البرمجة الحاسوبية – تقنيات تعليم وتطوير اللغة العربية

La Programmation Informatique: Les Techniques d'Enseignement et de Développement de la Langue Arabe

الدكتورة: بختة تاحي

banatahi44@gmail.com

جامعة حسيبة بن بوعلي بالشلف – الجزائر

الملخص:

تعدّ البرمجة الحاسوبية من أهم التقنيات لتعليم وتطوير اللغة العربية، حيث شكلت نظرية المعلوماتية، ولاسيما الحوسبة، تحديا معرفيا بالنسبة للغة منذ نضج هذه النظرية النسبي في أربعينيات القرن العشرين، فقد قامت هذه النظرية بتطوير الدرس اللغوي المعاصر، بتعاضدها مع المناهج المعرفية الحديثة.

فقد نهض البحث العلمي الذي أفضى إلى حوسبة اللغة على مميزات متعددة لابد من مراعاتها والأخذ بها مثل العلاقة بين المنطوق والمكتوب، وقد ارتبطت هذه الأخيرة أيضا بالذكاء الاصطناعي وكذلك، تقنيات البرامج الحاسوبية، في معالجة الأخطاء اللغوية بصفة عامة، ونحن في بحثنا هذا سنحاول الحديث عن الأخطاء الإملائية والصرفية بصفة خاصة.

الكلمات المفتاحية: البرمجة الحاسوبية– اللغة العربية– تقنيات– معلوماتية– أخطاء إملائية– صرفية.

Résumé:

La programmation informatique est l'une des techniques les plus importantes pour enseigner et développer la langue arabe. La théorie informatique, en particulier informatique, pose un défi cognitif pour la langue depuis la maturité relative de cette théorie dans les années 1940. Cette théorie a développé une leçon de langue contemporaine, conjointement avec des approches cognitives modernes.

La recherche scientifique qui a conduit à l'informatisation de la langue présente un certain nombre d'avantages, tels que la relation entre l'opérateur et l'écrit, cette dernière étant également associée à l'intelligence artificielle ainsi qu'aux techniques informatiques dans le traitement des erreurs de langage en général. Dans notre recherche, nous tenterons de parler notamment des fautes d'orthographe et morphologiques.

Les mots clés: Programmation informatique - langue arabe - techniques - information - fautes d'orthographe - morphologie.

ارتبط علم اللغة الحاسوبي أيما ارتباط بالذكاء الاصطناعي، حيث أصبح الحاسوب يحاكي عقل الانسان، من خلال الحوار الشفوي، وسنتبين ذلك من خلال بعض النماذج التي اعتمدت في الذكاء الاصطناعي.

لعبت شبكات الانتقالات دورا هاما في تطوير نماذج الحوار الشفوي بين الانسان والآلة وبالإضافة إلى نماذج Programmar، و Speechis، و Myrtille، التي كانت رائدة في تحليل الخطاب المنطوق وإنتاجه في نموذجين هامين ألا وهما:

- ✓ **نموذج: Lunar**

يعتبر جهاز Lunar (وودس Woods) أول جهاز يعتمد على سؤال وجواب انطلاقا من البحث في قاعدة للمعطيات الخاصة بالمعارف حول الأحجار المستقدمة من القمر ينتج التحليل تمثيلا داخليا على شكل مشجر تركيبي عميق يرتكز على تحليل لساني مؤسس على الأنحاء التحويلية، ولمعالجة الروابط اللغوية، يستعمل ميكانيزما خاصا ينبني على أساليب كشفية خاصة لتفادي اختيار كل التركيبات التي تفرضها عملية البحث، ولا يتدخل المكون الدلالي إلا بعد التحليل التركيبي الذي انجزته شبكة من الانتقالات المعززة، وهو ما يفيد – بالخصوص- في اقصاء التحليلات التركيبية الخالية من المعنى، وفي نموذج أحدث حاول " وودس" Woods تكثيف التفاعل بين التركيب والدلالة مع الحفاظ على استقلاليتهما مبررا ذلك بسببين اثنين:

أولا: إمكانية الحصول بهذه الكيفية على قالب تركيبي مستقل عن أي مجال للتطبيق.

