

# نبذة عن برنامج المحلل الصرفي للكلمات العربية

عز الدين مزروعي<sup>١</sup>، عبد الوافي مزيان<sup>١</sup>، عبد الحق لخواجة<sup>١</sup>، محمد ولد عبد الله ولد بيباه<sup>١</sup>

عبد الرحيم بودلال<sup>٢</sup>، رشيد بلحبيب<sup>٢</sup>

<sup>١</sup>جامعة محمد الأول، كلية العلوم، شعبة الرياضيات والإعلاميات، مختبر الأبحاث في الإعلاميات، فريق المعالجة الآلية للغات الطبيعية

<sup>٢</sup>جامعة محمد الأول، كلية الآداب والعلوم الإنسانية

<sup>٢</sup>جامعة قطر

[azze.mazroui@gmail.com](mailto:azze.mazroui@gmail.com)

## ١ - مقدمة

تتقسم عملية التحليل الصرفي في برنامجنا إلى المراحل التالية:

١- تقطيع الكلمة إلى مجموعة السوابق والجذوع واللواحق المحتملة كما في المثال التالي:

الكلمة: فهم

اللاحق	الجذع	السابق
#	فهم	# <sup>١</sup>
#	هم	ف

ومن المعروف أن عملية التقطيع هذه تفضي غالباً إلى نتائج متعددة. وقد اعتمدنا لهذا الغرض قاعدة بيانات بالزوائد التي تلحق بالكلمة العربية من سوابق ولواحق قمنا بتصنيفها على ثلاث مجموعات:

أ- زوائد لا تلحق إلا بالأسماء

ب- زوائد لا تلحق إلا بالأفعال

ج- زوائد مشتركة بين الأسماء والأفعال

كما أخذنا في الحسبان مسألة التوافق بين السوابق واللواحق التي قد تدخل على نفس الجذع، وقد كان الغرض من هذا الحد من عدد مخرجات البرنامج.

## ٢- تحديد الأوزان المحتملة للكلمة

بعد مرحلة التقطيع السالفة الذكر، تأتي مرحلة تحديد الأوزان المحتملة وذلك عن طريق تحليل الجذوع الناتجة ومقارنتها بقاعدة بيانات للأوزان العربية قمنا بإنشائها.

<sup>١</sup> يرمز الحرف # إلى غياب السابق أو اللاحق

إن التوصيف الذي اعتمدهنا في قاعدة بيانات الأوزان العربية ينطلق من اهتمامنا في مجال التشكيل الآلي بإنجاز محلل صرفي قادر على إعطاء جميع التشكيلات المحتملة للكلمة خارج سياقها من الجملة. حيث نعمل بعد ذلك على تطبيق نموذج إحصائي قمنا سابقا بتجربته بنجاح لغرض الاستخلاص الآلي لجذور الكلمات العربية داخل السياق.<sup>٢</sup>

وعلى سبيل المثال فإن توصيف الوزن تَفَعَّلْتُ في قاعدة البيانات لدينا كان على الشكل التالي

### تَفَعَّلْتُ: ف؛ م؛ غ؛ ت؛ تَفَعَّلْتُ؛ ث؛ ز؛ م؛ ف

ما يعني أن الوزن وزن فعل (ف)، ماض مبني للمعلوم (م م)، مسند إلى الغائبة (غ غ)، ومصرف من الفعل تَفَعَّلَ وهو فعل ثلاثي مزيد (ث ز)، مبني على الفتح (م ف). وعموما فإن توصيف أوزان الأفعال لدينا ينطوي على المعلومات التالية:

الوزن (مُشَكَّلًا تشكيلا تاما)	الزمن	البناء للمعلوم من عدمه	الإسناد	التجرد أو الزيادة	علامة الإعراب أو البناء
-----------------------------------	-------	---------------------------	---------	----------------------	-------------------------------

كما أن توصيف أوزان الأسماء لدينا يهدف لإعطاء التشكيل التام للوزن ونوع الاسم من حيث الأفراد أو التثنية أو الجمع وعلامة الإعراب أو البناء.

