

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة أبو بكر بلقايد
UNIVERSITÉ DE TLEMCEN



كلية الآداب واللغات
قسم اللغة العربية وآدابها

مذكرة مقدمة لنيل شهادة الماستر في اللغة العربية وآدابها

تخصّص: لسانيات تطبيقية
رمز المذكرة: 32/017/ل ت

الموضوع:

تطبيقات لسانيات المدوّنة الحاسوبية على اللغة العربية

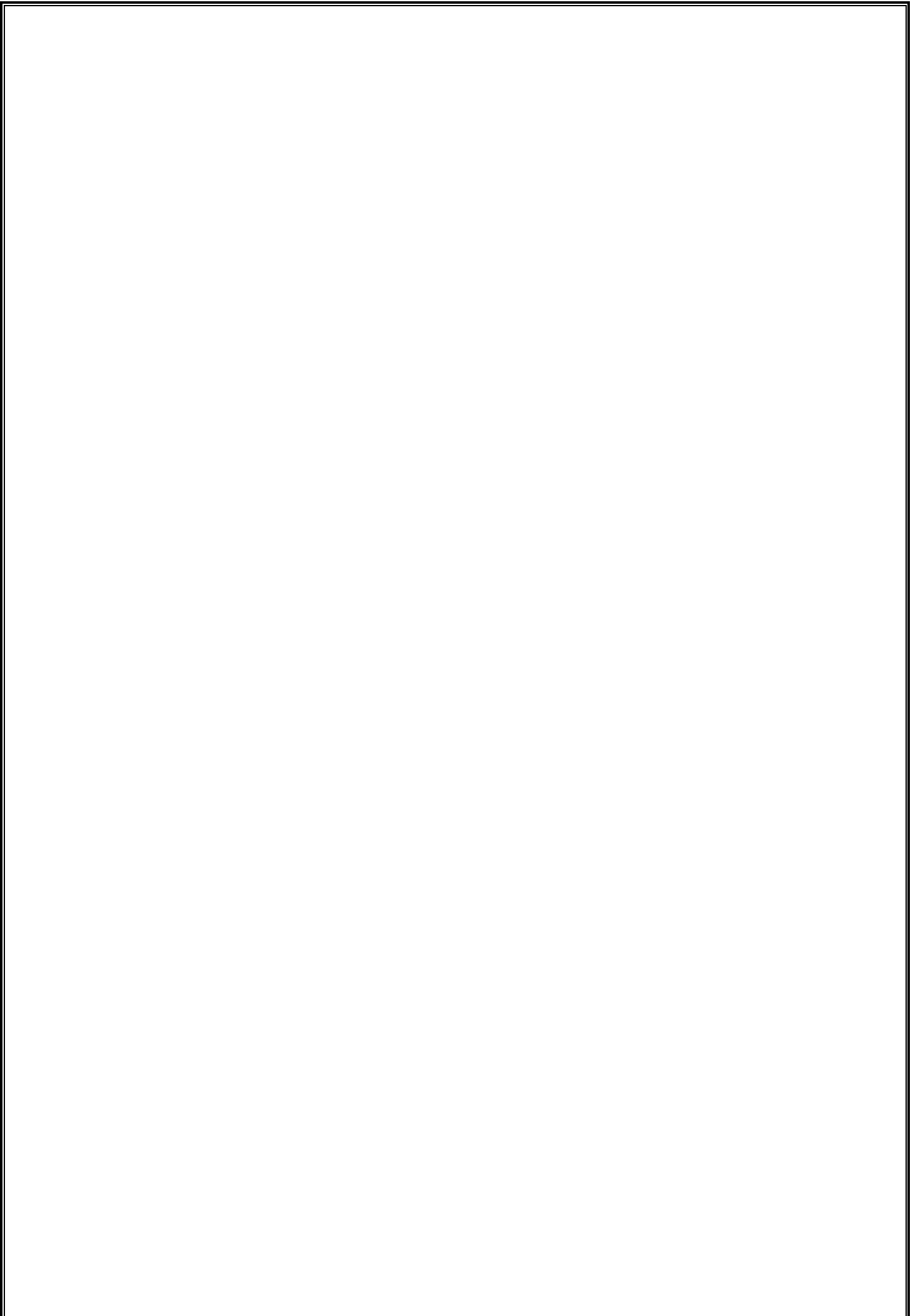
إشراف:
أ.د الهادي شريفي

إعداد الطالبة:
مامة كروش

لجنة المناقشة

لجنة المناقشة		
رئيسا	خالدي هشام	أ.الدكتور
ممتحنة	شيخي نورية	أ.الدكتورة
مشرفا مقررا	الهادي شريفي	أ.الدكتور

العام الجامعي : 1439-1440هـ/2017-2018م



إهداء

يسعدني ويسرني أن أهدي هذا العمل إلى والديّ

الكريمين برًا بهما وتزيّدًا في طاعتهما و دعائهما.

إلى من هم أنس عمري و نور حياتي : أخواتي ، و

أخي الغالي :علي و أولاد أختي، وعمّاتي وأعمامي.

إلى جميع أفراد عائلتي و أولادهم .

إلى صديقتي وزملائي دفعة " السنة الثانية ماستر "

تخصص "لسانيات تطبيقية " .

إلى أستاذي الفاضل " الهادي شريقي "

إلى كل من له الفضل بعد الله عز وجل في الأخذ بيدي

إلى طريق العلم والمعرفة .

شكر و تقدير

الحمد لله رب العالمين والصلاة والسلام على المبعوث
رحمة للعالمين سيدنا محمد وعلى آله وصحبه أجمعين .

الحمد لله أشكره و أحمده على إنجاز هذا العمل .

منتهى الشكر والتقدير لأستاذي الفاضل " الهادي شريقي "

الذي لم يبخل عليّ بالجهد والوقت ولا بالنصيحة والتوجيه التي

ذلت لي كل الصعوبات ، وكذلك دعمه النفسي والمعنوي قبل

الأكاديمي .

وشكرا طيبا وتقديرا خالصا للأستاذين الفاضلين عضوي

لجنة المناقشة الأستاذ الدكتور " خالد هاشم " و الأستاذة

الدكتورة " شيخي نورية " لتفضلهما بقبول مناقشة هذه المذكرة .

كما أشكر كل من ساندني بدعواته الصادقة، وتمنياته

الخالصة .

تمهيد

تنتمي العربية إلى فصيلة اللغات السامية، ولقد كُتِبَ لهذه اللغة الخلود بنزول القرآن الكريم، قال الله تعالى: ﴿إِنَّا أَنْزَلْنَاهُ قُرْآنًا عَرَبِيًّا لَعَلَّكُمْ تَعْقِلُونَ﴾¹.

ومن فضل القرآن الكريم على اللغة العربية أيضا أنّ عمل على توحيدها وطبعها بطابع خاصّ فيه العموميّة والشمول، بحيث تيسّر لهذه اللغة أن تكون لغة العرب عامّة، وأن تتغلب على الكثير من اللهجات، ثمّ استأنفت اللغة العربية بعد ظهور الإسلام مسيرتها التاريخية وعاشت دورها في تطوّر ونماء².

وقد تهيّأت للعربية في العصر الحديث عوامل جديدة للتطور والرقى، منها انتشار التعليم، وظهور الصحافة ووسائل الإعلام المرئية والمسموعة وإنشاء مجامع اللغة العربية في معظم البلدان العربية، والاهتمام بعقد اللجان والمؤتمرات التي تبحث مشكلات اللغة وتصنع لها الحلول المناسبة، وجعل العربية لغة التعليم في جميع المدارس والمعاهد وفي كثير من الجامعات العربية إضافة إلى أنّها أصبحت اللغة الرسمية في جميع الأقطار العربية، ولغة رسمية في الكثير من المحافل الدولية³.

وعلى سبيل المثال، فإنّ الشبكة العالمية، والتي دعانا ما تشهده من مآس لغويّة إلى فتح هذا الموضوع، نرى في طياتها من المبشرات ما ينبئ بأنّها قد تكون هي نفسها وسيلة إلى نهضة اللغة العربية السليمة وانتشارها، فالغالبية العظمى من المواقع العربية تستخدم اللغة الفصحى لا العامية، وهناك الكثير من المواقع التي تعنى بالشعر والأدب والتراث، وهناك مواقع مخصّصة للدفاع عن اللغة العربية، ومواقع تساهم في توحيد تعريب المصطلحات، ومواقع تميّز بلغتها الرّاقية وأسلوبها الجذاب، بل هناك مواقع وجهود تسعى إلى تعريب أسماء النّطاقات نفسها (أي عناوين المواقع)، لتصبح المواقع العربية على الشبكة العالمية عربيّة من بابها إلى محرابها، فلغتنا ستبقى خالدة

1- سورة يوسف، الآية 2.

2- ينظر: - جاسم محمود الحسون، د. حسن جعفر الخليفة، طرق تعليم اللغة العربية في التعليم العام، جامعة عمر المختار، دار الكتب الوطنية، بنغازي، 1996، ص 3.

3- ينظر: المرجع نفسه، ص 35.

بإذن الله، الذي يحفظها بحفظ كتابه، وبجهود المخلصين من أبنائها الذين لن يرضوا لسيادتها في أوطانهم بديلاً¹.

إذا كانت الشبكة العالمية قد أظهرت حجم المأساة التي تعاني منها اللغة العربية على أيدي أبنائها، وكانت وسيلة لتفشي لغة الضعفاء منهم، فهي في الوقت نفسه وسيلة لنشر الوعي اللغوي، وللعودة بالعربية إلى عصورها الزاهرة.

يأتي على رأس المواقع التي تقدم خدمات جليّة للغة العربية، موقع مكتب تنسيق التعريب التابع للمنظمة العربية للثقافة والعلوم، وعنوانه على الشبكة العالمية www.arabization.org.ma وهذا الموقع كنز لغوي عربي ثمين، لما يحويه من مواد ونشرات وأبحاث ومعاجم، وأهم خدمة يؤدّيها هذا الموقع للغة العربية ولأبنائها هي مساهمته في توحيد ترجمة المصطلحات العلميّة الأجنبيّة إلى اللغة العربية، وفي تقريب المسافات بين العاملين في التعريب، إلا أنّ وجود عدّة مجامع كان من سلبياته عدم وحدة المصطلحات العلميّة العربية، وذلك لعدم وجود اتصال وتنسيق بين المجامع المختلفة، فتداعى المسؤولون عن المجامع اللغوية العربية إلى تأسيس مكتب تنسيق التعريب في الرباط عاصمة المملكة المغربية، ومنذ تأسيسه قام المكتب بنشر خمسين قاموساً للمصطلحات العربية الموحدة التي اتفقت عليها المجامع اللغوية العربية، ولنا أن نتصوّر مدى أهميّة هذه الخدمة العظيمة، ونعني بذلك توحيد المصطلحات العلميّة في البلاد العربية²، وقد وظّف المكتب كل الوسائل التّقنية الحديثة في أداء هذه الخدمة النبيلة.

فقد بلغ التّقدم التّكنولوجي أوجّه، بسبب التّقدّم الهائل الذي أحرزه العلم، في بناء أجيال متطوّرة من الحواسيب، وقد كان لهذا التطّور أن يدخل في مجالات الحياة كلّها، وأن يعمل على تجديد النظر العلمي، والأساليب التي يطبّقها العلماء في ميادين العلم المختلفة.

1- فداء الحندي، العرب والعربية في عصر الثورة الحاسوبية، دار الفكر، دمشق، 2003، ص 145.

2- ينظر: المرجع نفسه، ص 146-147.

تعمل الحواسيب على أساس مشابه للعمل الذي يؤديه العقل الإنساني، ولكن لا يستطيع أحد أن يقول إن العمل هنا مطابق للعمل هناك، غير أنه من المؤكد أن أوجه تشابه كثيرة بينهما، هي التي جعلت الحاسوب يتفوق على كل الإنجازات العلمية السابقة، بل إن الإنجازات العلمية والحضارية اللاحقة كلها ما كانت لتنال حظها من الوجود لو لم يكن الحاسوب قد عمل على حل كثير من مشكلاتها المعقدة، وتجاوز عقباتها.

وقد كان واضحا لدى العلماء أن حلقة الوصل بيننا وبين الإبداعات التي يمكن أن يقدمها لنا الحاسوب، هي اللغة التي يتعامل بها الحاسوب، وبها تتعامل معه، من هنا توجهت أنظار العلماء إلى الإفادة من حركية اللغات الإنسانية في بناء النظام اللغوي للحاسوب¹.

وقد أدى هذا إلى ولادة فرع من فروع اللسانيات هو علم اللسانيات الحاسوبية:

²Computational linguistics.

وهي من بين المجالات اللسانية التي حصل فيها أيضا تقدم، فالثورة المعرفية الحديثة والثورة الآلية مكنتنا من التوصل إلى معالجة اللغة آليا في عدد من جوانبها، وتوظيف عدد من التقنيات التي يصعب القيام بها دون الآلة، مثل الإحصاء والتخزين الذي يتطلب ذاكرة ضخمة، ... إلخ. وقد طرح هذا الوضع الجديد عددا من الأسئلة وأفرز متطلبات جديدة، وكان لابد أن تفتح اللغة العربية على غرار غيرها من اللغات، ميدان الآلة، وتدخله من الباب الواسع. إلا أن هناك شروطا لتهيئ الأدوات الوصفية والصورية الضرورية لجعلها قادرة على اقتحام الميدان. وليس المشكل أساسا مشكل "أشكال الحروف" التي يجب أن توضع على لوحة المفاتيح، كما يعتقد البعض. فهذا المشكل بسيط إذا ما قورن بالمشاكل الجوهرية في معالجة اللغة العربية آليا. وعلى رأس هذه المشاكل

1- سمير شريف إستيتيه، اللسانيات المجال، والوظيفة، والمنهج- عالم الكتب الحديث- 2005- ص 527.

2- ينظر: المرجع نفسه، ص 528.

البّحث في قواعد اللّغة (نحوها وصرفها)، والبّحث في مفردات اللّغة (أو معجمها)، وفي أصواتها (الصّوتيات)¹.

فاللّسانيات مصطلح أتى من اللّسان، واللّسان يعني اللّغة، وتسمّى أيضا الألسنيّة، وعلم اللّغة، وتعرّف اللّسانيات بأنّها: "الدراسة العلميّة للّغات البشريّة".

وتنقسم اللّسانيات إلى فروع متعدّدة، يختلف تعدّدها من مرجع إلى آخر، حسب جهد المؤلّفين في هذا الميدان، منها علم اللّسانيات البيولوجي، وعلم اللّسانيات الاجتماعي، وعلم اللّسانيات النفسي، وعلم اللّسانيات الأثنوبولوجي، وعلم اللّسانيات الرّياضي، واللّسانيات العامّة، واللّسانيات الوصفية، واللّسانيات التّاريخية، واللّسانيات النظريّة، واللّسانيات التّطبيقية، حيث تعدّ اللّسانيات الحاسوبية أحد فروعها، التي نحن بصدد دراستها².

ومن الاهتمامات الأخرى في مجال اللّسانيات التّطبيقية: التّخطيط اللّغوي Language planing وتعلّم اللّغة بالحاسوب Computer-assisted language learning وعلاقة اللّغة بالتّربية، والتّرجمة، والتّرجمة الآلية Machine aided translation، والذكاء الاصطناعي Artificiel Intelligence، واللّسانيات الحاسوبية Computational linguistics³، والتي موضوعنا هذا هو أحد ميادينها الفرعية.

فمن الأمور التي لا يجهلها أحد أنّ الحاسوب هو أداة القرن الحالي دون منازع، فقد دخل تقريبا كلّ بيت ومؤسسة عامّة وخاصّة، وأدّى استخدامه في مجالات الحياة العلميّة والعملية إلى تطوّرات كبيرة في هذه الحقول المختلفة. ومن هذه المجالات ميدان اللّغة التي تميّز بها الإنسان

1- ينظر: منظمة الأمم المتحدة للتربية والعلوم والثقافة "يونسكو" - تقدم اللسانيات في الأفطار العربية- وقائع ندوة جهوية أبريل 1987- دار

الغرب الإسلامي- 1991- بيروت- لبنان- ص 25.

2- إيمان صبحي سلمان دلول- معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللّغة العربية- رسالة ماجستير- كلية الآداب- قسم اللّغة العربية- الجامعة الإسلامية- غزة- 2014- ص 01.

3- ينظر: محمّد محمّد يونس- مدخل إلى اللّسانيات- دار الكتب الوطنية- بنغازي- ليبيا- 2004- ص 15.

عن سائر المخلوقات. ويمكننا الاستفادة من الحاسوب في النشاط اللغوي بوجه متعدد، يمكننا تلخيصها في ثلاث صوّر رئيسية هي:

- الاستعمال العام،
- الاستعمال الخاص،
- والبحث اللساني الحاسوبي المتخصّص.

غير أنّ أحدث وأهمّ مجال خاض فيه الحاسوب في خدمة البحث اللغوي هو ما يعرف بلسانيّات المدوّنات (اللغويّة) Corpus linguistics، وهو موضوع الدّراسة التي بين أيدينا¹.

فالمدوّنة اللغوية Corpus هي مجموعة ضخمة من النصوص اللغوية تمّ جمعها من مصادر مختلفة: كتب ومجلات وحصص إذاعية وتلفزيونيّة مفرّعة ونشريات ودوريات، تمّ تنسيقها وفهرستها وتخزينها حاسوبياً لإتاحة استغلالها من قبل المستخدمين لأغراض متعدّدة، منها ما هو للاستشهاد ومنها ما هو لتدريب برمجيات حوسبة اللّغة، أو تطبيقات حاسوبية أخرى.

أمّا الحوسبة: فهي تحديد مجموعة من العمليّات التي تجعل من تنفيذ مهمّة (ما) تنفيذاً آلياً بواسطة الحاسوب، مقابل التنفيذ اليدوي للبشر².

وبالتالي فلسانيّات المدوّنات الحاسوبية يصبح معناها دراسة اللّغة على ضوء النصوص اللغوية المدوّنة والمخزّنة حاسوبياً.

1- محمود إسماعيل صالح/ أستاذ اللسانيات التطبيقية- السبت 05 أبريل 2014. Dr-Mahmoud-ismail-saleh-
blogspot.2014.

2- إيمان صبحي سلمان دلول- معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللّغة العربية- رسالة ماجستير- كلية الآداب- قسم اللّغة العربية-
الجامعة الإسلامية- غزة- 2014- ص قائمة المصطلحات العربية.

مقدّمة

بسم الله الرحمن الرحيم

الحمد لله الذي خلق الإنسان، علّمه البيان، وهبه التّمييز والحكمة وكرّمه على سائر مخلوقاته، ومن آياته اختلاف الألسنة والألوان، والصّلاة والسّلام على رسول الله أفصح من تكلم بالعربيّة وأبان، وبعد:

فدّول العالم تنقسم جغرافيا إلى دّول متقدّمة، ناميّة، ومتخلّفة أي حسب ثرواتها ومواردها الماليّة والطبيعيّة، لكن علميّا تصنّف إلى صنفين لا ثالث لهما أي إلى دّول متقدّمة تكنولوجيا ودّول متخلّفة تكنولوجيا، وهذا ما يؤكّد عليه التنافس الدّولي على المبتكرات التكنولوجية وعلى اللّحاق بالثورة التّقنية والمعلوماتيّة.

فالتكنولوجيا قد مسّت علوم عديدة ومتنوّعة، ومن بين هذه العلوم: "علم اللّغة" الّذي عرف قفزة نوعيّة وتحولات ملحوظة، مصطلحاته من "اللّسانيات" إلى "الألسنيّة" إلى "علم اللّسان" إلى "علم اللّغة الحديث"؛ إلّا أنّها تصبّ في وعاء واحد وهو دراسة اللّسان البشري بطريقة علميّة وقد تفرّع عن "اللّسانيات" فروع عديدة أهمّها ما يعرف بـ "اللّسانيات التّطبيقية" الّتي يعدّ علم اللّغة الحاسوبي أحد مجالاتها؛ والّذي أحدث ثورة معلوماتيّة خاصّة بعد ظهور ما يعرف بالتّخزين الآلي أو الإلكتروني للنصوص، والّذي أطلق عليه أهل الاختصاص بـ "المدوّنات الرقميّة"، وهذا ما أدّى بدوره إلى بروز "لسانيّات المدوّنة"، الّتي أصبحت محل اهتمام اللّغويين والحاسوبيين على حدّ سواء في الدّول العربيّة، أمّا في دولنا العربيّة، فقد عرف هذا الميدان تأخرا ملحوظا.

وانطلاقا ممّا ذكرنا سالفًا، جاءت ضرورة هذه الدّراسة لتضيف لبنة تساعد في بناء صرح هذا العلم الّذي بدأ يستقلّ بذاته "علم لسانيات المدوّنات الحاسوبية"، وتطرح هذه الدّراسة الإشكاليّة التّاليّة الأساسيّة:

- كيف تطبّق الأدوات الحاسوبية على المتون اللّغويّة؟

وقد تفرع عن هذه الإشكالية مجموعة من التساؤلات وهي كالتالي:

-ماذا نعني باللسانيات الحاسوبية؟

-ماهي المدونة وما هي أهم خصائصها؟

-كيف تبنى المدونة؟

-وما هي أبرز وأهم الأدوات الحاسوبية التي تطبق على المدونات؟

فبالرغم من أنّ هذا الموضوع لم يكن من اختياري شخصياً، وإتما من اقتراحات الإدارة ومن عرض الدكتور "الهادي شريف" الأستاذ المشرف على مذكري، إلا أنه كان حسب ميولي ورغبتي، والحمد لله الذي وقّني لما أردت، فكان اختياري إذن لهذا الموضوع عن رضا وقناعة. ومن هذا المنطلق عزمنا على توسيع أفكارنا، وهيكله بحثنا برسم الخطة النهائية التي تبينهاها في نهاية المطاف بعد أن رسمنا عدّة خطط أولية.

