اللسانيات الحاسوبية والمعالجة الآلية للغات الطبيعية

فطوم القربيش
أستاذة التعليم الثانوي التأهيلي
عضوة مع جمعية هندسة اللغة بالمدرسة المحمدية للمهندسين بالرباط
Fettoum.krieche@gmail.com

ملخص:
نهدف من خلال هذا المقال إلى التعريف باللسانيات الحاسوبية (تاريخها، موضوعها، منهجها، روادها)، ثم سنميز بين اللغة الطبيعية واللغة الاصطناعية ونعرف بآليتي التوليد والتحليل باعتبارهما عمليتان مهمتان في المعالجة الآلية للغات الطبيعية كما سنقدم طرق معالجة مستويات الدرس اللساني بطريقة آلية، وهي المستوى الصوتي والمستوى الصرفي والمستوى التركيبي والمستوى الدلالي والمستوى التداولي. مع إبراز أهم المشاكل التي تعاني منها المعالجة الالية في كل مستوى على حدة. وسنشير إلى مجالات استثمار المعالجة الآلية للغات الطبيعية، مع تذييل كل مبحث بخلاصة تركيبية.

الكلمات المفاتيح: ***اللسانيات الحاسوبية-المعالجة الالية- اللغة الطبيعية-اللغة الاصطناعية- الدرس اللساني-آليتي التوليد والتحليل.***

أولا: مدخل عام حول اللسانيات الحاسوبية

لقد تنوعت مناهج دراسة اللغة الطبيعية بتنوع الخلفيات المعرفية لكل دارس من جهة، وبتنوع النظريات اللسانية من جهة أخرى؛ فقد كانت دراسة اللغة فرعا من فروع الفلسفة في المراحل الأولى، حيث كان الغرض من دراستها هو التقعيد لها وضبطها، وبعد اكتشاف "وليامز جونز" التقارب الحاصل بين اللغة السنسكريتية واللغات اللاتينية، انتقل البحث في اللغة إلى المنهج التاريخي المقارن الذي يقوم بالأساس على المقارنة بين اللغة الأم واللغات البنات لتحديد المميزات والصفات المشتركة بين كل اللغات.

وستعرف مناهج دراسة اللغات الطبيعية طفرة نوعية مع نشر محاضرات في علم اللغة العام "Cours de Linguistique Générale" للساني فيردناند دو سوسير سنة 1916، والتي قدم فيها منهجا جديدا لدراسة اللغة وهو المنهج البنيوي الذي يقوم على فكرة أساسية وتكمن في اعتبار أن اللغة بنية مغلقة مكتفية بذاتها، ولذلك وجب دراستها في ذاتها ولذاتها دون ربطها بمحيطها الخارجي أو بيئتها أو تاريخها أو مقارنتها بلغة أخرى.

ثم سينتقل البحث في اللغة من اعتبارها بنية مغلقة ومكتفية بذاتها إلى دراسة الجهاز المسؤول عن اكتسابها وإنتاجها وذلك مع المدرسة التوليدية التحويلية لرائدها نعوم تشومسكي، هذا الأخير الذي اتخذ من اللغة موضوعا ومن الرياضيات والمنطق منهجا ومن علم النفس أداة ووسيلة لدراسة اللغة، وقد ترجم أبحاثه في مجموعة من البرامج والنظريات طورت لاحقها سابقها.

إلا أنه مع تنامي وتطور الوسائل التكنولوجية وتداخل اللسانيات مع مجموعة من العلوم (الذكاء الاصطناعي وعلم النفس والمنطق والرياضيات والعلوم المعرفية....)، ظهر منهج جديد في دراسة اللغة

وهو المنهج اللساني الحاسوبي؛ وقد جمع هذا المنهج بين نتائج اللسانيات الصلبة ونتائج تطور التقنيات المعلوماتية في معالجة اللغات الطبيعية.

فمتى ظهرت اللسانيات الحاسوبية؟ وما هو مفهومها؟ وما هو موضوعها ومنهج اشتغالها؟

1 نبذة تاريخية عن اللسانيات الحاسوبية

اكتشف المعلوماتيون أن الحاسوب يمكنه أن يقوم بالترجمة الآلية للنصوص بجانب مهمة الحساب الآلي، وقد تحدث عن هذا المعلوماتي ورن وفر "WarrenWeaver" سنة 1949 في مقال له بعنوان "الترجمة" حيث عالج فيه مجموعة من المشاكل المتعلقة بالترجمة الآلية من قبيل البنية المنطقية للغات، والنحو الكلي الذي تتقيد به كل اللغات. ورغم أن فكرة الترجمة كانت قد ظهرت في سياقات ثقافية وعلمية من قبل، فإن مقال وفر جعل المعلوماتيين يهتمون بتقديم حلول لمعالجة الجوانب التركيبية والدلالية للغة [1].

إلا أن البداية الحقيقية للسانيات الحاسوبية كانت مع ظهور الترجمة الآلية في سنة 1954، وفي سنة 1962 قدمت أول محاضرة حول الترجمة الآلية من قبل "Y. Bar-Hillel"، وسيعرف مجال المعالجة الآلية تقدما ملحوظا مع تطور النظريات اللسانية (هاريس 1951/ 1954، وتشومسكي 1957)، أضف إلى ذلك ظهور أبحاث بخصوص الذكاء الاصطناعي سنة 1956، حيث أنجز "Winorgrad Terry" أول برنامج قادر على التحاور باللغة الانجليزية مع إنسان آلي، وتوالت الأبحاث وتطورت المقاربات خاصة المرتبطة بالمستوى الدلالي.[2] لتتطور بالموازاة مع ذلك تقنيات المعالجة الآلية للغات الطبيعية.

وقد انعقد أول مؤتمر حول اللسانيات التطبيقية ب "Nancy" في أكتوبر 1964، ونظم العالم فكواس "Vouquois" في أكتوبر 1967 محاضرة عالمية ثانية في اللسانيات الحاسوبية في جرونوبل "Grenoble" بعنوان: "المحاضرة الثانية في المعالجة الآلية للغة"[3].

