مشروعٍ كبيرٍ يضمُّ صُفَوةَ الباحثين لتطوير التقنيات التي توصلنا إليها وتلقي الأخطاء التي تقعُ فيها هذه التقنية.

٢) طريقة بناء الجملة العربية.

طريقة بناء الجملة العربية تحتاجُ إلى عمليةٍ مُعَقَّدةٍ من الفهم ومَعْرِفَةِ سِيَّاراتِ المفرداتِ التي تناسبُ الوظيفة النحوية، بالإضافة إلى أنَّ الجملة العربية تحدثُ لها عوَارِضُ تركيبٍ من الحذف والإضافة والتقديم والتأخير. وهذا يُمثل تحدياً كبيراً للذكاء الاصطناعي.

٣) تنوع اللهجات.

تنوعُ اللهجات العربية يتحدى تقنياتِ التعرُّف على الصوتِ، خاصةً فيما يتعلق بإبدال الحروف (السين مكان الثاء والجيم مكان القاف أو الغين مكان القاف، أو الظاء مكان الضاد)، هذا فضلاً عن اختلافِ النبر والتنغيم وأداء الكلام، وهذا التنوعُ اللهجيُّ ليس موجوداً في اللغات الأخرى بهذه الكثرة والتنوع.

١٢. مقترن لمنصة تعليمية ذكية.

بعد هذا العرض فإنني أقترح منصة تعليمية للغة العربية تجمع بين كل هذه التقنيات، لتكون نواة لمدونة تعليمية كبرى تحتوي على أصوات اللغة ومفرداتها وتراثها، وهذه المنصة يمكن بناؤها من خلال صناعة مدونة لأشهر المستعمل في اللغة العربية.

ويمكن تطبيقها أيضاً من خلال تدريب كل التقنيات الحاسوبية على هذه المدونة، بمعنى أننا سنحاول إنشاء نصوص موجهة باستخدام مفردات هذه المدونة ومسكوكاتها فقط؛ حتى تتجنب المشكلات التي تقع فيها التقنية، وحتى نضمن دقة عالية في المعلومة التي تصل إلى المتعلم.

وإطار شكلي يمكن تصور هذه المنصة على النحو التالي:

كما في (الشكل ٣) نختار جملةً من النصوص ثم نبدأ في معالجتها لغويًّا بشكل كامل، حيث نبدأ بالتهيئة الحافرة التي قد تكون قصةً أو مقطعاً مرئياً أو فكرةً مبتكرة عند المعلم، والمنصة تتيح بعض هذه التهيئة للاستعانة بها، وتترك المجال للمعلم كي يضيف فكرته الإبداعية في هذه التهيئة التي ستضمُّ تجارب المعلمين وخبراتهم، فيحدث احتكاكًّا وتتوالد الخبرات التي تفيد المتعلمين، ثم تبدأ المعالجة اللغوية للنص، فيقرأ النص بصوت جميل، ويستمع الطالبُ لأداء نوعيًّا للنص، ثم يسمح لهم بقراءةٍ جهريَّةٍ على إثرها تُستخدم تقنية (التأثر على الصوت) حيث يستمع الطالبُ لصوته، وتبدأ التقنية في تضليل الكلمات التي أخطأ في نطقها صوتيًا أو صرفيًا أو نحوها، ثم بالإشارة إلى الكلمة أو الجملة تظهر الرسائل التوضيحية والتعليمية.

وبعد هذا التدريب على القراءة يكون الطالبُ قد استوعب فكرة النص، ويكون جاهزاً للإجابة عن أسئلة الاستيعاب التي تأتي في هذه النافذة؛ حيث يمكن استئمار تقنية التلخيص في أن يُطلب تلخيصُ النص الذي قرأه أو التعبيرُ عنه بأسلوبه.

كما تتيح هذه النافذة أن يمسح النص، ثم تبدأ التقنية بقراءته بشكل بطيء يسمح للطالب أن يكتبه إملاءً، وبعد أن ينتهي يقوم بادخال خطه مصوّراً، ومن خلال تقنية التعرف الضوئي يمكن أن يكتشف الحاسوب مواطن الخطأ ويعرّض له النص الأصلي مع النص المُمْلَأ.