وعلى هذا لنحو، جزأنا بحثنا إلى ثلاثة فصول وكلّ فصل ينقسم إلى ثلاثة مباحث، وكلّ مبحث يتفرّع بدوره إلى مطلبين:

ففي الفصل الأول، حاولنا من خلاله إعطاء نظرة شاملة عن اللسانيات الحاسوبية وكان عنوانه كالتالي: اللغة العربية بين اللسانيات والحوسبة، وتفرّع عنه ما يلي:

المبحث الأول: اللسانيات الحاسوبية؛

المبحث الثاني: مجالات علم اللغة الحاسوبي ومراحل نشأته وتطوره؛

المبحث الثالث: دواعي حوسبة اللغة العربية وبعض الجهود الرائدة فيها؛

أمّا في الفصل الثاني، فقد سلطنا الضوء على المدونات اللغوية الحاسوبية وتضمّنت مباحثه مايلي:

المبحث الأول: المدونة اللغوية وأنواعها؛

المبحث الثاني: أهمية المدونات اللغوية الحاسوبية؛

المبحث الثالث: إنشاء المدوّنات الحاسوبية، وأهمّ ما وجد منها في المواقع الإلكترونية .
أمّا بالنسبة للفصل الثالث، فقد احتوى على تطبيقات لسانيات المدوّنة. وقد تناولت مباحثه
مايلي:

المبحث الأول: الكشّافات السياقية؛

المبحث الثاني: عرض المدوّنة التي اشتغلنا عليها؛

المبحث الثالث: أعمال تطبيقية على المدوّنة.

افتتحنا بحثنا بمقدّمة، ويليهها مدخل وختمنا دراستنا بعرض أهمّ النتائج المتوصّل إليها.
ومن أهمّ المصادر والمراجع التي اعتمدنا عليها في إعدادنا في بحثنا، هي كالتالي:
مقدّمة في علم اللّغة الحاسوبي لـ"عصام الدّين أبو زلال"، المدوّنات النصّية ودور اللّغة العربيّة
في التّعامل معها لـ"سلوى حمادة"، استثمار المدوّنة في تدريب المترجم لـ"فاطمة الزّهران
توتاوي".

إنّ طبيعة الدّراسة تستلزم الاعتماد على المنهج الوصفي -التّحليلي .
كما لا يخلو أيّ بحث في بدايته من الصعوبات، خاصّة في جمع المادّة العلميّة لكن مع
الوقت وتوجيهات الأستاذ المشرف تبدّدت هذه العوائق، واستطعنا أن نجتمع قدر الإمكان
المصادر والمراجع التي تتناسب والموضوع.
وفي الأخير أحمد الله سبحانه وتعالى الذي وفقني لإتمام بحثي كما أتقدّم بشكري الخالص
ومع كل احتراماتي إلى أستاذي الفاضل "د. الهادي شريفني" الذي أرشدني ونصحني وصوّب
أخطائي.

في تلمسان: 03 شعبان 1439هـ/18 أبريل 2018 م.

كروش مامة.

توطئة

يدرس اللسانيون اللغة من جوانب مختلفة وفقاً لأغراضهم المتنوعة، واهتماماتهم المختلفة، وقد نتج عن ذلك نشأة فروع مختلفة للسانيات ومنها: اللسانيات الحاسوبية التي تهدف إلى دراسة اللغة من وجهة نظر حاسوبية.

فما هي اللسانيات الحاسوبية؟

المبحث الأول: اللسانيات الحاسوبية

أولاً: مفهوم اللسانيات الحاسوبية

يعدّ علم اللغة الحاسوبي من أحدث فروع علوم اللغة، وبالتحديد هو أحد فروع علم اللغة التطبيقي، ويتأسس على الاستفادة من معطيات الحاسوب في دراسة القضايا اللغوية المتعددة مثل: رصد الظواهر اللغوية بمستوياتها الصوتية والصرفية والنحوية والدلالية والبلاغية والعروضية، ويهدف إلى معالجة عمليات تحليل الظواهر اللغوية والنصوص التراثية وإجراء العمليات الإحصائية وصناعة المعاجم والترجمة الآلية وتعليم اللغات وإجراء التحليل على السياقات اللغوية المتعددة، كما يدرس علم اللغة الحاسوبي وضع اللغات الطبيعية البشرية في صيغ رياضية، من أجل دراسة اللغة بوصفها ظاهرة حاسوبية معلوماتية يمكن معالجتها آلياً من خلال الحاسوب، بما يحقق السرعة والدقة العلمية في البحوث اللغوية، ومن أجل ترجمة النصوص اللغوية ترجمة آلية فورية¹.

ويعدّ هذا المجال أحد فروع علم الذكاء الاصطناعي (Artificial Intelligence-AI) وهو ما يتم فيه الاستفادة من مزايا معطيات الحاسوب، لتوظيف التطبيقات الذكية، ومنها معالجة اللغات

1- عصام الدين أبو زلال - مقدمة في علم اللغة الحاسوبي - دار الوفاء لدنيا الطباعة والنشر - الإسكندرية - 2016م - ص 11.

الطبيعية، والتي مفادها تصميم وإنجاز أنظمة حاسوبية تحاكي قدرة الإنسان في التعامل مع اللغة الطبيعية تحليلاً وتركيباً¹.

ونعني بالتحليل (Analysis): توظيف القواعد والقوانين الحاسوبية في معالجة اللغة العربية، وتطبيقها على المادة اللغوية صوتاً ونصاً.

أما التركيب (Synthesis): فهو توظيف القواعد والقوانين الحاسوبية، لإنتاج المادة اللغوية صوتاً ونصاً².

وهناك بعض الأمثلة على الأنظمة الحاسوبية التي تحاكي قدرة الإنسان في التعامل مع اللغة الطبيعية:

(1) - مثال على التحليل والتركيب: الترجمة الآلية (Machine Translation)

(2) - مثال على التحليل:

* التدقيق الآلي (Automatic correction check) ومن أمثلته:

أ. التصحيح الإملائي (Automatic Spelling correction)

ب. التصحيح النحوي (Automatic Grammer check)

* التشكيل الآلي (Automatic Diacritization)، ومن أمثلته:

أ. التلخيص الآلي (Automatic summarization)

ب. التعرف الآلي على الخط المكتوب (Optical character recognition-OCR)

1- إيمان صبحي سلمان دلول - معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللغة العربية - رسالة ماجستير - كلية الآداب - قسم اللغة العربية -

الجامعة الإسلامية - غزة - 2014 - ص 02.

2- المرجع نفسه، ص 02.

ج. التعرف الآلي على الكلام (Speech recognition)

(3)- مثال على التركيب:

- التوليد الآلي للكلام (Speech synthesis)

- التوليد الآلي للمعجم (Automatic Lexicon generation)¹.

كما أنّ اللغة العربية تمتاز بجملة من الخصائص التي شكّلت سمات بارزة اقترنت بهذه الأبعدية السّاحرة، وكانت علامة جليّة فيها.

والواقع أنّ اللسانيات الحاسوبية هي حقل جديد في اللسانيات التطبيقية والنظرية فهي بشكل رئيس حقل يعالج اللغات البشرية وبرمجتها من خلال الحاسب الإلكتروني (الكمبيوتر)².

ولحسن الظنّ أن عددا من الأعمال اللسانية العربية أصبحت توفّر نماذج للدراسة التحوّية والمعجمية الكافية والمصوّنة التي يمكن اعتمادها وتعميمها لبناء قواعد للمعطيات في الآلة.

فهي توفّر فرصة للساني الحاسوبي العربي لتوظيفها في تحليله الآلي للغة، ونستغرب إذ نجد مهندسي اللغة يوظفون في قواعد معطياتهم المعلومات والضوابط التقليدية، رغم كونها غير كافية وغير صالحة، مع أنّهم يستعملون (أحيانا) أحدث التقنيات البرمجية أو تقنيات المعالجة. وهذا يخالف ما يجري في اللسانيات الحاسوبية غير العربية حيث نجد المحلّلات في الدول المتقدمة تعتمد التحاليل اللسانية الحديثة للغات، وعلى الأخصّ الأنحاء التوليدية.

1- المرجع السابق ص 03.

2- معالي هاشم علي أبو المعالي - الاتجاه التوافقي بين لسانيات التراث واللسانيات المعاصرة - الدكتور عبد الرحمن الحاج صالح أ نموذجاً - رسالة

ماجستير - جامعة بغداد - 2014 - ص 375.

وإذا كان الحوار بين اللساني والحاسوبي العربيين في بدايته، فإنّ التّقدّم في اللّسانيات الحاسوبية متوقّف على التّعاون والتّحاور المتينين بين الطّرفين، بعيداً عن الأهداف التجاريّة المحضة التي غالباً ما تقف عائقاً دون وجود حوار فعلي وصريح¹.

وتعدّدت تعريفات علم اللّغة الحاسوبي لدى الدّارسين، ومنها:

1. تعريف عبد العزيز عبد الله ميهوبي بجامعة الإمام محمد بن سعود بالمملكة العربيّة السعوديّة:

هو دراسة الجوانب الحاسوبية للّغة والمشاكل الشّائعة التي تواجه المعالجة الحاسوبية للّغة في صورتها المكتوبة أو المنطوقة.

2. تعريفات حمادي الموقت: هو ذلك الإطار التقني الذي تتصهّر داخله تجلّيات اللّغة الطبيعيّة

وتمظهراتها، في تفاعل يحاول ربط هذه الأخيرة (أي اللّغة) بالحاسوب وأنظمتها، أو الفرع الذي يبحث في مختلف طرق وكيفيات الاستفادة من قدرات الحاسوب في تحليل اللّغة ومعالجتها وتعليمها وتعلّمها، أو محطة تقنيّة تدرس بنية اللّغة الطبيعيّة في ارتباطها بجهاز الحاسوب، على اعتباره أساس المنظومة المعرفيّة الأولى، في تطويع هذه اللّغة كمادّة تكنولوجيّة ترمج في الحواسيب الإلكترونيّة.

3. تعريف مازن الوعر: هو الإطار العلمي الذي يبحث في اللّغة البشريّة كأداة طبيعة لمعالجتها

في الآلة ... وتتألّف مبادئ هذا العلم من اللّسانيات العامّة، بجميع مستوياتها التحليليّة الصوتيّة والنّحوية والدّلالية، ومن علم الحاسبات الإلكترونيّة، ومن الذّكاء الاصطناعي Artificial intelligence، والمنطق logique، ثمّ الرّياضيات، مشكّلة بذلك مبادئ اللّسانيات الحاسوبية².

1- منظمة الأمم المتحدّة للتّربية والعلوم والثّقافة "يونسكو" - ندوة جهويّة أبريل 1987 الرباط- دار الغرب الإسلامي - بيروت - لبنان-

1991- ص 26.

2- عصام الدّين أبو زلال- مقدّمة في علم اللّغة الحاسوبي- دار الوفاء لدنيا الطباعة والنشر- الإسكندرية- 2016م- ص 12.

ويُستنتج من التعريفات السابقة أنّ علم اللغة الحاسوبي علم يبيّن يجمع بين علمي اللغة ومجاله الرئيسة اللغات الطبيعية الإنسانية، والحاسوب ومجاله الرموز: من أجل تحويل اللغة إلى رموز رياضية، أي لإعداد اللغة الطبيعية الإنسانية لتصبح لغة تخاطب وتُحاور مع الحاسوب، لإنجاز المناشط اللغوية، وتوفير الجهد والوقت والتكلفة، وإتاحة الفرصة للتعلّم الذاتي والتعلّم عن بعد¹.

ثانياً: المصطلحات الدالة على علم اللغة الحاسوبي وفروعه:

تعددت المصطلحات العربية المعبرة عن علم اللغة الحاسوبي لدى المتخصصين من العرب المحدثين والمعاصرين، ولعلّ أشهر ما يأتي:

1. علم اللغة الحاسوبي،
2. علم اللغة المحوسب،
3. علوم اللغة الحاسوبية،
4. اللغويات الحاسوبية،
5. اللسانيات الحاسوبية،
6. اللسانيات الحاسوبية،
7. الهندسة اللسانية،
8. الألسنية المعلوماتية،
9. الهندسة التقنيّة للغة،
10. الهندسة التكنولوجية للغة².

ولقد اتفق علماء اللغة على تقسيم علم اللغة إلى شعبتين رئيسيتين هما:

1- المرجع السابق، ص 13.

2- المرجع السابق ص 14.

1- علم اللغة النظري أو ما يطلق عليه اسم علم اللغة العام:

Theoretical Linguistics Or General Linguistics

يقوم علم اللغة العام برسم الأسس المنهجية للتحليل اللغوي من جوانبه الصوتية والصرفية والنحوية والمعجمية. فإذا كانت أصوات اللغات تبدو لأول وهلة مختلفة متنافرة فإن كل أصوات اللغات تصدر من الجهاز الصوتي الإنساني، وهو مشترك عند كل البشر،¹ فالمكوّن النظري يُعنى بقضايا في اللسانيات النظرية، تتناول النظريات الصورية للمعرفة اللغوية التي يحتاج إليها الإنسان لتوليد اللغة وفهمها.²

كما يتناول القضايا النظرية المطروحة في علم اللغة الحديث من أجل استجلاء قدرات العقل على توليد المعرفة اللغوية، ثم صياغة هذه القدرات برموز رياضية منطقية عن طريق الخوارزميات المتنوعة.³

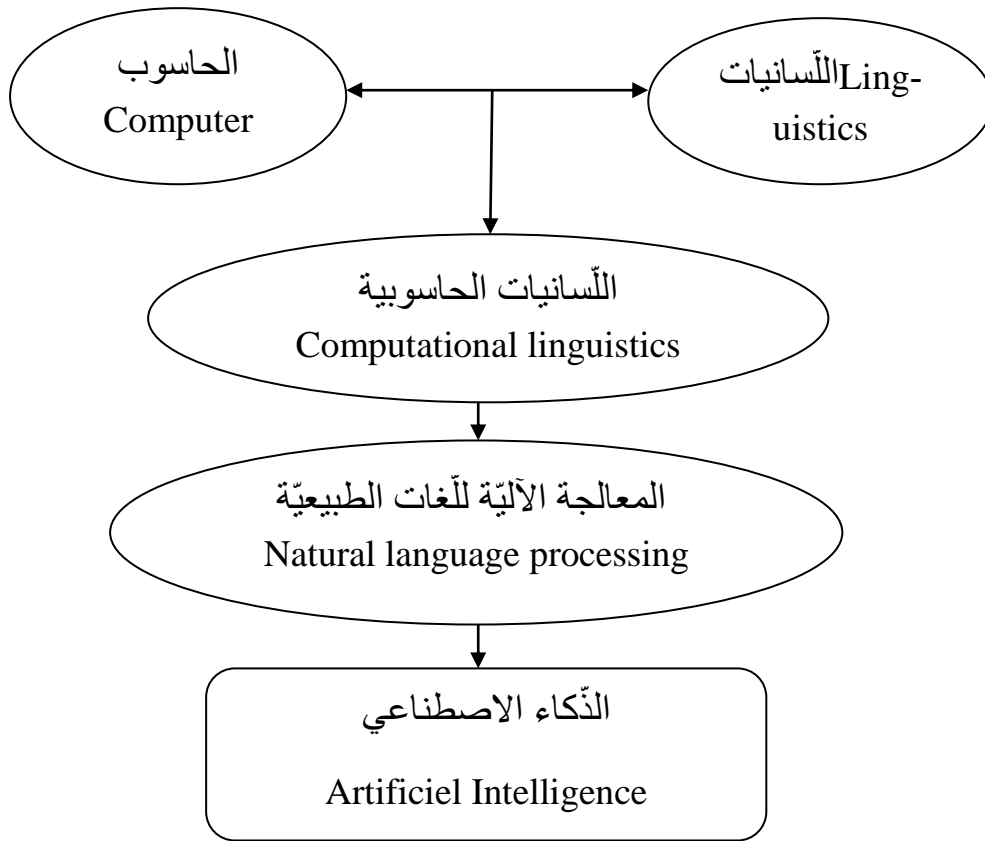
2- علم اللغة التطبيقي: علم مستقل بذاته له إطاره المعرفي الخاص به، وله منهج خاص به ومن فروعها ما يلي: علوم التربية، علم النفس، التخطيط اللغوي، علم الاجتماع، التعليمية، الترجمة والمعجمية، كما يصب اهتمامه على الناتج العملي لنمذجة الاستعمال اللغوي: بهدف إنتاج برامج ذات معرفة باللغة الطبيعية الإنسانية⁴، لتحسين التفاعل بين الإنسان والآلة، إذ أنّ العقبة الأساسية في طريق هذا التفاعل بين الإنسان والحاسوب إنما هي عقبة التواصل.⁵

-
- 1- محمد فتحي فهمي عبود- بناء مكنز عربي في مجال علم اللغة (المنهج والتطبيق)- الدار المصرية اللبنانية- 2004- ص 50.
 - 2- إيمان صبحي سلمان دلول- معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللغة العربية- رسالة ماجستير- كلية الآداب- قسم اللغة العربية- الجامعة الإسلامية- غزة- 2014- ص 03.
 - 3- عصام الدين أبو زلال- مقدّمة في علم اللغة الحاسوبي- دار الوفاء لدنيا الطباعة والنشر- الإسكندرية- 2016م- ص 14.
 - 4- المرجع نفسه ص 15.
 - 5- إيمان صبحي سلمان دلول- معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللغة العربية- رسالة ماجستير- كلية الآداب- قسم اللغة العربية- الجامعة الإسلامية- غزة- 2014- ص 04.

إذن فاللسانيات الحاسوبية تتضمن توجيهين رئيسيين:

- أولهما: لغوي حيث يرى اللغويين أنّ الحاسوب يقوم بالمعالجة الآلية للغة.
 - وثانيها: حاسوبي يرى الحاسوبيين: أنّ اللغة هي المادة التي تطبق عليها التقنيات الحاسوبية.
- وبالتالي يمكننا القول أنّ العلاقة بينهما علاقة تكاملية ولا نستطيع الفصل بينهما، لذا وجب التعاون فيما بينهم للخروج بنتائج وحلول تقضي على الصعوبات والعوائق التي تواجهه التحليل الحاسوبي للغة الطبيعية.

توضيح العلاقة بين الجانب النظري والتطبيقي لمعالجة اللغات الطبيعية من خلال الشكل التالي:¹



الشكل رقم 01: الإطار النظري والتطبيقي لمعالجة اللغات الطبيعية.

1- معتصم فتحي سليم الحمدان- نموذج محوسب محلّل نحوي للجمل الإسمية غير المشكولة في اللغة العربية- رسالة ماجستير- جامعة آل البيت

2002 على الرابط: <http://www.mgtasem.net/master-thesis>

وبالتالي لا نستطيع القول إنّ اللسانيات الحاسوبية تقتصر على جانب نظري بحت، أو دراسة لسانية محضة، بل هي مرحلة أولى تمهيدية للهندسة، لذلك يستطيع المتخصص اللغوي أن يتوقف عندها ويكتفي بها، ثمّ يواليها ميدان المعالجة الآلية للغة، يتمثل به دور المتخصص الحاسوبي، فيقوم بتصميم وإنجاز نظام آلي لمعالجة اللغة تحليلاً وتركيباً.

المبحث الثاني: مجالات علم اللغة الحاسوبي ومراحل نشأته وتطوره:

أولاً: مجالات علم اللغة الحاسوبي:

من أهمّ مجالات علم اللغة الحاسوبي ما يلي:

1. المعالجة الآلية الحاسوبية للأصوات اللغوية وتقنية التعرف الصوتي؛
2. التحليل الصرفي الآلي الحاسوبي بهدف المساعدة في الترجمة الآلية والتدقيق الإملائي الآلي وغير ذلك، من خلال المحلل الصرفي؛
3. المعالجة الآلية الحاسوبية للقواعد النحوية، من خلال المحلل النحوي التركيبي؛
4. صناعة المعاجم الآلية الحاسوبية؛
5. معالجة النصوص آلياً للمساعدة في صنع فهرس، وفي التصنيف والتلخيص والترتيب والترجمة؛
6. معالجة المصطلحات لصناعة بنوكها الآلية؛
7. استعادة المعلومات؛
8. التعرف الضوئي للحروف المكتوبة؛
9. تعلّم اللغات وتعليمها؛
10. التعلّم الذاتي؛

11. التّعلم عن بعد¹.

ثانيًا: مراحل نشأة اللّسانيات الحاسوبية وتطوّرها:

بدأت اللّسانيات الحاسوبية بالظهور عندما عقدت عدّة ورش عمل ومؤتمرات وندوات، وكانت أوّلها الندوة الدّولية لجمعية اللّسانيات بالمغرب، الرّباط 21 / 24 أبريل 1987م²، شاركت فيها جامعة الكويت، والتي نظّمتها الجمعية بالتعاون مع كلية الآداب والعلوم الإنسانيّة بالرّباط، جامعة محمد الخامس ومنظمة الثقافة والعلوم الإسلاميّة "إيسيكو"³. حيث ناقشت الندوة موضوعات تخصّ اللّسانيات النّظرية والتّطبيقية، وكذلك حصيلة البّحث اللّساني في اللّغة العربيّة، ومشاكلها الواقعيّة والنّظرية وأبرزت الندوة أيضًا الدور الذي أدّته اللّسانيات المقارنة واللّسانيات العامّة من خلال إبراز جوانب متنوّعة من ظواهر اللّغات الوطنيّة المختلفة.

فالمرحلة الأولى للّسانيات الحاسوبية، نشأت مع الحرب العالميّة الثانية وأوّل ما بدأت به هو التّرجمة الفوريّة بواسطة آلة (أَنْجَمًا) ENIGMA، وهي آلة مزوّدة بنظام التّرجمة السّريّة⁴.

