مصطلحات هندسة اللغة العربية بين التعريب والتأثيل

أ.د. محمد الحناش
الوكالة الدولية لهندسة اللغات الطبيعية

إن التطور الأسي Exponential[1] الذي تعرفه التقانة الحاسوبية، خاصة هندسة اللغات الطبيعية التي تقوم على تطبيقات الذكاء الصناعي Artificiel intelligence بنوعيه القوي والضعيف، وفروعه المعرفية مثل التعلم العميق Deep learning وتعلم الآلة Machine learning التي أدت إلى تشكل عصر الانفرادية Singularity، هذا التطور أفرز عدة مصطلحية غاية في الدقة والصرامة المعرفية، مما أصبح يتطلب ضبطها وتفكيك معانيها المضغوطة Compressed تيسيرا لاستخدامها في مختلف التطبيقات المتوقع تطويرها بلغة الضاد دفعا بها إلى الانخراط بشكل مكثف في بناء مجتمع المعرفة. مواكبة لهذا التطور التقاني أفرزت لسانيات الجيل الرابع منظومة من القواعد اللسانية الجديدة لاستخدامها في إعادة توصيف بنية اللغات الطبيعية Natural languages انطلاقا من أرضية عرفانية Cognition جديدة ظلت مغيبة في البحث اللساني التقليدي بأجياله الثلاثة السابقة. ولإلقاء الضوء على هذا التفاعل اللساني – الهندسي الذي أفرز هذا العلم الجديد الخاص بمعالجة اللغات الطبيعية، سنقوم بعرض موجز عن المنجزات التي حققتها لسانيات الجيل الرابع

1 – مصطلح يوصف مراحل التطور العلمي في أزمنة متقطعة، وهو عكس التطور الخطي Linear الذي يتتبع التطور عبر خط سير الزمن بطريقة مستمرة، مثال ذلك شركات الهواتف الذكية التي قلصت مدة الإعلان عن نتيجة أبحاثها كل 18 شهرا، بدل أن تعلن عنه باستمرار، وقد تقلصت هذه المدة بفعل التطور العلمي وفق قوانين موور المنبثقة عن تطبيق فيزياء الكوانتم Quantum physics على العلوم الصلبة.

التي نقلت البحث في ماهية اللغات الطبيعية في الكفاية البشرية، إلى إعادة توصيفها وصورنتها Formalization بهدف نقلها إلى الآلة، بعد ذلك سنتولى تقديم نماذج من المصطلحات الضرورية لبناء معجم مصطلحات هندسة اللغات الطبيعية، مركزين على منصة Nooj التي توظف آليات الذكاء الصناعي في تطوير مختلف التطبيقات المعتمدة على معاجم إلكترونية وأنحاء محلية Local grammar، سنختم هذا البحث بتطبيق عملي على أحد المقالات العلمية التي صدرت حديثا في موضوع تطبيق تقانة التعلم الآلي على اللغة العربية، حيث تبين أن 70 من هذه المفاهيم يتقاسمها مجالان متكاملان في هذه التقانة: حاسوبيات لسانية، وذكاء صناعي.

تعريفات أساسية:

سنوظف مفهومين يبدوان متعارضين ظاهريا:

1. هندسة اللغة Language engineering تنطلق من الهندسية لمعالجة اللغة بوصفها منظومة من الخوارزميات Algorithms، ومن مصطلحاتها الأساسية: text analysis, Corpora, Artificial intelligence, Machine learning, Text maining, Big data, data maining, Platform، Automatom، Sigularity، Deep learning.

2. اللسانيات الحاسوبية Computaional linguistics تنطلق من النظام اللغوي لبناء أنساق صورية، بهدف ملاءمة التوصيف اللغوي مع متطلبات الهندسة، أي التحكم في النظم بشكل عام، ومن المفاهيم الأساسية التي تقوم عليها: morphological Generation, Morphological analysis, Santactical analysis, Speech recognition, Database, Electronical dictionaries, Ontologies, Matrix...

1. المقدمة:

سواء انطلقنا من هندسة اللغة أو من اللسانيات الحاسوبية، فإن النتيجة واحدة، تتمثل في تطوير تطبيقات تسهم في انتقال الشعوب العربية إلى مجتمع المعرفة بلغتها الأم، وهو ما يتطلب إعادة توصيف هذه اللغة لتلائم متطلبات التقانة الجديدة، في زمن أصبحت فيه الآلة تحاكي Simulation قدرات البشر، والسبيل الوحيد أمام لغة الضاد، لتقوم بهذا الدور، هو توظيف العتاد المصطلحي الذي تشتغل به التقانة الجديدة، علما أن كل مصطلح علمي يعد جملة مضغوطة تلخص تجربة معرفية محددة، تختصر الكلام الطويل في متوالية على شكل كلمة محاطة ببياض من الجهتين، مثال ذلك مصطلح: Platform الذي أصبح يلخص مجالا فرعيا في التطور التقاني تعرفه العلوم بمختلف مجالاته، ومنه هندسة اللغات الطبيعية، حين تستخدمه مركبا كما يلي: Linguistic platform، ونحن نستخدمه في بحوث هندسة العربية بصيغة: لسانيات المنصات. هذه اللسانيات الجديدة يؤطرها ما أصبح يعرف اليوم بلسانيات الجيل الرابع G' 4 linguistics. الذي جاء بعُدة مُصطلحية جديدة لسانيا وحاسوبيا، تختلف عما كان متداولا إلى عهد قريب في الدراسات اللسانية النظرية التي كانت تقوم على عدة مصطلحية ذات دلالات أصبحت متجاوزة في أغلبها، وحتى في الحالة التي يتم فيها استخدام المفاهيم الإجرائية التقليدية المؤطرة لأجيال اللسانيات السابقة، فإنه يتم شحنها بدلالات جديدة مزدوجة، تجمع بين المعنى اللساني والمعنى الهندسي، خاصة بعد انصهار العِلْمين في منظومة معرفية واحدة هي التي يطلق عليها هندسة اللغة، مثال ذلك مصطلح: Operator الذي كان يستخدم في النحو بمعنى الفاعل، وفي الرياضيات بمعنى الرابط بين مكونات المعادلة: (+)، و(−)، ولكنه أصبح يستخدم في هندسة اللغات الطبيعية بمعنى: العامل المنطقي Logical operator، والمقصود به في هذا العلم نوعية العلاقة التي تربط بين

مكونات التركيب، الفاعل يختار فاعله وفق علاقة عاملية تولّدها بنيات الجمل السليمة على شكل علامة: +، بمعنى جملة مقبولة، و(–) بمعنى جملة غير مقبولة أو غير ممكنة الوجود [2]. هذه الرموز مأخوذة من جَبْر جورج بوول (1815 – 1869)، الذي كان يرى أن الذكاء البشري يقوم على ازدواجية العوامل المنطقية، وهما: Not, or تتولد بهما البيانات الصحيحة: Yes، والخطأ: Not. هكذا يتم شحن المفاهيم القديمة بمعان جديدة فرضها التزاوج المعرفي بين العلوم الإنسانية والعلوم الصلبة. هذا هو ديدن أغلب المصطلحات العلمية التي تواكب تطور العلوم.

2. التكامل المعرفي بين اللسانيات والحاسوبيات:

سنتناول، فيما يلي، طريقة التعامل مع العُدة المصطلحية التي تشتغل بها هندسة اللغة العربية وتعتمدها في صياغة أدواتها الإجرائية. علما أن مصطلحات هذا العلم يصعب حصرها لكونها تعدّ مجالا هجينا Hybride يجتمع فيه علما الهندسة واللسانيات، وكلاهما يصوغ مصطلحاته انطلاقا من مرجعيته الخاصة، وإن كانا معا يلتقيان في الصَّرامة المعرفية في صياغة المصطلح العلمي الذي تشتغلان به. ففي اللسانيات مناهج ونظريات، ولكل منها منظومته المصطلحية التي يشتغل بها، تذهب أحيانا إلى حد التناقض في مضامينها، مثل ما هو حاصل بين البنيوية ولسانيات الجيل الرابع، مثال ذلك مفهوم syntax الذي أصبح محلل تركيبي Syntactical analysis، وهما مصطلحان يوظفان اليوم في هندسة اللغة لكن بمضامين مختلفة، وللهندسة منظومتها المصطلحية الخاصة بها التي تشتغل بها، مثل المسوقة Automatom، وهو مصطلح يعادل مفهوم النحو grammar في اللسانيات النظرية، لكن بمعنى يختلف عن معناه عند اللسانيين،

2 – **انظر الشكل رقم: 1 في هذا البحث**. حيث يتم توظيف هاتين القيمتين في بناء أجروم Matrix البنيات التركيبية للغة العربية.

عربا وغربيين. حيث يتبين أن كثيرا من المصطلحات الجديدة أصبحت تأخذ معناها القديم وتشحنه بمعنى جديد نحتاج معه إلى توضيح، أو يكتسب معنى لا علاقة له بمعناه الكلاسيكي، مثل مصطلح مُتون / مُدونات Corpora الذي أصبح له معنى جديداً يقع في مفترق الطرق بين المتن / مدونة Corpus بمعناه القديم وقاعدة البيانات Database الذي تستخدمه مجموعة من المؤسسات العالمية المهتمة بمتون اللغة، مثل PDTB[3]، ويعني قاعدة البيانات الشجرية الذي أنجز في جامعة بنسلفانيا، وCATiB الذي بنته جامعة كولومبيا، وأخيرا مدينة الملك عبد العزيز للعلوم والتقنية التي بنت قاعدة بيانات ضخمة أطلقت عليها المدونة اللغوية العربية Arabic Corpus. هذا العمل يقع في صلب لسانيات الجيل الرابع الذي تؤطره هندسة اللغة بمعناها الجديد، وإن كان واضعوه يُنظّرون له هندسيا أكثر منه لسانيا.