ومما تجدر الإشارة إليه، أن الباحثين الذين اهتموا بالبحث في الترجمة الآلية وفي اللسانيات الحاسوبية في السنوات الأولى من ميلادها، كانوا علماء متخصصين في الهندسة الرياضية التطبيقية أو فلاسفة متخصصين في دراسة اللغات الطبيعية، أما اللسانيون فلم يولوا اهتماما كبيرا لهذا الجانب باستثناء جهود بعض اللسانيين التوزيعيين ومنهم: اللساني هاريس "Haris" واللساني هوكيت "Hokitt" بشراكة مع بعض الأنتربولوجيين اللسانيين مثل: الباحث فوجل "Voegeling" والباحث جرافين "Gravin" الذين نشروا سنة 1954 عددا خاصا حول الترجمة الآلية في الجريدة العالمية للسانيات الأمريكية، كما اشتغل اللغوي تشومسكي"Chomsky" أيضا مع الباحث يونجف "Yngve" سنة 1955 في مشروع للمعالجة الآلية بصحبة ثلاثة لسانيين آخرين وهم: يوزف أبلييت "Yoseph Appleyate" وفريد لكوف " Fred Lukoff" وبتي شفت "Betty Shefts"[4]

وبهذا تكون فكرة الترجمة الآلية هي التي شكلت اللبنة الأساسية لظهور اللسانيات الحاسوبية، وتطور هذه الأخيرة نتج عنه تطور مواز في النظريات اللسانية وفي التقنيات التكنولوجية المعلوماتية.

2 مفهوم اللسانيات الحاسوبية

يتكون مصطلح "اللسانيات الحاسوبية" من كلمتين:" اللسانيات" و"الحاسوبية"؛ ينضوي الأول تحت العلوم الإنسانية والثاني تحت العلوم الحقة؛ ومنه يمكن أن نستنتج أن الباحث في هذا الجانب ينهل من حقلين معرفيين ليدرس اللغة الطبيعية بكل مستوياتها اللسانية باعتماد آليات الحاسوب، وبمعنى آخر، دراسة اللغة بمنهج علمي حاسوبي مع ضرورة صورنة اللغة الطبيعية لتصير مرنة في التطبيق الحاسوبي على اعتبار أن الحاسوب يشتغل باعتماد معارف منطقية ورموز رياضية. فاللسانيات الحاسوبية -من هذا الجانب- تقوم بتقييس المعارف اللغوية التي توجد في الدماغ البشري بكيفية اشتغال الحاسوب.

3 موضوع اللسانيات الحاسوبية

تعد اللسانيات الحاسوبية "La Linguistique Calculatoire" جزءا من الدراسات اللسانية التطبيقية[5]، فهي مجال لبحث يجمع بين ما هو لساني لغوي وما هو معلوماتي رياضي، لأنها علم يهتم بدراسة مختلف الخصائص والظواهر اللسانية للغات الطبيعية بطريقة حاسوبية، وذلك بابتكار وخلق وتطوير أنظمة حاسوبية تسمح بمعالجة اللغات الطبيعية في كل مستوياتها اللسانية (الصوت، والصرف، والمعجم، والتركيب، والدلالة).

إن معالجة اللغة آليا تقتضي بالضرورة معالجة منطقية صورية ثم بعدها حاسوبية؛ فلا تتم حوسبة اللغة بدون صورنتها وذلك اعتمادا على مناهج رياضية وتقنية خاصة مثل: قواعد إعادة الكتابة ونظرية الأطوماطة، وآلية التشجير ولامدا الحساب... ومنه فإن كل الجوانب التي تخرج عن نطاق تركيب العبارات اللغوية لا يمكن صورنتها، ونقصد بذلك المستوى التداولي لأنه لا يرتبط بما هو لساني كلي بقدر ما يرتبط بما هو ثقافي خاص وبالواقع الخارجي، ولذلك يصعب صورنة التداولية. ومن خلال تعرفنا على موضوع اللسانيات الحاسوبية توصلنا إلى أن حوسبة اللغة تمر عبر معالجة منطقية صورية تتمثل في مجموعة من الآليات الصورية المنطقية. (والتي سنبينها في المبحث الثالث وبالضبط في المحور الأول منه: "اللسانيات الحاسوبية والدراسات المعجمية مقاربة معرفية_").

4 منهج اللسانيات الحاسوبية

إن منهج اللسانيات الحاسوبية هو منهج لساني، ورياضي معلوماتي، باعتبار أن هذا القسم من اللسانيات التطبيقية يهدف بالدرجة الأولى إلى خلق وابتكار أنظمة حاسوبية لمعالجة اللغة الطبيعية، وذلك باتباع خطوات إجرائية أساسية تتمثل فيما يلي:

* تحديد المستوى/الجانب اللغوي موضوع الدراسة.
* منطقة الجانب المحدد(اللغة) وتحويله من رموز لغوية إلى رموز رياضية (اللغة الصورية).
* ابتكار نظام معلوماتي/حاسوبي حسب ما تم صورنته.

خلاصة المبحث الأول

لقد حاولنا من خلال هذا المبحث أن نقدم بعضا من المعارف التي ترتبط بمجال اللسانيات الحاسوبية؛ بحيث استخلصنا ما يلي:

- ارتبط تاريخ ظهور اللسانيات الحاسوبية بظهور الترجمة الآلية.
- يتحدد موضوع هذا العلم ببناء أنظمة آلية لمعالجة اللغات الطبيعية.
- يتعلق منهج اشتغال اللسانيات الحاسوبية بالضرورة بالدراسة المنطقية الصورية للغات الطبيعية.

وقد هدفنا من خلال ما سبق التعريف بمفهوم اللسانيات الحاسوبية وما يرتبط بها من أجل توضيح الإطار النظري الذي اعتمدناه في بحثنا هذا.

ثانيا: المعالجة الآلية للغات الطبيعية

يتميز كل نسق لغوي بتضافر وانصهار خمس مستويات للدرس اللساني المكون لهذا النسق: الصوتي والصرفي والتركيبي والدلالي والتداولي، بحيث لا يتأتى لأي دارس لغوي أن يتعرف على مميزات هذه اللغة دون أن يلم بكل أو ببعض مستوياتها اللسانية. هذا فيما يخص الدراسات اللسانية النظرية أو ما يعرف باللسانيات الصلبة، أما فيما يخص اللسانيات التطبيقية وخاصة منها الحاسوبية، فيجب على معالج اللغة أن يكون ملما بكل الحيثيات المرتبطة باللسانيات الصلبة (النظرية) حتى يتسنى له منطقة اللغة لتكون طيعة أثناء معالجتها آليا. وبذلك سنقدم في هذا المبحث بعضا من مميزات كلا من اللغة الطبيعية واللغة الاصطناعية من جهة، وسنتحدث عن المعالجة الآلية لكل مستويات الدرس اللساني من جهة ثانية، وذلك باعتبار أن أية دراسة أو معالجة لغوية تقوم على عمليتين إجرائيتين أساسيتين هما: آليتا التحليل.

