# إزالة اللّبس عن بعض مسائل حوسبة الزّوائد العربية *

# Removing Ambiguity in some issues of Computing Arabic Affixes

**أ. أسماء عكاشة**
كلية الآداب والإنسانيات منوبة (تونس)
Asma.akacha6@gmail.com

**ملخص:**
تعتبر المعالجة الآلية للّغات الطّبيعيّة من الدّراسات البينيّة الحديثة التي ما يزال جهازها المفاهيميّ فوضويّا وغير متّفق على الكثير من مصطلحاته. ناهيك عن أنّ المشتغلين فيها ينتمون إلى خلفيّات معرفية مختلفة، فإن غاب التعاون بينهما نشأت العديد من مواطن اللّبس والغلط مثل التّشتت الوسميّ أي تعدّد المصطلحات لنفس المسمى، وهو ما نعتبره من ألحّ الموضوعات التي يجب الفصل فيها لأنّ وسم الأشياء جوهر معرفتها وإدراكها. ولا يقتصر اللّبس على المصطلحات فحسب، بل يمكن أن يمتدّ إلى الآليّات والكيفيّات المتعلّقة بحوسبة الظاهرة اللّغويّة. لذلك اهتممنا بفكّ اللّبس عن عدة مسائل في حوسبة الزّوائد العربيّة، وهو مبحث يطرح عدّة إشكاليّات، يُعدّ توضيحها تدقيقا وتسهيلا لمعالجة اللّسان العربيّ بشكل كبير.

**الكلمات المفاتيح:** زوائد عربيّة / لبس / لسانيات حاسوبية / معالجة آلية للّغات الطّبيعيّة / مستوى صرفيّ /تجذيع/ ألسن سامِيّة / وسم.

**Abstract:**
Automated natural language processing is considered one of the most modern interdisciplinary studies whose conceptual apparatus remains chaotic and disagreeable to many of its lexis. This is mainly due to the fact that the operators belong to different cognitive backgrounds. Hence, the absence of cooperation between them has created many areas of confusion and error, such as the toxic dispersion of the same terminology, which refers to the multiple terms used for the same concept, which we consider to be one of the most pressing subjects to be adjudicated because the marking of objects is the essence of their knowledge and perception. The confusion is not limited to terminology, but can also

*

| تاريخ استلام البحث: 2024/02/23 | تاريخ قبول البحث: 06/ 04 /2024 | تاريخ النشر:2024/05/13 |
|---|---|---|

extend to the mechanisms and processes related to computational linguistics. We have therefore been interested in unblocking several issues in the computerization of Arabic excess, a research that poses several problems, Therefore, we have focused on unraveling the confusion surrounding several issues in the computation of Arabic diacritics, a topic that raises several problematic issues which is considered to be a scrutiny and a facilitation for treating the Arabic tongue significantly.

**Keywords**: Arabic diacritics / confusion / computational linguistics / natural language processing / morphological level / derivation / Semitic languages / labeling.

تعدّ مسألة الزّيادة في الكلمة العربيّة من أهم المباحث اللّغوية المطروحة في المعالجة الآلية للّغات الطبيعيّة، توطئة للبحث في ذلك نشرح بداية المقصود بمفهوم اللّبس، ثمّ نعرّف الزّوائد حسب آراء النحاة العرب. فيم بعد نوضّح مختلف المجالات المعرفيّة التي ترتبط بحوسبة الزّوائد العربيّة، ذلك ما يقودنا إلى تفسير ماهيتها لدى الحاسوبيّين، وتبيين كيفيّة التفريق بين المزيد والمجرد آليّا، متطرّقين إلى أشهر تقنيّات المجذعات، ومشيرين إلى مختلف المستويات اللّغوية التي تندرج ضمنها جميع أنواع الزّوائد العربيّة.

نحاول من خلال هذا الطرح توضيح مسائل تتعلّق بمدى فهم المبرمج النّظام الذي تقوم عليه الزّيادة في العربيّة وما تمتاز به من خصائص لغويّة منها المسائل المختلف فيها وما يعتري بعض الآراء من الأخطاء المغفلة التي نسعى إلى تجاوزها بطرح عدّة حلول ناجعة لها في سبيل إزالة اللّبس عنها، وهو ما يساهم في تجويد مخرجات البرامج التي تحلّل الكلمة العربيّة عامة، وتدقيق نتائجها.

## 1- تحديد المفاهيم الأساسيّة وتعريفها:

### 1-1- مفهوم اللّبس وماهيّة الزّوائد العربيّة:

تدلّ كلمة [اللَّبْسُ] بفتح اللام «على معنى الخلط والاختلاط، ومنه لَبس عليه الأمر، إذا اختلط»[1]، ويمكن أن يقود اِلتباس الشيء بشيء آخر إلى صعوبة التفريق بينهما إلى حدّ الوقوع في الخطأ. والشيء الذي نسعى لفكّ اللّبس عنه هو حوسبة الزّوائد العربيّة. فلا يقتصر اللّبس على المسمّيات فحسب، إنّما قد يحدث أيضا عندما تقبل الكلمة نفسها أكثر من تحليل صرفيّ واحد حسب اختلاف السياقات النّصيّة، مثل دلالة الكلمة على أكثر من معنى وظيفي صرفي ومنه دلالة الضّمير غير المشكول المتّصل بالفعل [ـت] على معنى المخاطب أو المتكلّم، والضّمير المتّصل من الزّوائد العربيّة حاسوبيّا لأنّه ملتصق بالكلمة خطيّا وليس من أصولها، بالتّالي يمثّل تحديد الزّوائد ومعرفة أنواعها سبيلا ناجعا لتوضيح غموض معنى الكلمة وإزالة اللّبس عنها.

تجدر الإشارة إلى أنّ كلمة [زَوَائِد] «جمع لغير العاقل»[2]، مفردها [زَائِدَةٌ]، وهي «مشتقّة من فعل [زَادَ] ومصدره [الزِّيَادَةُ]. زاد الشيء: جعله ينمو ويكبر ومنها [يَزِيدُ] و[زِيدًا] وَ[زِيَادَةً] وَ[زِيَّادًا] و[مَزِيدًا]، أي [اِزْدَادَ]»[3]، وتشترك جميعها في جذر [ز ي د].

أمّا في الاصطلاح اللّغوي فتعرّف الزّيادة بأنّها «إلحاقُ الكلمة من الحروف ما ليس منها»[4]، أي ما ليس من أصولها، أي من غير الجذر، والإلحاق بمعنى الإضافة. بالتّالي فإنّ الحروف غير الأصول، هي المكوّنات الزّائدة في الكلمة العربيّة، وهي حروف تزاد في المستوى الصرفي أساسا، اتّفق النّحاة العرب على أنّها «عشرة حروف، وهي: الهمزة، والألف، والهاء، والياء، والنون، والتاء، والسين، والميم، والواو، واللاّم»[5]، يجمعها قولك: «[اليَوْمَ تَنْسَاهُ]، أو [أَتَاهُ سُلَيْمَانُ]، أو [سَأَلْتُمُونِيهَا]، أو [السِّمَانَ هَوَيْتُ]»[6] أو «[لاَ أَنْسَيْتُمُوهُ] (...) وقالوا: [أَسْلَمَنِي وَتَاهَ]»[7]، وأكثرها تداولا هو لفظ [سَأَلْتُمُونِيهَا].