ثانيا: القدرة على تفكيك مشكل التعلم حسب مهمات خاصة، وهو ما يمكن أن ييسر المعالج الأوتوماتية.[1]

- ✓ **نموذج GUS**

يتمثل الهدف من نموذج GUS بوبرو Bobrow وآخرون: في تدبير حوارات واقعية بين الإنسان والآلة عندما تكون موجهة للقيام بمهمة محددة مثلا: حوار بين عامل في شركة للأسفار وزبون يرغب في حجز تذكرة طائرة.

الفكرة المركزية لبرامج معالجة طلبات الزبائن هي القابلية، حيث تطابق قوالب الصيغ الصرفية والتركيبية، والدلالية...، التي تتحقق في استقلال عن بعضها البعض، وهذا لا يعني أنها تنطلق بشكل متتابع، بل تتم مراقبتها بواسطة مذكرة تخزن فيها مختلف المهمات الواجب تنفيذها.[2]

من بين أهم الإجراءات التي يشغلها هذا النموذج ما يلي:

- التحليل المورفولوجي الذي يعالج سلسلة الحروف في المدخل، وينتج مجموعة الإمكانات لكل كلمة، منظمة على شكل جدول.
- التحليل التركيبي الذي يوظف مفهوم شبكات الانتقالات المتكررة والجدول الذي ينتج عن التحليل الموفولوجي.
- التحليل الدلالي، حيث يولد النحو بنية من الأدوار تحول إلى بنية من الحالات.
- الاستنتاج والتفسير الدلالي: حيث تحلل الجملة حسب المهمة، وتركب عناصر للجواب.
- التوليد الذي يترجم عناصر الجواب إلى اللغة الانجليزية المنطوقة.[3]

فإذا كانت علوم الحاسوب وتقنياته استطاعت إيجاد حلول لمشكلات في العلوم الإنسانية فإن المجال اللغوي يعد من المجالات الأساس الأولى التي استفادت من تقنية المعلومات بشكل أكبر إذ تمكن اللغويون والحاسوبيون، جنبا

إلى جنب، من سبر أغوار القدرة اللغوية المخزنة في دماغ الإنسان من خلال الخوارزميات الصورية التي تتحكم في إنتاج النظام اللغوي الطبيعي توليدا وتحليلا.

إن التطور المعرفي والتقني الذي شهده العصر الحديث ساعد على معالجة أنظمة اللغات الطبيعية على المستويات الصرفية والنحوية والدلالية والمعجمية باستخدام علوم الحاسب وتقنياته، هذه المعالجة لم تعد تقف عند حدود الوصف الظاهري للموضوع اللغوي الطبيعي، بل تعدته إلى التفسير، وأحيانا أخرى إلى التعرية عن الإواليات التي تتحكم في عملية الإنتاج اللغوي، وما كان هذا ليتحقق لولا استهلاك اللغويات الحديثة عامة واللغويات الصورية للجهاز المفهومي للعلوم الدقيقة وعلوم البيولوجيا الحيوانية والنباتية وعلم الأعصاب وعلم هندسة الوراثة وهندسة المعرفة والذكاء الاصطناعي ...وغيرها.[4]

فمن بين الاشكاليات التي تقع فيها اللغة العربية نجد، الأخطاء الإملائية وسنتبين ذلك من خلال الفقرات الأتية:

تعد قواعد الإملاء من قضايا اللغة العربية المعاصرة التي تقتضي دراستها ومعالجتها اهتماما بالغا، كي تبقى اللغة العربية حية، تستعمل كتابة وحديثا، وتعلما وتعليما، وتواكب التطور التقني، وتلبي حاجات الناطقين بها، وتكون لغة العلم والمعرفة في جميع المجالات.[5]

- **المعايير اللغوية لتقييم برامج التدقيق الإملائي:**

1- مدى توافر عدة مستويات لعمل المدقق الإملائي:

إذ ينبغي أن تتحقق جليا في تطبيقات التدقيق الإملائي مستويات التدقيق، ومن الممكن اقتراح ثلاثة مستويات قابلة للزيادة على النحو التالي:

أ- التقيد بقواعد الإملاء، ويعني: التقيد بهمزة الألف الابتدائية أو بنقطتي الياء النهائية على سبيل المثال.