1 اللغة الطبيعية واللغة الاصطناعية

يدل مصطلح اللغة من جهة على اللغة الطبيعية بشكليها المنطوق والمكتوب، ومن جهة أخرى على اللغة الاصطناعية التي تستعمل في الرياضيات، والفيزياء، والكيمياء، والمنطق، والمعلوميات.

(1) اللغة الطبيعية

تتميز اللغة الطبيعية (كالعربية والفرنسية) -من وجهة نظر لسانية حاسوبية- بمجموعة من السمات منها:

(أ) *عنصر اللبس:*

هناك بعض الحروف في اللغة العربية مثلا لها شكل واحد في الكتابة وأشكال عدة في النطق، كحرف الواو، والياء، والألف، فتكون حروف مد (صام، يصوم) وحروف لين (قوْل، بيْع)، وتكون حروف علة إذا كانت متحركة (وَجد، يَبس)، وفي اللغة الفرنسية مثلا حرف "I" الذي يتخذ عدة صور في النطق(maison/lit/boire).

كما تعتمد اللغة العربية واللغة الفرنسية على مبدإ الاقتصاد في الجمع؛ فمثلا الحرف "S" في اللغة الفرنسية يدل على الجمع في الأسماء (les portes) والنعوت (les belles) وعلى المفرد المخاطب في بعض الصيغ الفعلية(tu es- tu as) . أما فيما يخص الجمع في اللغة العربية فله عدة صور (جمع المذكر السالم، جمع المؤنث السالم، جمع التكسير...).

(ب) *عنصر التضمين:*

تتسم اللغة الطبيعية بالتضمين باعتبارها ترتبط بموقف المتكلم/المستمع، وقصديته وبيئته، مثل:

[1] ***فلان كثير الرماد*** (مثل عربي يدل على الكناية)

هذه العبارة تدل على أن فلانا كريم لكثرة رماده (بقايا النار بعد إنطفائها) وذلك بسبب كثرة ضيوفه.

وكمثال في الثقافة الفرنسية:

[2]***Jean a cassé sa pipe***. (جون كسر غليونه)

فهذه العبارة تدل ضمنيا بأن جون قد مات.

إن عنصر التضمين يرتبط بالمجاز اللغوي من استعارة وكناية وغيرهما بالنسبة للغة العربية، وكذلك في اللغة الفرنسية.

(ج) *عنصر المرونة:*

تتميز اللغة الطبيعية بالثراء، لأنها تمكن مستعمليها من تحقيق التواصل كونها تسمح بالتوليد والاشتقاق والنحت، لذلك تتسم بالمرونة.

(د)*عنصر الترادف:*

تتسم اللغة الطبيعية بكثرة المفردات اللغوية؛ فمثلا **في اللغة العربية**:

- الجود يرادف السخاء والعطاء والسماحة والكرم.
- نقول: رأيت الشيء وأبصرته وعاينته وشاهدته.
- نقول: عام، سنة، حول.

وفي اللغة الفرنسية:

كلمة "le travail" لها أكثر من عشرون مرادفا، منها:

Labeur, emploi, fonction, métier, place, poste, profession, situation, boulot (populaire), turbin (populaire), ouvrage, tâche, œuvre….[6]

* البنية الصرفية المعقدة (مسألة الشكل ومسألة إفراغ الجذر في الصيغة الصرفية) بالنسبة للغة العربية فقط.

(2) اللغة الاصطناعية

تستعمل اللغة الاصطناعية في العلوم الحقة مثل: الرياضيات، والفيزياء، والمنطق...كصورة مجردة عن تحولات العالم الخارجي، وهي لغة جافة نوعا ما.

وتقوم المعلوميات أيضا على لغة اصطناعية تسمى بلغة البرمجة والتي تتأسس على لغة الخوارزمات، ومصطلح الخوارزم (نسبة إلى واضعه موسى الخوارزمي) هو مصطلح رياضي ومنطقي محض، قدمت له تعريفات عدة ومن أهمها:

- "جاء العالم موسى الخوارزمي الذي قدم طريقة رياضية رائعة في حل المسائل أطلق عليها اسم الخوارزمات (Algorithme) نسبة إلى اسمه، والتي تعتمد على مبدإ تقسيم المسألة قيد الحل إلى عدة خطوات (Steps) مترابطة منطقيا فيما بينها، لكي يؤدي تطبيقها إلى الوصول إلى الحل المناسب"[7].

- إن الخوارزم نظام من المبادئ والقوانين الصورية التي تتحكم في عمليات الإنتاج اللغوي على مستوى التحليل (الإدخال) والتوليد (الإخراج)"[8].

وتتميز الخوارزمات ببعض الخصائص من بينها:

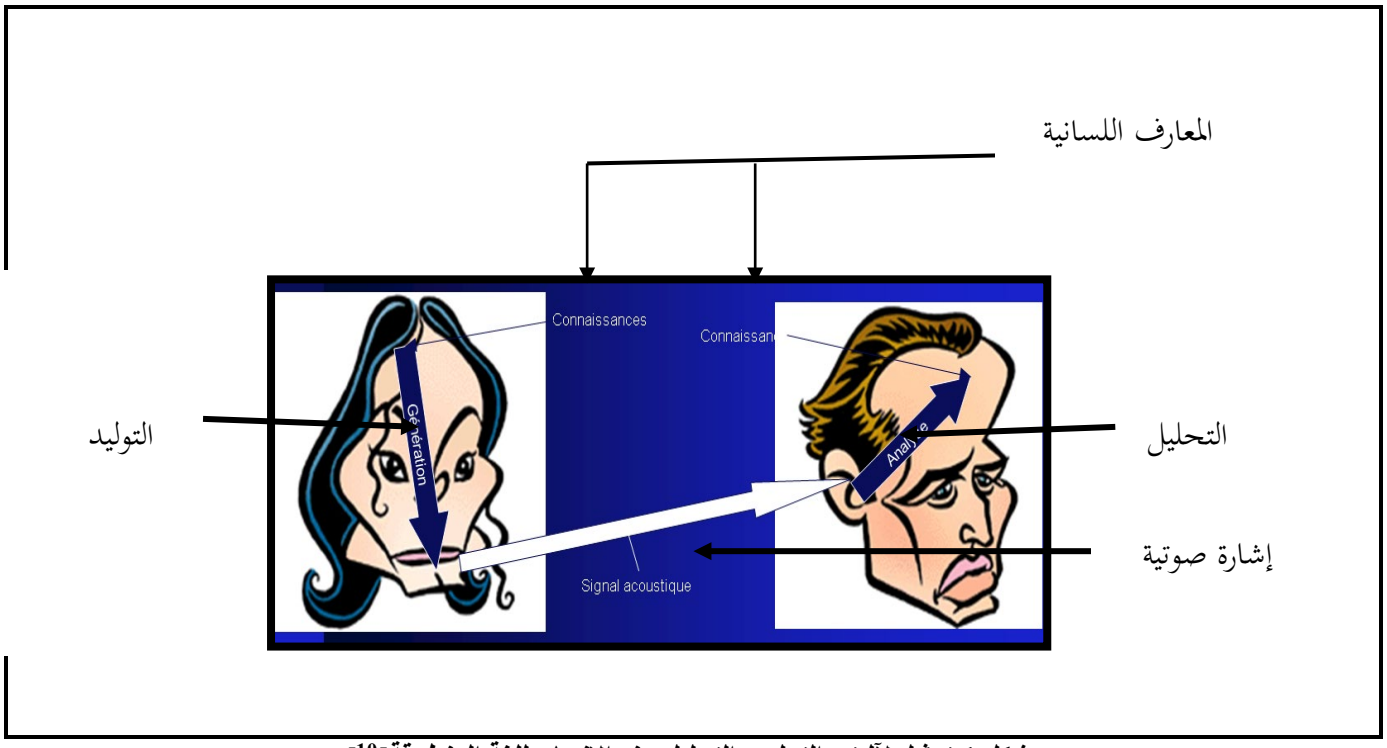
- الدقة (خطوات الخوارزم يجب أن تكون معروفة وواضحة)
- الفعالية والانتهاء (الوصول إلى حل صحيح للمشكلة بعد عدد معين ومحدود من الخطوات)[9].

إن لكل من اللغة الطبيعية واللغة الاصطناعية مميزاتها الخاصة إلا أن هذا لا يعني أن كلا منهما مستغن عن الآخر، بل على العكس، فكلاهما يتوسل بالآخر لتفسير ظواهر العالم الخارجي ومعالجتها.

2 آليتا التوليد والتحليل

ينطلق مستعمل اللغة مما هو مخزن في دماغ الإنسان أو ذاكرة الحاسوب-من معارف لسانية- لإنتاج جملة، وهذه العملية تسمى بالتوليد؛ أي أنه ينطلق من تصورات أو تمثلات ذهنية توجد في دماغ الانسان أو مخزنة في ذاكرة الحاسوب لتفعيلها في الواقع عن طريق رسالة منطوقة أو مكتوبة، وقد ينطلق مستعمل اللغة مما هو مولد (مجموع الكلمات والحروف والجمل)، لتحليل أجزائه اللسانية ثم الوصول إلى التصور أو التمثل الذهني (عند الانسان أو الحاسوب)، وهو ما يعرف بالتحليل. ومما تجدر الإشارة إليه أن هاتين العمليتين متكاملتين؛ حيث ننطلق من التوليد لنصل إلى التحليل، وقد ننطلق من التحليل للوصول إلى التوليد.

ويقوم الإنسان المستمع/المتكلم بهاتين العمليتين بناء على معارفه اللسانية الذهنية، ويمكن أن نمثل لذلك بما يلي:



شكل 1 تمثيل لآليتي التوليد والتحليل عند الإنسان للغة المنطوقة[10]

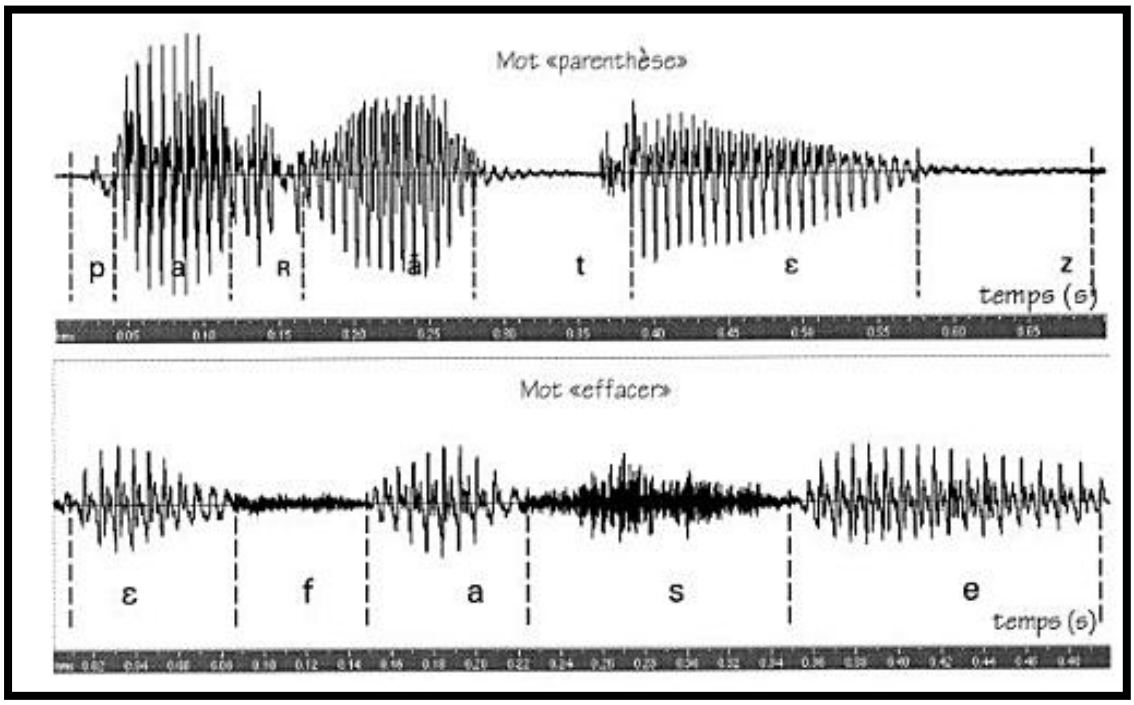
قد قام المهتمون بالعلوم المعرفية بعملية تقييس المعارف اللسانية بالحاسوب، لأن للحاسوب ذاكرة يمكن أن تخزن فيها هذه المعلومات كما يمكن استرجاعها لتحليلها ثم الاعتماد عليها لتوليد معارف لسانية ما، ولتوليد أو تحليل هذه المعارف اللسانية المخزنة في الحاسوب نعتمد على دخل وخرج، ومنه فهناك حالة أولية E1(دخل)، وحالة نهائية E2(خرج)، ونمثل لذلك كالآتي:

$$E_1 \rightleftharpoons E_2$$

3 المعالجة الآلية للأصوات

يهتم علم الأصوات بدراسة الأصوات الكلامية من ناحية سماتها وصفاتها ومخارجها، ودراسة أيضا القواعد التي تتحكم في تآلفها وتجاورها...إلا أن اللسانيات الحاسوبية اهتمت بالجانب الأكوستيكي والنطقي للصوت؛ أي بالذبذبات الصوتية دون الاهتمام بمكوناته وخصائصه الفيزيائية؛ وذلك ببرمجة الصوت في الحاسوب ثم محاولة التعرف على خصائصه بتحديد أجزائه ودراستها قطعة، قطعة، إلا أن المعلوماتي يواجه بعض المشاكل في هذا الجانب، وترتبط بالخصوص بطريقة نطق الصوت كالظواهر اللسانية فوق المقطعية مثل: النبر، والتنغيم...

وتتم معالجة التذبذب الصوتي كالآتي [11]:



شكل 2 صورة توضح طريقة معالجة (التحليل) الكلام في الحاسوب.

ومن المشاكل التي قد تعترض المعلوماتي في هذا الجانب نذكر:

(1) ظاهرة الإدغام **[12]:**

إلا: إن + لا

مما : من + ما

عما : عن + ما

ادتفع ⟵ ادفع

(2) ظاهرة القلب

- ازدهر وأصلها ازتهر؛ بحيث قلبت التاء دالا لأن التاء أخذت سمة الجهر من الزاي فأصبحت دالا.

- اصطبر والتي أصلها اصتبر، وقلبت تائها طاء لأن التاء أخذت صفة التفخيم من الصاد وأصبحت طاء.

- مضطرب ← مضترب

- قائل ← قاول

- مصطلح ← مصتلح

لكن ما هو الأصل الثلاثي بالنسبة ل: مضترب و مصتلح؟

(3) ظاهرة التضعيف:

مدَّ ← مدد

والسؤال المطروح هنا: هل سيقبل الحاسوب مدَّ أم مدد؟

4 المعالجة الآلية للصرف

يهتم الصرف بدراسة البنية الصرفية للكلمة في معزل عن البنية التركيبية وعن السياق الدلالي لجملة ما، ومن القضايا التي يدرسها الصرف العربي: دراسة الصيغ الصرفية، والجذوع الفعلية، والاسمية، والزوائد ودلالتها، ومظاهر الاشتقاق، والتصريف، وظواهر صرفية أخرى مثل: الإعلال والقلب والإبدال وغيرها....، بينما ترتكز المعالجة الآلية لعلم الصرف على دراسة وتحليل أجزاء الكلمة بتحديد نوعها، وجذرها، ولواصقها، ووزنها الصرفي انطلاقا مما تمت برمجته في الحاسوب، ولتفصيل ذلك أكثر نورد المثال أدناه:

(1)في اللغة العربية:

[3] نجح الطالب في الامتحان

تحلل الجملة على الشكل الآتي:

الوحدة 1: نجح (فعل، ماض، مذكر، مفرد.....)

الوحدة 2: الطالب (اسم، فاعل، مذكر، مفرد...)

الوحدة 3: في (حرف، يدل على المكان والزمان....)

الوحدة 4: الامتحان (اسم، مجرور، مذكر، مفرد...)

(2)في اللغة الفرنسية:

[4]Fatima a acheté une robe(اشترت فاطمة فستانا)

Unité1: Fatima (Nom propre. Féminin. Singulier…)

الوحدة الأولى: فاطمة (اسم علم، مؤنث، مفرد...)

Unité 2: a acheté (verbe. Passé composé. 3^{éme} Personne du singulier…)

الوحدة الثانية: اشترت (فعل، ماض، ضمير الغائب المفرد...)

Unité 3: une (article indéfini. Singulier. Féminin…)

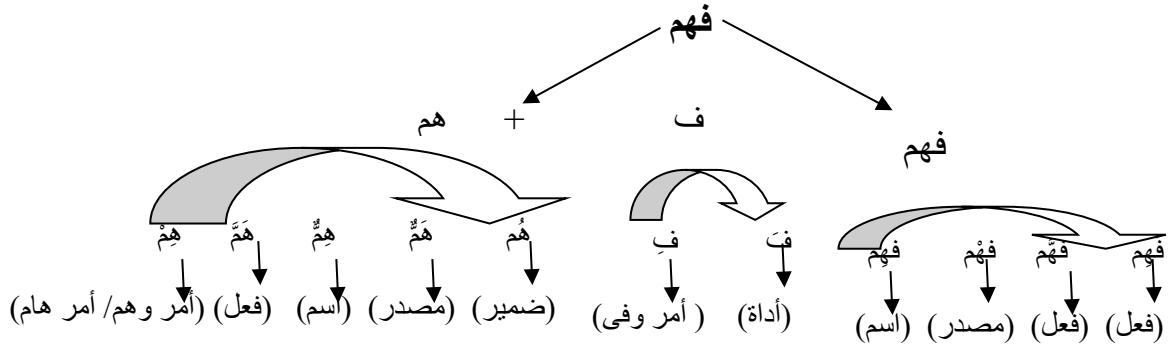
الوحدة الثالثة: تخص أداة نكرة بالنسبة للغة الفرنسية وهي في اللغة العربية حذف"ال" التعريف وتعويضه بتنوين الفتح في هذا المثال (أداة نكرة، مفرد، مؤنث....)