وتتيح النافذة أيضاً أسلة تقييم تُستخدمُ فيها الوسائلُ التربويةُ في وضعِ الأسئلةِ من تحليلٍ وتركيبٍ وتقعيدٍ.



الشكل ٣: نموذج المنصة التعليمية - نافذة مُعطيات رئيسية

تكون بالمنصة مدونة مصنفة ومرتبة حسب شيوع كلماتها، وبناء على المدونة، نختار مثلاً الآيات القرآنية ذات الكلمات المتداولة والأحاديث ذات الجمل البسيطة والشعر الذي يستخدم كلمات مألوفة للأذن وهكذا في كل النصوص. وهناك إمكانية في المنصة ليضيف المعلم والمتعلم ما يرود لهما من نصوص، وهذه المختارات يمكن أن ندرس طبيعتها وتوجهاتها وتكون قيمة مضافة للمنصة.



الشكل ٤: نموذج المنصة التعليمية - نافذة اختيار النص

وفي هذه المنصة توجد نافذة لتعلم المهارات التي نوظف فيها التقنيات التي ذكرناها في هذا الفصل، فمثلاً إذا دخل على نافذة الإملاء يظهر له نموذج ليحاكيه، ثم يكتب بالقلم العادي على ورقة تصورُ بشكل آلي وتقارن خطوطها بخطوط النموذج، فيكتشف المعلم أخطاءه الإملائية، كما يتعرفُ أيضاً على طريقة رسم الحرف بشكله المعياري سواء في الرُّقعة أو النسخ باعتبارهما من أشهر الخطوط المستعملة في العربية.



الشكل ٥: نموذج المنصة التعليمية - نافذة المهارات

وتشتمل المنصة أيضاً على مجموعة من معارف لغوية تضم خلاصاتٍ في الأصوات والصرف والنحو والدلالة، كما تضم معارف ثقافيةً، ويمكن أن تُمثل مدونةً يستطيع المستخدم أن يُضيف إليها، مع وجود آلية لدراسة مُختارات المُتعلمين واعتراضها ضمن المنصة.



الشكل ٦: نموذج المنصة التعليمية - نافذة المعرف

وتشمل هذه المنصة أيضاً وجدانياتٍ ترقي بمشاعر المتعلم، وتبني مدونتها من خلال مختاراتٍ تناسبُ المستويات والأعمار والثقافات، ولا تقتصرُ على الثقافة العربية، بل تمتدُّ لثقافاتٍ أخرى، ويتاحُ للمتعلم أنْ يُضيفَ عليها، أو أنْ يُبْدِي رأيه من خلال قاعدةٍ بياناتٍ يمكن استثمارُها في دراسَةٍ بحثيَّةٍ تبيَّنُ معرفةً طريقة التفكير وأسس الاختيار التي ستحتَّلُ حسبَ البيئات والثقافات التي تتعددُ في وطننا العربي، ويمكن استثمار هذه النافذة في معرفة العلاقات بين الألفاظ في ذهن المتلقي، بحيثُ تُنشئُ شبكةً من الكلمات التي تتميَّز بحلق دلالي واحد، وما بينها من علاقاتٍ دلالية كالترادف والتضاد والمشترك اللغظي وغير ذلك من العلاقات.

إنَّ الطَّلَابَ وَالْمُعَلِّمِينَ يُمْكِنُهُمْ أَنْ يُنْشِئُوا قاعدةَ بياناتٍ تفيدُ الباحثين، وهذه القاعدةُ تَنْشَأُ بِشَكْلٍ طَبِيعِيٍّ وَفِي بَيَّنَاتٍ مُخْتَلِفَةٍ، فَتَمْثِلُ مَادَةً خَصْبَةً يُمْكِنُ دراستها وَالبناءُ عَلَيْها.



الشكل ٧: نموذج المنصة التعليمية – نافذة وجدانيات

وتشمل أيضًا هذه المنصة وسائل التقييم حسب مستوى المتعلم، بدايةً من المرحلة الابتدائية، ولكن في صورة مستويات. فمن الممكن أن يكون الطالب في الصف الخامس، لكنه يظل في المستوى الأول الذي قد يشتراكُ فيه مع طالب في الصف الثالث، وهذه المنصة ستحسنُ من مستوى الطالب، وسترفعُ عنه الخجلُ من مرافقه طالبٌ أصغر منه سنًا.