وبما أن مجالي هندسة البرمجيات Software Engineering واللسانيات Linguistics أصبحا يتكاملان من حيث توظيف أدوات العمل، لينصهرا مؤخرا في جيل جديد يطلق عليه لسانيات الجيل الرابع GL4، وهو جيل اللسانيات الصورية التي تسعى إلى نقل الكفاية البشرية إلى الحواسيب، كما أصبحا يتعاونان معا في صياغة برامج تمكن الآلة من محاكاة العقل البشري، فإن مهمة واضع المصطلح أصبحت تتعقد أكثر فأكثر، خاصة وأن العلمين معا، الهندسة واللسانيات، قدما إلى العربية من عوالم بحثية غير عربية في معظمها، وبما أن الباحث في هندسة العربية أصبح لزاما عليه أن يوظف المفاهيم ذات القواسم المشتركة بين الهندسة واللسانيات، أولا، وبين علوم كثيرة ذات منحى صوري ثانيا، فإن مجهوده أصبح مضاعفا في صياغة مصطلحات علمية تسعف المتلقي في

3 – بنك البيانات الشجرية Data tree bank، يربط البيانات ببعضها عن طريق مبدأ الاشتقاق في المستويين الصرفي والتركيبي، معتمدا الإطار اللساني التوليديالذي يصنف في إطار لسانيات الجيل الثالث، أي ما قبل جيل هندسة اللغات الطبيعية.

مسايرة ركب التطور الجديد الذي تعرفه هندسة الحواسيب التي أصبحت ذات كفاءة عالية تقترب في عملها من كفاءة الدماغ البشري، حيث تجاوزت مهمة الحاسوب مساعدة الباحث في تنظيم المعلومات إلى تلقينه أمورا كثيرة لم يكن في إمكانه أن يصل إليها بأدوات علمية أخرى، ولنأخذ كمثال مصطلح الذكاء الصناعيArtificial intelligence، الذي يعدّ صلب أي عمل هندسي لساني في أي بحث علمي في هندسة اللغات الطبيعية، أو مصطلح التعلم الآلي Machine learning، الذي ظلّ إلى عهد قريب أحدث ما توصلت إليه التقانة الهندسية – اللسانية، لكنه أصبح شبه متجاوز اليوم مع ظهور مصطلح الشبكة العصبية Neurones network، الذي يفوق إمكانات الدماغ البشري في مجال التعامل مع المعلومات المبرمجة في دماغ البشر، تخزينا واسترجاعا، على شكل تطبيقات Applications حاسوبية ذكية تتولى إعادة برمجة نفسها آليا، في إطار ما أصبح يعرف عالميا بعلم الانفرادية Singularity، الذي يشتغل بترسانة من المفاهيم الجديدة التي تتجاوز مرجعياتها ما كان عليه الأمر في الهندسة اللسانية التقليدية، خاصة المفاهيم المستخدمة في فيزياء الكوانتيوم Quantum Physics التي أصبحت تغزو مجال العلوم، خاصة ذات المنحى الصوري، ومنها هندسة اللغات الطبيعية التي أصبحت تسهم في تطوير مجالات صناعية كثيرة، في مقدمتها صناعة الروبوتيك Robotics وتقوم أساسا على مسوقات من صلب تقانات الذكاء الصناعي الضعيف، مثل: التعلم الآلي والتعلم العميق، والشبكة العصبية، وهلم جرا.

هذا، وغيره، يبين أن ترسانة من المصطلحات العلمية الجديدة أصبحت تظهر بشكل متسارع في مجال هندسة اللغات الطبيعية، والمطلوب من الباحث في هندسة اللغة العربية ضبط معانيها، أولا، في مظانها المعرفية الأصلية، قبل الشروع في توظيفها في معالجة لغة الضاد، مصطلحات من المفروض فيها التفاعل إيجابيا مع المنظومة اللسانية العربية ذات التكوين الانصهاري

Fusionsit، من حيث وضع خوارزميات Algorithms تمكّن من فهم تركيبة الدماغ المجهز، أساسا، بقواعد لغة تختلف عن قواعد اللغات الإلصاقية Concatinative، وقواعد الهندسة القائمة على مبدأ التحكم في النظام اللغوي توليدا Genaration وتحليلا Analysis، حيث لا يمكن الاستغناء عن أحدهما دون الآخر في أي تطبيق هندسي على اللغة الطبيعية، ومن ضمنها لغة الضاد.

وهنا يبرز سؤال: هل ما نستخدمه من مصطلحات باللغة العربية لمعالجة اللغة العربية هندسيا يعكس بشكل معرفي كامل ما يريد قوله واضعو النظريات الهندسية اللسانية غير العربية، أم أننا نتجنى عليها بترجمتها إلى لغة الضاد ؟.

في محاولة للإجابة عن هذا السؤال المركب، يجب الانطلاق من المربع المعرفي الجديد الذي أجمع عليه علماء الذكاء بنوعيه الطبيعي والصناعي[4]، وهو يلخص التطور الذي وصلت إليه العلوم اليوم، ويلخصونه في الرموز التالية NBIC وتعني: Nano technologie, Biotechnologie, Informatique, et Cognition) النانو تكنولوجي، والبايوتكنولوجي، وعلوم الحاسوب، والمعرفية. هل يمكننا ترجمة هذه المصطلحات، وهي مداخل أساسية لفهم هندسة اللغة؟ في تقديري، كل ما نستطيع مقاربته هو تعريب المصطلح مع شرح له يوضح المقصود منه، أولا لأننا لم نصل بعد إلى دراسة هذه العلوم بالعربية، وثانيا لأننا لم نطوّر لغتنا لتمكينها من ترجمة هذه المصطلحات بدقة، ولذلك سنبقى دائما في حضن التعريب، وهو الوسيلة الوحيدة التي تمكننا من فهم المعارف الأربعة المشار إليها أعلاه، خاصة وأن كل واحد منها يشتغل بترسانة من المصطلحات التي تمكن من فهمه وطريقة تطبيقه في الواقع، لأننا لا يجب أن نحصر عملنا في دراسة العلوم في فهمها، ولكن من أجل تطبيقها، وهذه تحتاج أكثر إلى مصطلحات علمية تمكننا من بلوغ درجة معينة من النضج المعرفي لاستيعاب

4 - Dr Laurent Alexandre, La guerre des intelligences : Intelligence artificielle versus Intelligence humaine, Ed. JCLattès, France 2017.

هذه المفاهيم، أما الحديث عن أي علم دون تطبيقه، فهو كمن يحكي عن أمور لم يشاهدها في حياته، ومع الأسف هذا هو حال أمتنا العربية.

إنَّ التخلف الذي يعرفه تطور البحث اللساني العربي مقارنة مع التطور السريع الذي تعرفه الهندسة، يوقعنا في حرج استخدام المصطلح اللساني المناسب، مما يجعل الباحث، في هذا الميدان، يلجأ أكثر إلى المصطلح الهندسي ليحل به القضايا التي يطرحها العمل في أنساق اللغة، وهكذا وجدنا مصطلح المعجم، يصبح المعجم الإلكتروني Electronical dictionary، هل فعلا تنقل هذه الترجمة محتوى المصطلح الإنجليزي ؟، لا أظن، وقد عانينا كثيرا في إبلاغ الباحثين هذا المعنى كما هو مضمن في المصطلح الإنجليزي، الأمر نفسه، مع مصطلحيMorphological analysis, syntaxical analysis، حيث ترجمناهما بـ: المحلل الصرفي والمحلل التركيبي، علما أن العربية فيها أيضا المحلل التصريفي Congugaison، مما يحتاج معه الأمر إلى تعديل المصطلح الإنجليزي نفسه لأن مصطلح تصريف يرتبط بقواعد المستوى الفونولوجي Phonological ruls الخاص بالعربية بوصفها لغة انصهارية، خلافا للإنجليزية التي تعد إلصاقية. عندما نترجم المصطلح فإننا نقارب دلالته المعرفية، دون أن نتمكن من النفاذ إلى عمقه المعرفي، لأن وضع أي مصطلح علمي متخصص يستند إلى منظومته الإبستمولوجية التي تطورت في أحضانها، ولذلك فإننا، قبل ترجمتها، نحتاج إلى الإحاطة بها في أصلها المعرفي الذي أفرزها، وهو ما يبدو صعب المنال في غالب الأحيان، لأننا نوظف المصطلح في تبادل المعارف الدقيقة، ولا نوظفه فقط لننقل جانبه السطحي، ومثال مصطلح العامل Operator، الذي سبقت الإشارة إليه، خير دليل على ما نذهب إليه، فعلى الرغم من وجود مصطلح عامل في العربية، فإن ترجمة المصطلح الإنجليزي سيكون ضربا من السخرية، الأمر نفسه وقع فيه مترجمو العلوم العربية إلى اللغات الأجنبية، خير مثال على ذلك مصطلح Computer التي اختلف التراجمة في نقله إلى لغة الضاد. هذه ليست دعوة