### 1-2- فرضيّة الزّيادة بالحركات في العربيّة:

يعتبر أصلُ حروف الزّيادة في النّحو العربيّ «حروف المدّ واللّين التي هي الواو والياء والألف، وذلك لأنّها أخفُّ الحروف، (...) مأنوسٌ بزيادتها، إذ كلُّ كلمة لا تخلو منها، أو من بعضها، (...) وغيرُ حروف المدّ من حروف الزيادة مشَبَّهٌ بها، ومحمولٌ عليها. (...) فلمّا اجتمع فيها ما ذُكر من شَبَه حروف المدّ واللّين اجتمعت معها في الزّيادة»[8]، بالتّالي تمثّل حروف المدّ واللّين وهي [و] [ي] [ا] أصل الزّوائد العربيّة وما سواها من حروف "سألتمونيها" زوائد فروع لأنّها تشبه الأصل في عدد من الخصائص اللّغويّة الصوتيّة، من ذلك أنّ الهمزة «مُجاورةُ الألف في المَخرج (...) [والميم تشبه] الواو لأنّهما من مَخرج واحد، وهو الشّفة، وفيها غُنَّةٌ تمتدّ إلى الخيشوم، فتناسبت بغُنّتها لين حروف اللّين (...) وأمّا النّون، ففيها أيضًا غنّةٌ، ومخرجُها إذا كانت ساكنة من الخيشوم»[9]، نلاحظ أنّ جميع تلك الخصائص تصف طبيعة نطق الحروف لأنّ تقعيد اللّغة لدى النّحاة الأُوّل يعتمد على منهج السّماع الذي يحتكم إلى ما نقل بالمشافهة عن الأفحاح العرب.

وقد وصف ابن يعيش العلاقة بين حروف المدّ واللّين، والحركات بالاعتماد على منهج آنيّ (Synchronic)، يقول «إنّما رأى النّحويّون صوتا أعظم من صوت، فسمّوا العظيم حرفا والضعيف حركة، وإن كانا في الحقيقة شيئا واحدا»[10]، وفي ذلك إقرار بأنّ الحركات «زوائد لا محالةَ»[11] لأنّ أصل الحركات حروف المدّ واللّين التي هي أصل حروف "سألتمونيها". أمّا ابن جنيّ فقد وصف العلاقة بين الحروف والحركات عامة، فقال «الحركات أبعاض حروف المدّ واللّين»[12]، لعلّه يشير إلى أنّ الحركات أضيفت إلى نظام الكتابة العربيّة في فترة زمنيّة لاحقة، معتمدا في تفسيره على منهج تاريخيّ

(Diachronic). بالتالي يمكن اعتبار التنقيط في الكتابة من الزّوائد أيضا لأنّه وُجد في فترة زمنيّة لاحقة لكتابة الحرف العربيّ.

فإذا تمّ اعتماد الاستدلالين السابقين تكون جميع الحركات القصيرة زوائد. وقد قال في ذلك سيبويه «الفتحة والكسرة والضّمة زوائد، وهنّ يلحقن الحرف ليوصل الى التكلّم به»[13]، يعني أنّ الحركات من مسهّلات نطق الحرف، والإلحاق بمعنى الإضافة أيضا.

وإذا لم يُقَرّ بذلك، فالحركات القصيرة زائدة عندما تتبع الحروف الزّائدة، وهي أصليّة عندما تتبع الحروف الأصليّة، مثل الضمّة الأولى في كلمة [مُحَمَّدُ] التي جذرها [ح م د] تبعت حرف الميم وهو زائد، مع استثناء علامات الإعراب لأنّها تزيد معنويا الوظيفة التركيبيّة للكلمة.

ذلك فيما يخص اختلاف النّحاة العرب في مسألة الزّيادة بالحركات، أمّا الحاسوبيّون فقد اعتادوا حذفها في أغلب برامج معالجة العربيّة، ويُصطلح عليها في مؤلّفاتهم عناصر "التشكيل"(Diacritization)[14]، أو عناصر "التصويت"(Vocalization) [15] وقد اعتبروها مجرّد «علامات صوتية تساعد على قراءة النصوص وفهمها، فالكلمة العربية يمكن أن تكون مشكّلة بالكامل أو شبه مشكّلة أو غير مشكّلة على الإطلاق»[16]. لذلك وُصفت الحركات بأنّها "ضجيج خطيّ" (Orthographic Noise)[17]. بالتّالي يعتمد أكثر الباحثين الحاسوبيّين النّصوص العربيّة غير المشكّلة، إمّا لظنهم بأنّ دور الحركات هامشيّ فلا يُحتاج إليها في عمليّة المعالجة، وإمّا لغاية منهجيّة تجريديّة بالأساس تساهم في التقليل من أبعاد المستند المُعالج.

وفي كلّ الحالات تمثّل «تعرية الحروف من حركاتها عائقا مهمّا في فهم النّص المدخل»[18]. وهو موطنٌ من مواطنِ اللّبس، منه أن يكون لعدد من «الكلمات ذات المعنى المختلف نفس الشّكل (...) لعدم اقتصار الكلمة العربية على معنى واحد محدّد»[19] على سبيل المثال، الكلمة [حسب] إذا أضفنا إليها التشكيل يمكن أن تكون مثلا [حَسِبَ] بمعنى [ظَنَّ] أو [حَسُبَ] بمعنى [شَرُفَ نَسَبُهُ] أو[حَسَبَ] بمعنى [عَدَّ]. لذلك تشكيل الكلمات يساعد كثيرا في فكّ اللّبس الصرفيّ، فتُميّز الحروف عن الحركات الطّويلة بسبب رسمها المطابق للحروف من خلال الأوزان المشكولة، مثل الزّائدة الحرفيّة [+وَ+] في وزن /فَعْوَلَةُ/ التي تختلف عن الزّائدة الحركيّة [+وُ+] في وزن /فَعُولَةٌ/.

تُعتبر الحركات إذن، عناصر لغويّة مهمّة، فهي عمدة الوزن ومكياله. تساعد على إخراج نتائج أكثر دقّة من نتائج المعالجات التي تُهمّش التشكيل. لكنّ الواجب علينا عدم إغفال حقيقة أنّ «معظم النّصوص العربيّة الحديثة تُكتب غير مشكولة»[20]، لعلّ ذلك يجعلنا نضطرّ إلى الاعتماد على مفهوم جديد من الأوزان والصيغ محذوف الحركات، وهو رأي يحتاج في حدّ ذاته إلى مقال آخر وربما كتاب

ضخم، نؤلّفه نحن أو غيرنا. وسنقتصر في هذا المقال على فكّ اللبس عن حوسبة الزّوائد العربيّة الحرفيّة فحسب.

## 2- مظاهر اللّبس في حوسبة الزّوائد العربيّة:

### 2-1- منزلة الزّوائد في مباحث اللّسانيات الحاسوبيّة:

تندرج "الحَوْسَبَةُ" (Computing) عامة ضمن ميدان "الذّكاء الاصطناعي" (Artificial Intelligence)، الذي يُعرّف بكونه «العلم الذي يسعى إلى جعل الآلة تؤدي ما يؤديه البشر من الأعمال بتمكينها من مهارات ذهنيّة ذكيّة لها قدرات الذّكاء التي للذهن البشري، (...) يقوم علم الذّكاء الاصطناعي على ركيزتين هما البرمجيّات الحاسوبيّة والآلة»[21]، تبعا لذلك فإنّ المسألة المبحوث فيها تتنزّل ضمن مجالات علميّة كبرى وصغرى، نعرضها وفق الهرم (1):[22]

(1) حوسبة الزّوائد العربيّة ضمن المباحث العلميّة

يحدّد الهرم (1) مراتب مختلف المباحث المتّصلة بحوسبة الزّوائد العربيّة، ويعتبر الذّكاء الاصطناعي المجال الذي يجمع كلّ العلوم المذكورة تحته. وقد تأسّس ذلك العلم على منهج محاكاة