والمصة أيضاً بها قصص تثري معارف الطالب، كما أن بها مقاطع مرئية ومسموعة لتنوع أدوات التعليم، وكل هذا في مكان واحد وبأسلوب سهل الاستخدام.



الشكل ٨: نموذج المنصة التعليمية – نافذة التقييم

إن هذه المنصة تُعدُّ واجهة يمكن من خلالها توظيفُ التقنيات التي أشرنا إليها؛ لتكون بيئةً تعليميةً مثاليةً يمكن استخدامها في المدرسة أو البيت أو أي مكان.

الخلاصة.

خلاصة القول في هذه القضية أن برامج معالجة اللغات الطبيعية حققت طفرةً لا يمكن إنكارها، وهذه الطفرة يمكن استثمارُها في تقنياتِ تعليم اللغة العربية، ولكنَّ الأمر يحتاج إلى مشروعٍ كبير يضمُّ صفوة الباحثين في الوطن العربي، كما أنَّ هذا المشروع الذي يوفر منصة تعليمية متصلة بالمدارس في مختلف أنحاء الوطن العربي يمكن أن يُستثمر في تعليم الطلاب وفي تكوين قاعدة بياناتٍ ضخمة، ويمكن أن يُكون تغذياتٍ راجعةً تفيده في تطوير التقنيات التي يوظفها، حيث يتاح نافذة لإبداء الرأي في التقنيات والمشكلات التي تواجه المتعلمين، وبهذه الطريقة يمكن أن تكون هذه المنصة منهجاً تعليمياً مستقلاً يوفر جهود طباعة الكتب ويعطي نتائجَ نوعية في تعليم الطالب اللغة العربية.

ببليوجرافيا مرجعية.

١. ابن جني، أبو الفتح عثمان، (٢٠٠٦) الخصائص، دار الكتب المصرية، تحقيق، محمد علي النجار.
٢. جرين، جوديث، (١٩٩٢)، التفكير واللغة، الهيئة المصرية العامة للكتاب.
٣. جمعة، عمرو (٢٠١٦)، تقنيات اللغة العربية الحاسوبية، معايير التقييم ورؤى التطوير، دراسة لغوية حاسوبية، مركز الملك عبد الله لخدمة اللغة العربية.
٤. حامد، صلاح الدين، (٢٠٠٤)، خوارزمية التعرف على أخطاء التلاوة، دكتوراه، ٢٠٠٤، كلية الهندسة جامعة القاهرة.
٥. دي سوسيير، فردينان دي سوسيير (١٩٨٥)، دروس في الألسنية العامة، تعریب محمد شاوش وآخرون، الدار العربية للكتاب، تونس، د.ط.
٦. السعيد، المعتر بالله (٢٠٠٧) مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، ماجستير، جامعة القاهرة.
٧. السعيد، المعتر بالله (٢٠٠٩)، المعجم التكراري لألفاظ القرآن الكريم، «المنهج والنموذج» ندوة القرآن الكريم وتقنية المعلومات، المدينة المنورة، السعودية.
٨. السعيد، المعتر بالله (٢٠١٦)، توظيف المُدَوَّنات اللُّغَوِيَّةِ في تطوير مُقرَّرات اللغة العربية لراحل التعليم العام، مجلة التخطيط والسياسة اللغوية، مركز الملك عبد الله لخدمة اللغة العربية، الرياض، العدد ٣، ٢٠١٦.
٩. صبح، إبراهيم (٢٠٠٩)، التلخيص الآلي، ماجستير، كلية الهندسة بجامعة القاهرة.
١٠. طعيمة، رشدي (٢٠٠٦)، معاهد تعليم اللغة العربية لغير الناطقين بها، اتجاهات التطوير، معايير الاعتماد، مؤشرات الجودة، مجلة العربية للناطقين بغيرها، جامعة أفريقيا العالمية، العدد الثالث.
١١. عبد الغني، نعيم محمد (٢٠١٥) تفاعل المفردات مع الوظائف النحوية في الجملة القرآنية، دار النابغة، مصر، ط: ١.