ويمكن توضيح نشأة اللّسانيات الحاسوبية وتطوّرها من خلال المراحل التّالية:

1- مرحلة الإرهاص: تزامنت البدايات الأولى لظهور العمل بعلم اللّغة الحاسوبية

ولو في شكله الأوّل مع اكتشاف الحاسوب نفسه، بحكم طبيعة العمل الذي اقتصر

- 1- عصام الدّين أبو زلال- مقدّمة في علم اللّغة الحاسوبي- دار الوفاء لدنيا الطباعة والنشر- الإسكندرية- 2016م- ص 13.
- 2- المجلة العربيّة للعلوم الإنسانيّة، تصدر عن جامعة الكويت، مجلة فصلية محكمة، تقدّم البحوث الأصليّة والدّراسات الميدانية والتّطبيقية في شتّى فروع العلوم الإنسانيّة والاجتماعية باللّغتين العربيّة والإنجليزيّة- ع 28- م 7- خريف 1987م- رئيس التحرير: د. عبد الله أحمد المهتّا- سكرتير التحرير: محمود بركات- ص 277.
- 3- المرجع نفسه ص 277.
- 4- ينظر: موسوعة ويكيبيديا الحرّة مادّة، (أنجما والحرب العالميّة الثانية)- دراسات جيوسراتيجية 17/

<http://etudesgeostrategiques.com/2013/01/17/enigma-le-renseignement-au-cour-.2001/01-de-la-seconde-guerre-mondiale>.

آنذاك على معالجة البيانات العددية والكلمات المطبوعة والمنطوقة والأصوات والرسومات ... إلخ، لبعض المؤسسات الحكومية في إطار سياستها¹ الداخلية والخارجية، وذلك منذ الأربعينيات من القرن العشرين الميلادي تقريبا؛

2- مرحلة البداية: تأسس هذا العلم أو الاختصاص في نهاية الأربعينيات من القرن العشرين الميلادي المنصرم، فعندما اخترع الحاسوب بدأ السياسيون والعسكريون ورجال المخبرات يفكرون في كيفية الاستفادة من هذه الآلة في مجال الترجمة الآلية، وكان الاعتقاد السائد أنّ الأمر سهل من قبل المعلوماتيين في ذلك الوقت، فبدأت الترجمة الآلية كلمة كلمة عبر استعمال القواميس الآلية.

فاللغويين وعلماء الألسنية فقد بدأوا في التفكير في كيفية ترميز (Symbole) اللغة شكلا باعتمادهم على علوم الرياضيات، والمنطق الرياضي الذي وضعت أسسه مع بوول (Boole)² في النصف الثاني من القرن التاسع عشر الميلادي.

3- مرحلة الازدهار: ازدهر هذا العلم الجديد في منتصف القرن الماضي تقريبا على يد العالم اللغوي الشهير أفرام نعم تشومسكي Avram Noam Chomsky الذي قدّم نظريته الشهيرة النظرية التوليدية التحويلية وبعض أعماله من أهم الأعمال التي تطوّرت في مجال معالجة اللغة، إذ وضع أسس الكتابة الشكلية للنحو، وبلغت نظريته اللغوية درجة لم تبلغها الدراسات اللغوية السابقة، لما قدّمته من إسهامات عميقة في الدراسات اللغوية الحديثة، فقد أجاب من خلال نظريته عن نظام العقل البشري وطبيعته، حيث كانت

1- عصام الدين أبو زلال- مقدّمة في علم اللغة الحاسوبي- دار الوفاء لدنيا الطباعة والنشر- الإسكندرية- 2016م- ص 16.

2- المرجع نفسه، ص 17.

هذه الجوانب قد أقيمت من الدرس اللغوي لدى مدارس لغوية كثيرة، خاصة المدرسة البنيوية والسلوكية¹.

ويبدو أنّ دراسات تشومسكي انفتحت على مجالات معرفية جديدة، مثل: علاقة اللغة بالرياضيات، والدلالة والبيولوجيا وعلم النفس، لدرجة أنه عدّ علم اللغة فرعاً من فروع علم النفس المعرفي، كما هيّا تشومسكي لعلماء اللغة إيجاد النظريات والأساليب المختلفة والمتنوعة في صياغة اللغة شكلياً، وذلك من خلال نظريته الشهيرة، حيث تتضمن صياغة غير نمطية للغة أساسها علاقات شكلية ومنطقية بسيطة، ويبدو أنّ هذه النظرية كانت تمهيداً لاستخدام تلك الصياغات في تصميم أنظمة حاسوبية تتعامل بلغة البشر، حيث أفاد المختصون في الحاسوب من الصياغات الشكلية للغة التي وضعها علماء اللغة لتطبيقها حاسوبياً، لأنّ وجود الصياغة الشكلية الرياضية شرط لبناء التطبيقات الحاسوبية².

المبحث الثالث: دواعي حوسبة اللغة العربية وبعض الجهود الرائدة فيها

أولاً: ضرورة حوسبة اللغة العربية:

كانت العربية عبر قرون عدّة وما تزال تسير الحضارة وتعبّر عن منجزاتها هائلة التّنوع: فقد ظلّت قروناً طويلة بمثابة اللغة العالمية الأولى بلا منازع على مستوى العالم؛ تلك اللغة التي تعبّر عن مظاهر الحضارة العربية الإسلامية نتيجة انتشار الإسلام في شتى أرجاء العالم ودخول الناس في دين الله أفواجا؛ إذ وصلت حدود الدولة الإسلامية إلى جنوب أوروبا شمالاً، وأواسط إفريقيا جنوباً، والصين وأواسط آسيا الصغرى شرقاً، والمغرب وموريتانيا ومالي غرباً، وفي ظلّ الدولة الإسلامية مترامية الأطراف، نشطت حركة الترجمة إلى العربية ومنها نشاطاً متميّزاً فانتشرت العربية انتشاراً واسعاً وصارت لغة عالمية للعلوم والفنون وشتى ألوان الثقافة ومظاهر الحضارة في بقاع

1- المرجع السابق، ص 18.

2- المرجع السابق، ص 19.

العالم قاطبة، وقد ظلت العربية إلى ما يزيد على سبعة عشر قرناً سجلاً أميناً لحضارة أمتها في ازدهارها — وانتكاسها، وشاهداً على إبداع أبنائها وهم يقودون ركب الحضارة، ودليلاً على تبعيتهم وقد تخلفوا عن هذا الركب، ويؤكد هذا أنه لم يكن خروج العربية من تلك الأمصار مرافقاً لخروج العرب منها، إذ لا تطابق بين التاريخ الإسلامي وتاريخ العربية في مثل هذه الحال، وخير مثال على ذلك بقاء العربية أو آثارها اللغوية في كثير المناطق التي فتحها العرب إلى اليوم، مع أنّ قرونًا طويلة مضت على خروج العرب منها¹.

ورغم ما يعمّ الثقافة اللسانية العربية من السمات السالبة، وخصوصاً على مستوى المؤسسات، فإنّ هناك، على مستوى الأفراد على الأقل، ما يدعو إلى الشعور بالتقدم، أو بضرورة التقدم في عدّة مناح، لأنّ نتائج البحث اللساني لم يعد بالإمكان تجاهلها دون أن نقف في موقف من يتخلف عن الركب أو تتجاوزه الأحداث².

إلا أنّ العربية بالرغم من تعرّضها لعدّة هجمات استعمارية على صعيد العالم العربي والعالم الإسلامي، مازالت إحدى اللغات العالمية المعتمدة في المنظمات الدولية كهيئة الأمم المتحدة وهيئة اليونسيف والاتحاد العالمي للبريد، وهذا لا يعني أنّ العربية لا تواجه تحديات ومخاطر عديدة، وأهمها العولمة ومخاطر الهيمنة والتهميش، والبعث عن المشاركة الفعالة في إنتاج التّقنيات والبرمجيات، ولذا لا بدّ من محور الألفية الحاسوبية من المجتمع العربي (وإنتاج المعرفة بصفة عامّة).

وإذا كانت اللسانيات في الأقطار العربية علماً لم يكتب له أن يتقدّم على غرار تقدّمه في أقطار وحضارات أخرى، وإذا كانت الجهود الفردية الواعدة (على قلّتها) قد ظلّت على هامش

1- المرجع السابق، ص 91.

2- منظمة الأمم المتحدة للتربية والعلوم والثقافة "يونسكو" - تقدّم اللسانيات في الأقطار العربية - وقائع ندوة جهوية أبريل 1987 - دار الغرب

الإسلامي - 1991 - بيروت - لبنان - ص 21.

المؤسسات التي أنيطت بها مهام الاضطلاع بالبحث اللغوي، وتخطيط السياسة اللغوية، ووضع أدوات لتلبية الحاجات اللغوية، فلا بدّ من البحث عن مسببات هذا الوضع السالب وأسبابه، وواضح أنّ هناك عددا من الأسباب "الخارجية" تتحكّم في إقرار وضع غير شفاف اتّجاه القضية اللغوية¹.

ثانياً: جهود وإنجازات:

من خلال البحث والتنقيب في مجال اللسانيات الحاسوبية يمكننا أن نذكر الباحثين العرب الذين ولجوا هذا المجال تنظيراً وإنجازاً، حيث قمنا بتقسيم الباحثين حسب أسبقيتهم:

- 1- الدكتور "نبيل علي": ومعه الدكتور "تمام حسّان"، وكذلك الدكتور "الحاج عبد الرحمن صالح" و"محمد زكي خضر"، كلّهم لهم الفضل في وضع المناهج والتنظير للسانيات الحاسوبية.
- 2- الدكتور "نهاد الموسى"، والدكتور "وليد العناتي": عملوا على توصيف وتبسيط المفاهيم للسانيات الحاسوبية.
- 3- الدكتور "مروان البواب" برع في حوسبة المعجم، والدكتور "منصور الغامدي" برع في المستوى الصوتي.
- 4- أمّا في عصرنا الحالي، فهناك من الباحثين في مختلف الأقطار العربية برعوا في مجال اللسانيات الحاسوبية ومنهم: الدكتور "مصطفى جرار" من فلسطين²، وإضافة إلى ذلك الدكتور "الهادي شريفني" الذي شرف الجزائر داخل الوطن وخارجه أستاذي الفاضل والمشرف على مذكرتي، وهذا شرف لي.

1- منظمة الأمم المتحدة للتربية والعلوم والثقافة "يونسكو" - ندوة جهوية أبريل 1987 الزباط - دار الغرب الإسلامي - بيروت - لبنان -

1991 - ص 12.

2- ينظر: إيمان صبحي سلمان دلول - معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللغة العربية - رسالة ماجستير - كلية الآداب - قسم اللغة

العربية - الجامعة الإسلامية - غزة - 2014 - ص 08.

فالدكتور "الهادي شريفني: " برع في الإعلام الآلي، وكذلك في التخصص اللغوي وخاصة في المستوى الصوتي والمعجمي، ومن أهم إنجازاته: "التقعيد للمعالجة الآلية لعلم الأصوات " الذي عمل عليه حوالي تسع عشرة سنة (19).

أما بالنسبة لميدان لسانيات المدونات اللغوية، وهو موضوع مذكرتنا: هذا العلم الذي يدرس الظاهرة اللغوية من خلال مجموعة كبيرة من النصوص التي يمكن قراءتها آلياً، في نصوص مقتبسة من العالم الحقيقي والواقع الفعلي للاستعمال اللغوي، والهدف من لسانيات المدونة تحليل البيانات والنصوص بعد الحصول عليها من المدونة¹. وسنوضح ذلك في الفصل الموالي.

1- ينظر: كمال لعناني- دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح- تيزي وزو- ص 04.

توطئة

انبثقت لسانيات المدونة كاتجاه في البحث اللغوي عن اللسانيات التطبيقية وعرفت تطوّرًا واسعًا، كما سهّلت المدونات المقروءة آليا على الباحث كيفية التعامل معها وفحصها وفق طرق بسيطة، فالمدونات أو المتون اللغوية تصمّم لأغراض وأهداف معيّنة.

فما هو مفهوم المدونة اللغوية؟

وما هي أنواع المدونات اللغوية؟

وما هي أهمية المدونة اللغوية؟

وما هي كيفية بناء المدونة اللغوية، وما هي أهمّ المدونات الموجودة؟

سنفصّل في هذا من خلال ما يلي:

المبحث الأول: المدونة اللغوية وأنواعها

أولاً: مفهوم المدونة اللغوية:

اسم مشتق من (دَوَّنَ يُدَوِّنُ تَدْوِينًا) بمعنى كَتَبَ، والفعل (دَوَّنَ) مشتق بدوره، من كلمة فارسيّة معرّبة هي (ديوان).

يقول الجوهري: الدّيوان أصله دِوَانٌ، فَعُوْضٌ من إحدى الواوین یاء لأنّه یجمع دواوین، ولو كانت الیاء أصليّة لقالوا دیاوین، وقد دُوِّنَت الدّواوین. قال ابن بري: وحكى ابن دريد وابن جنيّ أنّه يقال دياوین. وفي الحديث: لا یجمعهم ديوان حافِظٌ، قال ابن الأثير: هو الدّفتر الذي یكتب فيه أسماء الجيش وأهل العطاء، وأوّل من دَوَّنَ الدیوان عمر رضي الله عنه، وهو فارسي معرّب¹، ومنه (دَوَّنَ) الدّيوان: أنشأه أو جمعه، ودَوَّنَ الكتب، جمعها ورَتَّبَها، وهي كلمة معرّبة².

نجد في معظم اللغات الأوروبية أنّ مفهوم المدونة اللغوية يقابله كلمة "Corpus" وهي كلمة مشتقة من اللاتينية بمعنى "الجسد" وترجمة الكلمة اللاتينية الأصل "Corpse" التي تقابلها في الإنجليزية "body" وجمعها في اللاتينية هو "Corpora"، وهي عبارة عن مجموعة من النصوص المتاحة للاستخدام الحاسوبي، ويطلق على مجموعة "Corpus" المدونات بالإنجليزية لفظ "Corpora" ونظريًا يجوز أن نطلق على أكثر من نصّ مصطلح (مدونة).

وقد عرّفها قاموس أكسفورد للغة الإنجليزية بأنّها جسم من المادّة المكتوبة أو المنطوقة يبنى عليه التّحليل اللّغوي³.

1- ابن منظور- لسان العرب- دار صادر للطباعة والنشر- بيروت- لبنان- م 5- ط 4- 2005- ص 333.

2- إبراهيم أنيس- عبد الحليم منتصر- عطية الصوالحي- عبد الحليم منتصر- محمد خلف الله أحمد- دار الفكر- ط 2- ج 1- ص 305.

3- ينظر: سلوى حمادة- المدونات النصية ودور اللغة العربية في التعامل معها- www.thowala.com

تعدّ المدونة اللغوية تقنية حاسوبية متقدمة، يتمّ بمقتضاها إجراء مسح لغوي مكثّف لمواد نصّية وصوتية أو نقول: "مخطوطة ومنطوقة"، تمثّل اللغة العربيّة عبر مختلف أحقابها أصدق تمثيل¹. وتشمل كذلك المعاجم والموسوعات الثّقافية والصحف اليوميّة الرّاقية، ومجموع المقالات المتميّزة المنشورة في مجلّات ثقافية أو متخصصة، أو تلك المجموعة في كتاب، ومصادر الأدب العربي على تفاوت عصورها،² أي جميع المصادر اللغوية والثّقافية بما فيها: الحصص الإذاعيّة والتلفزيونية المفرّغة ونشرية ودوريات يتمّ فهرستها وتخزينها حاسوبياً. فهذه المدونة تجمع بين المادّة اللغوية الفعلية والدراسات العلميّة الإحصائية التي ساهمت في تحديد الشائع والمشهور³.

فالمدونة الحاسوبية: عبارة عن مجموعة مهيكلة من النصوص اللغوية الكاملة المكتوبة أو المنطوقة التي تقرأ إلكترونياً. وكثيراً ما تكون هذه النصوص مصحوبة بالشارات الشارحة لمكوّناتها اللغوية، وتمدّدنا المدونة بالأدلة والأمثلة على كيفية استعمال اللغة في سياقات طبيعية بحيث يستطيع اللغوي إجراء بحوثه عليها ويستطيع المعجمي أن يختار مداخل معجمه ويكتب موادّه بصورة علمية ودقيقة⁴.

ثانياً: أنواع المدونات اللغوية الحاسوبية:

تجدر الإشارة إلى أنّ مصطلح المدونة في اللغة العربيّة يقابله مصطلحين مختلفين في اللغات الغربيّة (كالفرنسيّة والانجليزيّة): Blog (بلوغ) و Corpus (كوربوس). يشير المصطلح الأوّل

1- إيمان صبحي سلمان دلول- معجم محوسب لمعاني الأفعال الثلاثية المجردة في اللغة العربيّة- رسالة ماجستير- كلية الآداب- قسم اللغة العربيّة- الجامعة الإسلامية- غزة- 2014- ص 132.

2- وليد لعاني- العربيّة في اللسانيات التطبيقية- دار كنوز المعرفة للنشر والتوزيع- عمان- 2012- ص 249.

3- مجلة تصدر عن الجمعية المغربية للدراسات المعجمية- الدراسات المعجمية- الرباط- المغرب- السابع والثامن- 2009- ص 239.

4- كمال لعاني- دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح- جامعة تيزي وزو- ص 03.(pdf)

(بلوغ) إلى تلك المواقع الشخصية على الشبكة والخاصة بالأفراد وهي عبارة عن دفتريوميات إلكتروني. أما مصطلح (كوربوس) فهو يعني تلك المتون اللغوية والتي هي موضوع بحثنا وهي مجموعة ضخمة من النصوص المخزنة إلكترونياً.

- تصنيف المدونات الشخصية (بلوغ)

يعدّ تصنيف المدونات مطلباً بعد انتشارها بهذا الشكل الهائل، على مستوى العالم، الذي يتزايد بطريقة متسارعة كل عام، إذ يعكس هذا التصنيف خصائص المدونين والمشاركين من جانب، وخصائص المحتوى من جانب آخر، ويعكس أيضاً اهتمام المدونة وربطها مع المدونات الأخرى، فمن هذا المنطلق يصعب إيجاد تصنيف متفق عليه¹.

ومعظم تصنيفات المدونات تحتل التصنيف تحت فئات متعدّدة أو يغيب معيار فيها مثل التصنيف إلى الأنواع التالية:

المدونات الإلكترونية التي تحتوي على الروابط التشعبية؛

المدونات الإلكترونية التي تحتوي على المقالات؛

المدونات الإلكترونية التي تحتوي على الصور؛

المدونات الإلكترونية التي تحتوي على مقاطع بث إذاعي؛

المدونات الإلكترونية التي تحتوي على مقاطع بث مرئي؛

المدونات الإلكترونية المنوعة التي تضم مزيجاً من الأنواع المذكورة أعلاه².

بالإضافة إلى هذا التصنيف هناك شكل آخر من التصنيفات وهي كالتالي:

1- ينظر: مطلق سعيد العميري- تأثير المدونات الإلكترونية الكويتية على اتجاهات طلبة قسم الإعلام في جامعة الكويت- رسالة ماجستير- كلية

الإعلام- جامعة الشرق الأوسط- 2011- ص 36.

2- المرجع نفسه- ص 36، 37.

1. حسب الموضوع:

- * مدونات تحتوي على روابط تشعبية: هي المدونات التي تحتوي على وصلات تشعبية وتعدّ أول أنواع المدونات التي تمّ نشرها على الشبكة، ويحتوي هذا النوع من المدونات على العديد من الروابط لمواقع الانترنت.
- * المدونات التي تحتوي على المذكرات اليومية: تتناول هذه المدونات الحياة اليومية للمدوّن من بداية اليوم إلى نهايته، ولا تحتوي بالضرورة على روابط لمواقع إلكترونية¹.
- * المدونات التي تحتوي على المقالات: وتحتوي على عرض تقارير وتعليقات على مختلف الأخبار والأحداث².
- * المدونات العلمية المتخصصة: هي المدونات التي تكون مهمة بالجانب العلمي المحض مثل التكنولوجيا والعلوم الدقيقة... أو غيرها من العلوم المتخصصة، وغالبا ما تنطلق من منتديات يحوّلها أصحابها إلى مدونات.
- * مدونات تحتوي على صور: تحوي هذه المدونات على الصور حيث يخصّصها صاحبها لنشر صور التقطها بنفسه أو أعجب بها أو أيّ موضوع آخر.
- * مدونات تحتوي على مقاطع بثّ إذاعي: وهي تحتوي على مقاطع بثّ إذاعي تكون في كثير من الأحيان مسجّلة من طرف المدوّن نفسه في شكل برامج إذاعية قصيرة مع إمكانية تحميلها من طرف القارئ³.

1- أمنا أمينة جاري- فاطمة الزهراء سعداوي- المدونات التعليمية ودورها في تنمية التحصيل العلمي لدى طلاب جامعة قاصدي مرياح ورقلة-

مذكرة ماستر - جامعة قاصدي مرياح ورقلة- 2015- ص 31.

2- دراسات ومقالات في المدونات التعليمية <http://mabob537-blogspot.com/blog-page-html> على الساعة 15:00

نتائج المؤتمر الدولي الثاني للتعليم الإلكتروني والتعلم عن بعد، جريدة الرياض.

3- أمنا أمينة جاري- فاطمة الزهراء سعداوي- المدونات التعليمية ودورها في تنمية التحصيل العلمي لدى طلاب جامعة قاصدي مرياح ورقلة-

مذكرة ماستر - جامعة قاصدي مرياح ورقلة- 2015- ص 32.

* مدونات تحتوي على مقاطع بث مرئي: وتعدّ من أحد أشكال المدونات، مثلها مثل مدونات البثّ الإذاعي إذ تعرض مقاطع فيديو يعدّها المدون في كثير من الأحيان في مختلف المواضيع¹.

- المدونات المتنوعة: هي المدونات التي تجمع بين مختلف أشكال المدونات
- المدونات الجماعية: يعني يشترك في كتابتها مجموعة من الأشخاص.

2. حسب الهدف:

- * مدونات اليوميات الشخصية: هي المدونات التي يكتسب فيها المدون تجاربه الشخصية التي لا تهم العامة، ولكنها تدخل في إطار نشر الحياة الخاصة، والسيرة الذاتية وجمهورها غالباً ما يكون من الأقارب والأصدقاء وبعض الفضوليين.
- * مدونة الجمهور الخاص: هي مدونات موجهة لفئة معينة من القراء في إطار نوع خاص من التواصل، مثل: مدونة مدير شركة يخاطب فيها الموظفين.
- * مدونات التأثير في الرأي العام: هي مدونات موجهة لجميع الناس، وتهدف إلى توجيه قرائها في اتجاه معين، إزاء موضوع أو قضية يتبناها المدون، يعرف هذا المدون بالمواطن الصحفي أو المواطن الإعلامي².

