الاسترجاع الآلي للصيغ الصّرفيّة من المعجم الوسيط الإلكتروني

أ. راديّة حجبار*

تاريخ الإرسال:23-04-2019 تاريخ القبول:27-11-2019

الملخّص: إنّ رقمنة مدوّنات اللّغة العربيّة هي بدايّة لمشروع الذخيرة اللغويّة العربيّة، وكذا لمشروع المعجم التّاريخي للغة العربيّة؛ وذلك لأجل التّعامل معها آليا، واسترجاع المعلومات والتّنقيب في البيانات، إلاّ أنّه حين أردنا التّعامل مع إحدى هذه المدوّنات، والمتمثلة في المعجم الوسيط الذي ألفه مجمع القاهرة، في طبعته الرّابعة، وذلك باسترجاع أو بالقراءة الآليّة للصيغ الصّرفيّة المشتقة من الفعل الثلاثي المجرد الواردة كمداخل معجميّة، وكذا التّعريف المعجمي الذي يأتي بعده، واجهتنا عدّة إشكالات، بعضها متعلقة بطريقة رقمنة مدوّنة المعجم الوسيط، وبعضها متعلقة بخصائص اللغة العربيّة وبعضها بواضعي المعجم؛ ما جعلنا مجبرين على تصفيّة النّتائج المحصّل عليها بعد كلّ قراءة آليّة لصيغة صرفيّة معيّنة من مدوّنة المعجم الوسيط الرّقمي أو الإلكتروني، والدعوة إلى إعادة النّظر في رقمنة المدوّنات العربيّة؛ لأجل التّعامل معها بلغات البرمجة دون إشكالات.

الكلمات المفاتيح: الصّيغة الصّرفيّة؛ الاسترجاع الآليّ؛ المدوّنة الرّقميّة؛ البرنامج؛ الخوارزميّة.

Abstract: The digitization of the corpora of the Arabic language is the beginning of the "thesaurus" project, as well

* جامعة امحمد بوقرة بومرداس، الجزائر، البريد الإلكتروني: radiahadjebar@gmail.com

as the Arabic "historical lexicon" project. This digitization makes it possible to apply certain automatic processing procedures to real linguistic samples such as: automatic search, information retrieval, etc. ... As far as we are concerned, we have adopted as a digital corpus the lexicon "Al muajem Al Wassit " conceived by a team from the Arabic Academy of Cairo, in its fourth edition, and we tried to apply an automatic search mechanism in all the lexical entries of the different schemes (morphological forms) derived from the trilitary mujarrad verbs, as well as the lexical definition that follows. We have encountered several difficulties, some of which relate to the digitization of our adopted lexicon (Al Muajem Al Wassit), others relate to the features of the Arabic language, and finally, others relate to the obsolete lexicographical techniques of this lexicon. For these reasons, we had to filter the results obtained after each automatic extraction of a scheme from the electronic version of our lexicon. Our conclusion is an appeal to all actors in the lexicological and lexicographical field of the

Arabic language, to question their techniques and methodology of corpora digitization of the Arabic language.

Keywords: scheme (morphological form); automatic extraction; digital corpus; program; algorithm.

المقدّمة: تكون القراءة الآليّة أو الاسترجاع الآلي من المدوّنات الرّقميّة؛ بمعنى أن تكون مكتوبة على محرّر النّصوص الوورد (word)؛ لأنّه لا يستطيع أي برنامج تشفير المدوّنة التي تكون على شكل بي دي ف (pdf) أو التي تمّ مسحها ضوئيا، وإنما يجب أن تكون مكتوبة على الوورد (word) والتي تحوّل بنمط (utf8)(Gérard swinnen, 2009, p37 .) إلى نسخة ثانيّة مكتوبة على (bloc– notes) أي ذات امتداد (txt) حتى يتمكن مترجم البرنامج من تشفيرها أو القراءة منها، وهي تمثّل قاعدة البيانات، ولهذا نرى أنّ البلدان العربيّة بدأت بتشكيل فرق خاصّة، تقوم برقمنة المدوّنات اللّغويّة العربيّة التي دوّنت في مختلف العصور، والتي تنتمي إلى مختلف الاختصاصات؛ لأجل استغلالها آليا، وإنجاز مشروع الذخيرة اللّغويّة العربيّة الذي تشرف عليه الهيأة العليا التّابعة لجامعة الدّول العربيّة، وكذا مشروع المعجم التّاريخي للغة العربيّة، الذي بدأت الدّول العربيّة بإنجازه، ونلاحظ في عصرنا المنفتح على العلوم، أنّ عماد ال ممّارسة اللّسانيّة، ينبني في شقه الإجرائي على النّمذجة؛ لأنّ اعتماد نهج النّمذجة يشكّل سبيلا، للانتقال من الاعتماد على تراكم المعارف في العلوم اللّغويّة، وأبحاثها إلى صوغ أنحاء منضبطة بقيود وعمليات خوارزميّة لتوليد خصائص البنى اللّغويّة وسماتها واشتقاق هذه الخصائص وتمثيلها، التي تمثّل قمة بلاغة الخطاب اللّساني