للتوقف عن ترجمة المصطلح العلمي إلى لغة الضاد، لكن القصد هو أن يتمكن المترجم من معنى المصطلح في بيئته الأصلية في اللغتين (المنطلق والهدف) قبل الإقدام على ترجمته. يحضرني، في هذا الصدد، مصطلح أجروم الذي يستخدم عادة في ترجمة مصطلح Matrix، ظاهريا لا علاقة بين معنييهما، لكنك عندما تتأمل وظيفة كل واحد منهما في بيئته الأصلية ستسلم بهذه الترجمة. البديل الذي لا يثير أي نقاش مصطلحي هو تعريب المصطلح، لكن بشروط نعتقد أنها ذات طبيعة لسانية أكثر منها هندسية، حيث يتطلب الأمر من المعرب الإحاطة بالجانب الصواتي للغتين: الأصل والهدف، لكي يسهل الرجوع بالمصطلح إلى أصله الروماني عند الضرورة، من ذلك مثلا مصطلح ميتالغة الذي يجمع بين التعريب والترجمة، ميتا وتعني وصف، وهي ترجمة لنصف المصطلح Meta-language، أما نصفه الثاني فمترجم: لغة، وقد اجتمع التعريب والترجمة في مصطلح واحد، ومعناه اللغة الواصفة، أي المصطلحات التي تستخدم في معالجة أي ظاهرة لسانية. إلا أن التعريب الكامل للمصطلح يكثر في ميدان الطب والصيدلة، خاصة أسماء الأدوية التي تنسب، في الغالب، إلى أسماء مخترعيها، أو أسماء الأمراض التي تعالجها، وفي هذه الحالة تستحيل الترجمة، خلافا لمجال هندسة اللغات الطبيعية التي تلتقي مع علوم كثيرة في المفاهيم الإجرائية التي يوظفانها معا كما سنرى.

3. لسانيات المنصات:

سنقدم، فيما يلي، الطريقة التي توظف بها الصورنة اللسانية المصطلحات العلمية التي يتم توظيفها حاليا في هندسة اللغات الطبيعية Natural Language Processing. يندرج هذا العمل في إطار مشروع علمي كبير، ينجز تحت مظلة لسانيات الجيل الرابع linguistique de la 4ème génération أو لسانيات المنصات Linguistique des plateformes، التي تعد قوام مجتمع المعرفة Knowledge society. هذا الجيل الذي أصبحت فيه اللغة تتولّى مهمة فلترة المعارف

وصورنتها وفق أحدث المعايير المعرفية، تتفاعل مع أحدث التّقانات الرقمية من أجل الوصول إلى إنتاج معرفي تجريبي Emperical، يبتعد عن التجريد Abstraction الذي طبع البحث اللساني على مدى القرون التي خلت من عمر النبش في بنية اللغة، أي لغة، فأصبحنا نتعامل اليوم بمفاهيم جديدة لم يفكر فيها دارسو اللسانيات بالمعنى التقليدي، لأن الأدوات الإجرائية التي كانوا يوظفونها في مقاربة مصدر اللغة في الدماغ البشري، لم تلامس في مجملها جوهر إنتاج العملية اللسانية في الدماغ البشري، مما جعلها تعجز عن نقل الكفاية البشرية إلى خارج محيطها البيولوجي Biologic، وهو أحد مطالب البحث العلمي في عصر الانفرادية Singularity والرؤية الثالثة Therd vision[5]، التي تعتمد تقانات الذكاء الصناعي Artificial Intelligene لمّا جعلت من إنسان هذا العصر كائنا مهجّناً بامتياز Hybride، حسب رأي صاحب كتاب "حرب الذكاءات"[6].

تسهم لسانيات الجيل الرابع في إنتاج المعارف الجديدة من خلال مصاحبة النظريات الرقمية الجديدة في موطنها المعرفي الأصلي نظراً ومنهجاً، إنها تشتغل في المنطقة نفسها التي يشتغل فيها مهندسو الذكاء الصناعي، في مختلف مجالات تطبيقه في الحياة البشرية، خاصة في القطاع الصناعي الذي يعتمد اللغة الطبيعية في بناء نفسه وفق متطلبات التقانة الرقمية الجديدة. كما أنها تصاحب الذكاء الصناعي في عمله داخل المربع المعرفي الذي يرمز إليه بـــ NBIC[7]. ومع التطور الحاصل في هذه المجالات، فإن البشرية ما زالت إلى اليوم تجتهد لتجاوز مرحلة الذكاء الصناعي الضعيف IA faible، المنشغل بتقانة التّعلم العميق Deep learning[8]، في انتظار استكمال التّوصيف الدقيق للبيانات، وجمعها وفق

5 - Le mythe de la singularité, Jean-Gabriel Ganascia, Seuil 2017

6 - La guerre des intelligences : Intelligence artificielle versus Intelligence humaine, Dr. Laurent Alexandre, JCLattès, 2017.

7 - N : Nano technology, B : Bio technology, I : Informatics, & C : Cognition

8 - Comprendre le Deep Learning, Jean Claude Heudin, Science e-Book, 2016

مقتضيات لسانيات الجيل الرابع، لبلوغ ما يطلق عليه البيانات الكاملة Full data، التي ستؤهله للانتقال إلى الذكاء الصناعي القوي IA forte وقَدْ ضربت له البشرية موعداً سنة 2030م، حينها سيشرع العلماء في تطوير الجيل الثالث من الذكاء الصناعي الذي يطلق عليه جيل الانفرادية[9]، وقد وضع خبراء الذكاء الصناعي برنامجا يصل مدى تنفيذه إلى سنة 2045م، وهو الذي سيؤدي إلى انتقال برنامج التطوير كليا من البشر إلى الآلة. آنذاك لا ندري ما الذي سيصبح عليه الإنسان المهجّن كلياً حسب رأي علماء الذكاء الصناعي القوي[10]. في كل هذه المراحل، نجد لسانيات الجيل الرابع تسهم في صياغة المفاهيم الجديدة التي توصف مصطلحيا هذا الانتقال المعرفي، وبدونها سيبقى العمل التّقاني مغيبا عن التداول.

هذه هي الأجواء العلمية التي تشتغل فيها لسانيات الجيل الرابع أو هندسة اللغة، إنها تواكب البحث الهندسي وتسهم في تطويره من حيث اشتغالها على الأطر الصورية التي يقوم عليها النظام اللغوي العام، وأن أي عمل هندسي في الذكاء الصناعي، من أي نوع أو درجة كان، لا يستقيم إلا من خلال ضبط العملية اللغوية صناعياً وبيولوجيا، وبدون هذا التفاعل الصناعي مع المكون اللساني سيكون الأمر ضربا من ضياع الوقت، ولن يمكن خبراء الذكاء الصناعي من نقل الكفاية المعرفيةCognitive competence البشرية كاملة إلى الآلة، وأي خلل في فكّ شفرة اللغة Linguistic code سينعكس سلبا على بقية المكونات. لهذا كانت هندسة اللغة في طليعة العمل الهندسي المستقبلي.

9 – أحدث أخيرا (2013) في جامعة الملك سعود كرسي علمي أطلقت عليه "كرسي الانفرادية والرؤية الثالثة" تطبق فيه تقانات الذكاء الصناعي، وقد نشر فريق البحث العامل في هذا الكرسي عدة أبحاث علمية تربط جميعها بين صورنة اللغات الطبيعية والذكاء الصناعي، مستخدمين مصطلحات لسانيات الجيل الرابع..

10 - Les vies secrètes de cerveau, Davis Eaglman, 2016

كل هذا يؤكّد أن اللّسانيات لم تعد ذلك العلم الذي يبحث في بنية اللغة ابتداء من الصوت إلى التركيب، مرورا بالمستوى الصرفي – الاشتقاقي، فقد اجتهد لغويونا القدماء في وضع الإطار العام لتحديد خريطة بنية اللغات بوصفها أداء للتواصل بين البشر، مما يعني التركيز على الجانب الشكلي، ثم انتقل البحث في اللغة بوصفها أشكالا منضدة خطيا، إلى البحث في اللغة بوصفها قواعد مضمرة في كفاية الإنسان، ونظرا لأن البحث انصب على الافتراضات وليس على البيانات اللغوية الحقيقية (التوليدية) نموذجا، فقد بقيت هذه اللسانيات في إطار التجريد الذي لم يؤد إلى إنتاج أي تطبيق يبرز دور اللغة في التطور المجتمعي، أي لم تتمكن من إنتاج تطبيقات تواكب التّطور المعرفي الحاصل في مجال المعرفة الصلبة. ومع تطور المعارف الهندسية (عتاداً وبرمجةً) التي أصبحت تبحث في عمق الدماغ البشري بهدف محاكاته بنقل قدراته إلى الآلة، أصبح لزاما على البحث اللساني أن يقود عملية البحث في هذا المجال، فانتقل السؤال من كيف بنيت اللغة في كفاية الإنسان، إلى كيف يمكن نقل هذه الكفاية إلى الآلة ؟ وللإجابة عن هذا السؤال بات لزاما على الباحثين في اللسانيات التعاون مع مهندسي الحواسيب لفهم السؤال أولا، ثم الشروع في تنفيذ برنامج العمل لاحقا. هكذا ظهرت لسانيات جديدة توظف معارف هجينة تجمع بين الفهم اللساني للغة، والفهم الهندسي الذي تقوم عليه صورنة البيانات اللغوية. هذا التزاوج بين الأساسين المعرفيين أنتج توجها جديدا أطلق عليه لسانيات المنصات، وهو جيل صورنة البيانات تنفيذا لمبدأ الاستقراء الشامل، في إطار منصة حاسوبية تقوم على تقانات الذكاء الصناعي في مختلف أوجهه. بهذا المبدأ تمكنا من بناء معاجم إلكترونية مشفرة تستند على تطبيقات جديدة متعددة الأغراض، فضلا عن تطوير أنحاء صورية تقوم على مسوقات Automatom، تتكون من عصبونات Axones ومن تفريعات عصبونية Dendrites تربط بينها لواقط Transducers تحاكي تفريعات الأعصاب في الدماغ البشري، وتقوم بمهمة إنتاج البيانات اللسانية في الدماغ الطبيعي.