3. حسب المحتوى:

- * المدونات الشخصية: يكتب المدون فيها تجارب حياته اليومية.
- * المدونات السياسية: هي أكثر أنواع المدونات انتشاراً، بها وصلات لمواقع إخبارية ويضيف المدون فيها تعليقاته على الموضوع السياسي.

1- نهي نجم الزامل - المدونات الإلكترونية، مركز التدريب التربوي - السعودية - 2013 - ص 02.

2- أمنا أمينة جاري - فاطمة الزهراء سعادوي - المدونات التعليمية ودورها في تنمية التحصيل العلمي لدى طلاب جامعة قاصدي مرياح ورقلة -

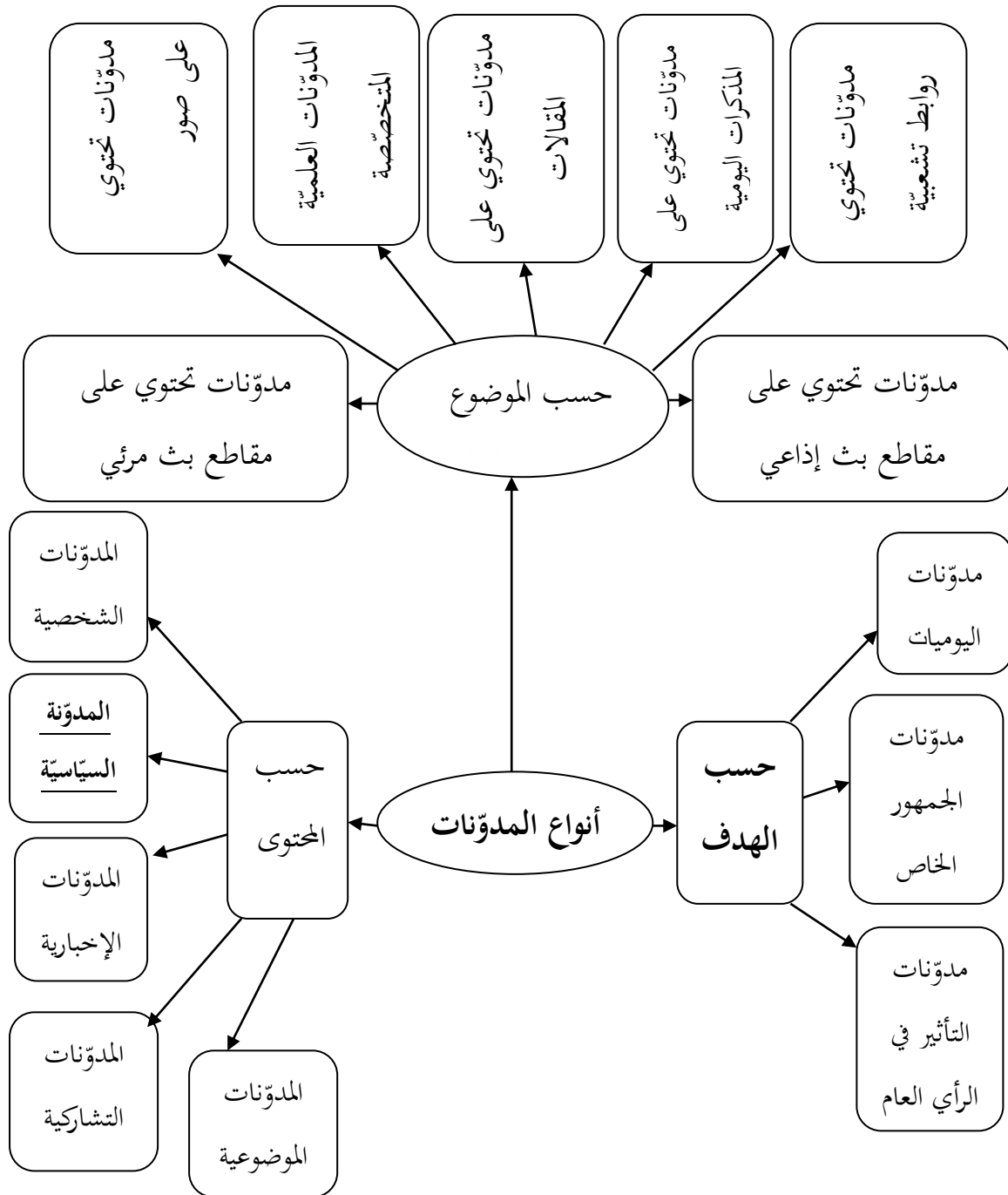
مذكرة ماستر - جامعة قاصدي مرياح ورقلة - 2015 - ص 32.

- * المدونات الإخبارية: توفّر الكثير من المدونات موجز للأخبار عن موضوع معيّن.
 - * المدونات التشاركية: تكون غالباً مفتوحة لمشاركة الجميع وبعضها أو أقلّها يكون محدداً لمجموعة من الأشخاص¹.
 - * المدونات الموضوعية: هي المدونات التي تركز على موضوع معيّن، وجلّ المدونات تسمح للمستخدم بإيجاد موضوع ما داخل الموقع، لذا فالمدونة العامة من السهل أن تصير مدونة تخصّ موضوع معيّن وذلك بواسطة المستخدم².
- وفي عالم المدونات نجد مدونات للأخبار، مدونات للتكنولوجيا، ومدونات للقضاء، ومدونات للموسيقى، ومدونات للأفلام، وأخرى للأدب، وغيرها للفنون، وللصحافة، وللإعلام، وللخياطة، والصيد، فقد أصبح للمدونات عالم واسع³.

1- المرجع السابق، ص 33.

2- دشن الفحطاني- المدونات الإلكترونية وحرية الرأي والتعبير- مؤتمر تقنيات الاتصال والتعبير الاجتماعي- جامعة الملك سعود- الرياض- مارس- 2009.

3- ينظر: زكي حسين الوردی- صحاف المدونات الإلكترونية على الانترنت عرض وتحليل- كلية الإعلام- جامعة بغداد- العدد 311- حزيران 2007.



مخطط رقم 01: أنواع المدونات الشخصية (بلوغ)

-تصنيف المدونات اللسانية (كوروبوس)

أما المدونات اللسانية - والتي هي موضوع بحثنا - فيمكن تصنيفها كما يلي:

من حيث التحليل اللغوي:

- 1) متون نصية خام "Raw Corpora": يقصد بها مجموعات كبيرة وشاملة من النصوص متوفرة إلكترونيا جمعت دون ترتيب أو منهجية.
- 2) متون نصية موسومة "Annotated Corpora": مجموعات كبيرة وشاملة من النصوص محللة تحليلًا لغويًا بأية صورة من الصور أو على أي مستوى من المستويات اللغوية، وهذا التحليل تم يدويًا عن طريق متخصصين لغويين.
- 3) المدونات النصية أو المتون النصية المعلّمة "Tagged Corpora": مجموعات كبيرة وشاملة من النصوص محللة تحليلًا لغويًا على مستوى أجزاء الكلام، وهذا التحليل تم إما يدويًا أو آليًا أو نصف آلي.

من حيث آلية البحث:

ويُقصد بها طريقة استرجاع ما نبحث عنه من مفردات أو صيغ تركيبية أو حتى جمل عن طريق:

- 1) مطابقة القالب الشكلي "Template/pattern keyword": ونصوص المتن وفي نفس الوقت تقوم آلية الاسترجاع بعد تكرار هذا القالب في المتن وعرض الأجزاء النصية التي يوجد بها كلمات تحدّد قبل وبعد القالب المطلوب.¹

1- فاطمة الزهراء توتاوي- استثمار المدونة في تدريب المترجم-رسالة ماجستير-كلية الآداب واللغات- الترجمة-جامعة وهران-2009-ص24

(2) البّحث الصّرفي: وهو البّحث عن كلّ جذور الكلمات موضوع البّحث بجميع المشتقات. وهذا التّوع مازال يعاني قصورا شديدا إذ إنّهُ يحتاج لقاعدة صرفية كبيرة، وقد قام أستاذنا المشرف بعدّة بحوث في هذا الميدان عبر عدّة دراسات من بينها "الاستخراج الآلي للصيغ الصّرفية المشتقة من الفعل الثلاثي المجرّد"، ومشروع "الاستخراج الآلي للصيغ الصّرفية المشتقة من الفعل الثلاثي المزيد بحرفين" وأخيرا مشروع "الاستخراج الآلي للأبنية الإسمية للمشتقات العربيّة".

من حيث اللّغة: (وحيدة اللّغة وثنائية اللّغة ومتعدّدة اللّغات):

(1) النّص في المدوّنة أحادية اللّغة "Monolingual corpus": وهو نصّ مكتوب بلغة واحدة فقط.

(2) النّص في المدوّنة ثنائية اللّغة "Bilingual corpus": وهو نصّ مكتوب بلغتين إحداها اللّغة المصدر، والأخرى اللّغة الهدف.

(3) النّص في المدوّنة متعدّدة ومتوازية اللّغات " Multilingual corpus or Parallel corpora": هو نصّ مكتوب بلغات متعدّدة بحيث يتمّ عمل تنظير بين الجمل المتقابلة في كلّ اللّغات. فالمدوّنة المتوازية Parallel corpora هي إذن مجموعة من أزواج من النّصوص بحيث، بالنسبة لزوجين من النّصوص، يكون كلّ واحد منهما ترجمة للآخر. ومن المهمّ جدا أن يتمّ محاذاة النّصين، يعني مقابلة كل وحدة من النّص الأوّل (النّص باللّغة المصدر) بالوحدة من النّص الثاني (النّص باللّغة الهدف)، وتكون المقابلة على مستوى الفقرات والجمل والكلمات، لكي نتحصّل على توليفة من البيانات الثنائية اللّغة والتي توظّف في أنظمة التّرجمة الآلية¹.

1- ينظر: المرجع السابق -ص 25.

من حيث العموم:

تمت عملية تصنيف المدونات منذ التسعينات، والآن يوجد العديد من هذه الأنواع:

1) المدونات العامة "General corpora": وهي متعددة الأغراض "كمدونة براون"، الموجهة لميدان صناعة المعاجم وتصميم برامج معالجة النصوص، بالإضافة إلى استخدامها لأغراض تعليمية.

2) المدونات الخاصة "Specialized corpora": وتستخدم لغرض معين "كمدونة كولبيد" التي تستخدم لصناعة المعاجم التعليمية

من حيث نوعية النصّ / المادة النصّية:

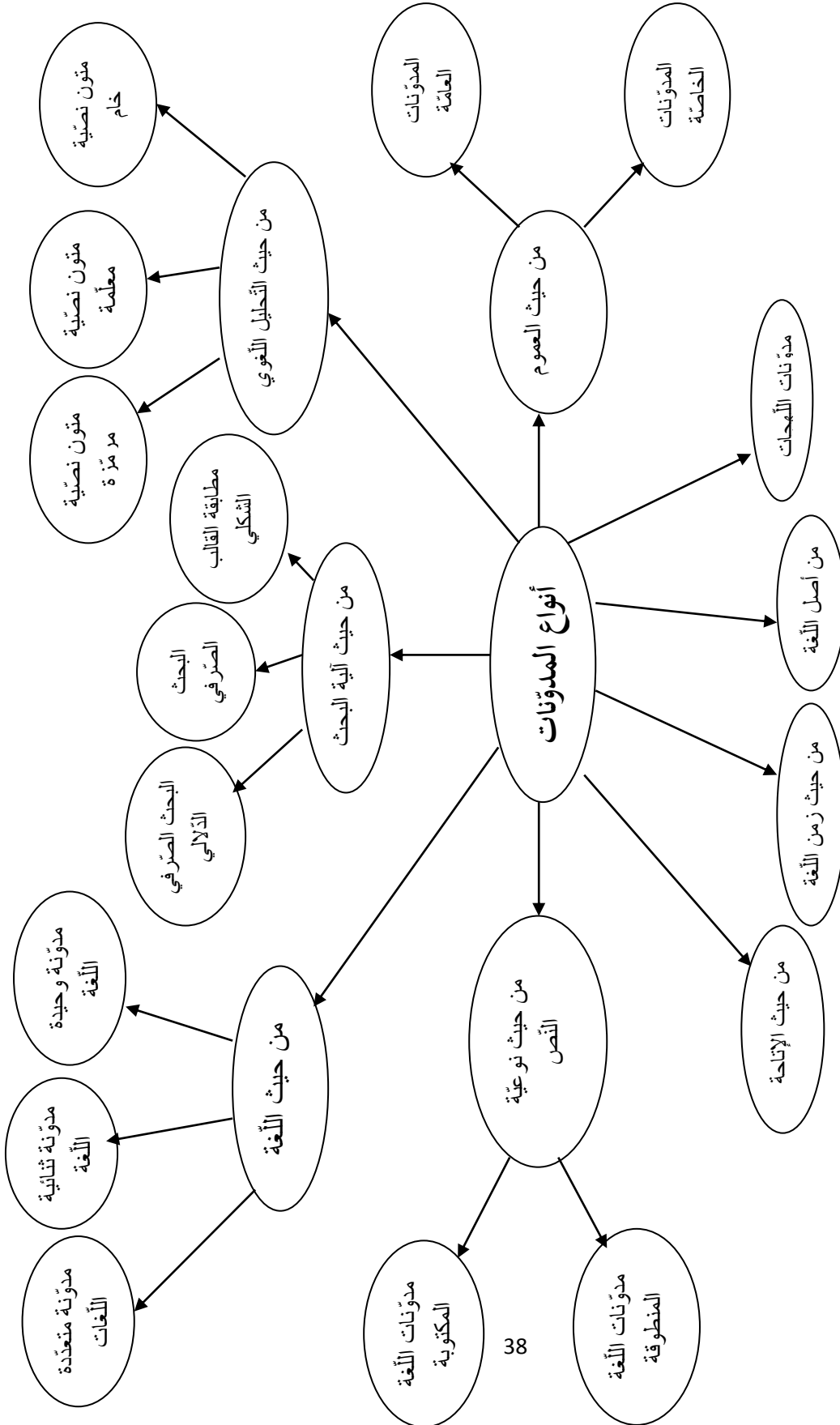
1) مدونات اللغة المكتوبة "Written corpora"

2) مدونات اللغة المنطوقة "Spoken corpora": مثل مدونات البرامج الإذاعية أو المكالمات التليفونية باللغة العربية والمستخدمه بجامعة بنسلفانيا.

-مدونات اللهجات "Dialect corpora": المدونات الخاصة بلهجة واحدة كمدونات اللهجة المصرية. - من أصل اللغة: إذا كانت المدونة أصلية Original أم تمت ترجمتها Translations .

- من حيث زمن اللغة: أي مدونة محددة بفترة زمنية محددة أم شاملة مختلفة الفترات.
- من حيث الإتاحة: هناك Standard corpora وهي مدونات مصممة وفقا لمبادئ خاصة ومدونة حرّة Free corpora مثل تلك المتاحة على شبكات الاتصال والتي تضم نصوصا ضخمة¹.

1- ينظر: سلوى حمادة- المدونات النصّية ودور اللغة العربية في التعامل معها- 2007. www.thowala.com



مخطط رقم 02: أنواع المدونات اللغوية

المبحث الثاني: أهمية المدونات اللغوية الحاسوبية

أولاً: أهداف المدونات اللغوية الحاسوبية:

المدونات اللغوية الحوسبة عملية وتجريبية، مبنية على نصوص حقيقية للاستعمال اللغوي، وتدرس نماذج واقعية للغة أو التركيب، كما أنها تخضع للبحث والتحليل، والتنوع المبني على أسس علمية، ويكون هناك حوار بين الحاسوب والباحث، حيث يقوم الحاسوب بالجانب الآلي، كالإحصاء والتنبه إلى ظواهر معينة كاللبس المعجمي أو التركيبي، ويقوم الباحث اللغوي بالجانب الذي يتطلب تدخلاً بشرياً مثل فكّ اللبس وتحديد الاسم الذي يعود إليه ضمير ما في النص، وعلى تحليل وتفسير النتائج.

كما يمكن الحصول على المدونات اللغوية الحوسبة بصورة مختلفة منها ما يلي:

- الأقراص المضغوطة CD؛
- على الشبكة (الانترنت)؛
- أو في صورة نصوص إلكترونية/ رقمية...¹

فالهدف من التدوين اللغوي ليس استخراج البيانات آلياً، ولكن دراسة وتحليل البيانات المستخرجة من المدونات. فالتدوين اللغوي هدفه دراسة اللغة والظواهر التي تتضمنها باستخدام المدونات الآلية، وهو يختلف عن الدراسات التقليدية بتوفير قاعدة نصية هائلة من العالم الحقيقي والاستخدام الفعلي للغة، مما يتطلب المعالجة الآلية التي قد لا نستغني عنها أحياناً.

1- محمود إسماعيل صالح/ أستاذ اللسانيات التطبيقية- السبت 05 أبريل 2014--Dr-Mahmoud-ismail-saleh

والتدوين اللغوي ليس مجرد استخراج بيانات لغوية بمساعدة الحاسب ولكنها تحليل لهذه البيانات كذلك كما أشرنا إليه.

كما تهدف المدونات الحاسوبية إلى ما يلي:¹

(1)- على المستوى الصرفي:

- الإحصاء الصرفي وما يمكن أن يضيفه للقواعد الصرفية؛

- التحليل والتوليد الصرفي الآلي؛

- تصحيح الأخطاء الإملائية الناتجة عن الصرف؛

- دعم التشكيل الآلي للنصوص؛

(2)- على المستوى النحوي:

- الإحصاء النحوي وتفكيك الجملة لعناصرها النحوية الأولية "مبتدأ، خبر، فعل، فاعل، ..."

- الإحصاء النحوي وما يمكن أن يضيفه للقواعد النحوية؛

- التحليل والتوليد النحوي للنصوص؛

- تعليم النحو للصغار ولغير الناطقين بالعربية باستخدام الحاسوب.

(3)- على المستوى الدلالي:

- الترجمة الآلية،

- الفهم الآلي للنصوص،

1- سلوى حمادة- المدونات النصية ودور اللغة العربية في التعامل معها- 2007. www.thowala.com

- دراسة تطوّر أساليب الكتابة من حيث الشكل والمضمون،
- دراسة الترابط النصّي في النصوص الحديثة،
- تحديد المفردات الأكثر شيوعاً،¹ والمتلازمات اللفظية،
- كما أصبحت المدونات اللغوية المحوسبة تعتبر كمصادر جديدة للمعلومات بشرط أن تكون هذه المعلومات صحيحة وموثوقة،² وكثير ما يستخدمها الطلاب في مجال تخصّصهم العلمي من خلال البحث عن مقالات، كتب، بحوث ودراسات خاصّة المتخصّصة.³

ثانياً: الفائدة من المدونات اللغوية المحوسبة وكيفية البحث فيها:

1- الفائدة من المدونات اللغوية المحوسبة:

تستخدم المدونة في بحوث تنتمي إلى مجالات لغوية شتى منها: اللسانيات العامة، واللسانيات التاريخية، واللسانيات الحاسوبية واللسانيات التطبيقية، أما عن خصائصها:

* السهولة،

* السرعة،

* الدقة.⁴

- فالسهولة: تتمثل في سهولة الوصول إلى الوحدات والتراكيب اللغوية المطلوبة.

1- المرجع السابق www.thowala.com

2- ينظر: مطلق سعد العميري- تأثير المدونات الإلكترونية الكويتية على اتجاهات طلبة قسم الإعلام في جامعة الكويت- رسالة ماجستير- جامعة الشرق الأوسط- كلية الإعلام- 2011- ص 99.

3- ينظر: أمنا أمينة جاري- فاطمة الزهراء سعداوي- المدونات التعليمية ودورها في تنمية التحصيل العلمي لدى طلاب جامعة قاصدي مرياح ورقلة- مذكرة ماستر- جامعة قاصدي مرياح ورقلة- 2015- ص 63.

4- كمال لعناني- دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح- جامعة تيزي وزو- ص 04.(pdf)

- أما السرعة: إذ يمكن البحث في نصوص تتألف من ملايين الكلمات وتحليلها بسرعة أكبر من استعمال العين المجردة.

- وأخيرا الدقة: فمعالجة النصوص أكثر دقة من معالجتها بالعين المجردة، فإذا كنا نبحث في مدونة كبيرة جدا عن كلمة معينة مثلا: يستطيع الحاسوب أن يعثر على جميع السياقات التي ترد فيها تلك الكلمة دون أن يفوته سياق واحد¹.

بالإضافة إلى اختصار الجهد والوقت، أي غزارة وتنوع المعلومات وسهولة وسرعة الولوج إليها وسلاسة البحث فيها، كما تعدّ مرجع مهمّ للمعلومات، وهذا لما تمتاز به من تنظيم وهيكلية ومحتويات مختصرة ومفيدة، وتنوع أشكال المعلومات ما بين النص والصورة والصوت².

وهناك جملة من الفوائد يمكن جنيها من المدونة اللغوية وتمثّل تلك الفوائد في:

- 1- تبسيط الإملاء العربي، والوقوف على أشهر المفردات الإملائية؛
- 2- تحديد المفردات الشائعة بالذخيرة اللغوية العربية؛
- 3- حصر المفردات الفصيحة والتراكيب الأكثر تداولاً بهدف تطوير تعليم اللغة العربية بالمؤسّسات التربوية³؛
- 4- استثمار المدونة في بناء المعاجم العامة والمتخصصة التاريخية، وبهذا الشكل يكون المعجم التاريخي عبارة عن عمل معجمي ضخم يؤثّق حقيقة للغة بكل أشكالها ومستوياتها واستعمالاتها ويقدم تاريخاً حقيقياً لحياة الألفاظ وتطور معانيها منذ أوّل

1- أحمد برماد- المدونات الحاسوبية وصناعة المعجم التاريخي العربي المختص- مجلة الذاكرة- تصدر عن مخبر التراث اللغوي والأدبي في الجنوب الشرقي الجزائري- العدد التاسع- جوان 2017- ص 04.