Unité 4: robe (nom. Singulier….)

الوحدة الرابعة: فستانا (اسم، مفرد....)

وقد تعترض المعلوماتي في معالجة هذا المستوى مشاكل عدة منها:

مثال: تحليل كلمة "فهم" آليا تحتمل أوجها عديدة هي:



شكل 3تحليل كلمة "فهم" آليا.

5 المعالجة الآلية للتركيب

يقوم المستوى التركيبي على دراسة المكونات التي تتشكل منها الجملة، كما أنه يهتم بتحديد مواقع الكلمات داخل الجملة؛ حيث إن لكل مقولة تركيبية نحوية(الفعل، والفاعل، والمفعول...) موقعها داخل الجملة، أما فيما يخص المعالجة الآلية للمستوى التركيبي في الحاسوب، فتقوم على أساس القواعد والمحددات التي تمت برمجتها في المعالجة الآلية للمستوى الصرفي، لكن إذا ما اعتمدنا فقط على هذا المنطلق (تحديد مواقع الكلمات في الجمل)، سيتم تحديد ما اصطلح عليه في الأدبيات اللسانية التوليدية التحويلية (نوام تشومسكي) بالمقبولية النحوية[13] دون اعتبار المقبولية الدلالية [14] فستكون الجملة:

[5] أكل الولد الحائط.

عبارة صحيحة ومقبولة نحويا (فعل، فاعل، مفعول به) في الحاسوب وسيحللها كالاتي:

الوحدة الأولى: أكل ← فعل

الوحدة الثانية: الولد ← فاعل

الوحدة الثالثة: الحائط ← مفعول به

بينما إذا ما تأملنا هذه الجملة دلاليا فهي خاطئة، لذلك سيكون من الضروري برمجة المحددات الدلالية لكل كلمة في اللغة المعالجة.

وقد يواجه المعلوماتي في برمجة هذا المستوى مجموعة من المشاكل من قبيل:

(1)مسألة التطابق بين الفعل والفاعل عددا وجنسا [15]

(أ) التطابق العددي

هو تناسب وتطابق من ناحية العدد بين الفعل والفاعل في الجملة الفعلية وبين اسمين كيفما كانت نوعيتهما النحوية في الجملة الاسمية.

[6-أ] الأولاد دخلوا.

[6-ب] دخل الأولاد.

ففي الجملة الثانية لا يوجد تطابق بين الفعل والفاعل من حيث العدد، وذلك يحدث مشكلا في التوليد الآلي كون أن "دخل" مفرد و"الأولاد" جمع، وعبارة "دخلوا الأولاد" خاطئة في اللغة العربية صحيحة آليا، لذلك اعتبر التطابق العددي في المثال [6-أ] ضروريا، بينما في المثال [6-ب] فالتطابق العددي غير ضروري.

(ب) التطابق الجنسي

هو تناسب وتطابق من ناحية الجنس بين الفعل والفاعل في الجملة الفعلية وبين اسمين كيفما كانت نوعيتهما النحوية في الجملة الاسمية.

[7] ذهبت فاطمة إلى المدرسة ← هناك تطابق.

[8-أ] جاءت ثلة من التلاميذ (في هذه العبارة تحقق التطابق الجنسي باعتبار أن لفظة ثلة مؤنث).

[8-ب] جاء ثلة من التلاميذ (لم يتحقق التطابق الجنسي بين الفعل "جاء" والفاعل "ثلة"، لكن تحقق هذا التطابق بين الفعل "جاء" والاسم المجرور"التلاميذ".

بالنسبة لهذين المثالين، فلا يجب أن يتحقق التطابق بالضرورة لأن كلا المثالين صحيح، لكن ذلك يحدث مشاكل في التوليد والتحليل الآلي، لأن الحاسوب لن يحدد الطريقة التي يجب أن يتعامل بها مع هذا النوع من الجمل؛ باعتماد ضرورة التطابق الجنسي أم بدونه.

(2)مسألة الضمائر

(أ) الضمير المستتر

إنها مسألة لم يوجد لها تمثيل في الحاسوب بعد، مثل:

[9-أ] خرج إلى المسجد.

[9-ب] خرج الإمام إلى المسجد.

إن الفاعل في الجملة الأولى [9-أ] ضمير مستتر تقديره "هو" ولا يظهر كتابة إنما يقدر وجوده، لأنه مستتر في الفعل، بينما في الجملة الثانية [9-ب] فالفاعل اسم ظاهر(الإمام).

(ب) الضمائر المتصلة

نجد في اللغة العربية مثلا كلمة واحدة تتكون من ثلاث مقولات كبرى: فعل+ فاعل+ مفعول مثل: "*أدخلونا*"، فهذه الكلمة تتكون من الفعل "أدخل" أما الواو والنون فهما ضميران متصلان؛ الأول في محل رفع فاعل والثاني في محل نصب مفعول به. فيؤدي هذا الأمر إلى مشاكل تخص التقطيع الصرفي والتركيبي للكلمات في الحاسوب.

6 المعالجة الآلية للدلالة

يهتم المستوى الدلالي بدراسة معنى الكلمات والتعابير والجمل ومدى تطابقها مع السياق ومنطق الكلام [16] ولذلك يتم معالجة الجملة السابقة [5] دلاليا على هذا النحو:

أكل: + فعل + حدث + يتطلب فاعلا للحدث + يتطلب مفعولا قابلا للأكل....

الولد: +إنسان + عاقل+ مذكر+.....

الحائط: + جمد+ شيء مرتفع+ غير قابل للأكل+....

ومن خلال هذا التحليل، يتبين أن الجملة غير مقبولة دلاليا، لذلك وجب على الحاسوب أن يعطي إشارة بعدم مقبولية الجملة من حيث الدلالة، ويتسنى ذلك من خلال برمجة السمات الدلالية لكل كلمة على حدة في الحاسوب.