وهكذا أصبح مفهوم النحو عبارة عن آلة تحاكي الدماغ الطبيعي، إلا أنه نحو لا يستقيم بدون توافر معجم إلكتروني يزوده بالبيانات المشفرة: صوتا، وصرفا، ونحو، ودلالة، كما لا يمكن لهذا النحو أن يشتغل خارج إطار المنصة الهندسية التي تؤطره، من هذا المنطلق نشير إلى أننا نستخدم المنصة الهندسية Nooj القائمة على مبادئ نظرية المعجم التركيبي Lexicon – grammar التي تعد الواجهة اللسانية للمنصة الحاسوبية، الأمر الذي يؤكد درجة التعاون بين اللسانيين والحاسوبيين في مسار بناء هندسة لسانية جديدة.

تطلب الوضع الجديد الذي أصبحت تشتغل فيه اللسانيات الجديدة استخدام مفاهيم جديدة تنبع من صلب التطور الهندسي- اللساني الجديد، من قبيل: التعلم الآلي Machine learning، والتعلم العميقDeep learning، والبيانات الضخمةBig data، والبيانات الكلية Full data، والتهجينHybridation، والانفرادية Singularityوالتنقيب عن النصوصText mining، والمنصة Platform، والرسوم الهندسية graphs، والنحو المحلي Local grammar، والمعجم الإلكترونيElectronical dictionary، والأنطولوجيا Ontology، تعرف الكلام البشريSpeech recognition، والشبكة العصبية Neurones-network وغيرها من المفاهيم الجديدة التي تجمع بين النظر الهندسي والنظر اللساني، إنها مفاهيم ذات دلالات جديدة لم يكن لها وجود في الأجيال الثلاثة السابقة[11].

بناء على هذا التطور الأسي Exponential الذي يعرفه البحث في أنظمة اللغات الطبيعية، فإن السؤال الذي يطرح نفسه اليوم هو: هل ما زال هناك مجال لمواصلة البحث في النظريات اللسانية Linguistic theories التي تصنف في عداد الجيل الثالث القائم على التجريد، بما فيها تلك التي تجتهد في وضع مفاهيم جديدة محاولة منها في اللحاق بالتطورات الجديدة التي تعرفها هندسة اللغات

11 – انظر مقدمة المجلد الخامس عشر من مجلة التواصل اللساني، 2015

الطبيعية[12]، لا أظن ذلك، لأن المنطلق المعرفي مختلف إبستيمولوجيا Epistimology، وبالتالي فلا مجال للالتفات إليها لجعلها تسهم في بناء معارف جديدة إلا في حدود ضيقة مع أخذ الحذر من نتائجها على تغيير مسار البحث العلمي والدفع به في اتجاه التقليد الذي يستنزف طاقات باحثينا، والشباب منهم بخاصة. كل ما تقدمه هذه اللسانيات للبحث العلمي يبقى مجرد مصطلحات عاجزة عن التفاعل مع أي تقانة حقيقية في الواقع.

هكذا تشكلت نقطة الالتقاء بين لسانيات الجيل الرابع، ممثلة في نظرية المعجم التركيبي Lexique-grammaire مع لسانيات المنصات Linguitic platforms، التي انبثقت من النموذج الرياضي لهاريس Mathematical model ويعد الأساس النظري والمنهجي الذي بنيت عليه إحدى أشهر المنصات الحاسوبية المتخصصة كليا في معالجة اللغات الطبيعية، ومن ضمنها العربية، ونعني بها منصة نووج Nooj platform بوصفها أحدث منصة هندسية–لسانية فاعلة في صورنة اللغات الطبيعية Linguages formalization في أدق تفاصيلها ومستوياتها، وقد أصبحت معيارا عمليا لتقدم البحث اللساني في جانبه الهندسي، حيث قدمت لها نظرية المعجم التركيبي المفاهيم الإجرائية اللسانية التي انطلقت منها، وطورتها بشكل جعلها تسابق الزمن الهندسي الذي يتطور بطريقة أسية progress Exponential وفق قوانين موور G. Moore[13]، فحققت، بذلك، قفزة نوعية في الصّورنة التي صاغتها هذه المنصة على شكل معاجم إلكترونية وأنحاء محلية، متخذة من الرمزية الرياضية– الهاريسية القائمة على مبدأ الطاكسينومي

12 – نقصد بها بعض الطروحات الجديدة في مجال البحث اللسانين من قبيل مصطلح: برنامج الحد الأدنى Minimalist program الذي ينظر لحوسبة اللغات الطبيعية في الكفاية اللسانية للمتكلمين، مثل هذه المفاهيم تفقد مصداقيتها بمجرد عرضها على التجريب Empirisme، أي بناء تطبيقات لسانية حاسوبية يمكن استغلالها في أي عمل وظيفي في الحياة.

13 – وضع غوردون موور مطور شكة إنتيل سنة 1965، قوانين جديدة للزيادة في سرعة الحواسيب، معتمدا على مبادئ فيزياء الكم، وتوظيف المنطق الضبابي Fuzzy logic، وقد أصبحت التقانة تتطور بموجب هذا القانون تتطور بمعدل 18 شهرا، بعد أن كان تطورها يستغرق أجيالا.

Taxinomy في بعديه التوزيعي Distribution والتصنيفي Classification أساسا لتطوير كافة التطبيقات الموظفة في تقانة التنقيب عن النصوص وتحليلها، والتعليم المبرمج تفاعليّاً، ومعالجة المدونات Corpora analysis على مختلف الصّعد اللسانية، وغيرها الكثير من التطبيقات العملية التي رأت النور مع تساكب السنين المنصرمة، لتوّها من عقم البحث اللساني الحاسوبي، وما زالت مستمرة في التطور الذي يواكب تشكل الجيل الثالث من الذكاء الصناعي، المعتمد أساسا على مفهوم البيانات الكلية Full data، الوليد الشرعي لتطور مفهوم البيانات الضخمة Big data، وهو ما تعول عليه الأبحاث الهندسية اللسانية لمعالجة هزال الاكتساب المعرفي الذي عمر طويلا مع النظريات اللسانية التجريدية Abstraite، والتي لم تنفع معها العمليات التجميلية الممثلة في مصطلحات موغلة في التجريد، إذْ غالبا ما تولد، وفي أحشائها أسباب موتها، لافتقارها إلى قوة التجريب المحوسب الذي يمثل الوقود العملي لاستمرار أي نظرية على قيد الحياة.

4. لغة الضاد هندسياً:

لقد أصبح العمل اللساني يقوم على مفاهيم تجمع بين النظر اللساني بمختلف توجهاته وتطبيقاته على جميع مستويات النظام اللغوي، والنظر الهندسي الذي يتطلب إعادة توصيف اللغات الطبيعية وفق منطق التقطيع الحاسوبي Shunking لمكونات بنية اللغة، حيث أصبحت المفاهيم اللسانية تكتسب دلالات جديدة، مثل مفهوم المعجم الذي أصبح مختلفا كليا عن مفهومه القديم. ينسحب مفهوم المعجم في هندسة اللغة ينسحب على كافة المستويات اللسانية، من الصوت فالصرف فالتركيب والدلالة، وانتهاء بالجانب التداولي، تجمع كل هذه المستويات في منظومة واحدة يطلق عليها المعجم الإلكتروني Electronic dictionary، وَقوامه مداخل لغوية مختصرة في الجذر والوزن، العمدتان الأساسيتان للغة الضاد في مستواها الصّرفي الرباعي التقسيم:

ف (الفعل) س (الاسم) مص (المصدر) صف (الصفات)، وقد تم ضبطها بطريقة تسمح بتطوير معجم تركيبي لنظام العربية، القائم بدوره على متوالية نظرية محدودة تتمثل في: ج=: ف س ك، تتفرع إلى عدد نهائي من الجمل الأساس تطبيقا لمقتضيات مبدأ الانتشار النهائي Maximal Projection، وتنتهي إلى معالجة البيانات التركيبية الكلية Full data الموضوعة في أجروميات Matrix، من خلال تطبيق أربع عمليات إجرائية نلخصها في: البناء المقلوب Passif والبناء التوصيفيAdjectivation والبناء الموسم Nominalization، والبناء المتهيكلRestructuration، ولكل واحدة منها نموذج أجرومي يتكون من قوانين اختيارية Selection يمارسها الفعل على بقية المكونات الصرفية – المعجمية لبناء متوالية لغوية مسبوكة بطريقة تكشف عن مكنون الكفاية اللسانية بطريقة رياضية – هاريسية. هكذا، يصبح الأجروم صورةً مصغرة عن التشكل الهندسي للغة في دماغ الإنسان، وهو ما تسعى منصة نوج إلى الكشف عنه من خلال قراءة البيانات المشفرة التي يتضمنها المعجم الإلكتروني بواسطة المسوقات Atomatom التي يبنيها وفق نحو محلّي قوامه عوامل Operators وخوارزميات Algoritms بنوعيها: التوليد Genaration والتحليل Analysis.