2- ينظر: أمنا أمينة جاري- فاطمة الزهراء سعداوي- المدونات التعليمية ودورها في تنمية التحصيل العلمي لدى طلاب جامعة قاصدي مرباح ورقلة- مذكرة ماستر- جامعة قاصدي مرباح ورقلة- 2015- ص 84- 85.

3- كمال لعناني- دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح- جامعة تيزي وزو- ص 04.(pdf)

- ظهور لها إلى آخر استعمال لها، وهذا الأمر يتوقف بطبيعة الحال على النصوص التي يتشكل على أساسها المعجم أو "المدونة اللغوية"¹؛
- 5- دراسة مفردات اللغة العربية وبنائها ونحوها؛
- 6- تيسير المعالجة الآلية للغة العربية تحليلاً وتوليداً؛
- 7- الترجمة الآلية لمفردات اللغة وتراكيبها بتوظيف النحو المقارن؛
- 8- تمكين الدارسين من المصادر العلمية المختلفة -غير أنه لا بد من ضرورة تحديث المدونة باستمرار لأنّ هناك كلمات تصبح أكثر شيوعاً أو تختفي في الاستعمال، وتظهر كلمات أخرى.²

المبحث الثالث: إنشاء المدونات الحاسوبية، وأهم ما وجد منها في المواقع الإلكترونية

أولاً: تصميم المدونة (Designing the corpus)

1- تجميع النصوص (Obtaining the texts)

- أ. الانتقاء: فعلية الانتقاء تقوم على معايير تتعلق أساساً بخاصية التمثيل والتنوع والثراء والتوازن وكذا الحجم، ويفصلها جون سنكلير (John Sinclair) في مجمل النقاط التالية:
- مراعاة الوظيفة التواصلية: على الباحث أن يتأكد من أنّ مجموعة النصوص التي انتقاها هي من صنع كاتب/ متكلم أصلي (Nature speaker/Writer).

1- ينظر: أحمد برماد- المدونات الحاسوبية وصناعة المعجم التاريخي العربي المختص- مجلة الذاكرة- تصدر عن مخبر التراث اللغوي والأدبي في الجنوب الشرقي الجزائري- العدد التاسع- جوان 2017- ص 104.

2- كمال لعناني- دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح- جامعة تيزي وزو- ص 104. (pdf)

- أمّا من جهة أخرى، فلا ينبغي على الباحث أن يختزل المادة المجموعة في أعمال كاتب/ متعلّم بعينه، بل يجب عليه أن يتحرّى سمة الموضوعيّة وطابع الشّموليّة ما أمكــــنه إلى ذلك سبيلا، حرصا منه على إضفاء صيغة الدّقة على نتائج بحثه¹.

- فضلا على ذلك، ينبغي أن يتضمّن مادّة التّحليل نصوصا بأكملها، بدل إدراج نصوصا مبتورة ...

- وفي سياق آخر، ليس للباحث أن يدرج ضمن مدوّنته نصوصا مستقاة عن الكتب والمجالات والمذكّرات وما شاكلها من الكتابات أو عن البرامج الحوارية والثّقافية ...، ما لم يطلب إذنا من الناشر، وذلك مراعاة لحقوق التّأليف.

يجب على الباحث أن يراعي شرط "الجدّة" أثناء عملية انتقاء النّصوص، لرصد جلّ المستجدّات اللّغوية والمفاهيميّة التي يمكن أن تطرأ.

* لبناء المدونة يجب توفّر العناصر الرئيسية التالية:

1- طبيعة النّص (Text mode): إن كان مكتوبا أو منطوقا؛

2- مصدر النّص (Text source): إن كان مأخوذا عن كتاب، جريدة، رسالة، أو رواية

2...
...

كما يمكن أن تحفظ على شكل: SGML, HTML, XML, rtf, doc

1- ينظر: سلوى حمادة- المدونات ودور اللّغة العربية في التّعامل معها- 2007. www.thowala.com

2- ينظر: فاطمة الزهراء توتاي- استعمار المدونة في تدريب المترجم- رسالة ماجستير- كلية الآداب واللّغات والفنون- قسم ترجمة- جامعة

وهران- السانية- 2009- ص 9- 11.

3- مصدر النصوص (Corpora source)

- أ. استخدام الماسح الضوئي (Scanner): إنّ أبسط المصادر المدوّنة، يتمثّل في المادّة المطبوعة على اختلاف ألوانها وتباين أشكالها.
- ب. شبكة الويب: تزخر الشبكة العنكبوتية بنصوص متباينة اللغات وهي تشكّل مصدرا هامًا لصناعة المدونات، كما تفرض على الباحث توخّي الكثير من الحذر، إذ ينبغي عليه أن ينتقي منها ما يراه مناسباً وملائماً وموثوقاً منه.

4- إعداد النصوص:

- أ. بيانات الوثيقة (Document informations): وتتعلّق بمصدر النصّ المراد إدراجه ضمن المدوّنة، نحو اسم المؤلف/ المتكلم، اسم الناشر (ضماناً لحقوق التأليف)، وكذا بيانات سياقية نحو: مكان النشر وتاريخه (أول التسجيل في حالة مدوّنة منطوقة) سنّ المؤلف/ المتكلم وجنسه، المستوى الاجتماعي والثقافي،... إلخ.
- ب. التّقيح (Markup): خاصّة فيما يتعلّق ببيانات حول بنية الوثيقة الأصل، على شاكلة: العناوين الرئيسيّة، الفقرات، بداية ونهاية كلّ صفحة من الوثيقة، وإعداد الصّفحة (الحروف الطباعيّة، قياسها، نمطها، ...) وما إلى ذلك من الخصوصيات الشّكلية للوثيقة.
- ج. التّوسيم (Annotation): ونعني به تحليل مادّة المدوّنة الخام (Raw corpora) تحليلاً لغويّاً بأيّة صورة من الصّور، أو على أيّ مستوى من المستويات اللّغوية¹.

3- حقل النصّ (Text domain): ويتجلّى ذلك مثلاً مع اللّغة الإنجليزيّة: لغة المملكة المتّحدة،

الإنجليزيّة الأمريكيّة، إنجليزيّة الهند، أستراليا....

4- اللّغة أو ضروب اللّغة التي تتضمّنهما المدوّنة.

-الفترة الزمنيّة للتّصوص أو تاريخها.

1- ينظر المرجع السابق- ص16 - 19.

وبالإضافة إلى ذلك يجب مراعاة المعايير التالية:

- أ. اللغة: يجب أن تكون النصوص أصلية، ومحررة في اللغة الأم للكاتب، إلا إذا كان متمكنا من اللغة الثانية مثله كمثل الناطق الأصلي بها.
- ب. الميدان أو الفروع: ينبغي أن يكون الميدان محددا ومحدودا، قبل أن يشرع الباحث في صناعة المدونة.
- ج. جنس النص (Text genre): يحدد مستوى التخصص ومؤهلات الكاتب وخبراته في المجال.
- د. مستوى التخصص (Specialization level): أي كل حسب المستوى الذي وظف فيه، على غرار: تقني جدا، تقني مختص، تبسيطي، تكوين عام، ولمعرفة مستوى تخصص ما، ينبغي أن يأخذ الباحث بعين الاعتبار كلاً من: نوع النص وكاتبه والمتلقي.
- هـ. مؤهلات الكاتب وخبراته: إن كان كاتباً مختصاً (Expert)، وآخر شبه مختص (Relative expert)، أو أستاذ (Teacher).
- و. المتلقي: تصنف النصوص أيضا حسب المتلقي، وهو عنصر لا تخفى مكانته في الوضعية التواصلية على أحد.
- ز. المرجع: ينبغي توثيق النص
- ح. التاريخ أو الفترة الزمنية للنصوص: يرتبط هذا المعيار بمشروع البحث وبنوع المدونة المراد صنعها (سانكرونية، دياكرونية، ...).
- ط. شكل الملف (File format): تحفظ النصوص عموما ضمن الشكل txt قصد تسهيل معالجتها وتحليلها بأدوات عدّة (المفهرسات الآلية أو ما يسمى الكشافات السياقية Concordançiers، المرمّزات الآلية étiqueteurs...)¹.

1- ينظر المرجع السابق: 15.

ثانياً: جرد أهم المدونات اللغوية الحاسوبية الموجودة في المواقع الإلكترونية

أول المدونات الحاسوبية ظهوراً كانت مدونة جامعة (براون Brown University) وذلك في عام 1961م، وكانت في بدايتها تتألف من مليون كلمة من نصوص اللغة الإنجليزية الأمريكية، ثم أُنشئت بعدها مدونات أخرى مثل: لانكستر (Lancaster)، أوصلو (Oslo)، برغن (Bergen) للغة الإنجليزية البريطانية، ومدونة اللغة الإنجليزية الهندية، فهذه المدونات مهّدت الطريق لتأليف مدونات أكبر كمدونة التراث الأمريكي¹.

ومن أشهر المدونات اللغوية العربية انتشاراً، هي "المدونة العربية"، والتي أُعتمدت كمرجع مهمّ في العديد من الدراسات اللغوية التقنية، وهذه المدونة تُعدّ من أكبر المشاريع الحضارية للغة العربية على الشبكة العالمية. وقد تمّ إطلاق الموقع الرسمي للمدونة اللغوية العربية التابع لمدينة الملك عبد العزيز للعلوم والتقنية بتاريخ 06 مارس من عام 2012م،² حيث احتوت في مرحلتها الأولى على سبعمائة مليون كلمة ممّا دوّن بالعربية ابتداءً العصر الجاهلي وحتى العصر الحديث، ومن مختلف المناطق والبلدان، مع تنوع أوعية النشر فيها (مخطوطات، صحف، كتب، مجلات، دوريات علمية، ...)، وفي شتى المجالات العلمية والفكرية المختلفة³.

كما قام فريق من جامعة ليدز (Leeds) ببريطانيا بإشراف من المستشرق "إيريك أتويل" بتصميم وإنجاز مجموعة من المدونات اللغوية العربية لتكون مصدراً مساعداً لإجراء الدراسات على اللغة العربية، ويمكن حصرها فيما يلي:

1- كمال لعاني - دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح - جامعة تيزي وزو - ص 03. (pdf)

2- ينظر: موقع صوت العربية - أكبر مشروع حضاري للغة العربية على الشبكة العالمية: إطلاق موقع المدونة اللغوية العربية التابع لمدينة الملك عبد العزيز للعلوم والتقنية، للكاتب: عبد العزيز حميد على الرابط:

http://www.voiceofarabic.net/inden.php?option=com.contentview=article&id=1038:2012-03-08-11-11-53&catid=159:2008-08-25-14-57-05&Itemid=327.

3- موقع المدونة العربية التابع لمدينة الملك عبد العزيز للعلوم والتقنية، على الرابط:

http://www.kastac.org.sa/page/about.aspx

- 1- مدونة تعليم العربية بواسطة الحاسب الآلي (Arabic By Computer ABC) وكان الهدف من هذا المشروع إنشاء مصدر لغوي لمتعلمي اللغة العربية، يحتوي على قاعدة بيانات للنصوص العربية، ومعجم للمفردات.
- 2- المدونة اللغوية للعربية المعاصرة (The corpus of contemporary Arabic : CCA) صممت هذه المدونة لتكون قابلة للمقارنة مع مدونة Bergen - Lancaster - Oslo (corpus) للإنجليزية البريطانية المعاصرة التي تشمل على مليون كلمة، وكذلك مدونة Brown للإنجليزية البريطانية المعاصرة وفيها مليون كلمة أيضا.
- 3- مدونة الانترنت العربية (Arabic internet corpus): في جامعة ليدز أُنبتت هذه المنهجية (أي باعتبار الأنترنت مدونة لغوية كبرى) لجمع مدونات من الانترنت للغة العربية، والصينية، والروسية، والبولندية، والفرنسية، والألمانية، والإيطالية، والإسبانية. وهي مدونات متاحة للجميع من خلال واجهة لكشاف السياقات والمتصاحبات (المتلازمات اللفظية) على الموقع التالي: <http://corpus.leeds.ac.uk/internet.html>
- 4- مدونة اللغة العربية حول العالم (World wide arabic corpus): تشمل مدونات فرعية تتألف كل واحدة من مائتي ألف كلمة من كل دولة، وذلك لدراسة اختلاف اللهجات العربية من بلد إلى بلد.
- 5- المدونة اللغوية لتدريس معلومات عن الإسلام (Corpus for teaching about Islam) وذلك من أجل تأليف موسوعة جامعية لاستخدامها في تدريس الطلاب عن الإسلام والمسلمين، وتشمل اللغة واللسانيات، ودراسة المناطق.
- 6- البنك الشجري للخطاب العربي (Arabic discourse tree bank): مدونة تشمل 537 نصًا إخباريًا قد وُسمت جميع أدوات الوصل فيها¹.

1- إيريك أتويل، عبد الله بن يحيى الفيبي - أبحاث جامعة ليدز في مجال لسانيات المدونات العربية - 2015/

E.s.Atwell@leeds.ac.uk/ayjfaifi@imamu.edu.sa.2016

7- المدونة اللغوية لمتعلمي اللغة العربية (Arabic learner corpus ALC): أنشئت

هذه المدونة لتكون مصدرا لغويا لأبحاث تعليم اللغة العربية، وكذلك للبحث في مجال

معالجة اللغة الطبيعية، ويمكن تنزيل نصوص المدونة من موقعها

الرئيس: <http://www.arabiclearnercorpus.com>

8- المدونة العربية للغة القرآن الكريم (Quranic Arabic Corpus): استخدمت المدونة

العربية نصوص القرآن الكريم كمصدر معياري لعدد من الأبحاث التي أجريت على اللغة

العربية الفصحى.

9- مدونة الإحالة الثنائية لضمائر القرآن الكريم (Quran Pronoun Anaphoric)

(Coreference Corpus): مدونة غنية من حيث التحشية، فقد سمت جميع الضمائر

الشخصية فيها بمعلومات عن العائد الذي يشير إليه الضمير، سواء أكان العائد سابقا

للضمير أو كان لاحقا له وهو الغالب في المدونة، كما شمل الوسم كذلك العائد نفسه،

أي الكلمة أو الجملة التي تحوي المعنى، أو الشخص، أو الكيان، أو الفكرة التي يشير إليها

الضمير.

10- مدونة الترابط الدلالي بين آيات القرآن الكريم (Quran Verse Similarity corpus):

نشرت هذه المدونة على الموقع التالي على شبكة الانترنت: Tentmining The

Quran.com حيث يستطيع المستخدم رؤية شبكة العلاقات المباشرة وغير المباشرة لأي

آية من آيات القرآن الكريم.¹

11- مدوّنة القرآن الكريم الموسومة بموضوع الآيات مع الترجمة الإنجليزيّة (Quran corpus)

(Annotated with english translations and verse topics): عبارة عن أداة ثنائية

اللّغة (إنجليزي- عربي) ذات قدرة على تحسين الدّقة والاستجابة عند البسّـحث

في موضوعات القرآن الكريم.¹

وقد استُخدمت هذه المدوّنة اللّغوية والأدوات الحاسوبية لإجراء العديد من

الأبحاث في مجال لسانيات المدوّنة العربيّة، ومن ذلك على سبيل المثال تعليم اللّغة العربيّة،

والمقارنة بين اللّغتين الإنجليزيّة والعربيّة. ويعدّ موقع المدوّنة العربيّة لنصوص القرآن الكريم أحد أبرز

الأمثلة على هذه الأبحاث، حيث يُستخدم على نطاق واسع من قبل الباحثين في اللّسانيات

العربيّة، وكذلك علماء الشريعة الإسلاميّة، وعامة النّاس في البلدان الإسلاميّة وغيرها من البلدان

حول العالم.

1- ينظر على الرّابط: E.sAtwel.leeds.ac.uk/ayjfaifi.imamu.edu.s

سنعرض في هذا الجدول أشهر المدونات العربية، تتفاوت من حيث الحجم سواء في شكلها المكتوب أو المنطوق

اسم المدونة	من صنع	الحامل اللساني	حجمها	الغرض منها	مصدر مادتها
- مدونة بوكوالتر العربية	- تيم بوكوالتر TimBuck walter	- مكتوب	-2.5- 3 بلايين كلمة	المعجماتية	-الموارد المتاحة على مستوى الشبكة
- مدونة لوفن Leuven /1990 /2004	- جامعة لوفن الكاثوليكية (بلجيكا)	- مكتوب ومنطوق	-3 ملايين كلمة (والمنطوقة منها 70000)	- في صناعة معجم عربي- هولندي/ هولندي- عربي لمتعلمي اللغة العربية	- الكتب المدرسية، الشبكة، الإذاعة والتلفزيون
- مدونة كلفرند (1995) Callfriend	- جامعة بنسلفانيا (LDC)	- محادثات	-60 مكالمات هاتفية	تطوير تقنيات المماثلة اللغوية	- متحدثو اللهجة المصرية
- مدونة نيميغن (1996) Nimegen	- جامعة نيميغن (هولندا)	- مكتوب	- أكثر من 2 مليون كلمة	- صناعة معجم عربي- هولندي/ هولندي- عربي	- مجالات وروايات
- مدونة كوهوم (1997) Callhome	- جامعة بنسلفانيا (LDC)	- محادثات	- 120 مكالمات هاتفية	- التعرف إلى الكلام، مصدره الخطوط الهاتفية	- متحدثو المصرية
- مدونة مصر (1999)	- جامعة جون هوكينز John Hopkins	- مكتوب		- الترجمة الآلية	- مدونة متوازية إنجليزية- عربية (القرآن الكريم)
- الإذاعية (2000)	- جامعة بنسلفانيا (LDC)	- منطوق	- أكثر من 110 برامج إذاعية	- التعرف إلى الكلام	- الأخبار المداعة على أمواج "صوت أمريكا"

<p>- منشورات المجلس الكويتي الوطني</p>	<p>- تدريس الترجمة والمعجمية</p>	<p>- 3 ملايين كلمة</p>	<p>- مكتوب</p>	<p>- جامعة نيميغنوسوسيتيل بالتعاون مع جامعة ليون 2</p>	<p>- مدونة دينار (2000)</p>
<p>- منشورات المجلس الكويتي الوطني</p>	<p>- تدريس الترجمة والمعجمية</p>	<p>- 3 ملايين كلمة</p>	<p>- مكتوب</p>	<p>- جامعة الكويت</p>	<p>- المدونة المتوازية إنجليزية- عربية (2003)</p>
<p>- مواقع تقنية المعلومات المتخصصة</p>	<p>- الترجمة</p>	<p>- 11.5 مليون كلمة (العربية منها 2.5 مليون)</p>	<p>- مكتوب</p>	<p>- جامعة مانشستر معهد العلوم والتقنيات Umist</p>	<p>- المدونة متعددة اللغات (2004)</p>
<p>- مواد أدبية أكاديمية وصحافية</p>	<p>- المعجمية</p>	<p>- 8 ملايين كلمة</p>	<p>- مكتوب</p>	<p>- سوتيتيل لتقنيات المعلومة (تونس) Sotetel-IT</p>	<p>- مدونة سوتيتيل Sotetel</p>

جدول رقم 01: أهم المدونات عربية¹

1- فاطمة الزهراء توتاي- استثمار المدونة في تدريب المترجم- رسالة ماجستير- كلية الآداب واللغات والفنون- قسم ترجمة- جامعة وهران- السانية- 2009- ص 175.

اسم المدونات	عدد الكلمات (حجمها)	الحامل اللساني	الموقع
BNC (British National Corpus)	100 مليون كلمة	90 % مكتوبة 10 % منطوقة	Simple searches : http://www.natcorp.ox.ac.uk/using/index-xm/?ID=simple . More advanced searches: http://bncweb.lancs.ac.uk/bncwebsignup/user/iogin.php .
COCA (the corpus of contemporary American English)	425 مليون كلمة	- مكتوبة ومنطوقة	http://www.AmericanCorpus.org/
ANC (American National Corpus)	20 مليون كلمة	- مكتوبة ومنطوقة	http://www.Idc.Upen.edu/catalog/catalogEntry.jsp?catalogId=LD2003T20 .

الجدول رقم 02: مدونات طلاب جامعة هيدلبرغر (Corpora of studentsuniversity of Heidelberg)¹

1 - ينظر: NadjaNesselhauf. Corpus linguistics :Apractical Introduction-October 2005-(last updatedSeptember 2001).

توطئة

تُستخدم المدونات المحوسبة في ميدانين كبيرين: الميدان الأول هو تدريب برمجيات المعالجة الآلية للغة، والميدان الثاني هو لَمَّا يقوم الباحث -ضمن مشاريع بحثية على اللغة العربية - بتطبيق أدوات حاسوبية على المدونات تنقيًا وتحليلًا. وفي هذه النقطة الأخيرة، يوجد مجموعة من الآليات يستعين بها الباحث للتنقيب داخل المدونة ونعني بهذه الآليات مختلف الطرق التي تتيح للباحث المعالجة الآلية للظواهر اللغوية ضمن المدونات، سواء من حيث استخراج البيانات وضبطها في شكل معطيات إحصائية، أو من حيث تحليلها ودراستها بدقة وسرعة هائلة وتمثّل هذه الأدوات في "الكشّافات السياقية".

المبحث الأول: الكشّافات السياقية concordancers

أولاً: مفهوم الكشّافات السياقية :

الكشّافات السياقية هي أداة (برمجية) التي تبحث في المدونات (المتون النصية)، وتعرض فهارس (كلمات، جذور، تصنيف حسب طول الكلمات، ترددات، بنيات نحوية أو معجمية). لا تقتصر الكشّافات السياقية في معالجة المدونات على الاستخراج الآلي للبيانات، بل إنّها تمكّن من ضبط القوالب المتكرّرة (Recurring Patterns)، وعرض الأجزاء النصية التي تنتمي إليها، ولا يقف البحث عند هذه المرحلة الكمية، بل يتعدّها إلى التحليل، بحيث تتولّى مهمتين رئيسيتين هما:

- تنظيم مكوّنات المدونة وهيئتها على نحو ييسّر على الباحث معاينتها ودراستها؛
- ضبط تلك المكوّنات في شكل معطيات إحصائية¹.