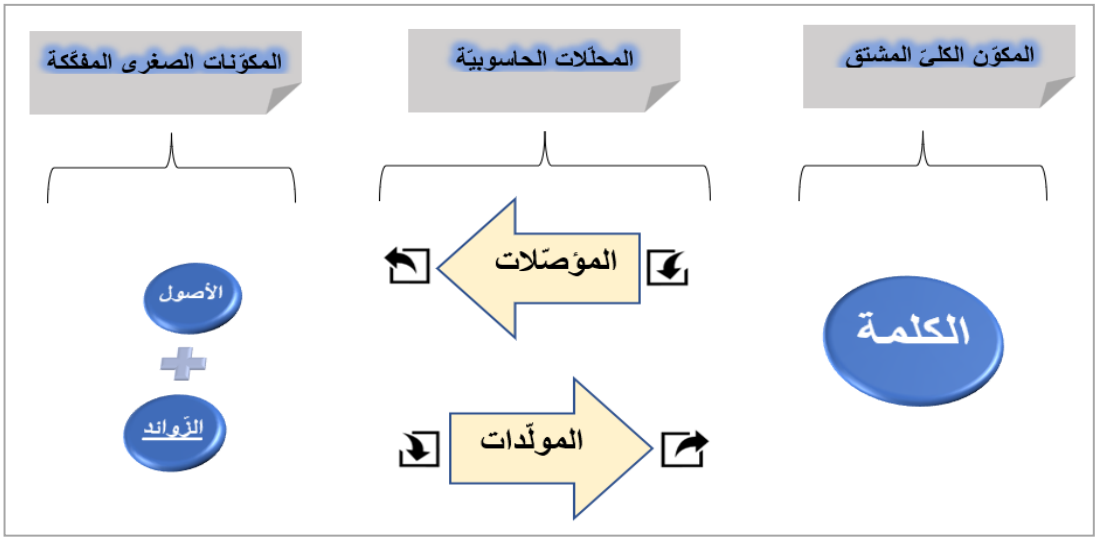
المصنوع الصّانع، أي محاكاة الآلة الإنسان، ويمكن أن نعتبره متفرعا عن علم الحوسبة أيضا التي تتفرّع بدورها من علم الرّياضيّات وأصله علم المنطق.

تصنّف الحوسبة إلى صنفين إثنين، الأوّل (Hardware) يهتمّ بصنع أجزاء الآلة وتركيبها بعضها ببعض، وذلك يحاكي جسم الإنسان. والثّاني (Software) يتعلّق بابتكار البرمجيّات الحاسوبيّة المتحكّمة في أنظمة تشغيل الآلات وإدارتها، وذلك يحاكي عقل الإنسان، ومنها الأنظمة المبرمجة على تحليل ظاهرة "التواصل" (Communication) لدى البشر التي ترتبط بجملة من المعارف أهمّها المعرفة اللّسانيّة اللّغويّة. وقد نشأ عن ذلك الارتباط ظهور علم بينيّ حديث هو علم "اللّسانيات الحاسوبيّة" (Computational Linguistics) يبحث في الطّرق والوسائل التي تجعل الآلة قادرة على تحليل ظواهر الألسن في جميع مستوياتها اللّغويّة.

وللسانيّات الحاسوبيّة جانبان، «الجانب النظريّ يهتم بتوصيف هذا الفرع من اللّسانيات ووضع نظريّات وقواعد تساهم في تحقيق غاياته التي وجد من أجلها. أمّا الجانب العمليّ فيطبّق هذه النّظريّات والقواعد على اللّغة ويربطها بالحاسب ويُظهر مدى اطّرادها وانضباطها ويُسهم أيضا في اكتشاف ثغراتها من أجل تطويرها وتعديلها»[23]. يجمع ذلك الميدان بين معارف اللّسانيّات بوصفها العلم الحديث المهتم بدراسة الألسن الطبيعيّة دراسة علميّة، وتقنيّات الحاسوب المُسهمة في بناء البرامج والنّظم الرّقميّة الخوارزميّة.

إذن يندرج مجال "المعالجة الآليّة للغات الطّبيعيّة" (Natural Language Processing: NLP)، ضمن الجانب التطبيقيّ الذي يهدف إلى تمكين الحاسوب من تحليل الظّواهر اللّغوية «تحليلا يضاهي تحليله الطبيعيّ دقّة ومهارة بالاهتداء الى حدود القطع والكلمات والجمل»[24]، وقد شملت التحليلات مختلف المستويات اللّغوية من أهمّها "المستوى الصرفيّ" (Morphological Level) الذي يحظى بأهميّة كبرى في البحوث الحاسوبيّة العربيّة مقارنة بالمستويات اللّغويّة الأخرى لأنّ الصرف «يعدّ أكثر الجوانب المدروسة في اللّغة العربيّة»[25]، بالتّالي تعتمد المعالجات الحاسوبيّة غالبا منهج «التركيز على وظائف التحليل الصرفي»[26] لأنّه محور دراسة الكلمة العربيّة وتحديد طبيعة مكوّناتها «مثل اللّواصق والسوابق واللّواحق والجذر أو الجذع»[27]، وهي ظواهر لغويّة تُعالج بواسطة "المحلّلات الصرفيّة" (Morphological Analyzers)، نستعين على توضيحها بالشكلين التّاليين:[28]

(2) الهندسة العامة لمختلف المحلّلات الآليّة الصرفيّة




(3) عمليّات التحليل الصرفي في المعالجات الآليّة

| برامج المحلّلات الآليّة الصرفيّة | المدخلات | | العمليّة | المخرجات | |
|---|---|---|---|---|---|
| | عناصرها | طبيعتها | | عناصرها | طبيعتها |
| التجذيع | الكلمة | مكوّن كلّي | تفكيك | أصول + زوائد | مكوّنات صغرى |
| التّولّد | أصول + زوائد | مكوّنات صغرى | تجميعٌ | الكلمة | مكوّن كلّي |

يشير الشكل (2) والجدول (3) إلى أنواع البرامج الحاسوبيّة المهتمّة بحوسبة الزّوائد العربيّة وغير العربيّة. منها ما يختصّ بـ"التجذيع" (Stemming) وفيه عمليّة "التقطيع" (Segmentation) أي فصل أصول الكلمة عن زوائدها التي تُحذف ليُستخرج الجذع، فتكون الزّوائد في هذه الحالة النّتيجة المحصّلة من عمليّة المعالجة وتسمى "مخرجات" (Outputs)، ومنها ما يختصّ بـ "المولّدات الصرفيّة" (Morphological Generators) وهي تفيد في تجميع الكلمة بالانطلاق من عناصرها المفكّكة التي من بينها الزّوائد وتسمى في هذه الحالة "مدخلات" (Inputs)، مثل الانطلاق من الجذر [س م و] للحصول على كلمة [السَّمَاوَاتُ] بإضافة الزّوائد [اَلْ-] و[+ـَا+] و[+ـَاتُ].

تعتبر الزّوائد إذن، عناصر مهمّة لا يمكن الاستغناءُ عنها في برامج معالجة الألسن، ففي كلّ المحلّلات الصرفيّة لا بدّ من المرور بمرحلة حوسبة الزّوائد من خلال تحديدها وبيان خصائصها، ثمّ

معالجتها سواء بفصلها عن الحروف الأصليّة للكلمة في العمليّة الأولى أو بدمجها مع الأصول في العمليّة الثّانية.