وعلى سبيل المثال فإن توصيف الوزن فَاعِلَانٍ في قاعدة البيانات كان على الشكل التالي

### فَاعِلَانٍ: س؛ ث؛ فَاعِلٍ؛ ر؛ ا

ما يعني أن الوزن وزن اسم (س)، مثنى (ث)، ناتج عن تثنية الوزن فَاعِلٍ وهو وزن لاسم مرفوع علامة رفعه الألف (ر ا).

ويقوم البرنامج، بعد تحديد الوزن، باستخلاص الجذر المحتمل للكلمة بناء على هذا الوزن والتحقق من وجوده في قاعدة بيانات للجذور العربية، كشرط ضروري لقبول هذا الحل.

كما أننا اعتمدنا توصيفا للحروف ولأسماء الإشارة والأسماء الموصولة وغيرها من الكلمات الوظيفية سعيا للحصول على التشكيل التام لها والتعرف على وظيفتها.

<sup>٢</sup> "A Markovian Approach for Arabic Root Extraction", ACIT08.

## ٢- مخرجات

في ما يلي بعض الأمثلة على نتائج التحليل الصرفي:

### المدخل: فرمت

التشكيل التام لللمة	نتيجة تحليل وزن الجذع	الجزر	الوزن	اللاحق	الجزع	السابق
فَرَمَتَ	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مجرد مبني على السكون	ف ر م	فَعَلَّتَ	#	فرمت	#
فَرَمْتُ	فعل ماض مبني للمعلوم مسند إلى المتكلم ثلاثي مجرد مبني على السكون	ف ر م	فَعَلْتُ	#	فرمت	#
فَرَمْتِ	فعل ماض مبني للمعلوم مسند إلى المخاطبة ثلاثي مجرد مبني على السكون	ف ر م	فَعَلْتِ	#	فرمت	#
فَرَمْتُ	فعل ماض مبني للمعلوم مسند إلى الغائبة ثلاثي مجرد مبني على الفتح	ف ر م	فَعَلْتُ	#	فرمت	#
فَرَمْتُ	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مجرد مبني على السكون	ف ر م	فَعَلْتُ	#	فرمت	#

فَرَمْتُ	فعل ماض مبني للمعلوم مسند إلى المتكلم ثلاثي مجرد مبني على السكون	ف ر م	فَعَلْتُ	#	فرمت	#
فَرَمْتِ	فعل ماض مبني للمعلوم مسند إلى المخاطبة ثلاثي مجرد مبني على السكون	ف ر م	فَعَلْتِ	#	فرمت	#
فَرَمْتُ	فعل ماض مبني للمعلوم مسند إلى الغائبة ثلاثي مجرد مبني على الفتح	ف ر م	فَعَلْتُ	#	فرمت	#
فَرِمْتُ	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مجرد مبني على السكون	ف ر م	فَعِلْتُ	#	فرمت	#
فَرِمْتِ	فعل ماض مبني للمعلوم مسند إلى المتكلم ثلاثي مجرد مبني على السكون	ف ر م	فَعِلْتِ	#	فرمت	#
فَرِمْتُ	فعل ماض مبني للمعلوم مسند إلى المخاطبة ثلاثي مجرد مبني على السكون	ف ر م	فَعِلْتُ	#	فرمت	#
فَرِمْتِ	فعل ماض مبني للمعلوم مسند إلى الغائبة ثلاثي مجرد مبني على الفتح	ف ر م	فَعِلْتِ	#	فرمت	#
فُرِمْتُ	فعل ماض مبني للمجهول مسند إلى الغائبة ثلاثي مجرد مبني على الفتح	ف ر م	فُعِلْتُ	#	فرمت	#
فُرِمْتِ	فعل ماض مبني للمجهول مسند إلى المتكلم ثلاثي مجرد مبني على السكون	ف ر م	فُعِلْتِ	#	فرمت	#
فُرِمْتُ	فعل ماض مبني للمجهول مسند إلى المخاطب ثلاثي مجرد مبني على الفتح	ف ر م	فُعِلْتُ	#	فرمت	#