1- ينظر : فاطمة الزهراء توتاوي : استثمار المدونة في تدريب المترجم - مذكرة ماجستير - كلية الآداب واللغات والفنون - قسم الترجمة -

جامعة وهران - السانية - 2009 - ص 42.

ثانيا: أهمّ الكشافات السياقية

1- aConcorde: بُني في قسم الحاسب الآلي بجامعة ليدز، وهو أداة مجانية ومفتوحة

المصدر، صمّمت في المقام الأول لمساعدة الباحثين في لسانيات المدونات العربية¹.

وهذا الكشاف السياقي يوفر دعما كاملا للنصوص العربية، ويتمثل هذا الدعم في عدة

جوانب منها:

واجهة المستخدم التي يمكن تحويلها بين اللغتين العربية والإنجليزية، وتبنيه لترميز الحروف المعياري

Unicode، إضافة إلى دعمه لمجموعة من الخطوط العربية، وكذلك اتجاه النص العربي من اليمين إلى

اليسار².

صُنع الكشاف السياقي (aConcorde) باستخدام لغة البرمجة (java) مما يجعله صالحا لجميع أنظمة

التشغيل، وأنظمة التشفير بما في ذلك اليونيكود (UTF16).

كما يمكن حفظ النتائج إما ضمن الصورة النصية (Plain text format) أو ضمن صورة لغة الربط

بين النصوص التشعبية HTML.

- يحلل تكرار الكلمات.

- يتضمن تشكيلة واسعة من خيارات البحث: "بالكلمات، بالجمل، بالتجاور، بالجذور، بالجدع،..."

أي:

. arabik stem/ rootqueriers/wileard/booleaun/proximity/phrase/Word-Key

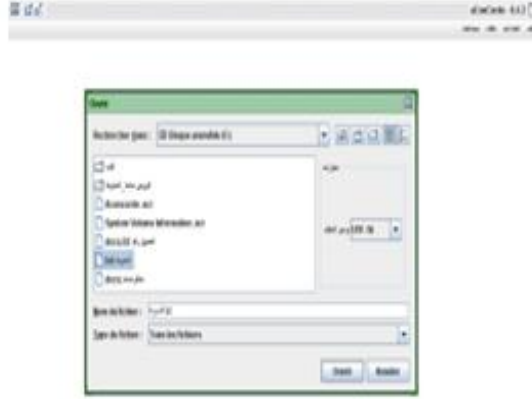
1-ايريك أتويل وعبد الله الفيبي - أبحاث جامعة ليدز في مجال لسانيات المدونات العربية :

/ ayjfaifi@imammu.edu.saE.S.Atwell@leeds.ac.uk

2-ينظر : فاطمة الزهراء توتواوي : استثمار المدونة في تدريب المترجم - مذكرة ماجستير - كلية الآداب واللغات والفنون - قسم الترجمة -

جامعة وهران - السانية - 2009 - ص 42.

- كما يتوفّر على واجهتين بحثيتين كما ذكرنا سالفًا، إحداهما بالإنجليزية والأخرى بالعربيّة، وهي واجهة تتضمّن ترجمة لكل القوائم (menus) والأزرار (buttons) تعرض جميع العناصر المكوّنة لها وفق اتجاه (يمين-يسار)، ومثال على ذلك ما يلي:



الشكل رقم 01: الواجهة الرئيسيّة للكشّاف السياقي أكونكورد

2- ويبكورب (webcorp): طقم من الأدوات التي تمكّن الباحث من الولوج إلى الشبّكة العنكبوتية ومعالجتها باعتبارها أضخم مدوّنة نصيّة على الإطلاق، وهو من إنتاج وحدة البّحث وتنمية الدّراسات الإنجليزيّة :

بجامعة برمنغهام ببريطانيا (RDUES) Research and development unit for english studies
ويبكورب بخلاف محرّكات البّحث البسيطة والمعهودّة، مصمّم بما يتوافق ومتطلّبات البّحث في مجال لسانيات المدوّنة.¹

وجميع الحقول المعرفيّة، التي تستثمر تلّكم اللّسانيات، إذ يتولّى عملية استرجاع المعطيات اللّسانية من صفحات الويب في شكل أسطر مفهرسة ومرتبّة ألفبائياً، حيث تتمركز وحدة البّحث (الكلمة المفتاحية مبرزة بلون أحمر) ضمن جميع سياقاتها المعتمدة، كما يمكن للويب كورب أن يصلك بجميع المواقع والصفحات الإلكترونيّة التي استقيت منها الكلمات المفتاحية .

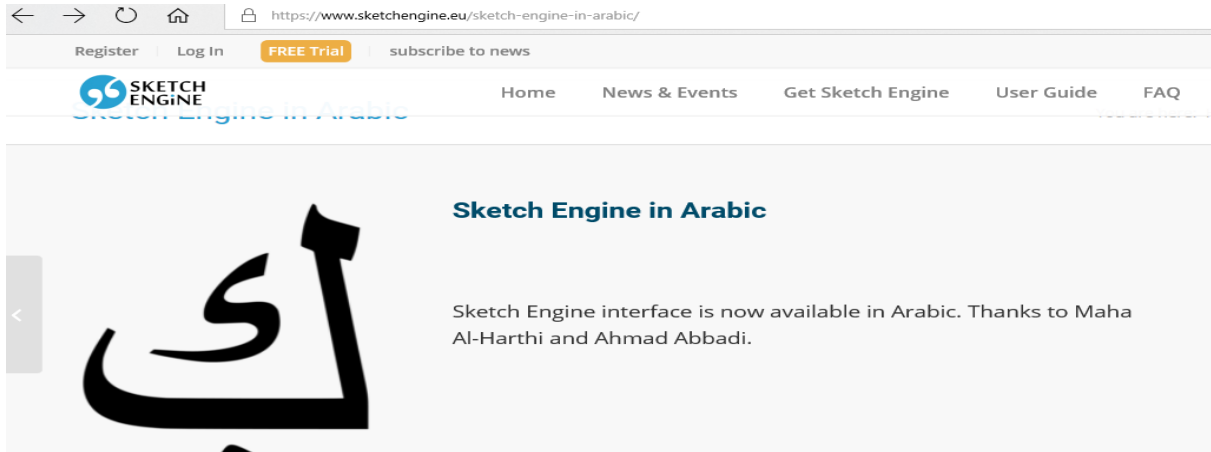
أي أنّ الويبكورب يستعمل الويب كمدوّنة باللّغة (الصينيّة، الإنجليزيّة، الألمانيّة، العربيّة) .

¹http://www.webcorp.uk: على الرّابط webcorp

3-سكتش انجائين:SketchEngine: وهذا أبسط برنامج يفيد الباحث في العثور على الكلمات في سياقات مختلفة وينبغي التنبه إلى أنّ البرنامج قد يكون بسيطاً بحيث ينظر إلى الكلمة بمعناها الحاسوبي (حسب الشكل فقط)، وهو البرنامج المتوافر حالياً مع مدوّنة اللّغة العربيّة التابع لمدينة الملك عبد العزيز للعلوم والتّقنية، أو يكون متطوراً بحيث يشمل تحليلاً تركيبياً وصرفياً حيث يمكن البحث بالجذر مثلاً أو بالجدع، ويورد البرنامج الكلمة في صورها المختلفة؛ وهذا النوع الأخير متوافر مع مدوّنة اللّغة العربيّة التابعة لجامعة برينغهام يونغ (Byu)

Brigham young universty في ولاية يوتا الأمريكية.¹ وهذا البرنامج متوفّر على الرّابط التّالي:

<https://www.sketchengine.eu/sketch-engine-in-arabic>



الصورة رقم 02: الواجهة الرئيسيّة لسكتش أنجائين.

1-محمود اسماعيل صالح -لسانيات المدونات اللغوية على الرابط :

Dr.mahmoud-ismail-salah- blogpost.com

4-أرضية "نوج" Nooj Plat Form

"نوج" هي بيئة تطوير هندسية لغوية مجانية تستخدم على المدونات اللغوية لتحليل جميع المستويات اللغوية: الإملاء، المعجم، الاشتقاقات الصرفية، التركيبية والدلالية، ولكل مستوى من هذه المستويات يوفر نوج تدقيقاً لغوياً والتوسيم Annotation. يميز "NooJ" في معالجته اللغوية الحاسوبية عن البرمجيات الأخرى التي تتعامل مع المدونات اللغوية بأن معالجته تغطي جميع الظواهر اللغوية وليس كأداة كشف سياقي فقط. باعتباره بيئة تطوير هندسية، يحتوي برنامج NooJ على أدوات للمساعدة في إنشاء مجموعات كبيرة من الموارد اللغوية واختبارها وتصحيحها وصيانتها وتجميعها، بالإضافة إلى أدوات لمعالجة النصوص الكبيرة والمحتويات. تم تطوير النظام منذ عام 2002 وتم استخدامه لبناء أكثر من 20 وحدة لغوية. وكأداة لمعالجة النصوص، يسمح NooJ للباحثين في مختلف العلوم الإنسانية والاجتماعية واللغوية والأدبية باستخلاص المعلومات من أي نص أو مدونة (مدونة خام أو موسومة) من خلال تطبيق استعلامات متطورة تعتمد على المفاهيم بدلاً من استمارات الكلمات المفتاحية فقط وإجراء التحليلات الإحصائية.

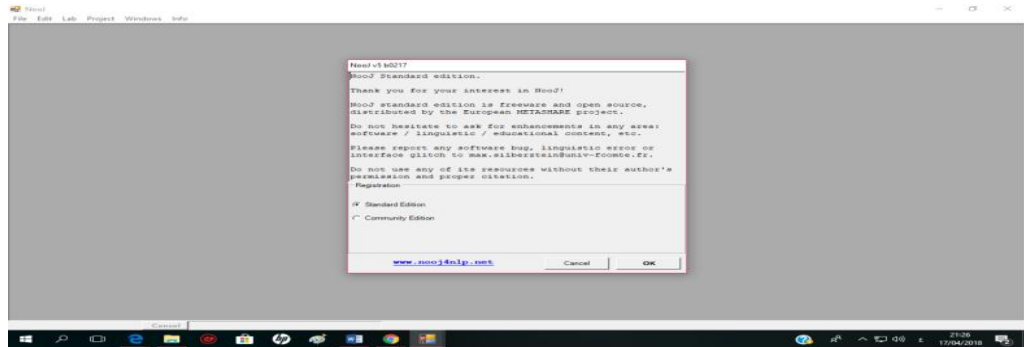
يمكن تنزيل آخر نسخة من بيئة نوج Nooj والتي تدعم اللغة العربية من الموقع الرسمي على الرابط التالي:

http://www.nooj-association.org/index.php?option=com_k2&view=item&layout=item&id=4&Itemid=615

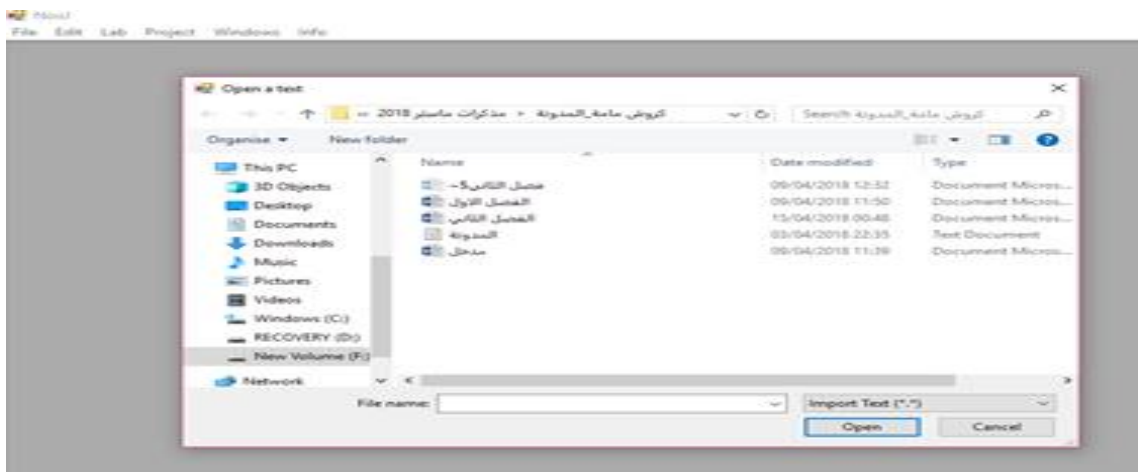
والواجهة الرئيسية للموقع الرسمي لنوج:



وبعد تنزيل نوج وتنصيبه، وبعد تشغيله تظهر الصفحة الرئيسية كالتالي:



فيمكن بعدها فتح أيّ مدونة لغوية انطلاقاً من بيئة نوج لكي نطبّق عليها الأدوات المتاحة في هذا الأرضية، كما هو مبين في الرسم التالي:



المبحث الثاني: عرض لمدونتنا

أولاً: جمع المدونة

قامت مجموعة من طلبة بقسم اللغة العربية وآدابها بجامعة تلمسان، بشعبة اللسانيات التطبيقية، سنة أولى ماستر دفعة 2018 بجمع ما تيسر من نصوص بتوجيه وإشراف من الأستاذ "الهادي شريف" تضمنت الشعر والنثر والنصوص الأدبية والخطب وغيرها من مختلف العصور، وقدرت صفحاتها بأكثر من ألفين وخمسمائة صفحة (2500) في برنامج الورد، احتوت على أكثر من ثلاثة ملايين ومئتي ألف حرف (3.200.00)، بما يقدر بأكثر من 500 ألف كلمة موزعة كالاتي:

-العصر الجاهلي: كان لدينا فيه نبذة عن الشعر بما فيه المعلقات السبع والأدب والنثر؛

- صدر الإسلام: كان في صدارة هذا العصر ذكر الله الحكيم أي القرآن الكريم وبعض الأحاديث النبوية الشريفة؛

-أما العصر الأموي والعباسي: فأهم ما جمع في هذا العصر بعض من النصوص الأدبية، القصائد الشعرية النثر، الخطابة والكتابة، بالإضافة إلى بعض التفاسير، مثل: تفسير ابن كثير وعدة مقالات فقهية مختلفة وخطب؛

-وأخيراً العصر الحديث، فقد ورد فيه بعض المقالات الفلسفية والبحوث اللسانية والعلمية بأنواعها، القصص والروايات والمقالات المنشورة في أممات الصحف والجرائد، والمجلات العربية سواء في ميدان السياسة، أو الرياضة، أو الاقتصاد، أو الطب وغيرها.

سنفصل ذلك في الجدول التالي:

نوع النص	
خطبة الوداع للرسول صلى الله عليه وسلم	01
- مقدمة في السيرة النبوية	02
- بحث عن المعلقات	03
قضايا الشعر الجاهلي	04
- خطب الحجاج بن يوسف	05
- قصيدة واحر قلباه للمتنبى	06
- بحث في اللسانيات الحديثة ومفهوم المصطلح	07
- كتاب مفاخرة الجواري والغلمان	08
- موسوعة عن السرطان أنواعه وأسباب علاجه	09
- شعر العصر الوسيط في جهود النقاد الباحثين المحدثين	10
- من نوادر أشعب	11
- مفهوم الكلمة	12
بحث في الخصائص النفسية والاجتماعية للتلاميذ	13
رسالة ماستر حول الموارد البشرية	14
أهم المدارس اللسانية العربية والغربية	15
اللغة واللسانيات	16
محاضرة حول مفهوم مصطلح اللسانيات التطبيقية وعلاقتها باللسانيات العامة	17
دراسة وصفية لأشهر الكتب التي تصدت للرد التفصيلي على كتاب في الشعر الجاهلي للدكتور طه حسين	18
الأدب الفلسطيني في العصر الحديث	19
- ظاهرة العولمة الاقتصادية	20
- الاحتباس الحراري	21
الوطن العربي دراسة طبيعية بشرية	22

23	- النمو الاقتصادي
24	- الأنترنت
25	- أساليب البحث العلمي
26	مقالات عن الاستشراق
27	مقالات عن السينما

ثانيًا: نماذج مقتطفة من المدونة

بسم الله الرحمن الرحيم

الم (1) ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ (2) الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنْفِقُونَ (3) وَالَّذِينَ يُؤْمِنُونَ بِمَا أُنزِلَ إِلَيْكَ وَمَا أُنزِلَ مِنْ قَبْلِكَ وَبِالْآخِرَةِ هُمْ يُوقِنُونَ (4) أُولَئِكَ عَلَى هُدًى مِنْ رَبِّهِمْ وَأُولَئِكَ هُمُ الْمُفْلِحُونَ (5) إِنَّ الَّذِينَ كَفَرُوا سَوَاءٌ عَلَيْهِمْ ءَأَنْذَرْتَهُمْ أَمْ لَمْ تُنذِرْهُمْ لَا يُؤْمِنُونَ (6) خَتَمَ اللَّهُ عَلَى قُلُوبِهِمْ وَعَلَى سَمْعِهِمْ وَعَلَى أَبْصَارِهِمْ غِشَاوَةٌ وَلَهُمْ عَذَابٌ عَظِيمٌ (7) وَمِنْ النَّاسِ مَنْ يَقُولُ آمَنَّا بِاللَّهِ وَيَالْيَوْمِ الْآخِرِ وَمَا هُمْ بِمُؤْمِنِينَ (8) يُخَادِعُونَ اللَّهَ وَالَّذِينَ آمَنُوا وَمَا يَخْدَعُونَ إِلَّا أَنْفُسَهُمْ وَمَا يَشْعُرُونَ (9) فِي قُلُوبِهِمْ مَرَضٌ فَزَادَهُمُ اللَّهُ مَرَضًا وَلَهُمْ عَذَابٌ أَلِيمٌ بِمَا كَانُوا يَكْذِبُونَ (10) وَإِذَا قِيلَ لَهُمْ لَا تُفْسِدُوا فِي الْأَرْضِ قَالُوا إِنَّمَا نَحْنُ مُصْلِحُونَ (11) أَلَا إِنَّهُمْ هُمُ الْمُفْسِدُونَ وَلَكِنْ لَا يَشْعُرُونَ (12) وَإِذَا قِيلَ لَهُمْ آمِنُوا كَمَا آمَنَ النَّاسُ قَالُوا أَنُؤْمِنُ كَمَا آمَنَ السُّفَهَاءُ أَلَا إِنَّهُمْ هُمُ السُّفَهَاءُ وَلَكِنْ لَا يَعْلَمُونَ (13) وَإِذَا لُقُوا بِالَّذِينَ آمَنُوا قَالُوا آمَنَّا وَإِذَا خَلَوْا إِلَى شَيَاطِينِهِمْ قَالُوا إِنَّا مَعَكُمْ إِنَّمَا نَحْنُ مُسْتَهْزِئُونَ (14) اللَّهُ يَسْتَهْزِئُ بِهِمْ وَيَمُدُّهُمْ فِي طُغْيَانِهِمْ يَعْمَهُونَ (15) أُولَئِكَ الَّذِينَ اسْتَرَوْا الضَّلَالَةَ بِالْهُدَىٰ فَمَا رَبَحَتِ تِجَارَتُهُمْ وَمَا كَانُوا مُهْتَدِينَ (16) مَثَلُهُمْ كَمَثَلِ الَّذِي اسْتَوْقَدَ نَارًا فَلَمَّا أَضَاءَتْ مَا حَوْلَهُ ذَهَبَ اللَّهُ بِنُورِهِمْ وَتَرَكَهُمْ فِي ظُلُمَاتٍ لَا يُبْصِرُونَ (17)

الشكل رقم 02: عينة من المدونة عن عصر صدر الإسلام.

لقصيدة كعب بن زهير المعروفة بالبردة، التي مطلعها:

بانّت سعاد فقلبي اليوم متبول متيم اثرها لم يغد مكبول

وما سعاد غداة البين اذ رحلوا إلا انحّت غضيض الطرف مكحول()

وعادة الشعراء في معارضته كانت في السير على نهجه، فيبدأون قصائدهم بالغزل. إلا أن البوصيري قد خرج عما هو مألوف في هذه المعارضة، فافتتح قصيدته المسماة(ذخر المعاد، في معارضة بانّت سعاد) في الوعظ والارشاد، وزجر النفس، وحثها على الزهد في الدنيا والاستعداد للأخرة. منتهياً إلى اظهار شوقه وتوقه إلى زيارة الضريح النبوي(). وهذا ما أشار إليه المحقق إلا أنه لم يذكر مطلع قصيدته ولعله يريد لاميته التي مطلعها:

إلى متى أنت باللذات مشغول وأنت عن كل ما قدّمت مسؤل

في كل يوم تُرجى أن تتوب غداً وعقد عزمك بالتسويف محلول()

الشكل رقم 03: عينة من المدونة عن عصر صدر الإسلام.

نقد اللغة

تعد اللغة ((الظاهرة الأولى في كل عمل فني يستخدم الكلمة أداة للتعبير))، فهي بذلك قبل كل شيء أساس من أسس الأدب الراسخة الذي لا نستطيع إن نفهمه فهماً تاماً ((ولا يحسن إن نستكشف كنوزه بغير فقه لخصائص هذه اللغة وأحاطة بما لاصواتها ومفرداتها وتراكيبها من مزايا وصفات.))

إن امسك الأدب بناصية اللغة وتطويعها لفنه الأدبي لاسيما الشعري ، يعد معياراً أساسياً في نجاحه أو اخفاقه في عمله الأدبي ؛ لان اللغة ((وسيلة وغاية في اداء المعنى والايحاء به والأيماء اليه وتأصيله بجو خاص ... واللفظة في الشعر كائن حي ينمو ويقوى ويزدهر بعلاقته الجديدة في الالفاظ الاخرى)) وقد أدت العناية باللغة وتقديسها من لدن النقاد والباحثين إلى العناية بالنقد اللغوي في عصور الأدب العربي كافة ، لأنه وسيلة من وسائل الحفاظ على اللغة ومكانتها العليا ، ولاسيما إن ((النقد العربي القديم في جملته نقد لغوي.))

الشكل رقم 04: عينة من المدونة عن العصر الحديث.