ب- المرونة في التقيد بقواعد الإملاء، ويعني التقيد بهمزة الألف الابتدائية في خيارات التدقيق الإملائي (في حالة الكتابة المصرية) على سبيل المثال، وكذلك التقيد بنقطتي الياء النهائية في خيارات التدقيق الإملائي (في حالة الكتابة الشامية) على سبيل المثال.

ت- مراعات الأخطاء الشائعة، ويعني عدم التقيد بهمزة الألف الابتدائية أو بنقطتي الياء النهائية على سبيل المثال.

2- مدى معالجة تطبيقات التدقيق الإملائي لأنماط الأخطاء الإملائية الشائعة في النصوص العربية ومن أهمها:

- الخلط بين أشكال همزتي القطع والوصل (أ، إ، ا، آ) مثل كتابة كلمة (استقبال) بهمزة الوصل هكذا: " إستقبال"، " أستقبال"، " آستقبال".
- الخلط بين الألف المقصورة (ى) والياء المنقوصة (ي)، مثل كتابة كلمة " مرتضى" بياء غير منقوطة هكذا: " مرتضي" بنقطتين.

- الخطأ برسم الهمزات المتوسطة والمتطرفة (ء، ئ، ؤ، أ، آ) مثل كتابة كلمة " شيء" بهمزة على السطر هكذا: " شيئ بهمزة على النبرة.
- الخلط بين التاء المربوطة (ة) والهاء (ه) مثل كتابة كلمة " رمادية" هكذا: " رماديه" بهاء وليست بتاء مربوطة.
- اتصال كلمتين دون مسافة بينهما، مثل كتابة كلمتي " اتصال كلمتين" هكذا: " اتصالكلمتين".
- وجود نقص في حروف الكلمة، مثل كتابة كلمة " المحامي" هكذا: " المحام".
- وجود تكرار أو زيادة في الكلمة، مثل كتابة كلمة " مشكور" هكذا: " مشكوووور".
- وجود قلب بين حروف الكلمة الواحدة، مثل كتابة كلمة " مشكور" هكذا: " مكشور".
- الحروف التي تنطق لهجيا بطريقة مخالفة لكتابتها، فحرف " الضاد" في بعض البلاد العربية ينطقها البعض " ظاء" ومن ثم يكتبونها " ظاء".
- الخلط بين الحروف المتجاورة في لوحة مفاتيح الحاسوب، حيث يخطيء البعض في كتابة بعض الحروف نتيجة لسرعة في الكتابة على لوحة المفاتيح، مثل كتابة كلمة " فاروق" هكذا: " فاروق" وذلك لتجاور حرفي " الفاء" و" القاف" على لوحة المفاتيح، ويخطئون كذلك في الحروف التي تكتب بطريقة مركبة على لوحة المفاتيح مثل كتابة كلمة " إيمان" هكذا: " غيمان" حيث تكتب الهمزة المكسورة عن طريق الضغط على " غ shift+"، وفي حال عدم الضغط على مفتاح " shift" تكتب " غين" على الفور.[6]