تنفيذا لهذا المخطط المعرفي الجديد في البحث اللّساني، أصبح الفعل جملة بسيطة، لأنه يتحكم في سائر العناصر المتوزّعة معه، وأصبح يتفرع إلى ثلاثة أقسام: 1. جملة فعل عاد، يتحكم في سائر مكونات الجملة، 2. وجملة فعل عماد لا يجمعه بالنوع الأول إلا الجذر والوزن، أما الوظيفة الفعلية فتضمحلّ بفعل عملية التوسيم المشار إليها أعلاه 3. وثالث أنواع الفعل هو الجملة المسكوكة Idioms، وهذه الأخيرة نوع من التعابير اللسانية التي ظلت عصية على الحصر والضبط في متاهات النظر اللساني التجريدي العتيق، بأجياله الثلاثة وقد ظهرت الحاجة إليها عندما اصطدمت تطبيقات الترجمة الآلية بهذا النوع من المتواليات اللسانية العصية على التأليف الدلالي المفتوح، كونها تنماز بخاصية اللاَّتأليف

Non compositionalité وبالتعتيم الدّلالي Opacity الناتج عن سك المكونات التأليفية اللسانية فيما بينها، مما أخرجها من دائرة المتن المعجمي التقليدي، المبني أساساً على الوحدات المفردة الدالة، في حد ذاتها، حتى قبل أن تدمج في شكلها الهندسي المضغوط تركيبيا، وعند البحث عن العراقيل التي تقف في وجه استكمال هذا التطبيق، تبين أن هناك أموراً كثيرة يجب أن تحل، قبل بناء نموذج لساني – هندسي جديد شامل لنظام لغة الضاد بجميع مكوناته ومستوياته التصاعدية، من الصوت المفرد إلى التأليفية التركيبية، فأنجزت دراسات جديدة صبت فيها خبرات صورية أكثر منها لسانية، ذهبت إلى حد بناء معجم آلي لهذه التعابير بلغ عددها ما يناهز الثلاثين ألف تركيب مسكوك، معظمه مستخدم في لغة الضاد الوظيفية اليوم. بعد الانتهاء من صورنة متن هذه التعابير، ووضعها في أجروميات، أصبح في الإمكان الشروع في حوسبتها وفق أحدث منصات البحث اللساني الهندسي الجديد، فكانت منصة نوج هي الأنسب له، بوصفها الأكثر ملاءمة لنظام لغة الضاد[14].

ولأن هذه التعابير مصمَّمة في دماغ متكلي، أي لغة متخدة صيغة أشكال هندسية غير قابلة للتفكيك إلا بشكل جزئي، بل إنها عصية على التأويل المتسرع، فإن رصيدها يختلف من لغة إلى أخرى، إذ من الصَّعب العثورُ على أحدها منسوخا حرفيا بين لغتين، لأن كل لغة تملك منها رصيدها الخاص الذي لا يتكرر لفظيا في لغة أخرى، ولو كانت من الفصيلة اللغوية نفسها، مما جعل الانتقال من لغة إلى أخرى عن طريق الترجمة، يزداد صعوبة كلما زادت الفرقة والاختلاف بين أرصدتها في كفاية المتكلمين. هذا الوضع الذي وجدت عليه هذه التعبيرات يطعن، بشكل أو بآخر، في المفهوم التقليدي للإبداعية اللغوية، خاصة عند دعاة لانهائية اللغة وكونيتها، إذ لو كان الأمر كذلك لما اختلف رصيدها من لغة إلى أخرى، وأحيانا كثيرة يتعذر العثور على المقابل المسكوك

14 – ينظر علي بولعلام، لسانيات المنصات، 2018

لأحد التعابير في لغة ثانية، إن لم يكن أكثر من لغة ثانية وثالثة ورابعة وهلم جرا، وذلك لأن عملية سكّ التعبير مسألة معرفية Cognitive أكثر منها عملية لغوية عادية، حيث إن عدد هذا النوع من التعابير يتكاثر في لغة ما حسب درجة توظيفها في مختلف السياقات التداولية Pragmatic. وبالنظر إلى ارتفاع نسبة ترددها في جميع اللغات، فإن أيّ عمل في هندسة اللغات الطبيعية يجب أن يركز على آليات بنائها من أجل إدماجها في منظومة البحث الهندسي المتعلق بلغة الضاد.

5. نماذج تطبيقية:

نقدم في هذه الفقرة بعض النماذج الأكثر ترددا في منشورات خاصة بهندسة اللغات الطبيعية والتي تبين ما سيكون عليه معجم أو موسوعة مفاهيم هندسة اللغات الطبيعية، مطبقة على لغة الضاد في إطار منصة نوج والمعجم التركيبي، من خلال :

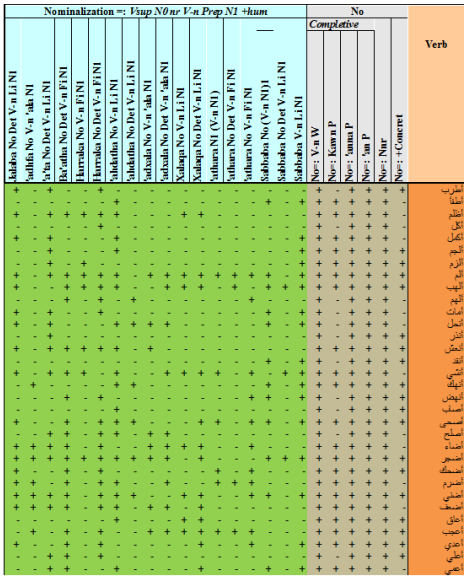
1. الأجروم Matrix؛
2. المعجم الإلكتروني Electronical dictionary؛
3. المسوقات Automata؛
4. الكشاف السياقي Concordancier.

1.5. الأجروم Matrix:

جدولٌ مصمّم لتصنيف البيانات اللغوية Language data من منظور طاكسونومي Taxonomy اعتمادا على خصائص لسانية تتمحور حول مكونين: ثابت Constant لغوي، غالبا ما يكون هو المحمول Predicate الذي تتمحور حوله بقية الموضوعات Arguments التي تتألف منها الجملة البسيطة، ومتغيرات (الأسماء) يتم اختيارها من قبل الثابت بشكل خطي Linear، حيث يوضع كل متغير في مكانه المناسب في المتوالية اللسانية التي تم شرح تكوينها الرياضي سابقا، أي من الفعل إلى الفاعل إلى المفعول، وهكذا.

يتكون الأجروم من أعمدة وأسطر وعوامل:

1. الأعمدة Columns ويحتوي على البنيات Structures التركيبية[15] التي تحدد العلاقات التوزيعية Distribution بين مكونات المتوالية اللسانية البسيطة، حيث يتولّى الفعل اختيار العناصر الاسمية التي تتألف منها الجملة، كما يتضمن بنيات العمليات الاشتقاقات التركيبية Syntactical derivation والعمليات التحويلية Transformation التي تخضع لها المتوالية اللسانية تطبيقا لمبدأ الإطناب التركيبي Syntactical paraphrase[16].



الشكل رقم: (1) أجروم التركيب[17]

2. الأسطر وتتضمن النتائج Output المتوصل إليها بتطبيق البنيات التركيبية، الواردة في الأعمدة، على الدخل Input وهنا الفعل، وَيتم التعبير عن هذه النتيجة بعاملي الإيجاب (+) والسلب (–)، كما هو ظاهر في وسط الأجروم.

3. العوامل Opertors وتمثل التطبيق الصحيح للأعمدة على الأسطر، وَتظهر إجباريا عند نقطة التقاطع بين المحورين: العمودي والأفقي، يمثل العامل (+) التطبيق الإيجابي للبنية اللسانية على الدخل / الفعل، بينما يمثل العامل (–) التطبيق السلبي، أي النتيجة الصفر، مقابل النتيجة واحد التي

15 – نميز بين البنية والجملة، البنية إطار شكلي عام، يولد عددا كبيرا من المواد اللغوية، مثل الوزن الذي يولد عددا كبيرا من المفردات، بينما الجملة تطبيق عملي للبنية، وهي لا تولد إلا نفسها.

16 – ينظر كتابنا: هندسة اللغة العربية، ج: 1، 2018

17 – نموذج من الأجروميات الواردة في كتاب الحناش: هندسة اللغة العربية ج: 2، 2018

يمثلها العامل الإيجابي. وقد تمّ الاقتصار على هذين العاملين تطبيقا لمقتضيات جبر جورج بول المشار إليه سابقا. وعليه فإن الأجروم يمثل نسخة مصغرة من الجهاز اللساني الموجود في الكفاية البشرية التي تشتغل بعاملين فقط هما: - Not and لا أو (نعم)، وهو ما يقوم عليه العتاد الحاسوبي الذي يوظف فقط عاملي: 0، 1. بهذه الطريقة يتم توصيف اللغة دونما حاجة إلى مدونات Corpora، أو الاعتماد على وضع قواعد افتراضية ينقصها التجريب.