المبحث الثالث: أعمال تطبيقية على المدونة

أولاً: إحصاء الأصوات وتحليلها

1. إحصاء الفونيمات: Phonèmes

بدأنا كأول تطبيق للأدوات الحاسوبية بإحصاء الأصوات الأحادية (الفونيمات) في المدونة وذلك باستخدام أداة سكتش أنجائن Sketch Engine التي عرضناها في المبحث السابق، فقمنا بالبحث في المدونة بواسطة هذه الأداة على موقعها الرسمي بعد التسجيل والدخول، وأول ما بدأنا به هو صوت الهمزة /ء/، الذي يتضمّن إشكالا، لأنّ هذا الصوت له عدّة تمثيلات خطيّة إملائيّة: ء –أ– إ– و– ئ – آ، فنقوم بالبحث عن كلّ هذه الصيغ الإملائيّة التي تتمثّل في نهاية المطاف أداءً نطقياً واحداً:

Query 37,861 (124.71 per million) ⓘ

Page 1 of 51,894 Go Next Last

ان يقنعني ب ان اجلب الي ه ممرضة .. لك ي يخفف على العيب ء يا الهي ل م اكن اريد ه يفكر ... و اوضحت ل ه س ب ب حزني .. و جاني و وسط , و الخلف قائم و ل ن يزال , لكن ب الهدو ء و التعلل و سعة الصدر , و تقوى الله ف ي السر و العطن و صدفة , و التوب صدفة , الاصوات و الاعاني , الضو ء و العمة , مجموعة من الصدق او الحز التي تلقى ب ها بد ان تحرض على الابتعاد عن الحسيبة و التحدث ب هذو ء و معالجة الموقف مع عدم التوسع ف ي ذكر التفاصيل <p> هل تتصورين شعور غواص نزل قاع البحر و هو ينزل ب هذو ء كانت الصور الجميلة و الرائحة تلف ه ل فا , كسبه كيف س يتعامل الخليج مع ها ? عرفت دول الخليج ب الهدو ء الدبلوماسي , و الذي ب النفس عن الحروب و التصعيد جعل تحويل الارصدة من ب تك الحظ . <p/> تحسس خطوأة لك ب هذو ء , او اضرب ب قدمك ما تفتاه , ل ا يهم ; ف انت على لك ل ل ا محمود البكر عنوان " ياكورة الجنى " . تم تسليط الضو ء على اهمية التراث الثقافي اللامادي ف ي ضوء صلى الله على ه و سلم : " من تواضعا ل الصلاة ف اسبح الوضو ء , ثم مشى الى المكتوبة ف صلاها مع الناس , او مع .. و ب الفعل ادى هذا الى نوع من الاستقرار و الهدو ء بين سكان الشرق و الغرب نوعا ما , لكن ه كان يعود بين ل هؤلاء الاطفال من الممكن ان يكون هناك مركز ايو ء ل ابناء اطفال الشوارع يتم ايداع الاطفال في ه مع الجلسة الى مسابقة ذهنية يربح في ها الاكثر هذو ء او احباطا او اشراقا ف ي الروح . ظواهر عديدة دخلت تحصل على ممتك الشخصية ف ي الفرحة على العالم ب هذو ء و سلام و طمأنينة دون ان يفتر احد هم داخل دماغك و الشباب ضد الفقر كما ان ها تساهم ف ي تخفيف العيب ء المالي عن الاسر المحتاجة مع بداية لك ل عام دراسي ماضية جرت و يمكن ان تحطى اماما عاما يلقي الضو ء على التحديات ف ي مصر ف ي ظل التغيير الخاص " . جاء هذا الى بيوتك . <p/> ايقسم المسبي و ارقسم السلام و الهدو ء على وجه ه , و ب سرور و اغتباط اجاب : <p/> - اطمنن يا سيدي ف ين منذ زواجنا " , ف ل م يخبرها ب شيء , ب ل قال ب هذو ء ان على ه ان ينتهي الليلة من تصحيح اوراق كلية اركان ان الانسان يرى الاشياء نتيجة ل انعكاس الضو ء عنها و دخول ه الى العين , و ب ذلك حضض نظريات بطليموس

الشكل رقم 05: البحث عن "ء" في المدونة بواسطة سكتش أنجائن.

ثم نفس الشيء بالنسبة للصيغة الإملائيّة للهمزة: "ؤ" فنحصل على:

Query 4,835 (0.60 per million) i

Page 1 of 242 Go Next Last

الدعارة في العالم الإسلامي . تم توزيع نسخ مجانيه في علي نفقة الأزهر الى كل من يصلي في جامع ايام الجمع ه
و داعية الاسلامي خالد الجندي في البيت بيتك في علي الفاهر اليوم ازاى هو بيوصف داخل الجوامع من
يصف الكتبي ه والمصلين في ها في النظافه والباس ؟ . كيف يصف الجامع من الوساخه ووساخة المصلين و
؟ . وكيف يصف الجامع من الوساخه ووساخة المصلين في الفوضى ؟ <p></p> صحیح الحیاة فائیه لكن الإسلام يعلم
والغدير . وب ذلك تكون قد قدمت عملا كاملا . ل . <p></p> 26 : 52 : 23 2009-07-06 <p></p> Anonymous <p></p> ميازل الشيوخ و
اكل . حلوياس في عمر . كتاب (ترميض) مرخص من الأزهر في كمان اظهرت ختم الأزهر الشر ... يا عمر هل تعلم ان 90
الشعر العربى بى ن التطور والتطور التدرى جى : <p></p> فى كذا احمد المعادى ان حركة تطور الشعر العربى
<p></p> PM <p></p> سونال خليفة الله في الأرض اله اكثر من جواب ديني رسالة عمان / 37 <p></p> احمد محمد <p></p> 1/1/2012 9 : 54 : 34 PM <p></p> فى في
في حالات صحية عملة ل ان العملية تحسن ل فترة قصيرة في من تم يموت كما كان <p></p> اوكرانيا <p></p> ابي <p></p> 21 : 15 06-12-2010
وتجرؤا على حبيبك و نبيك اللهم ارنا في هم ما تحب ه في رضاه اللهم اعز دينك وانصر نبيك / 584 <p></p> . اللهم صلى
والاعلامي الصانق ب خوش .. هذه التعاليمات و غيرها ت في كد على ان هناك حراك ثقافي هام و ان نبض الثقافة و
والحكومات في ه اثنيه ب جميعات خيرية غير ربحية , ت من ب فكرة العمل التطوعي و تخوض حروبها ل تقديم
الى اهلك . ف ان احبوا ان اقضى عنك كذبة له , و يكون ل ا في ل ي , فطت . ف نكرت ذلك بريرة ل اهل ها . ف ابوا . و
: ان شابه ان محتسب عليك ف ل فعل . و يكون ل نا و ل ا في . ف نكرت ذلك ل رسول الله صلى الله على ه و سلم . ف قال ل
مع صحيفة "نيويورك تايمز" في 14 ابريل 2007 عن ر فيك للور الذي تلعب ه القوات الامريكية الاضافية
.. و اعني ف شعرت اني منتب ... ل و ان ملطي ضمنني ! اشلا في كم متفورة ... و ننام مله الاخين ! و دماؤكم مهدورة
ل ا وجود ل ه في كتيبات الشاعر التظي رى ه , و هذا ما في كد عبرى ه نازك الملاكة فى صنع صورة الشاعر العرب
: ل ا يؤخذ العلم من اربعة , و يؤخذ ممن سوى ذلك و ل ا في كذم ن سفيه ملطن ب السف ه و ان كان اربى الناس و ل ا
و اخرها مثال المعنى ass 11 a اليوم / النهار u و udem الوجه ا ئ ا في اليرى <p></p> الصوتان ت و ذ : يتميز
تضع يدها على قلبها خوفا من تخير في اسعار النفط في نرف في الاقتصاد و الصناعة العربية !! <p></p> كانت امريكا ك

Page 1 of 242 Go Next Last

وهكذا بالنسبة للصيغ الإملائية المتبقية لصوت الهمزة، ولما نجمع الأعداد سنحصل على

عدد مرات تكرار صوت الهمزة في مدونتنا: 296949مرة.

ولإحصاء باقي الأصوات، نتبع نفس الطريقة بدون أي إشكال، لأن باقي الأصوات تمثلها

خطيا باقي الحروف وفق تقابل تام: فحرف الباء هو تمثيل لصوت /ب/، والتاء لصوت /ت/ وهكذا.

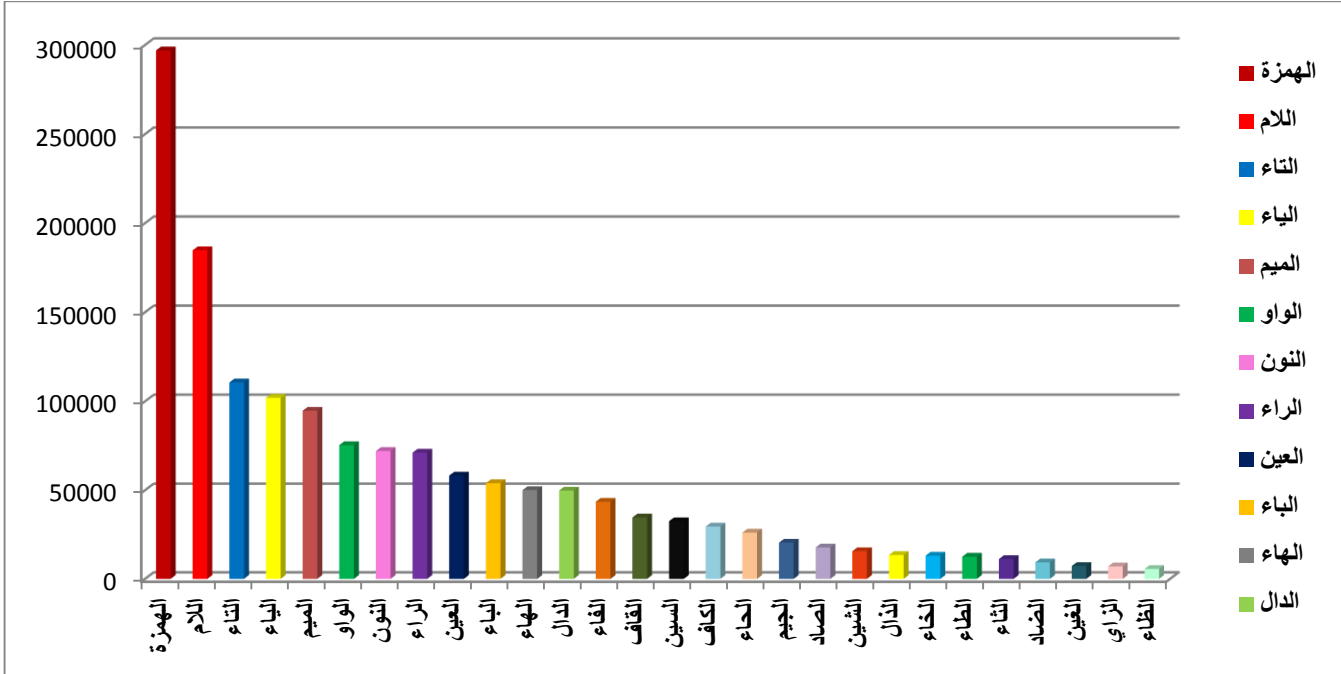
ولما نحصل من سكتش أنجائين على عدد تكرار الأصوات من المدونة نسجل النتيجة على

جدول في الإكسل ثم نقوم بترتيبها من أعلى قيمة إلى أصغر قيمة حسب ترددها:

الأصوات	العدد
الهمزة	296949
اللام	184788
التاء	110607
الياء	101985
الميم	94719
الواو	75397

72147	النون
71177	الراء
58242	العين
53963	الباء
50065	المهاء
49814	الذال
43506	الفاء
34681	القاف
32521	السين
29514	الكاف
26107	الحاء
20423	الجيم
17722	الصاد
15623	الشين
13478	الذال
13086	الحاء
12552	الطاء
11177	الثاء
9327	الضاد
7254	الغين
7032	الزاي
5605	الظاء

ورسمنا المنحنى البياني التالي انطلاقاً من الإكسل:



الشكل رقم 06 : منحنى بياني يمثل تكرار الأصوات الأحاديّة "الفونيمات"

فإذا عمدنا إلى تحليل هذه النتائج، سنرى أنّها متوافقة مع نظريات صوتيّة كانت معروفة لدى الدارسين، وما نتأجنا هذه انطلاقاً من المدونات الحاسوبية إلا تأكيداً عملياً لهذه المعارف النظرية. وفعلاً، لكل صوت من أصوات اللّغة وظيفة وما تكراره الأداء لهذه الوظيفة المنوطة به في المعنى، ومن خلال تحليلنا لهذا المنحنى نلاحظ:

1- صوت الهمزة: هو الأكثر تردداً وهذا ليس بالجديد على اللّغة العربيّة لأتّها وكما هو معروف لغة تهميز، ضف إلى ذلك الأسماء المعرّفة تُنطق "ال" التعريف همزة. كما أنّها من الأصوات الحلقية التي مخرجها من الحلق من ناحية الصّدر وهو أوّل المخارج الصّوتية عند سيويوه وابن جيّي وكذلك علماء الأصوات المحدثين.

2- صوت اللام: إذا كانت الكلمات المعروفة في لغة العرب تبدأ بهمزة، فإنّ هذه الهمزة لا يكتمل عملها إلاّ بمجانبتها لصوت اللام وهو ما يصطلح عليه "أل" التعريف، هذا ما جعله - أي صوت اللام - يحتلّ المرتبة الثانية بعدها.

3- صوت التاء: كانت التاء الصوت الثالث من حيث عدد التكرار لأنّ التاء تأتي مفتوحة في أول الأفعال وآخرها كما أتتها تأتي مربوطة في الأسماء المؤنثة.

4- صوت الظاء: كان آخر الأصوات تكراراً، حيث يقول الداني القرطبي (371هـ - 444هـ) - وهو أحد علماء اللغة- في كتابه الفرق بين الضاد والطاء: "قد أجمع علماء اللغة على أنّ العرب خصّت بحرف الظاء دون سائر الأمم، لم يتكلّم بها غيرهم، ولغرابتها صارت أقل حروف المعجم وجوداً في الكلام وتصرفاً في اللفظ، واستعمالاً في ضروب المنطق" ويضيف فيقول: "فهي لا توجد إلاّ في نحو مئة كلمة من جملة كلام العرب"¹.

2. إحصاء الأصوات الثنائية (الدايفونات) Diphones

الدايفون أو الأصوات الثنائية هي تتابع لصوتين لغويين، يعني تجاور لفونيمين. اتّبعتنا نفس طريقة إحصاء الأصوات الأحادية (الفونيمات)، لكن هذه المرّة أحصينا الأصوات الثنائية مثل: تجاور همزة وهمزة /ء+ء/، أو همزة وباء /ء+ب/... إلخ، وذلك باستعمال نفس البيئة الحاسوبية "Sketch Engine"، لنحصل في الأخير على المصفوفة التالية :

1- عثمان بن سعيد الداني أبو عمرو الأندلسي-الفرق بين الضاد والطاء في كتاب الله -تحقيق غانم قدوري-دار عمار-الأردن 2010-ص15.

تطبيقات لسانيات المدونة

الفصل الثالث

	الهمزة	الباء	التاء	الثاء	الجيم	الحاء	الخاء	الدال	الذال	الراء	الزاي	السين	الشین	الصاد	الضاد	الطاء	الظاء	العين	الغين	الفاء	القاف	الكاف	اللام
577	7859	6247	36038	1728	2805	3658	2780	8895	1562	9390	818	8491	1528	2101	1117	1476	351	5348	534	3757	3815	2165	
756	8036	409	2723	105	139	1357	109	1690	239	1811	52	409	853	345	94	693	20	2882	350	94	989	544	
761	4718	1866	858	212	966	2327	1091	1020	150	2492	331	944	1008	1337	361	888	142	2698	499	2212	2114	1150	
047	1287	317	412	2	0	0	0	1	0	1264	0	0	0	0	0	0	0	14	2	7	478	13	
987	3750	513	1641	3	17	97	0	1722	86	974	559	173	7	3	1	0	1	868	45	73	0	5	
190	2171	755	1342	831	288	14	0	3050	97	1204	2	1415	77	430	191	98	344	2	0	270	1117	674	
964	1214	370	1053	1	7	0	1	518	441	1529	75	10	34	876	515	750	750	2	0	1	1	1	
691	5576	1355	3285	25	49	163	384	751	11	3553	15	134	10	6	0	3	0	560	15	560	476	851	
086	3838	223	133	0	107	2	7	1	2	207	1	1	0	0	0	0	0	28	1	34	21	643	
40	10784	3129	5512	198	1401	787	4	1850	8	158	257	1183	166	214	1034	1	5	338	327	1721	938	1250	
347	1126	90	360	0	75	1	30	129	0	113	16	0	1	1	0	0	0	287	8	19	39	27	
457	3754	1761	5420	0	212	145	123	192	11	697	0	122	2	0	0	607	0	606	1	250	169	247	
44	2475	457	472	0	133	84	422	322	18	2133	1	2	2	1	0	85	4	3574	96	210	124	837	
180	2276	564	662	0	0	687	2	689	0	1619	0	0	0	337	0	197	0	292	75	1051	30	2	
351	1196	106	250	0	47	427	43	49	1	929	0	0	0	0	4	105	0	741	40	46	1	2	

تطبيقات لسانيات المدونة

الفصل الثالث

6	954	0	287	394	10	238	0	62	0	0	14	5	0	1396	0	5	4	37	3	0	349	978	1564	الطاء
8	247	3	0	51	1	14	0	0	0	0	0	0	0	1007	0	0	0	0	0	0	165	29	623	الطاء
1	10746	174	662	229	0	4	266	217	891	812	378	74	224	7089	2	2745	0	5	333	157	3160	1452	4756	العين
4	456	1	0	95	0	1	0	38	74	9	2	31	218	739	38	280	0	0	0	21	170	152	626	العين
4	1229	933	1381	90	17	1110	312	70	463	451	60	1357	88	2038	51	109	114	259	107	9	1666	88	3897	الفاء
2	1575	13	172	602	0	422	13	524	516	85	311	285	6	1650	30	4843	1	68	1	0	1230	1076	1702	القاف
7	2811	32	25	281	8	83	3	14	6	10	130	304	158	1742	459	277	11	34	8	1135	3136	622	4792	الكاف
#	4939	5396	5755	4398	2258	6836	289	1156	349	1635	4290	2957	493	2314	2178	3464	2069	3665	2588	1038	38185	4141	24808	اللام
5	3023	1278	1149	565	163	5022	119	415	236	655	656	2015	223	3161	183	1810	637	1467	1277	1105	4857	912	13714	الميم
9	148	443	1596	1502	48	660	2020	504	85	757	649	1515	5	145	186	1224	102	668	575	132	4048	1133	6742	النون
7	1576	118	16	70	1	19	1	6	75	3	59	3	132	1104	4091	1073	2	4	667	3	584	397	1025	الهاء
7	7170	1350	2728	1690	348	2340	376	409	1355	1174	397	1572	477	4251	350	1704	340	1164	1821	518	2801	2116	18372	الواو
2	2766	1565	3148	1376	189	2585	125	586	571	598	778	2095	903	5704	297	3505	1336	1247	1178	1146	14694	1606	7353	الياء

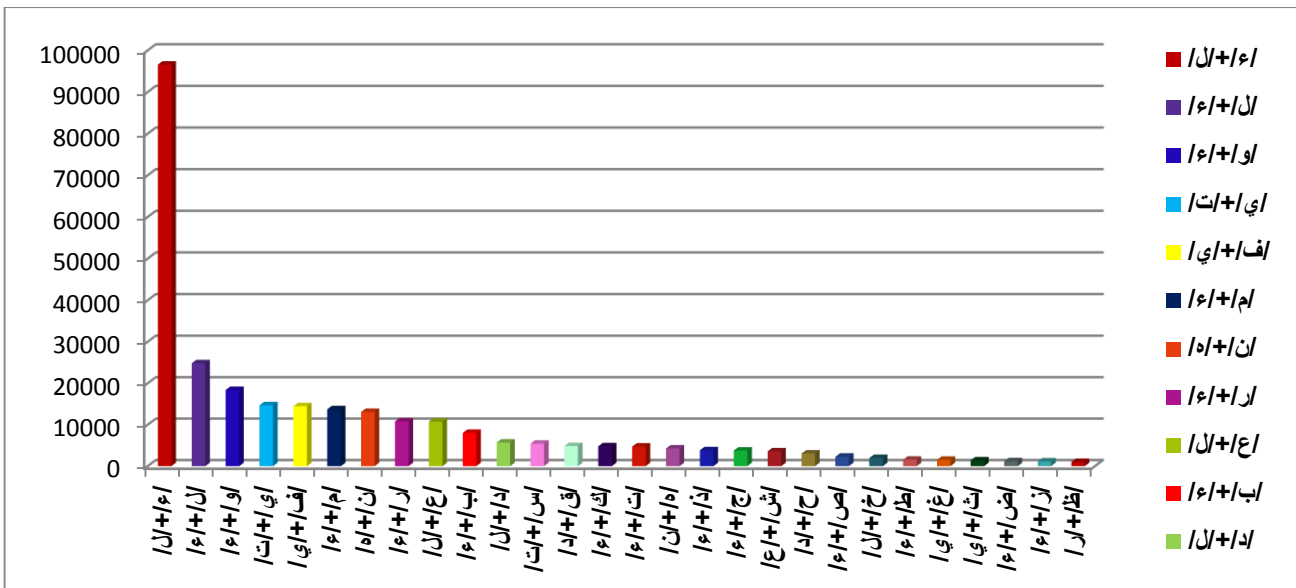
حيث كل خانة رقمية من المصفوفة تمثل عدد تكرار الدايفون المكوّن من فونيم السطر يتبعه فونيم العمود، فمثلا في أول خانة على اليمين (أول خانة رقمية ملوّنة بالأخضر) نجد العدد 7859 الذي يمثل عدد مرّات تكرار في المدوّنة الدايفون /ء/ + /ء/، وكمثال آخر، ثاني خانة ملوّنة بالأخضر والتي تحتوي العدد 1020 تمثل عدد تكرار الدايفون: /ت/ + /د/.