- **نجاح عملية التصحيح التلقائي على المستوى الإملائي:**

يطرح المدقق الإملائي بدائل عديدة مقترحة تصحيحا للخطأ الإملائي المكتشف في النص، من خلال نافذة صغيرة تحوي كلمات أخرى مقترحة لتصحيح الخطأ، الموجود في حالة الخطأ الإملائي، وعادة ما ترتب هذه الاقتراحات ترتيبا تنازليا بدءا بأوسعها انتشارا أو انتهاء بالأقل، ويعتمد التصحيح التلقائي على ترتيب هذه البدائل ترتيبا صحيحا، إذ يختار المصحح التلقائي عند تفعيله من قبل المستخدم البديل الأول، خيارا وحيدا للتصحيح في كل المواضع، ومن ثم فإن نجاح المصحح التلقائي في أداء مهمته يعتمد في الأساس على الدقة في ترتيب البدائل، ولذا يمكن اعتبار مدى نجاح المدقق الإملائي في التصحيح التلقائي من معايير تقييم هذا المدقق.[7]

- **برمجة اللغة العربية في الحاسوب:**

إن التطور السريع والمتنوع الذي تعرفه المعرفة في جميع ميادينها المختلفة يستوجب ضبطها واستعمالها والاستفادة منها عن طريق أجهزة ووسائل عصرية متقدمة تقنية وعلمية مثل الحاسوب الذي يعتبر ذاكرة العصر التي تختزن ما عجزت عنه ذاكرة الإنسان، وإذا كانت دراسة اللغة تستوجب استخدام منهج لساني معين، فإن تخزينه في ذاكرة الإنسان يبدو أمرا صعبا مما يفرض الاستعانة بالحاسوب لمزيد من سرعة العمل العلمي وتحقيق المنهجية والموضوعية وذلك من خلال مستويات (الدلالي، النحوي، الصرفي، المعجمي...).[8]

فقد شملت هذه المعالجة أيضا المستوى الصرفي، في تحليله من جميع الجوانب سواء داخل السياق أو خارجه.

حيث يتناول البحث اللغوي في هذا المستوى الكلمة خارج التركيب فيدرس صيغ الكلمات من حيث بناءها والتغيرات التي تطرأ عليها من نقص أو زيادة وأثر ذلك في المعنى والبحث اللغوي الحديث يتعامل مع المسائل الصرف على أساس صوتي بدلا من اعتماد القدماء على الكتابة اعتبرها القدماء كلمة، في حين يتعامل البحث اللغوي الحديث مع الوحدة الصرفية Morpheme.

أيضا من الحقائق التي أنجزها علم اللغة الحديث تحليل الأنماط الصرفية الخاصة بكل لغة أو لهجة، بقصد الوصول إلى أسلوب كل لغة في بناء كلماتها، كما يهتم علم اللغة الحديث بدراسة التغيرات الصرفية التي تطرأ على بناء الكلمة لاعتبارات صوتية.[9]

ارتبط هذا الأخير بالتكنولوجيا الحديثة ونقصد بذلك ارتباطه بالحاسوب، الذي اصبح الوسيلة الفضلى في معالجة الصرف آليا دون اهدار الجهد، وذلك من خلال معالجته من جميع النواحي وتقسيم الكلمة وتحليل بنيتيها الشكلية والعميقة.

1- معالجة الصرف في الحاسوب:

ذلك باستخدام تقنية الآلات منتهية الحالات، حيث ظهر تطبيق لوصف البنية الصرفية للكلمات يعرف باسم الصرف الثنائي المستوى: Two level MorphologY، ظهر هذا التطبيق للمرة الأولى على يد " كيموكو سكينيمي " الذي استخدمه لتوليد وتحليل الصرف الفكرة الأساسية وراء الصرف ثنائي المستوى هي أن نفكر في الصرف على أساس أنه عملية تحديد ترابط بين التمثيل السطحي والتمثيل العميق للكلمة، يتم ربط احد المستويين، بالآخر باستخدام الآلات منتهية الحالات، الحزمة المعجمية تتكون من مجموعة للوحدات الصرفية (المورفيمات) التي تظهر في المعجم، الحزمة السطحية، تقابل الصيغة (البنية) التي تظهر عند استخدامها للغة فعليا.[10]