وقد يواجه المعلوماتي في معالجته للمستوى الدلالي آليا بعض المشاكل، فمثلا نجد في اللغة العربية بعض الحروف تؤدي معاني متعددة وتسمى بحروف المعاني؛ فالواو تارة تكون حرف عطف في:"ذهبت إلى المكتبة واشتريت كتابا"، وقد تدل على المعية في: "ذهب علي وفاطمة إلى المكتبة"، وقد تدل أيضا على الحال مثل: "جاء الولد وهو يبتسم"، وهناك حروف أخرى تشغل وظائف عدة مثل: حتى (حرف جر وأداة عطف)، أو (للتخيير وللإباحة وللشك وللإضراب وللإبهام وللعطف) لا (العطف والنفي)....وهذا يسبب عائقا أمام المعالجة الآلية للغة العربية.[17]

7 المعالجة الآلية للتداول

إذا كان المستوى الدلالي يرتبط بدراسة معنى الكلمات الملفوظة أو المكتوبة، فإن المستوى التداولي يرتبط بمواقف المتكلم المستعمل لهذه الكلمات، ولتوضيح ذلك نورد المثال الآتي:

[11.أ] علي: هل ستأتي عند محمد هذا المساء؟

[11ب] يوسف: أظن أن مصطفى سيكون هناك.

فالجملة[11.ب] تدل على أن يوسف لن يأتي، وذلك إذا علمنا أن يوسف لا يحب مصطفى، ومنه فإن هذا التحليل ليس له علاقة بما هو دلالي بقدر ما له علاقة بموقف يوسف من مصطفى.

أضف إلى ذلك، أن المستوى التداولي للغة يرتبط أيضا بثقافة كل عشيرة لغوية على حدة، فمثلا إذا تأملنا هذه الجملة "إن فلانا كثير الرماد" نفهم من كلمة الرماد الوسخ، بينما تدل هذه الجملة في الثقافة العربية بأن فلانا هذا كثير الكرم، لأنه يستعمل النار بكثرة لطهي الأكل للضيوف مما ينتج عنه كثرة الرماد.

ومن خلال هذين المثالين نستنتج أن المستوى التداولي يرتبط ارتباطا وثيقا بالسياق، وبمواقف المتكلم المستعمل للغة وبثقافة كل لغة أيضا، ومنه يصعب على المعلوماتي برمجة المستوى التداولي لأنه يخرج عن نطاق البنية الصرفية للكلمة(اللغة). ولذلك لا يوجد إلى حد الآن معالجة آلية لهذا المستوى.

8 مجالات استثمار المعالجة الآلية للغات الطبيعية

يهدف الباحثون في مجال المعالجة الآلية للغات الطبيعية -من خلال ابتكار أنظمة معلوماتية- إلى إنجاز برامج آلية تمكن من خلق حوار إنسان-آلة، وبذلك تستغل نتائج المعالجة الآلية لكل مستويات الدرس اللساني (الصوتي، والصرفي، والتركيبي، والدلالي) فيما يلي:

- الترجمة الآلية.
- إنتاج وثائق تقنية وتقارير حول معطيات رقمية كحالة الطقس والميزانية.
- المساهمة في تطوير الدراسات الخاصة بالوجائهيات حوار إنسان–آلة "Homme-Machine Interfaces" باللغات الطبيعية.
- ترجمة الكلام المنطوق ومساعدة الأشخاص المعاقين على التواصل.
- الكتابة الآلية للنصوص كإيراد مجموعة من المفردات المترادفة أثناء الحاجة لها، وكذلك تصحيح الأخطاء الإملائية والدلالية أيضا.
- الفهم الآلي للنصوص من خلال تمييز نوعية النص المكتوب وتقديم تلخيص آلي ثم استخراج الكلمات المفاتيح التي يحتويها النص.
- تطوير آليات البحث الآلي في الشبكة العالمية للمعلومات، وغيرها.
- البحث الآلي.
- التلخيص الأوتوماتيكي للنصوص.
-.....

خلاصة المبحث الثاني

لقد حاولنا في هذا المبحث أن نقدم فكرة مقتضبة عن طريقة معالجة مستويات الدرس اللساني للغة العربية على أساس آليتي التحليل والتوليد، وباعتبار أن المعالجة الآلية للغات الطبيعية تقوم على لغة

اصطناعية، قدمنا تمييزا بين اللغة الطبيعية واللغة الاصطناعية، ومنه توصلنا إلى بعض الخلاصات وهي كالآتي:

- رغم أن اللغة الطبيعية ترتبط بالإنسان (مستعملها)، واللغة الاصطناعية (المعلوميات) ترتبط بالآلة إلا أنهما متكاملان، فكلاهما واصف للآخر.
- تعترض المعالجة الآلية لمستويات الدرس اللساني مجموعة من الصعوبات ويرجع ذلك إلى البنية المركبة والمعقدة نوعا ما للغة العربية، وكثرة مفرداتها أيضا.
- لم يتمكن بعد فريق المعالجة الآلية للغة العربية من برمجة المستوى التداولي كونه يرتبط بثقافة، وبيئة، ومحيط اللغة المستعملة، مع العلم أن المعالجة الآلية للغة تتخذ من بنية الكلمة القاعدة الأساس في المعالجة، لذلك فلا يمكن برمجة ما هو خارج عن نطاق البنية الصرفية للغة.

خلاصة عامة:

وتأسيسا على ما سبق فقد قدمنا في المبحث الأول من هذا المقال تعريفا موجزا للسانيات الحاسوبية: تاريخها، موضوعها، منهجها.

كما وقدمنا في المبحث الثاني طريقة معالجة مستويات الدرس اللساني للغة العربية باعتماد آليتي التحليل والتوليد، من خلال المقارنة بين اللغة الطبيعية واللغة الاصطناعية، وقد توصلنا من خلال تحليلنا المقتضب لمعالجة مستويات الدرس اللساني في اللغة العربية إلى عدة صعوبات التي قد تعترض المعلوماتي في ذلك، من خلال تقديم مجموعة من الأمثلة في اللغة العربية.

المراجع:

[1]- فطوم القريبش، كيفية إنشاء قاعدة معلوماتية صرفية للفعل المشتق من الجذر الثلاثي السالم في اللغة العربية، بحث لنيل شهادة الماستر، ص:13، جامعة سيدي محمد بن عبد الله، كلية الاداب والعلوم الانسانية سايس فاس، سنة 2007.