وبعدها حسبنا القيمة القصوى للأصوات الشائبة الأكثر تكرارا ووضعناها في جدول كالآتي:

حساب القيمة القصوى MAX لعدد تكرار الأصوات الشائبة:

عدد التكرار	ثنائية الأصوات
96577	/ء/+/ل/
24808	/ء/+/ل/
18372	/ء/+/و/
14694	/ء/+/ت/
14457	/ء/+/ي/
13714	/ء/+/م/
13075	/ء/+/ن/
10784	/ء/+/ر/
10746	/ء/+/ل/
8036	/ء/+/ب/
5691	/ء/+/ل/
5420	/ء/+/ت/
4843	/ء/+/د/
4792	/ء/+/ك/

4718	/ت+/ء/
4263	/ه+/ن/
3838	/ذ+/ء/
3750	/ج+/ء/
3574	/ش+/ع/
3050	/ح+/د/
2276	/ص+/ء/
1964	/خ+/ل/
1564	/ظ+/ء/
1532	/غ+/ي/
1393	/ث+/ي/
1196	/ض+/ء/
1126	/ز+/ء/
1007	/ظ+/ر/



الشكل رقم 07: المنحنى البياني الذي يمثل الأصوات الثنائية الأكثر تكراراً

تحليل المنحنى الثاني :

1- /ء/+ /ل/: هي أكثر الأصوات الثنائية تكراراً، فقد سبق وأن أشرنا إلى أنّ صوتي الهمزة واللام يقترنان بكثرة باعتبارهما معا أداة تعريف وكلّ منهما يكمل الآخر، ضف إلى ذلك أنّهما من مخرجين متباعدين فالهمزة صوت حلقي، أمّا اللام فمخرجه حافة اللسان الأمامية مع التصاقها بما يجاذبها من الأسنان، وكلّما تباعدت مخارج الأصوات تقاربت الصلة بينها.

2- /ظ/+ /ر/: هي أقل الأصوات تكراراً لأنّه رأينا فيما سبق أنّ صوت الظاء من الأصوات القليلة الاستعمال في اللغة العربية، وأيضاً كلما تقاربت مخارج الأصوات كلما كان التكرار قليلاً أو منعدماً فمخرج صوت الظاء هو ظهر اللسان مع رؤوس الثنايا العليا، ومخرج صوت الراء طرف اللسان مع أصول الثنايا العليا قريباً من الظهر وهي متقاربة جداً.

3- هناك أصوات ثنائية نادرة جداً أو لم تتكرّر نهائياً في المدونة مثل: /ث/+ /ذ/ وكذلك /ث/+ /ظ/، /ث/+ /د/، /ث/+ /ص/ وتحليلنا لذلك وكما سبق ذكره أنّه كلّما تباعدت مخارج الأصوات كان عدد التكرار أكبر والعكس كلّما كانت المخارج متقاربة كان التكرار أقل أو منعدم .

والثناء: صوت أسناني احتكاكي مهموس.

الذال : صوت أسناني احتكاكي مجهور.

الطاء : صوت أسناني احتكاكي مجهور ومفخم .

الذال : صوت أسناني لثوي .

الضاد صوت أسناني لثوي¹

فكلّ هذه الأصوات أسنانيّة وهذا ما يفسّر ندرتها أو عدم تكرارها.

ثانياً: توظيف الكشاف aConcorde

1- إحصاء أسماء الإشارة:

الصورة رقم 01: عينة بحثية عن أسماء الإشارة باعتماد أكونكورد.

The screenshot shows the aConcorde software interface. The main window displays a search for the word 'هذا' (this) and a list of results. The results are sorted by frequency, with the most frequent results at the top. The interface includes a search bar, a list of results, and a summary of the search results.

تكرار	كلمة
20	104
20	105
20	106
21	107
20	108
1	108,7
20	109
1	10دول
1	-10
123	11
1	11.209
19	110
18	111
16	112
15	113
15	114
14	115
1	115,7
1	115و
15	116
15	117
1	117,6
14	118
12	119
1	-11

الجدول رقم 01: أسماء الإشارة

النسبة المئوية	عدد تكراراتها	أسماء الإشارة
4.43%	1750	هذا
2.88%	1136	ذلك

1- حسام البهنساوي ، علم الأصوات ، مكتبة الثقافة الدينية ، القاهرة ، 2004 ، ص : 64-67.

هذه	1435	3.63%
تلك	262	0.66%
هذان	05	0.03%
هاتان	01	0.02%
هؤلاء	132	0.33%
هنا	133	0.34%
هناك	212	0.54%
هنالك	18	0.05%

- تحليل الجدول رقم (01):

قمنا بواسطة برنامج أكونكورد (aConcorde) بمجرد عدد تكرارات أسماء الإشارة التي تدل على مشار إليه معين و هي نوع من المعارف اللغوية.

وأسماء الإشارة هي:

- هذا، ذلك: للمفرد المذكّر؛
- هذه، تلك: للمفرد المؤنث؛
- هذان: للمثنى المذكّر؛
- هاتان: للمثنى المؤنث؛
- هؤلاء: لجمع الذكور وجمع الإناث؛
- هنا، هناك، هنالك: للمكان.

6.09%	2402	مَن
2.63%	1038	ما

- تحليل الجدول رقم (02)

كما قمنا بواسطة الكشاف أكونكورد (aConcorde) بحساب عدد تكرارات الأسماء الموصولة التي تعد نوع من المعارف، لا يتضح معناها، ولا يتعين إلا بجملة تأتي بعدها، وهي أنواع:

- الذي: للمفرد المذكّر، التي: للمفرد المؤنث؛
- اللذان: للمثنى المذكّر، اللتان: للمثنى المؤنث؛
- الذين: للجماعة، اللاتي، اللائي، اللواتي: للجماعة الإناث؛
- مَن: للعاقل، ما: لغير العاقل.

استخلصنا من هذه الدراسة أنّ علم اللغة الحاسوبي من أهمّ مجالات اللسانيات التطبيقية، وفرع من فروع الذكاء الاصطناعي، كما يسعى لمعالجة اللغة ودراستها من وجهة نظر حاسوبية.

ولقد تعددت المصطلحات الدالة عليه، والأكثر تداولاً بين أهل الاختصاص اللغوي هي اللسانيات الحاسوبية، التي تتضمن مكونين رئيسيين ومتكاملين لا يستقل أحدهما عن الآخر هما: المكوّن النظري والمكوّن التطبيقي.

"لسانيات المدونة الحاسوبية" هي أحدث فروع اللسانيات الحاسوبية، والتي مادتها الخام هي المدونة الحاسوبية التي تعتبر مجموعة ضخمة من النصوص مخزنة إلكترونياً، سواء في شكلها المكتوب أو المنطوق، والغاية منها دراسة البيانات اللغوية وتحليلها وليس استخراج البيانات فحسب.

فالمتون اللغوية لها تصنيفات وأهداف ومزايا متعددة خاصة فيما يتعلق بالسرعة والسهولة والدقة؛ التي تسعى من خلالها إلى فتح آفاق واسعة لتطوير البحث اللغوي.

كما يمكن إجراء مجموعة من التطبيقات للبحث والتنقيب داخل المدونة اللغوية، وهذا لوجود مجموعة من الأدوات الحاسوبية لتحليل المتون النصية ومن بين هذه الآليات التي اعتمدنا عليها في الجانب التطبيقي من دراستنا "الكشافات السياقية" وخاصةً "أكونكورد"؛ الذي يقوم بتنظيم مكونات المدونة لمعاينتها ودراستها وضبطها في شكل معطيات إحصائية، الذي يعتبر أحد التحديات الجديدة والهامة في مجال البحث في لسانيات المدونة اللغوية؛ وكذلك "سكتش أنجايين"، ولذا تعتبر "الكشافات السياقية" أداة متميزة لتحليل النصوص العربية، وخطوة جدّ سهلة وبسيطة وبإمكان كل باحث لغوي استخدامها والتحكّم في وظائفها، وذلك باتّباع مراحل تطبيقها بتمعن ودقّة. كما تسمح هذه الأدوات الحاسوبية

خاتمة

وتطبيقاتها على المدونات بالبحث في المستويات الأربعة للتحليل اللساني (الصوتية، المعجمية، الصرفية، النحوية)؛ وبالتالي فهي تخدم اللغة العربية عامةً.

ومن بين نتائج تطبيق هذه الأدوات الحاسوبية على مدونتنا، التأكيد على حقيقة شيوع صوت الهمزة في الكلام العربي، عكس صوت الظاء، أمّا بخصوص الـدايفونات (ثنائيات الأصوات)، فبتطبيقنا لنفس الأدوات الحاسوبية (Skecth Engine) فنجد تتابع صوتي الهمزة + اللام هي الأكثر تكراراً، وعكسها الـدايفونات غير موجودة في الكلام العربي مثل الظاء + الثاء (وكل الـدايفونات التي في خانتها قيمة 0 في المصفوفة السابقة).

أمّا فيما يخص أسماء الإشارة فنرى أنّ "هذا" هي الأكثر تداولاً بينما الأسماء الموصولة "من" حققت أعلى نسبة، وذلك بتطبيق الكشاف السياقي "aConcorde" على مدونتنا.

كان يمكن لنا أن نتحصّل على نتائج عديدة لولا ضيق الوقت، فاختصرنا الدراسة التطبيقية على هذه الأمثلة.

وفي الأخير، فإنّ لسانيات المدونات الحاسوبية تساهم في تقدّم البحث اللساني، وأصبحت أداة أساسية في تطوير البرمجيات اللغوية وبالتالي ترقية الدراسات اللغوية.

قائمة المصطلحات العربية/الانجليزية

linguistics	اللّسانيّات
Computational linguistics-cl	اللّسانيّات الحاسوبية
Artificial Intelligence	الدّكاء الاصطناعي
Natural language processing	معالجة اللّغة الطبيعيّة آليا
Language playing	التّخطيط اللّغوي
Computer –assisted-language learning	تعلّم اللّغة بالحاسوب
Machine translation	التّرجمة الآليّة
Automatic correction-check	التّدقيق الآلي
Automatic spelling correction	التّصحیح الإملائي
Automatic Grammar check	التّصحیح النّحوي
Automatic Diacritization	التّشكيل الآلي
Automatic Summarization	التّلخيص الآلي
Optical character recognition	التّعرف الآلي على الخطّ المكتوب
Speech recognition	التّعرف الآلي على الكلام
Speech Synthesis	التّوليد الآلي للكلام
Automatic lexicon generation	التّوليد الآلي للمعجم
Logic	المنطق

General Linguistic	علم اللّغة التّظري
Applied Linguistic	علم اللّغة التّطبيقي
Modelization	التّمذجة
Corpus Linguistic	المدوّنة اللّغويّة
Raw Corpora	متون نصيّة خام
Annotated corpora	متون نصيّة موسومة

قائمة المصطلحات العربية/الانجليزية

Tagged corpora	المدونات النصية المعلّمة
Pattem Key Word	مطابقة القالب الشكلي
Monolingual Corpus	المدونة أحادية اللغة
Bilingual Corpus	المدونة ثنائية اللغة
Multilingual Corpus	المدونة متعددة اللغات
Parallel Corpora	مدونة متوازنة اللغات
General Corpora	المدونات العامة
Specialised Corpora	المدونات الخاصة
Written Corpora	مدونات اللغة المكتوبة
Spoken Corpora	مدونات اللغة المنطوقة
Original Corpora	المدونة الأصلية
Free Corpora	المدونات الحرة
Pesigning The Corpus	تصميم المدونة
Optaining The Texts	تجميع النصوص
Text Mode	طبيعة النص
Text Sourse	مصدر النص
Corpora Sourse	مصدر النصوص
Document Information	بيانات الوثيقة
Makup	التنقيح
Annotation	التوسيم
Text Domain	حقل النص
Text genre	جنس النص
Spezialization level	مستوى التخصص
File format	شكل الملف

قائمة المصطلحات العربيّة/الانجليزيّة

coocordancers	الكشّافات السيّاقية
scanner	الماسح الضوئي

قائمة المصادر والمراجع

- القرآن الكريم ، رواية ورش عن نافع.

المعاجم :

1- إبراهيم أنس - عبد الحلیم منتصر - عطية الصوالحي - محمد خلف الله أحمد - دار الفكر - ط2 - ج1

2- ابن منظور - لسان العرب - دار الطباعة للنشر - بيروت - لبنان - م 5 - ط4 - 2005

المراجع :

3- أمنا أمينة جارب - فاطمة الزهراء سعداوي - المدونات التعليمية ودورها في تنمية التحصيل العلمي لدى

جامعة قاصدي مرياح ورقلة - مذكرة ماستر - جامعة قاصدي مرياح ورقلة - 2015

4- إيمان صبحي سلمان دلول - معجم محوسب لمعاني الأفعال الثلاثية المجزدة في اللغة العربية - رسالة ماجستير -

كلية الآداب - قسم اللغة العربية - الجامعة الإسلامية - غزة - 2014

5- جاسم محمود الحسون - حسن جعفر الخليفة - طرق تعليم اللغة العربية في التعليم العام - جامعة عمر المختار

- دار الكتب الوطنية - بنغازي 1996.

6- حسام البهنساوي - علم الأصوات - مكتبة الثقافة الدينية - القاهرة - 2004

7- سمير شريف استيتية - اللسانيات المجال والوظيفة والمنهج - عالم الكتب الحديث - 2005.

8- عثمان بن سعيد الداني أبو عمرو الأندلسي - الفرق بين الضاد والظاء في كتاب الله عز وجل وفي المشهور من

الكلام - تحقيق غانم قدوري - دار عمار الأردن 2010 - ص 15.

9- عصام الدين أبو زلال - مقدمة في علم اللغة الحاسوبية - دار الوفاء لدنيا الطباعة والنشر - الإسكندرية -

2016.

10- فداء الجندي - العرب والعربية في عهد الثورة الحاسوبية - دار الفكر - دمشق 2003

11- فاطمة الزهراء توتواوي - استثمار المدونة في تدريب المترجم - مذكرة ماجستير - كلية الآداب واللغات والفنون

- قسم الترجمة جامعة وهران - السانية - 2009

12- كمال لعناني - دور لسانيات المدونة الحاسوبية في ترقية ونشر المصطلح - تيزي وزو pdf

13- محمد مكنز عبود - بناء مكنز عربي في مجال علم اللغة (المنهج والتطبيق) الدار المصرية اللبنانية - 2004

قائمة المصادر والمراجع

- 14- محمد محمد يونس - مدخل إلى اللسانيات - دار الكتب الوطنية - بنغازي - ليبيا 2004
- 15- مطلق سعد العميري - تأثير المدونات الإلكترونية الكويتية على إتجاهات طلبة قسم الإعلام في جامعة الكويت - رسالة ماجستير - كلية الإعلام - جامعة الشرق الأوسط - 2001
- 16- معالي هاشم علي أبو المعالي - الإتجاه التوافقي بين لسانيات التراث واللسانيات المعاصرة - الدكتور عبد الرحمن الحاج صالح أتمودجا - رسالة ماجستير - جامعة بغداد 2014
- 17- منظمة الأمم المتحدة للتربية والعلوم والثقافة " يونيسكو " تقدم اللسانيات في الأقطار العربية - وقائع ندوة جهوية أبريل 1987 - دار الغرب الإسلامي - 1991 - بيروت - لبنان
- 18- نهي نجم الزامل - المدونات الإلكترونية - مركز التدريب التربوي - السعودية 2013
- 19- وليد العناتي - العربية في اللسانيات التطبيقية - دار كنوز المعرفة للنشر والتوزيع عمان - 2012

المؤتمرات :

20- دراسات ومقالات في المدونات التعليمية : الساعة

<http://malsob537blogposte.comblog/pagehtml01205> 15:30

نتائج المؤتمر الدولي الثاني للتعليم الإلكتروني والتعليم عن بعد - جريدة الرياض

21- دشن القحطاني - المدونات الإلكترونية وحرية الرأي والتعبير - مؤتمر تقنيات الاتصال والتعبير الاجتماعي

جامعة الملك سعود - الرياض - مارس 2009

المجلات :

22- أحمد برماد - المدونات الحاسوبية وصناعة المعجم التاريخي العربي المختص - مجلة الذاكرة - تصدر عن مخبر

التراث اللغوي والأدبي في الجنوب الشرقي الجزائري - العدد التاسع - جوان 2017

23- زكي حسين الوردی - صحافة المدونات الإلكترونية على الأنترنت عرض وتحليل - كلية الإعلام - جامعة

بغداد - ع311- حزيران-2007

قائمة المصادر والمراجع

- 24- المجلة العربية للعلوم الإنسانية - تصدر عن جامعة الكويت -مجلة محكمة - تقدم البحوث الأصلية والدراسات الميدانية والتطبيقية في شتى فروع العلوم الإنسانية والإجتماعية باللغتين العربية والإنجليزية -ع28م-7- خريف 1987- رئيس التحرير : الدكتور عبدالله أحمد المهنا - سكرتير التحرير: محمود بركات
- 25- مجلة تصدر عن الجمعية المغربية للدراسات المعجمية - الرباط -المغرب العدد السابع والثامن - 2009

المراجع الأجنبية:

- 26- NADJA NESSEHOUF-CORPUS LINGUISTICS:Apartical Introduction Octobre - 2005(LasteUpdated Septembre 2011) .

المواقع الإلكترونية :

- 26- ايريك أتويلوعبدلة العفيفي - أبحاث جامعةليدز في مجال لسانيات المدونات العربية 2015-2016 على الرابط :
- E.S.Atwell@leeds.ac.uk / ayjfaifi@imammu.edu.sa.
- 27- سلوى حمادة - المدونات النصية ودور اللغة العربية في التعامل معها -2007
- www-.thonala.com
- 28- محمود إسماعيل صالح - المدونات اللغوية على الرابط :
- mahmoud-ismail-salah-blogpost.com
- 29- معتصم فتحي سليم الحمدان - نموذج محوسب محلل نحوي للجمل الإسمية غير المشكولة في اللغة العربية - رسالة ماجستير - جامعة آل البيت 2002 على الرابط :
- http://www.motasem.net / master-thesis
- 30-- موسوعة ويكيبيديا الحرّة مادة (أنجما والحرب العالمية الثانية) دراسات جيوسراتيجية 17-01-2013 م
- http://edudesgeostrategique.com -17-01-2013enigma
- renseignement- au-cour-de-la-seconde-guerre-mondiale
- 31-- موقع صوت العربية ، أكبر مشروع حضاري للغة العربية على الشبكة العالمية ، إطلاق موقع المدون اللغوية العربية التابع لمدينة الملك عبد العزيز للعلوم والتقنية ، للكاتب عبد العزيز حميد ، على الرابط :
- http://www.voiceof arabic .net / index. Php option.com :content&view.article&id- 1038 :2012-03-08-11-11-53-&catid-159 :2008-08-25-14-57-05-&itemid-327.
- 32-- موقع المدونة العربية لمدينة الملك عبد العزيز للعلوم والتقنية على الرابط :
- http://www.kastac.org.sa/page/ab aout-asp.

فهرس الموضوعات

- إهداء
- شكر وتقدير
- مقدمة.....أ- ج
- مدخل: اللغة العربية تتحدّى التطوّرات التكنولوجية 12-06.....
- الفصل الأول :
- المبحث الأول: اللسانيات الحاسوبية 20-13
- المبحث الثاني: مجالات علم اللغة الحاسوبي ومراحل نشأته وتطوره 23-20
- المبحث الثالث: دواعي حوسبة اللغة العربية وبعض الجهود الرائدة فيها 26- 23
- الفصل الثاني :
- المبحث الأول: المدونة اللغوية وأنواعها 39-29
- المبحث الثاني: أهمية المدونات اللغوية الحاسوبية 44-40
- المبحث الثالث: إنشاء المدونات الحاسوبية و أهمّ ما وجد منها في المواقع الإلكترونية 53- 45
- الفصل الثالث :
- المبحث الأول: الكشّافات السياقية 61-54
- المبحث الثاني: عرض لمدوّنتنا 66-62
- المبحث الثالث: أعمال تطبيقية على المدونة 78-67
- خاتمة
- قائمة المصطلحات العربية / الإنجليزية.
- قائمة المصادر والمراجع.

الملخص:

المدونات (أو الذخائر أو المتون) اللغوية هي نصوص إلكترونية ضخمة جُمعت لغرض معين بناء على معايير محدّدة لكي تكون ممثّلةً للواقع اللغوي. رغم أنّ وجود المدونات اللغوية والأدوات الحاسوبية التي تسهّل استخدامها في الدّراسة اللغوية أو الحوسبة ليس أمراً جديداً، إلا أنّ الجهود العربية الخالصة التي تمّت بخصوص بناء المدونات وأدوات معالجتها لا تزال في بداياتها. والهدف من هذه المذكرة هو تقديم لمحة عامة عن هذا الموضوع.

الكلمات المفتاحية: لسانيات حاسوبية، مدونات لغوية، كشاف سياقي، الواقع اللغوي، متون و ذخائر.

Résumé :

Les copora linguistiques (ou thésaurus) sont des assemblages de textes électroniques compilés dans un but précis sur la base de critères spécifiques pour être représentatifs de la réalité linguistique d'une langue donnée. Bien que ce domaine (corpora et les outils informatiques associés) soit connu depuis plus deux décennies pour les autres langues, la langue arabe accuse un grand retard dans ce domaine, et les efforts déployés pour construire des corpora et leurs outils de traitement n'en sont qu'à leurs débuts. Le but de cette étude est de fournir un aperçu sur ce sujet.

Mots-clés : Linguistique informatique, Corpora, Concordanciers, fouilles, Thésaurus, Langue arabe.

Summary:

Linguistic copora (or thesaurus) are assemblies of electronic texts compiled for a specific purpose on the basis of specific criteria to be representative of the linguistic reality of a given language. Although this domain (corpora and associated computer tools) has been known for more than two decades for other languages, Arabic language is lagging behind in this area, and efforts to build corpora and their processing tools are just starting out. The purpose of this study is to provide an overview on this topic.

Key words: Computer linguistics, Corpora, Concordancers, excavations, Thesaurus, Arabic language.