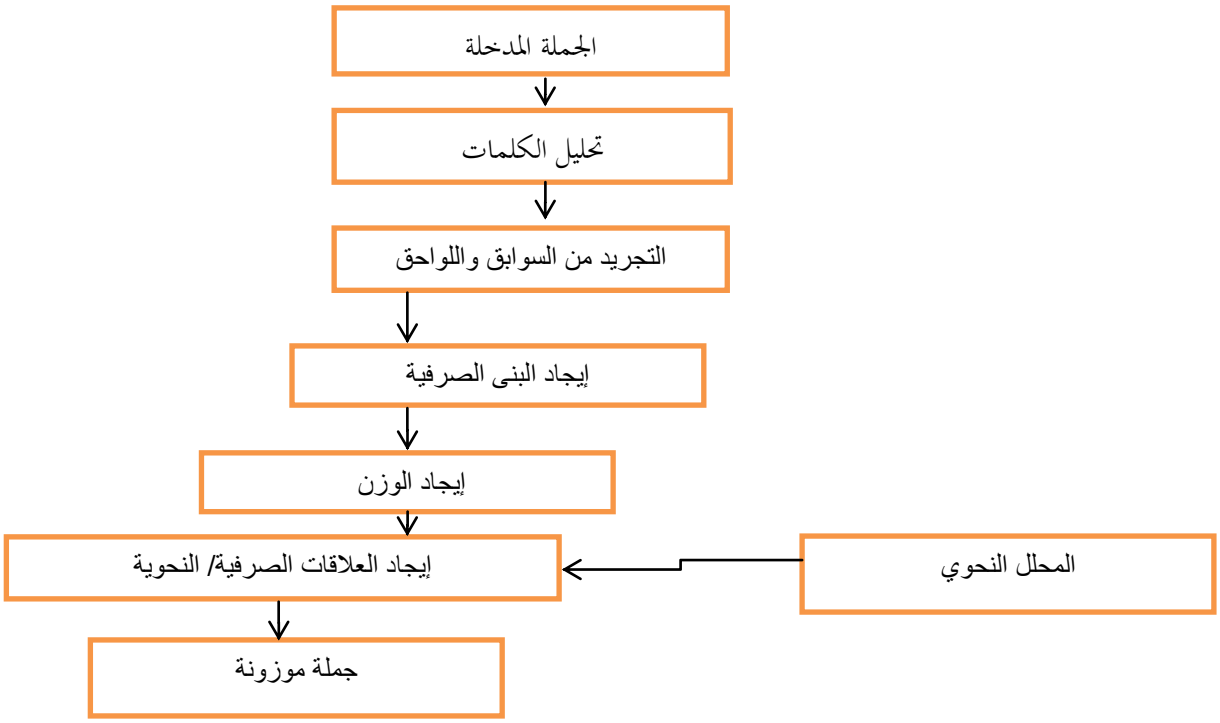
فقد سعى الحاسوبيين إلى إنشاء برامج ومحللات تخدم اللغة العربية ومن بين هذه المحللات ما نتحدث نحن اليوم عنه هو المحلل الصرفي والذي يخدم بدوره بحثنا هذا.

2- تصميم المحلل الصرفي:

يقوم المحلل الصرفي بتحليل كلمات الجملة المدخلة وتحديد بياناتها الصرفية، وذلك بعد إجراء عدة خطوات.

- تجريد الكلمة من السوابق واللواحق مثل "ال" التعريف والضمائر المتصلة وغير ذلك.
- إيجاد الوزن المقابل لكل كلمة من كلمات الجملة، وذلك من خلال بنى المعطيات.

إيجاد العلاقات الصرفية النحوية بين مركبات الجملة وذلك طبقا للقواعد النحوية التي تحدد الموقع الصرفي لكل وزن، ويبين المخطط هذه الأوزان من خلال المحلل الصرفي[11].