### 2-2- الخلط بين الأصول والزّوائد عند الحاسوبيّين:

تتنوّع المصطلحات التي تسم الحروف الأصليّة في البحوث اللّغويّة والحاسوبيّة على حدّ سواء، ويمكن حصرها في مسميين أساسيين هما "الجذر" الذي يقابله المصطلح (Root) و"الجذع" الذي يقابله المصطلح (Stem)، مع العلم أنهما لا يمثّلان نفس المفهوم لوجود عدّة اختلافات بينهما، منها أنّ الجذر «أصل افتراضيّ لأنه ليس من المفردات القائمة في الاستعمال، (...) ذو تكوّن صوتي صامِتيّ خالص لأن مكوناته هي الصوامت (Consonnes) دون الصوائت (Vowels)»[29] وتُكتب منفصلة بعضها عن بعض مثل [ف ع ل]. أمّا الجذع فهو «المفردة القائمة في الاستعمال، المتولّدة إما تولّدا لغويّا صوتيّا من أصل افتراضيّ هو الجذر، وإمّا تولّدا اشتقاقيّا من جذع آخر هو بدوره مفردة مستعملة، (...) ذو تكوّن صوتيّ صامتيّ وصائتي معا (Vocalique) لأنه متكوّن من الصوامت والصوائت»[30] وتكتب مكوّناته متصلة بعضها ببعض مثل [فَعَلَ]. مع العلم أنّ الجذع في أغلب المعالجات الحاسوبيّة يمثّل العنصر اللّغويّ المتبقّي من الكلمة دون حذف الدّواخل.

تعتبر الزّوائد المكوِّن الفارق لاتّصاف كلمة بـ«"مجرّدة" (Unaugmented) أو "مزيدة"(Augmented)»[31] وهي تمثّل معا "صنف المشتقّات"(Derivative Class)[32]. والمزيد «ما دخلته أحرف الزّيادة، عكسه مجرَّد»[33]، نفرّق بينهما بوجود حروف "سألتمونيها" وفق النّحاة العرب. أمّا في المؤلّفات المهتمّة بحوسبة اللّسان العربي، فنجد أنّ الإشارة إلى حروف "سألتمونيها" يكاد يكون منعدما، ومن المقالات القليلة التي ذكرتها لفظا مقال «توظيف قواعد النّحو والصرف في بناء محلل صرفي للغة العربية»[34]، وردت فيه إمكانيّة تحديد طبيعة الحروف العربيّة بتقسيمها إلى مجموعتين، نجد في «المجموعة الأولى حروف [سَأَلْتُمُونِيهَا]، وفي المجموعة الثانية بقيّة الحروف التي لا تنتمي إلى الأولى»[35]، غير أنّ تلك الدّراسة وغيرها من المقالات الحاسوبيّة تخلو من الإشارة إلى أنّ تحديد زّوائد "سألتمونيها" لا يكون بتعدادها في مجموعات أو لوائح، إنّما تزاد حسب الوزن الذي تصاغ وفقه الكلمة العربيّة. يعتبر هذا اللّبس الجسيم من أهمّ المؤاخذات التي يقع فيها الحاسوبيّون، وهو ما نجده مثلا في المعالجة (4):[36]

| Word | Letter(s) | Type of Matched Affix |
|---|---|---|
| فارس | ف | *Prefix* |
| لاعبون | ل | *Prefix* |
| بارد | ب | *Prefix* |
| بنات | ت | *Suffix* |
| متم | تم | *Suffix* |
| القرون | ون | *Suffix* |

(4) أمثلة من الزّوائد حاسوبيّا

نلاحظ في الجدول (4) أنّ معظم الحروف في العمود [Letter(s)] الموسومة بأنّها زوائد، هي في الحقيقة من الأصول، والظّاهر أنّها اُستخرجت بالاعتماد فحسب على المطابقة الخطيّة مع حروف "سألتمونيها". لكنّ الأصل في «معنى كونها زوائد أنّ كلّ حرف وقع زائداً في كلمة فإنّه منها، لا أنّها تقع أبدًا زوائد»[37]. وهي ترد على نحو يكون «موصولا ببنية الكلمة اتّصالاً وثيقًا فيغيّر معناها أو وظيفتها»[38]، فالزّوائد تتّصل ببنية (Structure) الكلمة العربيّة لا بالكلمة نفسها، والبنية أقرب إلى الوزن أو الصّيغة، مثل الزّائدة [مـ+] التي زيدت في وزن /مُفَعَّلٌ/ في كلمة [مُحَمَّد]، وهي تختلف عن الميم الثّانية في نفس الكلمة لأنّها تقابل عين الجذر أي من الحروف الأصليّة للوزن. وللصيغة العربيّة أهميّة كبرى لأنّها «تنفي عن المفردات في تكوّنها من دال ومدلول خاصيّة الاعتباط. فإنّ شكل الدّليل مُظهر لمدلوله ومُعبّر عنه. وإذن فإنّ مدلول المفردة قابل للتّبين من شكلها أي من الدّال»[39]، وقد تُرجم الوزن في الدّراسات الحاسوبية إلى (Template)، أمّا الصّيغة فقد اُصطلح عليها بعدّة تسميات منها "المقولة" (Categories)[40] أو "النّوع" (Type)[41].

بعد توضيح الفرق بين الزّوائد والأصول لابدّ لنا من معرفة جميع أنواع الزّوائد، وذلك بتصنيفها حسب المستويات اللّغويّة التي ينتمي إليها كلّ نوع لنعرف وظيفتها والعناصر الممكن أن تعادلها في الدّلالة النّحويّة، لإنشاء قاعدات بيانات تجمع أنواع الزّوائد وتُهيكل توزيعها على منوال مناسب.

### 2-3- المستويات اللّغويّة للزّوائد:

إنّ المبدأ الأساسيّ الذي تتأسّس عليه برامج معالجة الألسن حاسوبيّا في تحديد الزّوائد العربيّة عند الحاسوبيّين هو اعتبارها كلّ حروف الكلمة التي لا تمثّل أحد أصولها أي مكوّنات جذرها، من ذلك ما نجده في (5)[42]:

(5) أمثلة من السّوابق العربيّة ومعانيها

| Prefix | Meaning | Example |
|---|---|---|
| ب | with, in, by | بالسيارة |
| ك | same as, | كالدخان |
| س | will, | سأذهب |
| و | and | ورجالهم |
| ال | the | النساء |

يتكوّن الجدول (5) من ثلاثة أعمدة، يقدّم الأوّل الكلمات العربيّة المعالجة، ويحدد الثّاني ترجمتها بالإنجليزيّة، أمّا الثّالث ففيه الزّوائد المستخرجة التي تتضمّن حروفا لا تطابق سوابق "سألتمونيها" مثل الكاف والباء. نجد أيضا مجموعة أخرى من الزّوائد المعروضة تكون مطابقة لحروف "سَأَلْتُمُونِيهَا" ولكنّها لا تشترك معها في الدّلالة النحويّة، من ذلك أنّ الواو باعتبارها زائدة صرفيّة لا ترد في بدايات الكلمات مطلقا ومن أوزانها /فَوْعَلَ/ ومنه [حَوْقَلَ]. نهدف إلى إيجاد توصيف ملائم لها في الجدول (6):