١٢. علي، نبيل (١٩٨٧)، اللغة العربية والحاسوب، مجلة عالم الفكر، المجلد الثامن عشر، العدد الثالث.

١٣. علي، نبيل (١٩٨٨)، اللغة العربية والحاسوب، تعریب، القاهرة.

١٤. عمايرة، خليل عمايرة (١٩٨٥)، في التحليل اللغوي، مكتبة المنارة، ط: ١.

١٥. عمر، أحمد مختار (١٩٨٥)، علم الدلالة، عالم الكتب، القاهرة، ط: ٢.

1. Al-sulaiti, L. (2004). *Designing and Developing a corpus of contemporary Arabic*. Leeds University.
2. Chomsky, N. (2000). *New Horizon in the Study of language and mind*. Cambridge University Press.

الباحثون

الدكتور / محمد عطيّة محمد العربي



عمل «د. محمد عطيّة محمد العربي أَحمد» لمدة ربع قرن في مجال علوم الحاسوب، وكذلك علوم اللسانيات الحاسوبية، وحوسبة اللغة العربية، وبناء الموارد اللغوية. وقد نشر العديد من الأوراق البحثية باللغة الإنجليزية في الصف الأول من الدوريات العالمية المحكّمة، كما حَكَمَ العشرات من الأوراق في تلك الدوريات، وشارك في عدد من الكتب والدراسات والأوراق البحثية بالعربية حول المعجم، والصرف، والتركيب، والدلالة في اللغة العربية. وكذلك شارك وأدار العديد من التقنيات الحاسوبية الصناعية، والمشروعات البحثية في حوسبة اللغة العربية، والتعليم الإلكتروني، وكان خبيراً حاسوبياً لمشروع المعجم التاريخي للغة العربية بين عامي ٢٠١٤ و٢٠١٦ م. وكذلك كان أستاذ زائراً في عدة أكاديمياتٍ عربيةٍ مثل: الأكاديمية البحريّة للعلوم والتكنولوجيا والنقل البحري، وجامعة الإسكندرية.

الدكتور / المُعترَّ بالله السَّعِيد طه



أستاذ الدراسات اللّغويّة المساعد بجامعة القاهرة، وأستاذ اللّسانيات الحاسوبية المُشارك بمعهد الدّوحة للدراسات العليا، ومنسّق وحدة الموارد المُعجمية بمشروع معجم الدّوحة؛ وهو استشاريٌّ لعددٍ من المُؤسسات. نَشَرَ نحو ثلاثين ورقة علميّة محكّمة، بالإضافة إلى عددٍ من الكتب في حوسبة اللّغة وصناعة المعجم والدّرس اللّغوي المعاصر، وقام بتحكيم عشرات الأبحاث والأطروحتات العلميّة. شارك في العديد من المشروعات الدّولية المعنية بمعالجة اللغات الطّبيعية، وأشرف على بناء وتطوير الموارد اللّغويّة والحاوسيّة لعددٍ من هذه المشروعات. حصل على عددٍ من الجوائز العلميّة، منها: جائزة (الكسو ALECSO) للإبداع والابتكار في «المعلوماتيّة والمعالجة الآليّة للّغة العربيّة»، وجائزة راشد بن حميد للعلوم والثقافة.