تم تصميم بنى المعطيات للمحلل الصرفي التي تضمنت أوزان الأسماء والأفعال في اللغة العربية، حيث تم تمثيل المعلومات والقواعد الصرفية لهذه الأوزان بشكل معلنات، وتم وضع التعابير الإعلانية ذات السمات المشتركة ضمن هياكل SLOTS.[12]

- **المعايير اللغوية لتعامل محرك البحث مع كلمات البحث على المستوى الصرفي:**
  - ✓ مراعاة التدقيق اللغوي للمدخلات (input).
  - ✓ مراعاة الأشكال الكتابية.
  - ✓ مراعاة أشكال الكلمات المعربة.
  - ✓ مراعاة الأشكال اللهجية للكلمات العربية.
  - ✓ مراعاة أشكال لغة الفرانكو آراب.
  - ✓ مراعاة أشكال الكلمات المركبة من كلمتين بينهما مسافة.
  - ✓ مراعاة أشكال الكلمات التي تحوي حروفا تنطق ولا تكتب.
  - ✓ مراعاة أشكال زوائد الكلمة العربية (من السوابق واللواحق) (Affixes- Suffixs- Prefixes).

ينبغي لمحرك البحث أن يدقق كلمات المستخدم لغويا، تحسبا وتداركا لخطأ المستخدم ومن ثم ينبّه المستخدم إلى الخطأ ويقترح الكلمة الصحيحة لكلمة البحث الخطأ، ويقترح عرض النتائج الخاصة بالكلمة الصحيحة، وهو ما نلحظه في محرك البحث الشهير " جوجل" لكنه وإن اقترح الصورة الصحيحة للكلمة، لا يهمل الصور الخطأ كتابيا، فإن حصر المحرك بحثه على الكلمات الصحيحة فقط يقلّل من نتائج البحث ويتسبب في فقد الكثير من النتائج، فليس كل مستخدم عربي

على علم بقواعد الكتابة والإملاء اللغوية، وهو ما ينقلنا إلى ضرورة مراعاة الأخطاء الإملائية الشائعة (the common spelling errors).[13]

- **اللبس الصرفي على مستوى كلمات البحث**: نأخذ مثالا عن ذلك لكلمة " الدين"

الكلمة المشكولة	السابق	الجذع	نوع الكلمة	الوزن	الجذر	قسم الكلم	اللاحق
Volweled word	prefix	stem	Type	pattern	Root	postags	Suffix
الدّين	أل: التعريف	دين	اسم جامد	فَعْلُ	د ي ن	مفرد مذكر مرفوع في حالة التعريف	--
الدّين	أل: التعريف	دين	اسم جامد	فِعلُ	د ي ن	مفرد مذكر مرفوع في حالة التعريف	--
الدّين	أل: التعريف	دين	اسم جامد	فَعْلَ	د ي ن	مفرد مذكر منصوب في حالة التعريف	--
الدّين	أل: التعريف	دين	اسم جامد	فِعلَ	د ي ن	مفرد مذكر منصوب في حالة التعريف	--
الدّين	أل: التعريف	دين	اسم جامد	فَعْلِ	د ي ن	مفرد مذكر مجرور في حالة التعريف	--

*اللبس الصرفي على مستوى كلمة " الدين" في حالاتها الثلاثة النصب- الرفع- الجر.[14]

أهم النتائج التي استخرجناها من بحثنا في مجال حوسبة اللغة العربية:

- سهولة تعليم وتعلّم اللغة العربية عبر مستوياتها اللغوية (نحوية، صرفية، دلالية، معجمية، تركيبية،...إلخ).
- استثمار المعلومات وتوظيفها في معالجة الأخطاء اللغوية الشائعة.
- توفير الجهد والوقت للطلبة الباحثين والاساتذة وكذلك مختلف الفئات المثقفة.
- إمكانية تطوير اللغة العربية، بأحدث التقنيات لجعلها مواكبة للعصر.
- تزويد برامج تطوير اللغة العربية بمحركات بحث سهلة وسريعة لاستخراج المعلومات.
- إمكانية معالجة النصوص العربية بطريقة آلية دقيقة وسريعة، مع سهولة تصحيح الأخطاء اللغوية بشتى أنواعها.
- تزويد البرامج المحوسبة للغة العربية بمصححات تلقائية حتى تمكّن الباحث من توفير الوقت وتجنب الأخطاء التي من شأنها أن تنقص من قيمة البحث وجودته.