(6) مقارنة بعض الزّيادات بالأصول التي تماثلها

| الحروف المستخرجة | الاستعمالات الخاطئة | جذورها | الزّيادات المطابقة خطّيا | الاستعمالات الصّائبة | جذورها |
|---|---|---|---|---|---|
| فَ | فارس | ف ر س | [فَ+]: فاء الاستئناف | فَتَبَسَّمَ | ب س م |
| لَ | لاعبون | ل ع ب | [لَ+]: لام الامر | لَأَفْعَلَنَّ | ف ع ل |
| | | | [لِ+]: لام الملك | لِمَالِكٍ | م ل ك |
| بَ | بارد | ب ر د | [بِ+]: حرف الجر | بِقَلَمٍ | ق ل م |
| تٌ | بنات | ب ن و | [+ـاتٌ]: ضمير جمع المؤنث | شَجِيراتٌ | ش ج ر |
| تُمْ | متمّ | ت م م | [+تُمْ]: ضمير الجمع المذكر | أَكَلْتُمْ | ء ك ل |
| ونَ | القرون | ق ر ن | [+ونَ]: ضمير الجمع المذكر | تَمْرَحُونَ | م رح |

درسنا في الجدول (6) الحروف المعروضة في الشّاهد (4)، فاستخرجنا منها جذورها وما يطابق زوائدها التي تحمل دلالات نحويّة تركيبيّة، مدعومة باستعمالاتها الصّائبة. نلاحظ من خلال ذلك أنّ جميع الحروف المرصودة ليست من حروف "سألتمونيها" التي من شروطها الانتماء إلى البنية

الصرفيّة للكلمة، مثل لام الملك [لِـ+] ووظيفتها في الجملة [حرف جرّ]، ومنها ما ليس مذكورا في الحروف المجموعة في [سَأَلْتُمُونِيهَا] مثل كاف التشبيه [كَـ+].

تشترك جميع الزّوائد التي لا تنتمي إلى "سألتمونيها" في ثلاث خصائص، الأولى اتّصالها خطيّا بالكلمات، والثّانية عدم ورودها من الحروف الأصول، والثّالثة حملها وظيفة تركيبيّة مستقلّة عن بقيّة الكلمة الملتصقة بها مثل كلمة [فَتَبَسَّمَ] وجذرها (ب س م) التي تستقلّ فيها الفاء بوظيفة الاستئناف بينما لا يمكن أن تستقلّ تاء المضارعة عن الكلمة تركيبيّا. لذلك يمكن أن نسمي مبدئيّا ذلك النوع من الزّوائد التي تتّسم بتلك الخصائص الثّلاثة "الزّوائد التركيبيّة"، وهي تختلف عن الزّوائد الصرفيّة الاشتقاقيّة أي "سألتمونيها" في عدة نواح، نذكر بعضها في الجدول (7):[43]

(7) الفرق بين مستويات الزّيادة اللّغوية: التركيبيّة والاشتقاقيّة

| الخصائص | الزّوائد التركيبيّة | الزّوائد الصرفيّة |
|---|---|---|
| الأصناف | صنفان:<br>- لواصق قبليّة<br>- لواصق بعديّة | ثلاثة أصناف:<br>- سوابق<br>- دواخل<br>- لواحق |
| الموقع | أقرب إلى أطراف الكلمة | أقرب إلى الجذع |
| معيار تحديد الموقع | الرسم الخطيّ للكلمة | حروف الجذر |
| علاقة الزّائدة بالكلمة | علاقة معيّة | علاقة احتواء |
| الوظيفة الإعرابيّة | وظيفة مستقلّة عن الكلمة المتّصلة بها | نفس وظيفة الكلمة المتّصلة بها |
| المستوى اللّغوي | التركيبي | الصرفي |
| طبيعة الزّائد | وحدة من وحدات الجملة | وحدة من وحدات الكلمة |
| ترجمتها (غالبا) | تترجم بكلمة كاملة مرتبطة بكلمة أخرى معنويا | تترجم بتغيّر الكلمة المتّصلة بها |

تُشير محتويات الجدول (7) إلى الفروق التي تميّز الزّوائد التركيبيّة عن "سألتمونيها"، عددنا سبعة فروق بينهما، يمكن من خلالها أن نفصل بين أصناف الزّوائد العربيّة في المعالجة الحاسوبيّة. وقد اتّضح جزء كبير من الوعي بتلك المسألة لدى الحاسوبيّين في الجهاز الاصطلاحيّ للزوائد خاصة، جمعنا في الجدول (8) أبرز المصطلحات التي وُسمت (Tagging) بها أنواع الزّوائد المختلفة في مقالات الحاسوبيّين:[44]

(8) مصطلحات الزّوائد العربيّة المحوسبة

| الزّوائد بالمفهوم | المستوى اللّغوي | المصطلح | الأنواع | الأمثلة | الاستعمالات |
|---|---|---|---|---|---|
| النحويّ (حروف "سألتمونيها") | *المستوى الصرفي الاشتقاقي* | الزّوائد (Affixes) | سوابق (Prefixes) | همزة المضارعة | أَحْسَنُ |
| | | | دواخل (Infixes) | ألف الفاعل | مَالِكُ |
| | | | لواحق (Suffixes) | تاء تأنيث | قَالَتْ |
| الحاسوبيّ (غير حروف "سألتمونيها") | *المستوى التركيبي التصريفي* | اللّواصق (Clitics) | لواصق قبليّة (Proclitics) | حروف الجر | بِجِذْعِ |
| | | | لواصق بعديّة (Enclitics) | الضمائر المتصلة | أَبْصَارِهِنَّ |

في الجدول (8) سُميت الزّوائد التركيبيّة بالمصطلح (Clitics) الذي ترجمناه إلى "اللّواصق"، بينما احتفظنا بمسمى (Affixes) لتسمية الزّوائد الصرفيّة، أي حروف "سألتمونيها"، وهي تنقسم إلى سوابق (Prefixes) ودواخل (Infixes) ولواحق (Suffixes). وقد حدّدت أنواع اللّواصق بوصل المصطلح بالسّابقة الإنجليزية [Pro+] التي ترجمناها بـ [قَبْلِيَّةٌ] في (Proclitics) للدلالة على أنّ اللّاصقة في بداية الكلمة، ووصله بالسّابقة الإنجليزية [En+] التي ترجمناها بـ [بَعْدِيَّةٌ] في (Enclitics) للدلالة على أنّ اللّاصقة في نهاية الكلمة. وقد اخترنا مقابلا للمصطلح الأجنبيّ الأوّل هو "اللّواصق القبلية" والثّاني "اللّواصق البعديّة"، رغم أنّ تلك المصطلحات وجدت في بعض المؤلّفات بمعنى السّوابق واللّواحق أي «الزّوائد الصرفيّة الاشتقاقيّة»[45].

إذن هناك علاقتين تربط الزّوائد بـ"مقولات الكلم" (Class of words)"[46]، أوّلها علاقة التّساوي ونرمز إليها بـ [=]، وهي تدلّ على أنّ الزّائدة عنصر مقوليّ وحدها، وتتعلّق بمقولتين يمكن أن نزيد بهما وهما الضّمائر المتّصلة (Attached Pronoun) بنهايات الكلمات مثل ضمير النسوة الغائبات [+هِنَّ] في [نُحُورِهِنَّ]، وحروف المعاني (Particles) المتّصلة ببدايات الكلمات مثل حرف التسويف [سَ+] مع أفعال المضارعة المرفوعة ومنها كلمة [سَأَنْجَحُ]. أمّا العلاقة الثّانية فهي الاحتواء ونرمز إليها بـ [⊂]، وتتعلّق بالزّوائد الصرفيّة الاشتقاقيّة والسّوابق التصريفيّة فحسب من حروف "سألتمونيها"، باعتبارها لا تستقل بمقولة لوحدها مثل السّابقة [نْ+] في الكلمات التي على وزن /اِنْفَعَلَ/ ومنها [اِنْبَجَسَ] مع العلم بأنّ همزة الوصل في أوّل الوزن زائدة صوتيّة بالأساس جيء بها لعدم الابتداء بساكن.