المعاصر وتدخل بلورة النّماذج ضمن التّقانة العلميّة المعاصرة، حيث تعدّ جزءا لا يتجزأ من هندسة النّظريات العلميّة (محمد طعمة، 2017، ص104) كما أصبح بناء العلم غير منفك عن بنائه النّمذجي وتطور الحقول المجاورة للسانيات هو ما دفع الأبحاث اللغويّة إلى الآلات الواصفة والنّماذج الحاسوبيّة والرّياضيّة والعصبيّة والبيولوجيّة... إلخ. كما أنّ التّقانة الصّوريّة لها القدرة على إمداد اللساني بالوسائل الصّوريّة الكفيلة باختزال تعقيد أدوات وصف اللغات الطبيعيّة وتنوعها إلى مبادئ أوليّة ذات كفايّة تفسيريّة. (محمد طعمة، 2017، ص105)؛ يعني أن هناك إمكانيّة اعتماد علم المنطق وكذلك علم الحاسوب في الأبحاث اللغويّة.

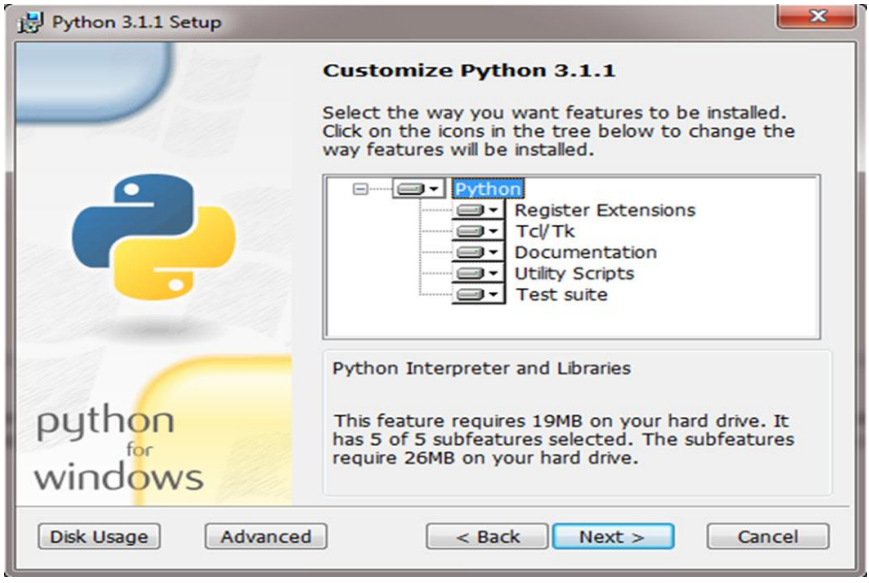
1- لغات البرمجة وأصنافها: إنّ القراءة الآليّة أو الاسترجاع الآلي للصيغ الصّرفيّة المصدريّة، يكون باعتماد واحدة من لغات البرمجة أو اللّغات الحاسوبيّة المختلفة، والتي تنقسم إلى ثلاثة أصناف (طارق المالكي، 2015، ص6) وهي:

1- لغات برمجيّة تخوّل للمستعمل صناعة برامج الحاسوب، تطبيقات الجوال، مواقع إلكترونيّة وأنظمة تشغيل ... ونمثل بلغة جافا، سي ++، بايثون،...

2- لغات الاستعلام التي تسمح باسترجاع المعطيات ومعالجتها من نظم حاسوبيّة تحتوي على بيانات، ونمثّل لها بلغة SQL ولغة QUERY ولغة SPARQL.

3- لغات واصفة تسمح بوصف المحتوى الرّقمي وفق قواعد محددة وأشهرها HTML و XML الذي اخترع لتسهيل نقل وتبادل المعلومات بطريقة موحّدة مهما اختلفت أنظمة التّشغيل، وقد تطورت هذه اللّغات الواصفة بشكل مذهل في العقد الأخير مستفيدة من المنطق الوصفي والذكاء الاصطناعي وأشهرها RDF و RDFs و OWL.