[2]- François Yvon , Une petite introduction au Traitement Automatique des Langues Naturelles.p 1.2.3 26 avril 2007.

[3] – فطوم القريبش، كيفية إنشاء قاعدة معلوماتية صرفية للفعل المشتق من الجذر الثلاثي السالم في اللغة العربية، بحث لنيل شهادة الماستر، ص:16. جامعة سيدي محمد بن عبد الله، كلية الاداب والعلوم الانسانية سايس فاس، سنة 2007.

[4]-Jacqueline Léon, La constitution du TAL étude historique les dénominations et des concepts Marcel Cori p: 26. HAL Id: halshs-00158854 https://halshs.archives-ouvertes.fr/halshs-00158854 Submitted on 29 Jun 2007

[5]- ومما تهتم به اللسانيات التطبيقية تطوير أنظمة معلوماتية خاصة بالمعالجة الآلية للغة، إنها بخلاف اللسانيات النظرية التي تهتم بدراسة اللغات الطبيعية من خلال مثلا الدراسة التاريخية التي تبحث في الأصول التاريخية للغة ومراحل تطورها أيضا، أو من خلال الدراسة البنيوية التي تقوم على دراسة المكونات الداخلية للغة باعتبارها بنية مغلقة مكتفية بذاتها.

[6] -http://www.larousse.fr/dictionnaires/francais/travail_travaux/79284/synonyme.

[7] – سالم الرامي، بلورة نسق لتوليد الصور المعجمية والكلمات في اللغة العربية، بحث لنيل شهادة الدكتوراه، ص، ص: 46-47. بحث لنيل شهادة الدكتوراه، سنة 2004-2005.

[8] – عمر مهديوي، توليد الأسماء من الجذور الثلاثية المقلة مقاربة لسانية حاسوبية، رسالة لنيل دبلوم الدراسات العليا ص:60. رسالة لنيل دبلوم الدراسات العليا، 1418-1419هـ/1998-1999م.

[9] – سالم الرامي، بلورة نسق لتوليد الصور المعجمية والكلمات في اللغة العربية، ص: 49. بحث لنيل شهادة الدكتوراه، سنة 2004-2005.

[10]-Guy Perrier,Initiation au traitement automatique des langues (http://www.loria.fr)

[11] - Aené Boite, Hervé Bourlard, thierry Dutoit, Joël Hancq , et Henri Leich ,Traitement de la parole, nouvelle édition, ISBN 2-88074-388-5 , 2000 presses polytechnique université romands, p : 6

[12]- Youssef Tahir, Modélisation A Objets d'une Base de Données Morpho-Syntaxiques pour la Langue Arabe, Thèse de Diplôme de Doctorat.P :58. Spécialité Informatique 17 juin 2006.

[13]– المقصود بذلك بأن الجملة المنتجة تتقيد بالقواعد الإعرابية والنحوية التي يقوم عليها تكوين تركيب معين لجملة ما.

[14]– المقصود بذلك بأن الجملة تتقيد بالسمات الدلالية المقترنة بكل كلمة على حدة في التركيب النحوي المنتج.

[15]-Youssef Tahir, Modélisation A Objets d'une Base de Données Morpho-Syntaxiques pour la Langue Arabe, Thèse de Diplôme de Doctorat.P 56:p5: Spécialité Informatique 17 juin 2006.Fes.

[16] انظر-Traitement automatique de la langue (texte et parole) PP 32 33[Rabat], DL 2006 Dr Yousfi abdellah

[17]– فطوم القريبش، كيفية إنشاء قاعدة معلوماتية صرفية للفعل المشتق من الجذر الثلاثي السالم في اللغة العربية، بحث لنيل شهادة الماستر، ص: 66. جامعة سيدي محمد بن عبد الله، كلية الاداب والعلوم الانسانية سايس فاس، سنة 2007.

كلمة شكر

أتوجه بالشكر الخالص لكل من سهر على تنظيم هذه التظاهرة العلمية القيمة، وأخص بالذكر اللجنة المنظمة والأساتذة الاجلاء المشرفين على قراءة هذه البحوث القيمة وتصويب أخطائها مضمونا وشكلا، والشكر موصول أيضا للدكاترة المشتركين في هذه المجلة المتميزة والتي تصبو على العموم إلى الرقي باللغة العربية لتساير الركب الحضاري والتطور التكنولوجي تمييزا للغة القرآن وتمجيدا لها.

السيرة الذاتية

فطوم القريبش، أستاذة السلك الثانوي التأهيلي بالجهة الشرقية بالمغرب، حاصلة على دكتوراه وطنية بميزة مشرف جدا في محور اللغة العربية تخصص اللسانيات الحاسوبية، بجامعة سيدي محمد بن عبد الله، كلية الآداب والعلوم الإنسانية، فاس، شاركت في عدة ندوات منها على سبيل المثال لا للحصر: ندوة وطنية حول موضوع: الكتاب الأمازيغي إشكاليات التأليف والقراءة،ومقال نشر في مجلة وقائع الندوة الدولية الرابعة للمعالجة الآلية للغة العربية CITALA 12 بالرباط، و مقال شاركنا به في رومانيا ونشر بجريدة Journal of Modern Education Review (usa) بعنوان:" Arabic wordnet :new content and new applications"، عضوة مع فريق ابتكارات بالمدرسة المحمدية للمهندسين بالرباط، عضوة مع منظمة الألكسو بتونس، عضوة في جمعية هندسة اللغة العربية بالرباط.

Computational linguistics and natural language machine processing

Fettoum Krieche
Professor of secondary education
Member of the Association of Language Engineering at the Mohammedia School of Engineers in Rabat
Fettoum.krieche@gmail.com

Abstract: ***We aim through this article to introduce computational linguistics (its history, subject, method, pioneer) then we will distinguish between natural language and artificial language and define the mechanisms of generation and transformation as two important processes in the automatic processing of natural language. Areas of investment in automatic processing of natural languages will be pointed with an appendix to each topic with synthetic summary.***

Key words: C***omputational linguistics, automatic language processing, natural language, artificial language, generative and transformative mechanisms, linguistic lesson.***