2.5. المعجم الإلكتروني Electronic dictionary[18]

المعجم الإلكتروني قاعدة بيانات شاملة للغة، مشفرة Crypted، موصفة بلغة عقلانية – منطقية تقوم عليها منصة حاسوبية معدة لمعالجة اللغة الطبيعية آليا، يتكون من دخلات لغوية على شكل مواد لغوية مباشرة، كل سطر يتضمن الأوصاف الصرفية والمعجمية والتركيبية التي تتعلق به، يتم فيه تداخل الجذر مع والوزن لإنتاج الكلمة العربية الصحيحة، جميع المفردات توصف صرفيا ونحويا من خلال الحركات الإعرابية المحتمل أن تأخذها في السياق التركيبي الذي تظهر فيه، ودلاليا حيث يظهر مع الكلمة توصيفها المعجمي الذي يحدد دلالتها في مختلف السياقات اللغوية. والنموذج التالي يبين الطريقة التي بني بها "معجم العرفان الإلكتروني" للغة العربية:

```
زَأْبVN+asl+فَعْل+FLX=Flex
زءب زَأْب NW+DRV=V_C26+DRV=010فِعْلَة:FlexT+DRV=010
فَعْلَة:FlexT+DRV=010مَفْعَل:Flex+DRV=010ف1:
Flexdrv+DRV=010:1مFlexdrv+DRV=010:2غFlexdrv+DRV=010
```

18 – المعجم الإلكتروني المصمم من خلال منصة حاسوبية يختلف عما بات يعرف بالمعاجم الإلكترونية التي توثق البيانات المعجمية على أقراص مضغوطة، مع تزويدها ببرامج استرجاع البيانات كما هي في مظانها الأصلية، فهذه الأخيرة يمكن أن يطلق عليها التوثيق الإلكتروني، بينما المعجم الذي نتحدث عنه يمثل نسخة هندسية لكفاية المتكلمين باللغة الطبيعية.

ض:Flexdrv+DRV=010؛Flexdrv+DRV=010:3ك3:
Flexdrv:6آFlexdrv+DRV=010
زَأْبFlexdrv+DRV=VN=صناعي1:FlexT

الشكل (2): نموذج من "معجم العرفان الإلكتروني"، يُقرأ من اليسار إلى اليمين، ويتضمن 80 مليون كلمة عربية.

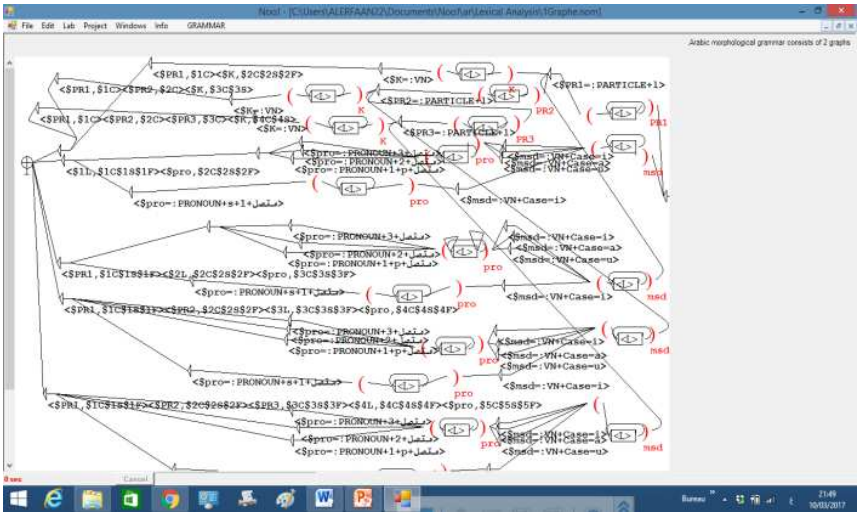
بهذه الطريقة سيكون مصطلح المعجم الإلكتروني تمثيلا هندسيا لكفاية المتكلمين باللغة الطبيعية القائمة على خوارزميات التوليد والتحليل، ويمكن أن يستخدم في مختلف التطبيقات الهندسية للغات الطبيعية، وبدون هذا المعجم فإن أيّ عمل في هندسة اللّغة سيجانبه الصّواب؛ ولذلك، فإن جميع المنصات الحاسوبية المعروفة في العالم تبادر، قبل الشروع في التطبيق على أي لغة طبيعية، إلى بناء معجم إلكتروني للغة المراد معالجتها آليا ([19]).

3.5. المسوقة Automatom

وَهي عبارة عن آلة تتولَّى توليد البيانات اللّغوية الموصفة في المعجم الإلكتروني السابق، تتخذ شكل مكونات الدّماغ البشري من حيث التغصنات Dendrites والمحاور العصبية Axones، ترتبط بواسطة عقد على شكل لواقط Transducers، وهي النقط التي تتجمع عندها الأعصاب لتتفرق إلى غصينات فرعية. يمثلُ الدخل الوحدة اللسانية الأساس التي تتولد منها مجموعة من المتواليات اللسانية الفرعية، مثل الجذر + الوزن =: التي يتولد منها الدخل كَتَبَ مثلا، حين تتولد انطلاقا منه مجموعة من المداخل الفرعية: كاتب، مكتوب، كتاب، إلخ. تتم عملية التوليد بتطبيق مجموعة من الخوارزميات Algoritms على الدخل، فينتج عن هذا التطبيق دخل فرعي جديد، تتوافر فيه جميع المواصفات الصرفية والدلالية والمعجمية التي تمّ وصفها في المعجم الإلكتروني السابق.

19 – ينظر الحناش 1988، و2018.

تعدّ المسوقة الواجهة المنطقية الخوارزمية للمعجم الإلكتروني، المبني بدوره بلغة واصفة Metalanguage من صميم المنصة الحاسوبية التي تتطلب التكامل بينهما، وبدون هذا التكامل سيصعب الحديث عن المعالجة الآلية للغات الطبيعية.

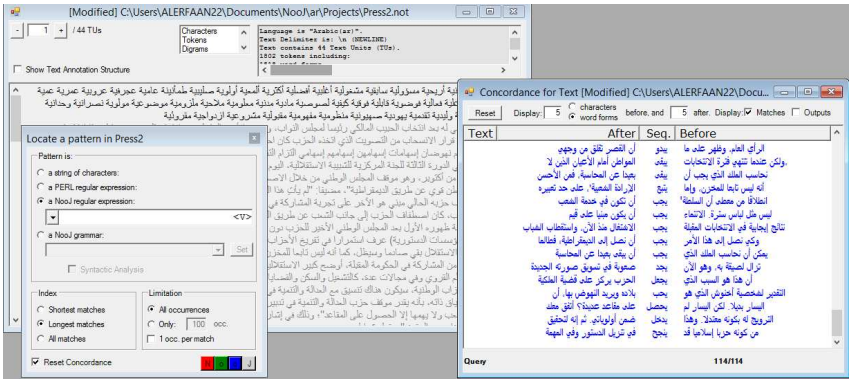


الشكل (3): مسوقة المفردات العربية الخاصة بمداخل "المعجم العربي الإلكتروني"[20]

4.5. الكشّاف السياقي Concordocier

هو نتيجة تطبيق المعجم الإلكتروني والمسوقة على أي مدونة لغوية مختارة عشوائيا، سواء من الشابكة مباشرة، أو من أي مصدر إلكتروني آخر. من خلال واجهة المنصة الحاسوبية يتم إدخال القيمة المراد البحث عنها في المدونة، ليتم عرضُ النتائج كما في الشكل التالي:

20 – ينظر الحناش 2018



الشكل (4): كشاف سياقي للبحث عن الأفعال <V> في نص مأخوذ مباشرة من الشابكة.

يقدم هذا الكشاف كما يلي: في خلفية الكشاف نص مختار عشوائيا طوله حوالي 15 صفحة، وعلى اليسار نافذة البحث التي وضعت فيها قيمة البحث (V>[21]<، أما على اليمين فتظهر نتيجة البحث على شكل قائمة من الأفعال التي عثر عليها الكشاف في النص وعددها 114، وقد تمّ وضعها في سياقها النصي الذي وردت فيه في الأصل. يسمح هذا الكشاف لدارس النصوص بالحصول على معلومات قابلة للتصنيف، خاصة وأن المنصة التي يرتبط بها تقدم إمكانية النتائج الإحصائية، وهو ما يساعد على معرفة دلالة النص. يشار أيضا، إلى أن الكشاف السياقي يستخدم بكثرة في استطلاعات الرأي التي تتأسس عليها السياسات الاقتصادية والاجتماعية للدّول.

نكتفي بعرض هذه المفاهيم الأربعة التي تستخدمها منصة نوج وقد أصبحت لها امتدادات كثيرة تجاوزت العلوم الإنسانية إلى مجالات العلوم الصلبة، مما يؤكد على أن أيّ عمل في هندسة اللغات الطبيعية، لا ينبني على منصة حاسوبية، تستند على منصة لسانية صورية سيكون جريا وراء المجهول.

21 – يمكن أن نضع في هذا المكان أي قيمة موصفة رقميا في المعجم الإلكتروني، بما في ذلك البنيات التركيبية، مثل: ف س° س1 أو غيرها.