ولتقصّي القراءة الآليّة أو الاسترجاع الآلي للصيغ الصّرفيّة المصدريّة، المشتقة من الفعل الثّلاثي المجرّد، من مدوّنة المعجم الوسيط الإلكتروني، فقد نزّلنا البرنامج الذي اخترناه من الشّابكة والذي يقوم بالقراءة، وهو برنامج البايثون (IDLE python 3,1) وتنصيبه على الحاسوب، وهذه صورة للبرنامج بعد تحميله.



2- القراءة الآليّة للصيغ الصّرفيّة: إنّ الصّيغ الصّرفيّة المصدريّة، المشتقة من الفعل الثلاثي المجرد التي نودّ استرجاعها آليا من المعجم الوسيط هي: صيغة فِعالة للدلالة على الحرفة، نحو: زراعة صناعة وصيغة فَعلان للدلالة على الحركة والاضطراب، نحو: غليان، دوران، صيغة فُعال للدلالة على المرض نحو: زكام، صداع، وصيغة فعيل وفُعال للدلالة على المرض والصّوت، نحو: صهيل، صراخ وصيغة فعَل للدلالة على عيب، نحو: عور، وصيغة فُعول للدلالة على المعالجة، نحو: خروج، لصوق وصيغة فُعلة للدلالة على اللون، نحو: خضرة، زرقة، وصيغة فُعولة للدلالة على معنى ثابت، نحو: ملوحة صعوبة، وصيغة فعْل نحو: سمع أكل، وصيغة فَعالة نحو: فصاحة، صراحة، وصيغة فِعال للدلالة على الامتناع، نحو: نفار، إباء. (رمضان عبد الله، 2006، ص81) والحقيقة أنها مصادر غير قياسيّة والأغلب فيها السّماع، والنّقل عن المعاجم اللّغويّة القديمة، إلاّ أنّ علماء الصّرف حاولوا وضع بعض الضوابط التي تنطبق على أنماط معينة منها.

ونبيّن في ما يلي الخوارزميّة المستخدمة لاستخراج الصّيغ الصّرفيّة المصدريّة، وما يتعلق بالتّعريف المعجمي بمقالاتها من المفردات في المعجم الوسيط، وفي قاعدة البيانات (11.txt) الموجودة على سطح المكتب، مع التّنبيه إلى أنّ في هذه الخوارزميّة تتوفر على ربط التّعليمات بمكتبة النّصوص (re) وكذلك مكتبة خاصّة بقراءة كل اللّغات (Tarek (codecs) ziadé,2009,p105)

- الخوارزميّة:

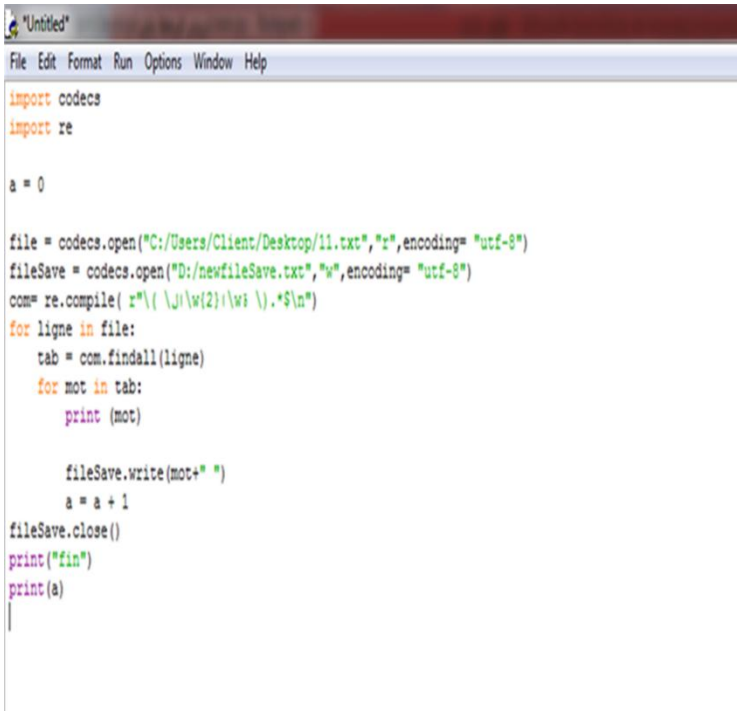
```
import codecs
import re
a = 0
file = codecs.open("C:/Users/Client/Desktop/11.txt","r",encoding= "utf-8")
fileSave = codecs.open("D:/newfileSave.txt","w",encoding= "utf-8")
com= re.compile( r"\( \).*$\n")ة\w\w{2}ال
for ligne in file:
tab = com.findall(ligne)
for mot in tab:
print (mot)
fileSave.write(mot+" ")
a = a + 1
```

```
fileSave.close()

print("fin")

print(a)
```