### 2-4- نظام الزّيادة في البنى اللّسانيّة السلسليّة وغير السلسليّة:

تجدر الإشارة إلى أنّ القليل من الدّراسات الحاسوبيّة لم تُغفل ذكر الدّواخل العربيّة، من ذلك مقال: "قواعد الخوارزميّات القائمة على التجذيع العربيّ" "A rule based Arabic stemming algorithm" الذي ورد فيه نموذج تقطيع الكلمة العربيّة التّالي:[47]

(9) تقطيع كلمة [الطَّالِبَاتُ] إلى أصول وزوائد

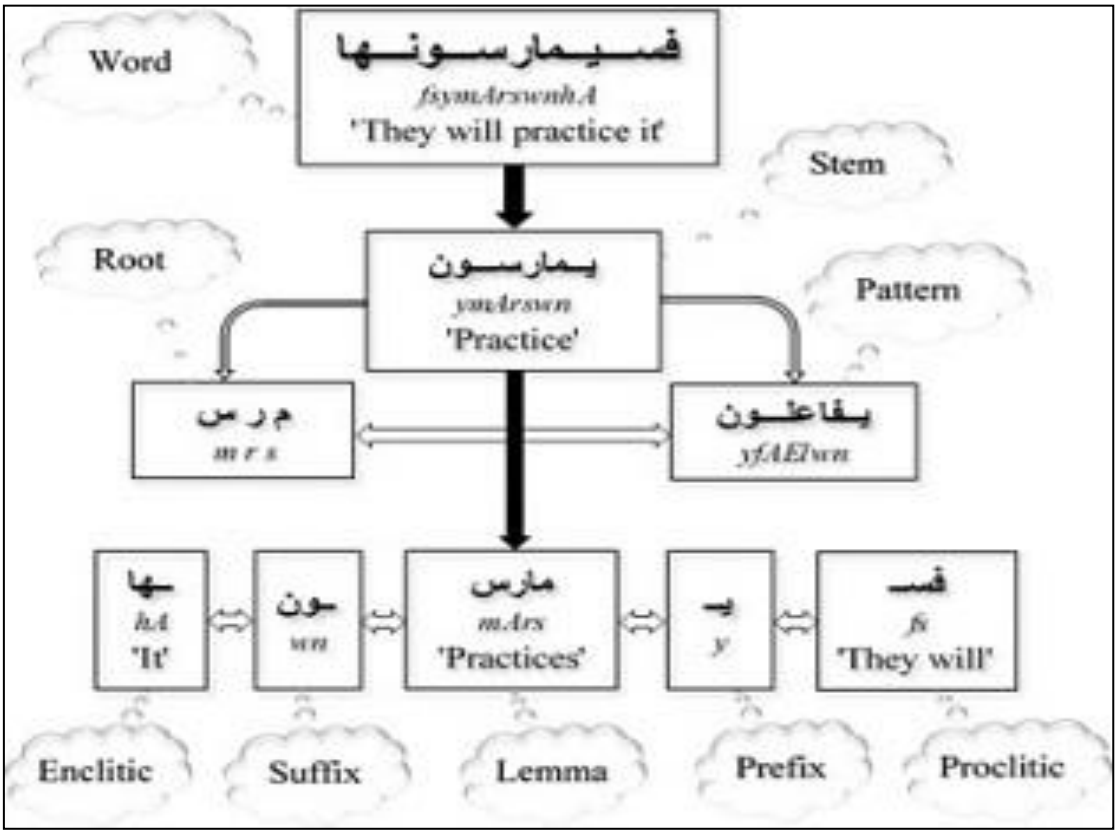
| Word | prefix | suffix | infix | root |
|---|---|---|---|---|
| الطالبات | ال | ات | ا | طلب |

يعرض الشّاهد (9) مكوّنات كلمة [الطّالِبَاتُ] التي تنقسم إلى أصول عُرضت في عمود الجذر، وإلى ثلاثة أصناف من الزّوائد عرضت في الأعمدة التّالية، وهي السّابقة [اَلْ -] والدّاخلة [+ـا+] واللّاحقة [+ـاتُ]. بالتّالي كلمة [طَالِبٌ] مزيدة لأنّ فيها زائدة داخليّة والمجرّد منها [طَلَبَ] وهو جذعها.

وتعتبر صيغ "جمع التّكسير" (Broken Plural)»[48] من أكثر البنى اللّسانيّة المحتوية الدّواخل في العربيّة، وتسمى أيضًا "جمع الأسماء السماعيّة أو الشّاذة" (The Plural Form Of The Irregular Nouns)[49] لأنّها لا تتقيّد بقواعد الجمع المتّفق عليها في "القياس"، إنّما يُعتمد على منهج "السّماع" لمعرفتها، مثل جمع المفردة [دَارٌ] التي تصاغ بتغير حرفها الأوسط [+ـو+] فتكون [دُورٌ] أو تضاف لها الدّاخلة [+ـا+] فتكون [دِيَارٌ]، و«ليس هناك إمكانية الجمع خطيًّا بين جذع [دَارٌ] وزائدة أخرى تدلّ على الجمع دون تغير جذع الكلمة»[50] مع الإشارة إلى أنّ الجذر في الشّاهد يجب أن لا تشكل جميع حروفه.

يفضي بنا ذلك إلى استنتاج مفاده أنّ اختلاف طبائع الأنظمة الصرفيّة يحتّم عدم «اعتماد التقنيات المخصّصة لمعالجة "الألسن الغربيّة"[51] حتى وإن ثبتت فعاليتها في التّعامل مع الثّراء المورفولوجيّ للعديد من اللّغات الطبيعيّة، لأنّ ذلك سيؤدي الى نتائج غير مرغوب فيها»[52]. ومن الملاحظ أنّ أغلب البرامج الحاسوبيّة تعتمد في تقطيعها الكلمة العربيّة على الرسم أدناه:[53]

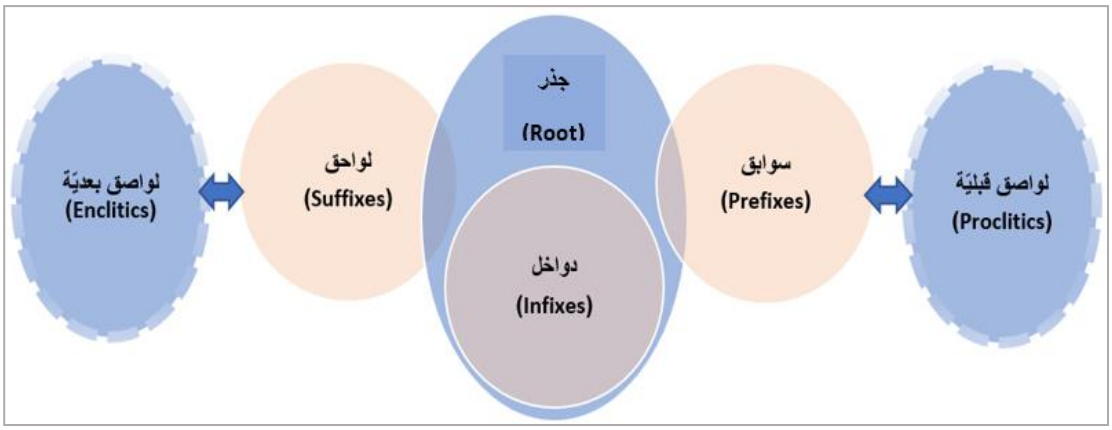
(10) مكوّنات الكلمة العربيّة




يعرض الرّسم (10) تقطيع كلمة [فَسَيُمَارِسُونَهَا] إلى أصول وزوائد، وتلك الكلمة نموذج فيه جميع أنواع الزّوائد العربيّة، إذ تتكوّن من [لواصق قبليّة + سوابق + لمة (مفهوم يشبه الجذع) + لواحق + لواصق بعديّة]، غير أنّ ذلك التمثيل ليس مناسبا لوصف بنية الكلمات العربيّة لأنّه يغيّب مفهوما مهمّا جدّا وهو "الدّواخل" التي تعدّ زوائد ذات نظام "غير سلسليّ" (Non-Concatenative) لأنّها تدخل بين حروف الجذر ولا تحافظ على تسلسل مكوّنات الكلمة، على نحو [يَتَكَاتَبُونَ] التي أُضيفت إليها الدّاخلة [+ـَا+] بين حروف جذرها.