فُرِّمَتْ	فعل ماض مبني للمجهول مسند إلى المخاطبة ثلاثي مجرد مبني على السكون	ف ر م	فُعِّلَتْ	#	فرمت	#
فَرَّمَتْ	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مزيد مبني على السكون	ف ر م	فَعَّلَتْ	#	فرمت	#
فَرَّمَتْ	فعل ماض مبني للمعلوم مسند إلى المخاطبة ثلاثي مزيد مبني على السكون	ف ر م	فَعَّلَتْ	#	فرمت	#
فَرَّمَتْ	فعل ماض مبني للمعلوم مسند إلى المتكلم ثلاثي مزيد مبني على السكون	ف ر م	فَعَّلْتُ	#	فرمت	#
فَرَّمَتْ	فعل ماض مبني للمعلوم مسند إلى الغائبة ثلاثي مزيد مبني على الفتح	ف ر م	فَعَّلَتْ	#	فرمت	#
فُرِّمَتْ	فعل ماض مبني للمجهول مسند إلى المخاطب ثلاثي مزيد مبني على السكون	ف ر م	فُعِّلَتْ	#	فرمت	#
فُرِّمَتْ	فعل ماض مبني للمجهول مسند إلى المتكلم ثلاثي مزيد مبني على السكون	ف ر م	فُعِّلْتُ	#	فرمت	#
فُرِّمَتْ	فعل ماض مبني للمجهول مسند إلى المخاطبة ثلاثي مزيد مبني على السكون	ف ر م	فُعِّلَتْ	#	فرمت	#
فُرِّمَتْ	فعل ماض مبني للمجهول مسند إلى الغائبة ثلاثي مزيد مبني على الفتح	ف ر م	فُعِّلَتْ	#	فرمت	#

ف	رمت	#	فَعَتُّ <sup>٣</sup> (فَعَلْتُ)	ر م ي	فعل ماض مبني للمعلوم مسند إلى الغائبة ثلاثي مجرد مبني على الفتح	فَرَمْتُ
ف	رمت	#	فَعَّتْ (فَعَلْتُ)	ر م م	فعل ماض مبني للمعلوم مسند إلى الغائبة ثلاثي مجرد مبني على الفتح	فَرَمْتُ
ف	رمت	#	فِعَتِ (فَعَلْتُ)	ر ي م	فعل ماض مبني للمعلوم مسند إلى المخاطبة ثلاثي مجرد مبني على السكون	فَرِمْتُ
ف	رمت	#	فُعْتُ (فَعَلْتُ)	ر و م	فعل ماض مبني للمعلوم مسند إلى المتكلم ثلاثي مجرد مبني على السكون	فَرَمْتُ
ف	رمت	#	فُعَّتْ (فَعَلْتُ)	ر و م	فعل ماض مبني للمعلوم مسند إلى المتكلم ثلاثي مجرد مبني على السكون	فَرَمْتُ
ف	رمت	#	فِعَتْ (فَعَلْتُ)	ر ي م	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مجرد مبني على السكون	فَرِمْتُ
ف	رمت	#	فُعَّتِ (فَعَلْتُ)	ر و م	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مجرد مبني على السكون	فَرِمْتُ
ف	رمت	#	فِعْتُ (فَعَلْتُ)	ر و م	فعل ماض مبني للمعلوم مسند إلى المخاطب ثلاثي مجرد مبني على السكون	فَرِمْتُ
ف	رمت	#	فُعَّتْ (فَعَلْتُ)	ر م م	فعل ماض مبني للمجهول مسند إلى الغائبة ثلاثي مجرد مبني على السكون	فَرَمْتُ