6.تقويم Evaluation

لتأكيد أهمية وضع موسوعة أو معجم موسوعي لمصطلحات هندسة اللغة التي يتم توظيفها حاليا في هذا المجال، قمنا بمسح مصطلحي لبعض البحوث العلمية المنشورة باللغة العربية في مجال هندسة اللغة العربية[22]، فتبين أنها تتوزع على مجموعة من المجالات الفرعية في هذا العلم، كل مجال فرعي يتطلب دراسة مفصلة تبرز طريقة تعامله مع أنظمة اللغات الطبيعية، لكل واحد منها مفاهيم علمية خاصة به. وقد لاحظنا أن النسبة الكبيرة من المفاهيم تستخدم في مجال اللسانيات الصورية Formal linguistics (23 ٪)، والمقصود بها الدراسة المنطقية لبنيات اللغة، حيث تركز على جانبها الرياضي، وهذا أمر طبيعي، لأن الهندسة اللسانية تتعامل بالأشكال الصورية Formal التي جاءت منها Formalization، أي الصورنة التي تقوم على منظومة من التوصيفات الشكلية للغة، وَتتأسس في مجملها على مفهوم البنية الصورية Structure، وهذا المصطلح لا علاقة له بالجملة كما قد يفهمه البعض، وبواسطته يتم إعادة توصيف المستويات اللغوية، بوصفها أشكالا منطقية قادرة على إنتاج المكونات اللسانية الطبيعية، فالدماغ البشري، حسب هذه النظرية، يقوم على منظومة من القواعد الصورية التي تنسج عليها مواد لسانية تستخدم في التواصل، وليس على مواد لغوية صوتية. يفيد هذا التحليل أن منطلق هندسة اللغة هو الصورنة اللسانية التي تبنى عليها تطبيقات حاسوبية، وبدون هذه المرحلة سيكون من الصعب العمل على اللغة انطلاقا من التوصيفات التقليدية. وعليه، فإن أي حديث عن هندسة اللغة يعني صورنة اللغة الطبيعية بداية، وتطوير التطبيقات الحاسوبية لاحقا، وأي خلل في الصورنة سيؤدي، حتما، إلى فشل التطبيق الهندسي، وهذا يفسر ارتفاع نسبة المفاهيم اللسانية الواردة في عينة المصطلحات السابقة.

22 – تم التطبيق على بحث الحناش المنشور في ديسمبر 2017، في الكتاب الجماعي الذي أصرته المندوبية الدائمة للملكة العربية السعودية في باريس، عنوانه: التعلم الآلي واللغة العربية: تطبيق منصة نوج على معجم العرفان.

تأتي نسبة تردّد المفاهيم الهندسية المتعلقة بالذكاء الصناعي في الدرجة الثانية (19٪)، لتؤكد العلاقة بين المجالين، وأغلبها يحمل دلالات رياضية لكن بشحنة لسانية، مما يدل على أن البحث موضوع الفحص يبني الفرضيات اللسانية على منظومة من مفاهيم الذكاء الصناعي المرتكز أساسا على ترسانة من الخوارزميات الرياضية. في الترتيب الإحصائي الثالث نجد مصطلحات الحاسوبيات العامة (14٪)، على اعتبار أن مفاهيمها ما تزال سارية المفعول، ولم تنطمس نهائيا بعد، مما يؤكد على استمرارية العمل بها وبمصطلحاتها التي تأسست عليها التطورات الهندسية الجديدة. في الترتيب الرابع (9٪) نجد مفاهيم الذكاء الصناعي الخاصة بمنصة نوج، بينما تحتل مصطلحات خاصة جدا بهذه المنصة الدرجة الرابعة (6 ٪)، مما يعني أنها منصة تتفاعل، كليا، مع مختلف المفاهيم الضاربة في عمق العلوم الصلبة، خاصة الحاسوبيات التي تشكل أحد أركان NBIC المشار إليها، سابقا في هذا البحث.

أما بقية المفاهيم التي استخدمت في هذا البحث، فتتوزع ما بين الإجراءات المتبعة في توصيف الموارد اللسانية، وبعض المفاهيم العامة في الدراسات اللسانيات الصورية، إلا أنها لا تخرج عن الإطار العام المرسوم للبحث، وهو تعلّم الآلة الذي يعدّ فرعا من الذكاء الصناعي، وصنو تقانات التعلم العميق والشبكة العصبية، وما يدخل في فلكها.

7. خلاصة البحث ونتائجه:

في هذا البحث الموجز، تمت معالجة قضية المصطلح العلمي الذي يستخدمه الباحثون في هندسة اللغات الطبيعية مع التركيز على اللغة العربية، وقد تم استعراض التطورات التي عرفتها اللسانيات التي تستقر اليوم عند الجيل الرابع، وهو الجيل الذي يمتزج فيه عمل اللغوي مع عمل المهندس، كما تم عرض التطورات الأخيرة التي عرفها البحث في الذكاء الصناعي وعلاقته مع اللغات الطبيعية، مع التركيز على المفاهيم الخاصة بمجال هندسة اللغة عامة. كما

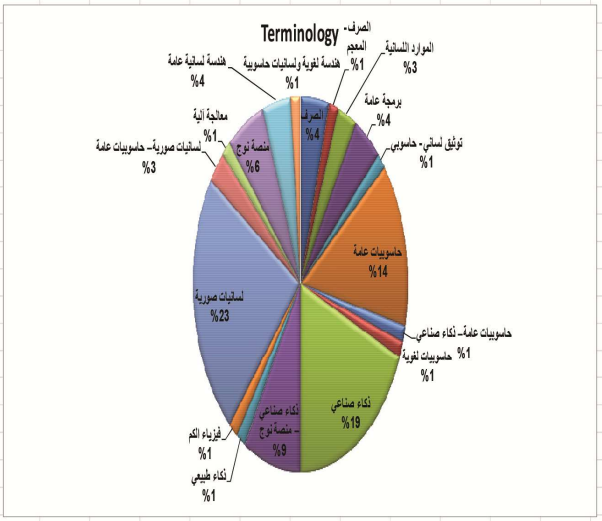
تم تقديم نموذج مصغر لما يمكن أن يكون عليه معجم المصطلحات العلمي التي يستخدمها هذا العلم وهو ينتج معظم مفاهيمه بالإنجليزية، كما تمت معالجة إحصائية لأحد المقالات الصادرة حديثا في هندسة العربية، حيث تمّ عرض النتائج في ملحق خاص إنجليزي – عربي، مع تحليلها وفق الطرق المتعارف عيها في هذا المجال.

8. ملاحق البحث

1.8.الملحق رقم: 1

رسم توضيحي بنتيجة إحصاء توزيع المصطلحات العلمية في بحث محمد الحناش 2017

Terminolog	Numbers
الصرف	3
الصرف - المعجم	1
الموارد اللسانية	2
برمجة عامة	3
توثيق لساني - حاسوبي	1
حاسوبيات عامة	10
حاسوبيات عامة – ذكاء صناعي	1
حاسوبيات لغوية	1
ذكاء صناعي	13
ذكاء صناعي – منصة نوج	6
ذكاء طبيعي	1
فيزياء الكم	1
لسانيات صورية	16
لسانيات صورية – حاسوبيات عامة	2
معالجة آلية	1
منصة نوج	4
هندسة لسانية عامة	3
هندسة لغوية ولسانيات حاسوبية	1
	70



2.8.الملحق رقم: 2

المصطلحات العلمية المستخدمة في بحثنا المنشور في الكتاب الجماعي الذي أصدرته المندوبية الدائمة للمملكة العربية السعودية لدى اليونسكو، بمناسبة اليوم العالمي للغة العربية 2017، تحت عنوان: اللغة العربية وتقانة تعلم الآلة: تطبيق منصة نوج على معجم العرفان، ص: 9 – 40.

رقم	إنجليزي	عربي	مجال استخدام المصطلح
1	Neurons network	شبكة عصبية	ذكاء صناعي
2	Lemmatization	تجذيع	الصرف
3	Fusion	انصهار	الصرف
4	Artificial intelligence	ذكاء صناعي	ذكاء صناعي
5	Machine learning	تعلم آلي	ذكاء صناعي
6	Text mining	تنقيب عن النصوص	ذكاء صناعي
7	Big data	بيانات ضخمة	موارد لسانية
8	Full data	بيانات شاملة	موارد لسانية
9	text analysis	تحليل النصوص	ذكاء صناعي – منصة نوج
10	Data base	قاعدة البيانات	حاسوبيات عامة
11	Root	جذر	صرف – معجم
12	Pattern	وزن	صرف
13	Tokinization	وسم	حاسوبيات عامة
14	Analysis	تحليل	هندسة لسانية عامة
15	Generation	توليد	هندسة لسانية عامة

16	Speech recognition	تعرف الأصوات	ذكاء صناعي
17	Machine translation	ترجمة آلية	ذكاء صناعي
18	Syntactical analyzer	محلل تركيبي	ذكاء صناعي – منصة نوج
19	Morphological analyzer	محلل صرفي	ذكاء صناعي – منصة نوج
20	Optical character recognition	تعرف المحارف آليا	هندسة لسانية عامة
21	Automat	مسوقة	ذكاء صناعي – منصة نوج
22	Graph	رسم هندسي	ذكاء صناعي – منصة نوج
23	Synthesis	توليف	حاسوبيات عامة
24	Electronical dictionary	معجم آلي	منصة نوج
25	Platform	منصة	منصة نوج
26	Summarization	تلخيص آلي	ذكاء صناعي
27	Retreival information	استرجاع آلي للمعلومة	ذكاء صناعي
28	Diacretization	تشكيل	ذكاء صناعي
29	Formalization	صورنة	حاسوبيات عامة