كذلك ننتبه في الرسم (10) إلى مفهوم آخر لا يقلّ أهميّة عمّا سبقه وهو "اللّواصق" (clitics) نسبة إلى الصّفة "الإلصاقيّة" (Agglutinative)[54]. وتعني «اتّصال الجذع بالأشكال النصيّة من اللّواصق»[55]، وتلك في الأصل صفة الألسن الهنديّة الأوروبيّة التي تتكوّن كلماتها بالانطلاق من الجذع خلافا للكلمات العربيّة التي ينطلق فيها من الجذر مع الاحتكام للوزن. على ذلك النحو تصنّف مكوّنات الكلمة العربيّة إلى ستّ حلقات نقترح لتمثيلها الشكل (11):[56]

(11) جميع أنواع الزّوائد العربيّة وفق المعالجات الحاسوبيّة




يمثّل الشكل (11) جميع أنواع الزّوائد التي توجد في البنى اللّسانيّة العربيّة، وفيه ترمز أوسط الحلقات إلى الجذر، بينما ترمز بقيّة الحلقات إلى المكوّنات الزّائدة التي منها ما يتصل ببداية الكلمات دائما، ومنها ما يتّصل بنهاية الكلمات دائما. وتعتبر حلقة الأصول المكوّن الوحيد الذي يلزم وجوده في الكلمات دائما. كذلك «كل وحدة من الوحدات يمكن تعويضها بشيء آخر دون أن يقع تغيير في الوحدتين الباقيتين»[57] مع شرط أن يكون العنصر المبدل به معادلا نحويّا دلاليّا للعنصر المُبدَّل، أي يحمل نفس دلالته الصرفيّة، مثل إبدال جذع بجذع آخر من ذلك [يَلْعَبُونَ] و[يَكْتُبُونَ]، أو إبدال سابقة بأخرى تشترك معها في وظيفة الدّلالة التّصريفيّة من ذلك حرفي المضارعة التّاء والياء في [يَلْعَبُونَ] و[تَلْعَبُونَ] أو في [يَكْتُبُونَ] و[تَكْتُبُونَ].

كذلك رسمنا حلقات الزّوائد الصرفيّة الثّلاثة بطريقة متداخلة مع الأصل، للإشارة إلى كونها عناصر من البنية الدّاخليّة للكلمة، بل وترسم "الدّواخل" داخل حلقة الأصول لأنّها تُقحم بين حروف الجذر. على خلاف اللّواصق التي تعتبر عناصر خارجيّة تمثّل وحدات من الجملة، وترتبط ببنية الكلمة تركيبيّا، وذلك ما عبرنا عنه بالأسهم الرّابطة بين اللّواصق وبقيّة المكوّنات. ووفق ذلك المنوال تتّخذ بنية الكلمة صورة سلسلة (Sequences) تُقحم فيها حلقة الدّواخل.

نخلص إلى أنّ الدّواخل تنتمي إلى البنى الصرفية غير السلسلية الموجودة في «الألسن السّامِيّة (Semitic Language) التي تنحدر منها العربية»[58]، أمّا البنى السلسليّة (Concatenative) فيمكن أن نجد فيها السّوابق واللّواحق فحسب وذلك ينطبق على الألسن اللّاتينيّة. بالتّالي تتميّز العربيّة بكونها «لغة اشتقاقية وليست لغة إلصاقيّة سلسليّة»[59]. إذن تُعالج الظّواهر اللّغويّة عامة بمعرفة خصائصها من الدّراسات وأمّهات الكتب النّحويّة، ثمّ معالجتها يدويّا واستخراج أنماط منطقيّة ومناويل شكليّة يمكن من خلالها إيجاد مراحل معالجة آليّة للظّاهرة اللّغويّة.

**خاتمة:**

درسنا في مقال "إزالة اللّبس عن بعض مسائل حوسبة الزّوائد العربيّة" بعض القضايا المهمّة المختلف فيها، وتكمن قيمة الطّرح الذي عرضناه في تركيزه على تفسير مكوّنات الكلمة عامة وتحليلها تحليلا علميّا دقيقا، وهي جوهر الألسن وسنامها، يقود الإحاطة بها وفهمها فهما صحيحا إلى امتلاك زمام اللّغة وقولبتها ضمن قواعد ومعدلات خوارزميّة تتّسم بالدّقة العلميّة.

ونستنتج ممّا سبق أنّ نظام الزّيادة في العربيّة ليس نظاما صوريّا يعتمد على ملء خانات فارغة (Empty Case) فحسب، لأنّ الزّوائد لا تضاف الى الكلمات العربيّة المستعملة وإنّما تصاغ وفق أوزان محدّدة، وقد فسّرنا طرق حوسبة الكلمة العربيّة بانطلاق من معرفة طبيعة أجزاءها وتقطيعها إلى وحدات صرفيّة صغرى منها الزّائد ومنها الأصلي، وصولا إلى تحديد صيغتها ومقولتها. بالتالي يقود التحليل الصرفي إلى التحليل التركيبي المهتمّ بمعرفة وحدات الجمل وأنواعها، مثلا إذا كانت الكلمة تنتمي إلى مقولة [فعل] فذلك يدلّ على أنّ الجملة فعليّة.

أشرنا كذلك إلى أنّ المعالجات الحاسوبية للّسان العربيّ أوجدت زوائد ترد ضمن حروف "سألتمونيها" ولكنّها ليست صرفية إنما تندرج في المستوى التركيبيّ، أُصطلح على تسميتها اللّواصق، معيارنا الأساسيّ في التفريق بين نوعي الزّيادة هو استقلال اللاّصقة بوظيفة نحوية في الجملة. وهي إمّا في بداية الكلمة فتكون لاصقة قبليّة، وإمّا في نهاية الكلمة فتكون لاصقة بعديّة.

ختاما لابدّ من الإقرار بأنّ توضيح ما سبق من المسائل اللّغويّة يستوجب دون ريب تدخّل الخبير اللّغوي الذي يختار ما يراه مناسبا من المراجع والمؤلّفات للاستعانة بها في تدقيق المعالجات التي تستوجب بطبيعة الحال جهد الخبير الحاسوبيّ لصنع البرامج الآليّة. بذلك تكون النّتائج موَثّقة وموْثقا فيها. وذلك من شأنه أن يجعل نتائج المعالجات الحاسوبيّة للكلام المنطوق والمكتوب سليمة ناجعة. وممّا لا شكّ فيه أنّ تطوّر المجتمعات العربيّة واستمراريتها رهين بمدى القدرة على حوسبة لغتها.

---

1 ابن منظور 2004، 202/6.

2 عمر 2008، 1015/2.

3 الأسمر 1993، 285.

4 ابن يعيش 2001، 314/5.