وهذه الصّورة المواليّة واجهة توضّح الخوارزميّة المعتمدة للقراءة الآليّة للصيغ الصّرفيّة من المعجم الوسيط الرّقمي، وفي هذه الحالة قراءة الصّيغة الصّرفيّة فِعالة.



ونشير إلى أن هذه الخوارزميّة تمثّل القراءة الآليّة للصيغة الصّرفيّة (فعالة) كما هو وارد في التّعليمة التي جاءت في السّطر السّادس من الخوارزميّة وهي:

\).*$\n")ةw \ \{2} \wل com= re.compile(r"\(\

وتمثّل قراءة المداخل المعجميّة التي نجدها واردة بين قوسين سواء في المدونة الورقيّة أم الرّقميّة والتي يكون شكلها أو بنيتها ألف ولام ثم حرفان ثم ألف ثم حرف ثم تاء ، وبعدها يقرأ التّعريف المعجمي الذي يليه إلى غايّة نهايّة السّطر. لتظهر النّتائج على الشّاشة، مثل: زراعة، زمالة، سقايّة، سمّاعة، نفايّة نقاهة، وقاحة... كما تظهر هذه النّتائج كذلك في ملف خاص للتخزين (fileSave).

وتبيّن هذه التّعليمة قراءة المداخل المعجميّة، والتي نلاحظ أنّها واردة بين قوسين، سواء في المدوّنة الورقيّة أم الرّقميّة، والتي يكون شكلها أو بنيتها ألف ولام ثم حرفان ثم ألف ثم حرف ثم تاء. ويلي ذلك قراءة التّعريف المعجمي الذي ينتهي بنهايّة السّطر، كما تظهر على الشّاشة نتائج البحث التي منها: زراعة، زمالة، سقايّة، سمّاعة، نفايّة، نقاهة، وقاحة...إلخ. وتظهر هذه النّتائج كذلك في ملف خاص للتخزين (fileSave).

ونطبّق نفس الخوارزميّة بالنّسبة إلى بقيّة الصّيغ الصّرفيّة المشتقة من الفعل الثّلاثي المجرد، وما نغيّره هو فقط التّعليمة الواردة في السّطر السّادس من الخوارزميّة. ونفترض أنّ الصّيغ بمثابة لواصق صرفيّة، وأنّ معظم ما سمي صيغة مركّب من جذع ولاصقة.

– القراءة الآليّة للصيغة الصّرفيّة فَعالة:

تكون القراءة الآليّة للصيغة الصّرفيّة فَعالة، كما في التّعليمة التّاليّة:

```
\).*$\n”)ةw \ ا{2} \wال com= re.compile( r”\( \
```

ونلاحظ في هذه التّعليمة، أنّها تشبه التّعليمة المعتمدة لقراءة الصّيغة الصّرفيّة فِعالة. وهذا يعود إلى عدم وجود الحركات في المدوّنة الرّقميّة ومن ثَمَّ إلى خلوّ الكلمات من الضّبط الحركي (الإعجام). وبالتّالي نتحصل على النّتائج نفسها التي ظهرت في قراءة الصّيغة الصّرفيّة فِعالة.

– **القراءة الآليّة للصيغة الصّرفيّة فُعال وفِعال:** تكون القراءة الآليّة كما في التّعليمة التّاليّة:

```
w \).*$\n”) \ ا{2} \wال com= re.compile( r”\( \
```

لقد حافظنا على التّعليمة السّابقة الخاصّة بالصّيغة الصّرفيّة فعالة، ونقوم فقط بحذف التّاء المربوطة لأجل قراءة الصّيغة الصّرفيّة فُعال وفِعال، وعلى نفس التّعليمة لقراءة الصّيغتين؛ لأنّ البرنامج لا يميّز بين فُعال وفِعال، نتيجة عدم وجود الحركات وعلامة التّضعيف. والنّتائج التي تظهر: جثام، جواد، خزان، دماع، دماغ، دوار، ذراب، قشاط، نهار، نواس، وفاة...إلخ.