30	Verbe support	فعل عماد	لسانيات صورية
31	Completive	مصدر مؤول	لسانيات صورية
32	Named entity	أسماء الأعلام	ذكاء صناعي – منصة نوج
33	Three Bank	قاعدة البيانات الشجرية	توثيق لساني – حاسوبي
34	Hybride	هجين	ذكاء صناعي
35	Processing	معالجة آلية	معالجة آلية
36	Algorithm	خوارزم	حاسوبيات عامة – ذكاء صناعي
37	Mathematical Model	نموذج رياضي	لسانيات صورية
38	Model	نموذج	حاسوبيات عامة
39	Input	دخل	حاسوبيات عامة
40	Matrix	أجروم	لسانيات صورية – حاسوبيات عامة
41	Source code	أصل البرنامج	حاسوبيات عامة
42	Open source	مفتوح المصدر	برمجة عامة
43	Compilation	تجميع برمجي	برمجة عامة
44	Distribution	توزيع	لسانيات صورية

45	Transformation	تحويل	لسانيات صورية
46	Singularity	انفرادية	ذكاء صناعي
47	Exponentiel developemnt	تطور أسي	فيزياء الكم
48	Framwork	إطار	لسانيات صورية
49	Argument	موضوع	لسانيات صورية
50	Simulation	تقييس	حاسوبيات عامة
51	Concordoncier	كشاف سياقي	منصة نوج
52	Transducer	لاقط	منصة نوج
53	Interactivity	تفاعلية	حاسوبيات عامة
54	Linguistic engineering	هندسة اللغة	حاسوبيات عامة
55	Computational linguistics	لسانيات حاسوبية	حاسوبيات لغوية
56	Lexicon grammar	معجم تركيبي	لسانيات صورية
57	Automatic indexation	فهرسة آلية	ذكاء صناعي
58	Human intelligence	ذكاء بشري	ذكاء طبيعي
59	Competence	كفاية	لسانيات صورية
60	Performance	إنجاز	لسانيات صورية
61	Idiom	تعبير مسكوك	لسانيات صورية

62	Lineary	خطية	برمجيات عامة
63	Taxonomy	تصنيف الظواهر	لسانيات صورية وحاسوبيات عامة
64	Corpus / Corpora	متن / متون	لسانيات صورية
65	Hierarchy	سلمية	لسانيات صورية
66	Classification	تصنيف	لسانيات صورية
67	Large scale / coverage	مسح شامل للبيانات	هندسة لغوية ولسانيات حاسوبية
68	Conversion	إدماج	لسانيات صورية
69	Deep learning	نعلم عميق	ذكاء صناعي
70	Predicate	محمول	لسانيات صورية

المراجع والمصادر:

أولا: المراجع العربية:

1. هندسة اللغة العربية: مدخل نظري، محمد الحناش، منشورات العرفان، فاس، 2018

2. هندسة اللغة العربية: الاشتقاق التركيبي، محمد الحناش، منشورات العرفان، فاس، 2018

3. لسانيات المنصات واللغة العربية، على بولعلام، فاس، 2018

4. مدخل إلى اللسانيات الحاسوبية، تنسيق عبد الله بن يحي الفيفي، مركز الملك عبد الله للغة العربية، الرياض 2017 (كتاب جماعي)

5. التعليم والعالم العربي، تحديات الألفية الثالثة، مركز الإمارات للدراسات والبحوث الإستراتيجية، أوظبي 2000

6. لغويات المدونة الحاسوبية، المنهج والنظرية والتطبيق، طوني ماك إينري، وأندريو هاردي، ترجمة د. سلطان بن ناصر المجيول، دار جامعة الملك سعود للنشر، 2016

7. المدونات اللغوية العربية، بناؤها وطرق الإفادة منها، تنسيق د. صالح بن فهد العصيمي، مركز الملك عبد الله للغة العربية، 2015

8. المعالجة الآلية للغة العربية، المشاكل والحلول، دة. سلوى حمادة، دار غريب، القاهرة، 2009

9. الحوسبة السحابية، الأساسيات والأنظمة، كتاب جماعي مترجم، دار جامعة الملك سعود، 2016

10. لغويات المدونة الحاسوبية، تطبيقاتها تحليلية على العربية الطبيعية، د. سلطان المجيول، مركز الملك عبد الله للغة العربية، 2016

ثانيا: المراجع الأجنبية:

[1] Beesley, Kenneth R & Karttunen, Lauri, 2000. Finite-states non concatenative morphotactics. In Procedings of the 38th annaul meeting of the association for comutational linguistics (ACL-00), 191-198.

[2] Diab, Mona & alii. 2007. Automatic processing of modern standard arabic text, In Soudi Abdelhadi (editor), Springer

[3] Gross Maurice, 1975. Métodes en synatxe, Hermann, Paris

[4] El Hannach, Mohamed. 1988. Sytaxe ses verbes psychologiques de l'arabe. Thèse de doctorat d'Etat, Université Paris VII.

[5] El Hannach, Mohamed, 2001. Syntaxe des verbes qualitatifs de l'arabe. Synergie monde arabe, Vol. I

[6] Elghamry, Khaled, 2004. A constraint-based algotithm for the identification of arabic roots. In proceeding of the Midwest computational linguistics colloquium.

[7] Goldsmith, John A 2001. Unsupervised learning of the morphology of a natural language. Computaional linguistics, 27 (2) : 153-198

[8] Goldsmith, John A 2006. An algorithm for the unsupervised learning of morphology. Natural language engineering, 12 (4) : 353-371

[9] Harris, Zellig S, 1972. Structure mathématique du langage, Duno, Paris

[10] Khaled Shaalan, Amin Allam, and Abdallah Gomah. Towards Automatic Spell Checking for Arabic. In *Conference on Language Engineering*, ELSE, Cairo, Egypt, 2003. 36

[11] Kilani-schoch, Marianne 1988a. Discontinuité ou continuité de la base morphologique en arabe classique et en arabe tunisien ? Zeitschrift fur arabische linguistik, 19 : 91-92

[12] Kilani-schoch, Marianne 1988b. Introduction à la motphologie naturelle. Berne : Peter lang

[13] Farghaly, Ali; 2003. Handbook for language engineers, CSLI Publications

14] Columbia Arabic treebank. Proceeding of the ACL-IJCNLP Conference Short Papers 221-224

[15] Kenneth R. Beesly 2009. Arabic finite-state Morphological analysis and generation. Bank Xerox research center, Gonoble

[16] Levy Mary M. 1971. The plural of noun in modern standard arabic. Thèse de doctorat. University of Michigan

[17] Buckwalter, Tim 2002. Backwalter arabic morphological analyzer, Version 1.0. Linguistic data consortium, Philadelphia

[18] Cantineau, Jean, 1950. La notion de Schème et son altération dans diverses langues sémitiques. In semitica, 3 : 73-83

[19] Ratcliffe, Robert R 1998. The broken plural problem in arabic and comparativ semitic. Cueent issues in linguistic theoty 198. Amsterdam : Jphn Benjamins

[20] Sasson, George T. 1992. The application of Sukhotin's algorithm to certain non-english languages. Crypologia, 16 (2) : 165-173

[21] Nizar Habash, 2010. Introduction to Arabic Natural Language Processing (Synthesis Lectures on Human Language Technologies) Morgan & Claypool

[22] Soudi , Abdelhadi & alii, 2007. Arabic Computational Morphology: Knowledge-Based and Empirical Methods, Spriger

[23] Mesfar, Slim, 2008. Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en arabe standard. Thèse, Université de Fanche Comté, France

[24] Misfar, Slim & Silberztein Max 2008. Transducer minimization and information compression for Nooj dictionaries.

Proceedings of the FSMNLP 2008 Conference. Frontiers in Artificial Intelligence and appliscations, IOS Press, Pays-Bas.

[25] Féhri, H. Haddar K. Ben H. 2010. Integration of transliteration process into an automatic translation sysrem for named entities from Arabic to Frensh. Proceeding of the Nooj 2009 International Conference and Workshop, Sfax, Center de Publication Universitaire, p. 285-300.

[26] Rodrigo, Andrea 2011. Tratamiento automatico de textos : el sintagma adverbial nùcleo, ediciones juglaria, Rosario, Argentina.

[27] Saleh Najim, 2007. Inheritance-based Approach to Arabic Verbal Root-and-Pattern Morphology. In Soudi A, Springer

[28] Siberztein, Max & al. 2015. Atomatic Processing of Natural-Language Electronic Texts with Nooj; Sptinge.

[29] Misfar, Slim 2008. Analyse morpho

[30] Silberztein, Max 2015. La formalisation des langues : l'approche NoojIste editions, London

[31] Silberzteun, Max, 1993. Dictionnaires électroniques et analysr automatique de textes : Le système INTEX, Masson, Paris

[32] Xantos, Aris 2008. Apprentissage automatique de la morphologie : le cas des structures racine-schème, Peter Lang edition, Bern.

[33] Tellier, I, 2010. Apprentissage automatique pour le TAL; Préface, Traitement automatique des langues ATALA, 2009, Vol. 50 (3), pp : 7-21

[34] Dichy , J. Linguistic Knowledge integration in optical Arabic word and text recognition process, *Linguistica Communicatio journal*, Sprcial issues, 2013

[35] Hamdani, A. Researches on Arabic Language Engineering, *Linguistica Communicatio journal*, Special Issues, 2014

[37] Boumediane, A. Arabic language and information technology, *Linguistica Communicatio journal*, Special Issues, 2015.

[38] Isabelle, T. Apprentissage automatique pour le TAL. inria-00541535, 2010

[39] Delphine, B. Apprentissage de connaissances morphologiques pour l'acquisition automatique de ressources lexicales, inria- 00119257, 2006.

[40] El Hannach, M. *Erfan* electronic dictionary, Publisher Al-Erfan for educational consulting, Morocco, 2018