5 ابن يعيش 2001، 315/5.

6 نفسه، 315/5.

7 ابن الأنباري 2003، 177.

8 ابن يعيش 2001، 315/5.

9 نفسه، 315/5.

10 ابن يعيش 2001، 64/9.

11 نفسه، 315/5.

12 ابن جنّي 1985، 781/2.

13 سيبويه 1988، 315/1.

14 OBEID & al. 2020, 7027.

15 ZRIBI & BAGHOULI 2017, 916.

16 نفسه، 916.

17 OBEID & al. 2020, 7027.

18 نفسه، 7027.

19 AL-LAHHAM & al. 2018, 559.

20 الموسى 2006، 96

21 الزنّاد 2010، 18-19.

22 عكاشة 2022، 11.

23 العارف 2007، 48.

24 الزنّاد 2010، 19.

25 OBEID & al. 2020, 7022.

26 نفسه ، 7022.

27 صوالحه وأتوال 2009، 2.

28 عكاشة 2022، 21.

29 ابن مراد 1997، 142-143.

30 نفسه، 142-143.

31 BOUDLAL & al. 2010, 20.

32 صوالحه وأتوال 2009، 2.

33 عمر 2008، 1015/2.

34 صوالحه وأتوال، 2009.

35 نفسه، 3.

36 SEMBOK & al. 2011, 395.

37 ابن يعيش 2001، 314/5.

38 عمر 2008، 1015/2.

39 ابن مراد 1997، 12-13.

40 SYARIEF & al., 2019.

41 KURNIAWAN & al. 2019, 3.

42 SEMBOK & al. 2011, 935.

43 عكاشة 2022، 48.

44 عكاشة 2022، 45.

45 الأغبر 2021، 186.

46 حمزة 2021، 73.

47 SEMBOK & al. 2011, 975.

48 AL-LAHHAM & al. 2018, 556.

49 El KAH & ZEROUAL 2021, 41.

50 AL-LAHHAM & al. 2018, 556.

51 وردت في الشّاهد (Western Languages) ويقصد بها الألسن اللاّتينيّة.

52 El KAH & ZEROUAL 2021, 41.

53 BOUDLAL & al. 2016, 18.

54 ZRIBI & BAGHOULI 2017, 916.

55 نفسه، 916.

56 عكاشة 2022، 47.

57 بلقاسم 2017، 83.

58 El KAH & ZEROUAL 2021, 42.

59 المسعودي 2019، 143-144.

## المراجع المعتمدة:

### بالعربيّة:


- ابن الأنباري (أبو البركات): الإنصاف في مسائل الخلاف بين النحويّن البصريّن والكوفيّن، تحق: محمد محي الدين عبد الحميد، مطبعة حجازي، بيروت 2021.
- ابن جنّي (أبو الفتح): سر صناعة الإعراب، ط1، تحق: حسن هنداوي، دار القلم، دمشق 1985.
- ابن مراد (إبراهيم): "الصيغمية المعجمية"، مجلة المعجمية، جمعيّة المعجميّة العربية، العدد 13/12، تونس 1997، ص ص. 121-137.
- ابن منظور (أبو الفضل): لسان العرب، ط3، دار صادر، بيروت، 2004.
- ابن يعيش (أبو البقاء): شرح المفصل للزمخشري، تحق. إميل بديع يعقوب، ط1، دار الكتب العلمية، بيروت، 2001.
- الأزهر (الزنّاد): نظريات لسانيّة عرفنيّة، الدار العربية للعلوم ناشرون، بيروت 2010.
- الأسمر (راجي): المعجم المفصّل في علم الصرف، تحق. إميل بديع يعقوب، ط1، دار الكتب العلمية، بيروت 1993.
- بلقاسم (إيمان): المونيمات الوظيفية في المعلّقات السبع دراسة نحويّة لسانيّة، رسالة مقدمة لنيل شهادة الدكتوراه، جامعة أبي بكر بلقايد، كلية الآداب واللّغات قسم اللّغة العربية وآدابها، تلمسان 2017.

- حمزة (حسن): أقسام الكلم في الأسس النّظريّة للقسمة الثّلاثيّة، معهد الدوحة للدراسات العليا – جامعة ليون2، فرنسا 2021.
- سيبويه (قنبر): الكتاب، تحق. عبد السلام محمد هارون، ط3، مكتبة الخانجي، القاهرة، 1988.
- صوالحه (مجدي) وأتوال (ايريك): "توظيف قواعد النحو والصرف في بناء محلل صرفي للغة العربية"، في: معالجة ALECSO، ورشة عمل المنظمة العربية للتربية والثقافة والعلوم حول التحليل الصرفي العربي، دمشق، 2009، ص ص 1-17.
- عكاشة (أسماء): إشكاليّات حوسبة الزّوائد في العربيّة، بحث لنيل شهادة الماجستير في اللّغة والآداب والحضارة العربيّة، إشراف "وسام العربيي"، كليّة الآداب بمنوبة، تونس 2022.
- عمر (أحمد مختار): معجم اللغة العربية المعاصرة، ط1، عالم الكتب، القاهرة، 2008.
- المسعودي (عبد العزيز): التّطور اللّغوي بين المعجم والنحو: بحث لسانيّ في ظاهرة الانحاء، دراسات 15، ط1، مركز الملك بن عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرّياض 2019.
- الموسى (نهاد): العربيّة نحو توصيف جديد في ضوء اللّسانيّات الحاسوبيّة، المؤسسة العربيّة للدراسات والنّشر والتوزيع، الأردن، 2006.

**بغير العربيّة:**

- AL-LAHHAM (Yaser) & al.: "Conditional Arabic Light Stemmer: CondLight", In: The International Arab Journal of Information Technology, Computer Science Department, Zarqa University, Vol. 15, No. 3A, Special Issue, Jordan 2018, pp 559-564.
- BOUDLAL (Mohamad) & al.: "Alkhalil Morpho Sys". In: Proceedings of the 6th International Computing Conference in Arabic Information Technology. Hammamat, 2010, pp 19-21.
- EL KAH (Anoual) & ZEROUAL (Imad): "The effects of PreProcessing Techniques on Arabic Text Classification", IJATCSE 10, 1, 2021, pp 41–48.
- KURNIAWAN (Taufik) &al.: "Development of a Sharaf Table for Vocabulary in the Qur'an using Rule Based", Conference: 2019 7th International Conference on Information and Communication Technology (ICoICT), 2019, pp 1-5.
- OBEID (Ossama) & al.: "CAMeL Tools: An Open-Source Python Toolkit for Arabic Natural Language Processing", in Proceedings of the 12th language resources and evaluation conference, 2020, pp 7022–7032.
- SEMBOK (Tengku) & al.: "A Rule and Template Based Stemming Algorithm for Arabic Language", Issue 5, Volume 5, in: International Journal of Mathematical Models and Methods in Applied Sciences 5(5) 2011, pp 974-981.

- SYARIEF (Mochamad) & al.: "Improving Arabic Stemmer: ISRI Stemmer, in computer science", IEEE 5th International Conference on Wireless and Telematics (ICWT), IEEE 2019, pp. 1–4.
- ZRIBI (Chiraz Ben Othmane) & BAGHOULI (Bechir): "A syntactico-semantic method for Arabic Collocations Extraction", IEEE/ACS'S 14th International Conference on Computer Systems and Applications, 2017, pp. 915–921.
- ZRIBI (Chiraz) & BAGHOULI (Bechir): "A syntactico-semantic method for Arabic Collocations Extraction", IEEE/ACS 14th International Conference on Computer Systems and Applications, 2017, pp 915-921.