## فمنكم

<sup>٣</sup> هذا مثال على بعض الأوزان الاصطناعية التي اعتمدها لتسهيل التعامل الآلي مع الظواهر الصرفية من إعلال وقلب وتضعيف إلخ.. وعلى كل حال فقد اعتمدنا إرفاقها في قاعدة البيانات مع أوزانها الحقيقية (بين قوسين)

التشكيل التام لللمة	نتيجة تحليل الجذع	الجذر	الوزن	اللاحق	الجذع	السابق
فَمِنْكُمْ	حرف جر			كم	من	ف
فَمَنْكُمْ	فعل ماض، مبني للمعلوم، مسند إلى الغائب، ثلاثي مجرد، مبني على الفتح	م ن ن	فَعَّ (فَعَلَّ)	كم	من	ف
فَمَنْكُمْ	اسم مفرد على وزن فَعْل مرفوع بالضممة	م ن ن	فَعُّ (فَعْلُ)	كم	من	ف
فَمَنْكُمْ	اسم مفرد على وزن فَعْل منصوب بالفتحة	م ن ن	فَعَّ (فَعَلَّ)	كم	من	ف
فَمَنْكُمْ	اسم مفرد على وزن فَعْل مجرور بالكسرة	م ن ن	فَعَّ (فَعَلَّ)	كم	من	ف

**أخذنا** (طلبا للاختصار نورد نتيجة تحليل الكلمة على اعتبارها اسما ذلك أن نتيجة تحليلها على أساس أنها فعل تعطينا نتائج مشابهة لما أوردناه في تحليل الفعل فَرَمَتْ)

التشكيل التام لللمة	نتيجة تحليل وزن الجذع	الجذر	الوزن	اللاحق	الجذع	السابق
أَخَذْنَا	اسم مفرد على وزن فَعْل منصوب بالفتحة	أ خ ذ	فَعْلَ	نا	أخذ	#
أَخَذْنَا	اسم مفرد على وزن فَعْل مجرور بالكسرة	أ خ ذ	فَعْلِ	نا	أخذ	#
أَخَذْنَا	اسم مفرد على وزن فَعْل مرفوع بالضممة	أ خ ذ	فَعْلُ	نا	أخذ	#

### ٣ - ملاحظات

بقي أن نشير في الأخير إلى أن المقاربة التي اعتمدها في إنجاز محللنا الصرفي كانت في الحقيقة مدفوعةً باحتياجاتنا في مجال التشكيل الآلي للنصوص العربية.

ولهذا السبب، فإن جانباً مهماً من جهدنا انصب على ضبط تشكيل الأوزان العربية تشكيلاً تاماً. كما قمنا بالتعامل مع الظواهر الصرفية من إعلال وإبدال وغيرها باستحداث أوزان اصطناعية لتسهيل التعامل معها. فمثلاً كلمة من قبيل **ازدهار** يقابلها في قاعدة بيانات الأوزان الوزن **ازدعال**.

كما أننا نلاحظ أن مرحلة التحليل الصرفي خارج السياق في برنامجنا ينجم عنها عدد من المخرجات التي تستلزم، من أجل تدقيقها أكثر، مجهوداً من جانب اللغويين في اتجاه تعريف الأوزان التي يقبلها كل جذر عربي على حدة. فعلى سبيل المثال، تحليل الفعل **فرم** يقود إلى التشكيلات **فَرَمَ فَرِمَ و فَرَمَ**، ولا شك أن تثليث عين هذا الفعل يمثل تعميماً من قبل المعالج الآلي يستدعي المراجعة اللغوية على مستوى الجذر. أما في ما يخص توصيف أوزان الأسماء فإننا نعكف حالياً على إنجاز توصيف أعم من الذي ذكرناه آنفاً، يسمح بإضافة معلومات صرفية من قبيل النوع (اسم فاعل، اسم مفعول إلخ..)، والتذكير والتأنيث، وغيرها.