الدكتور / أَمْهَدْ رَاغِبْ أَمْهَدْ



أَسْتَاذُ الدِّرْاسَاتِ الْلُّغَوِيَّةِ الْمُشَارِكُ، وَرَئِيسُ قِسْمِ الْلُّغَةِ الْعَرَبِيَّةِ وَآدَابِهَا بِالجَامِعَةِ الْإِسْلَامِيَّةِ الْعَالَمِيَّةِ بِمَالِيْزِيَا. حَصَلَ عَلَى درَجَةِ الدُّكْتُورَاهُ بِمَرْتَبَةِ الشَّرْفِ الْأَوَّلِيِّ مِنْ جَامِعَةِ الْقَاهِرَةِ عَامَ ٢٠٠٩ وَنَسَّرَ عَدَدًا مِنَ الْأَوْرَاقِ الْبَحْثِيَّةِ فِي دُورِيَّاتِ عَلَمِيَّةٍ؛ كَمَا شَارَكَ فِي الْعَدِيدِ مِنَ الْمُؤْتَمِراتِ الدُّولِيَّةِ الْمُعْنَيَّةِ بِحُوْسَبَةِ الْلُّغَةِ؛ وَلَهُ عَدْدٌ مِنَ الْمُؤَلَّفَاتِ الْعِلْمِيَّةِ.

الدكتور / نَعِيمُ مُحَمَّدْ عَبْدِ الْغَنِيِّ



حَصَلَ عَلَى الدُّكْتُورَاهُ فِي النَّحْوِ وَالصَّرْفِ وَالْعَرْوَضِ بِمَرْتَبَةِ الشَّرْفِ الْأَوَّلِيِّ مِنْ جَامِعَةِ الْقَاهِرَةِ. يَعْمَلُ فِي الْبَحْثِ الْعَلَمِيِّ وَالْتَّدْرِيسِ الْجَامِعِيِّ وَالْإِعْلَامِ، وَهُوَ الْآنَ خَبِيرٌ لِغَوِيٌّ فِي الْمَنْظَمَةِ الْإِسْلَامِيَّةِ لِلتَّرْبِيَّةِ وَالْعُلُومِ وَالثَّقَافَةِ (إِيسِيسِكُو)؛ وَقَدْ عَمَلَ مُخَاضِرًا فِي جَامِعَةِ قَطْرٍ، وَبَاحِثًا غَيْرَ مُتَفَرِّغٍ فِي مُعْجَمِ الدَّوْحَةِ التَّارِيْخِيِّ، وَكَاتِبًا وَمُحَرِّرًا فِي بَعْضِ الْمَجَالَاتِ وَالْإِذَاعَاتِ الْعَرَبِيَّةِ. نُسِّرَتْ لَهُ عَشَرَاتُ الْمَقَالَاتِ وَالتَّقَارِيرِ فِي صُحُفٍ وَمَجَالَاتٍ مُخْتَلِفَةٍ، وَأَعْدَّ وَقَدَّمَ عَدِيدًا مِنَ الْبَرَامِجِ الإِذَاعِيَّةِ الَّتِي تُعْنِي بِالشَّانِ الثَّقَافِيِّ، وَلَهُ سَيِّنَةُ كُتُبٍ مَطْبَوعَةٍ، بِالإِضَافَةِ إِلَى عَدِيدٍ مِنَ الْأَبْحَاثِ الْعِلْمِيَّةِ الْمُحَكَّمَةِ.

هذا الكتاب

يُصدر مجمع الملك سلمان العالمي للغة العربية هذه السلسلة ضمن خطوة عمل مقسمة إلى مراحل تشمل مرحلتها الأولى ثلاثين عنواناً موضوعات علمية رأى المجمع - بعد الدراسة - حاجة المكتبة اللغوية العربية إليها، أو إلى بدء النشاط البحثي فيها ، ومهدف من وراء ذلك إلى تنشيط العمل في المجالات التي تنبه إليها هذه السلسلة، سواء أكان العمل علمياً بحثياً، أم عملياً تنهيذياً، ويدعو المجمع الباحثين كافة من أنحاء العالم إلى المساهمة في هذه السلسلة ويود المجمع أن يشيد بجهد السادة المؤلفين، وجهد محرر الكتاب على ما تفضلوا به من التزام علي لا يستغرب من مثلهم. والشكر والتقدير الوافر لسمو وزير الثقافة رئيس مجلس أمناء المجمع الذي يحث على كل ما من شأنه تثبيت الهوية اللغوية العربية، وتمتينها، وفق رؤية استشرافية محققة لتوجيهات قيادتنا الحكيمية والدعوة موجهة لجميع المختصين والمهتمين بتكييف الجهد نحو الصعود بلغتنا العربية، وتحقيق وجودها السامي في مجالات الحياة.



مكتبة
الملك سلمان