المصادر والمراجع:

1. بن عيسى زغيوش، مصطفى بوعناني، المعالجة الآلية للغة مسارات تحليل الجمل العربية باعتماد نماذج شبكات الانتقال، مجلة كلية الآداب والعلوم الانسانية، ظهر المهراز، فاس، المغرب، العدد 14، 2006م.
2. عمر مهديوي: اللسانيات الحاسوبية واللغة العربية إشكالات وحلول، دار كنوز المعرفة للنشر والتوزيع، عمان، ط1، 2018م.
3. محمد لهلال، عماد فخر الدين: تكنولوجيا تعليم اللغة العربية عن بعد لمعالجة الأخطاء االاملائية.
4. عمرو جمعة: تقنيات اللغة العربية الحاسوبية معايير التقييم ورؤى التطوير – دراسات لغوية حاسوبية في المدققات الإملائية والنحوية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، ط1، 2016م.
5. محمد الحناش: استخدام اللغة في تقنية المعلومات، مجلة التواصل اللساني، مجلد 01، منتدى سور الأزبكية، فاس، المغرب، طبعة 01، 1413هـ- 1993م.
6. محمد محمد داود: العربية وعلم اللغة الحديث، دار غريب للطباعة والنشر والتوزيع، القاهرة، دط، 2001م.
7. Al- najem , salah, an exploration of comptutional arabic morphology, APhd thesis, exex: university of exex, 1998, p 03.
8. كريم بوزوبع وعبد الفتاح حمداني: وقائع الندوة الثالثة الدولية حول المعالجة الآلية للغة العربية، المدرسة المحمدية للمهندسين، الرباط، المغرب، 2009م.

الهوامش:

[1]- ينظر: بن عيسى زغيوش، مصطفى بوعناني، المعالجة الآلية للغة مسارات تحليل الجمل العربية باعتماد نماذج شبكات الانتقال، مجلة كلية الآداب والعلوم الانسانية، ظهر المهراز، فاس، المغرب، العدد 14، 2006م، ص 66.

[2]- ينظر : نفسه، ص 67.

[3]- نفسه: ص 68.

[4]- عمر مهديوي: اللسانيات الحاسوبية واللغة العربية إشكالات وحلول، دار كنوز المعرفة للنشر والتوزيع، عمان، ط1، 2018م، ص 20.

[5]- نفسه، محمد لهلال، عماد فخر الدين: تكنولوجيا تعليم اللغة العربية عن بعد لمعالجة الأخطاء االاملائية، ص 133.

[6]- عمرو جمعة: تقنيات اللغة العربية الحاسوبية معايير التقييم ورؤى التطوير – دراسات لغوية حاسوبية في المدققات الإملائية والنحوية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، ط1، 2016م، ص 10-11.

[7]- نفسه، ص 14.

[8]- محمد الحناش: استخدام اللغة في تقنية المعلومات، مجلة التواصل اللساني، مجلد 01، منتدى سور الأزبكية، فاس، المغرب، طبعة 01، 1413هـ- 1993م، ص 31.

[9]- محمد محمد داود: العربية وعلم اللغة الحديث، دار غريب للطباعة والنشر والتوزيع، القاهرة، دط، 2001م، ص 106.

[10]- Al- najem , salah, an exploration of comptutional arabic morphology, APhd thesis, exex: university of exex, 1998, p 03.

[11]- كريم بوزوبع وعبد الفتاح حمداني: وقائع الندوة الثالثة الدولية حول المعالجة الآلية للغة العربية، المدرسة المحمدية للمهندسين، الرباط، المغرب، 2009م، ص 19.

[12]- ينظر : المرجع نفسه، ص 18.

[13]- عمرو جمعة: تقنيات اللغة العربية الحاسوبية معايير التقييم ورؤى التطوير – دراسات لغوية حاسوبية في المدققات الإملائية والنحوية، ص 17-18.

[14]- نفسه، ص 24-25.