– **القراءة الآليّة للصيغة الصّرفيّة فُعول:** تُقرأ الصّيغة الصّرفيّة فُعول كما في التّعليمة التّاليّة:

```
w \).*$\n”) \ و{2} \wال com= re.compile( r"\( \
```

إنّ ما قمنا بتغييره في هذه التّعليمة مقارنة بالقراءة الآليّة للصيغة الصّرفيّة فُعال، هو استبدال الألف واوا لأجل قراءة الصّيغة الصّرفيّة فُعول، فظهرت في النّتائج: أسود، برود، جنوب، حشوة، حلوى، خروف، زبون، طيور، عجوز، غسول، قهوة، كسوة... وغيرها من المفردات التي حرفها الثّالثّ واوا.

– **القراءة الآليّة للصيغة الصّرفيّة فُعولة:** تكون القراءة الآليّة كما في التّعليمة التّاليّة:

\).*$\n")ةw \ و{2} \wال com= re.compile(r"\(\

لقد أضفنا إلى هذه التّعليمة التّاء المربوطة لأجل قراءة الصّيغة الصّرفيّة فُعولة مقارنة بالصّيغة الصّرفيّة فُعول، فنرى على الشّاشة مفردات مثل: ألوكة، أمومة، مجوعة، خروبة، خنوثة، ذكورة، سهولة، أسودة طعومة، عمولة، فحولة، فلوجة، وعوعة... إلخ.

– **القراءة الآليّة للصيغة الصّرفيّة فُعلة:** تكون القراءة الآليّة للصيغة الصّرفيّة فُعلة بتطبيق هذه التّعليمة:

\).*$\n")ة \ }3{ \wال com= re.compile(r"\(\

تتم القراءة بتوصيف المدخل المكوّن من ثلاثة حروف تليها تاء مربوطة، فتظهر النّتائج: أجرة، أصدة، أفرة، جزعة، جزمة، جمسة، عروة، عضلة، عطلة، قهرة، كبوة، كفتة، كتلة، مؤنة... إلخ.

– **القراءة الآليّة للصيغة الصّرفيّة فعَلان:** تُقرأ الصّيغة فعَلان آليا كما في التّعليمة التّاليّة:

\).*$\n”)3{ \wال com= re.compile(r"\(\

نتحصل آليا على كلمات على وزن فعلان، بالكتابة في التّعليمة، أنها مداخل تتكوّن من ثلاثة حروف ثم ألف ونون. فحينها تظهر على الشّاشة الكلمات: بركان، بستان، ثعبان، ثوران، طغيان، فرقان، فنجان، قفطان، نسيان، نهدان، هيجان... إلخ.

– **القراءة الآليّة للصيغة الصّرفيّة فعَل وفعْل:** يكون الاسترجاع الآلي للصيغة فعَل وفعْل كما في التّعليمة التّاليّة:

\).*$\n”)3{ \wال com= re.compile(r"\(\ }

تتكون الصّيغتان فعَل وفعْل من ثلاثة أحرف؛ لذا نكتب في التّعليمة جميع المداخل التي تكون ثلاثيّة سواء كان الحرف الثّاني ساكنا أو متحركا؛ لأن العمليّة الحاسوبيّة لا تقتضي الشّكل نتيجة غياب الحركات في برنامج تخزين المدوّنة، فيظهر من النّتائج على الشّاشة: أثر، أجر، أجل، حول، رأي، رجع، صفن صمد، عذق، عرش، فكر، فلس، كشف، موت، نقد، نقس، هدر، هدف، وجه.

ويكون بهذا البحث الآلي عن الصّيغ الصّرفيّة الواردة في المعجم الوسيط الرّقمي، تبعا لقواعد البرمجة في (بايثون)، بتوصيف (طارق المالكي، 2015،

ص8) الصّيغ الصّرفيّة، أو بإحداث تعديلات على مستوى التّعليمة الواردة في الخوارزميّة، تستجيب لشكل الصّيغة وعدد حروفها، وهذا يمكننا من استخراج الصّيغ بدون مشقة وفي وقت وجيز، فنتوصل إلى تأليف معجم عربي محوسب يشتمل على محلل صرفي يسمح لنا بالتّعامل بسهولة ويسر مع المدوّنات الكبرى المكتوبة بصيغة وورد.

– إشكالات القراءة الآليّة للصيغ الصّرفيّة من المعجم الوسيط الرّقمي:

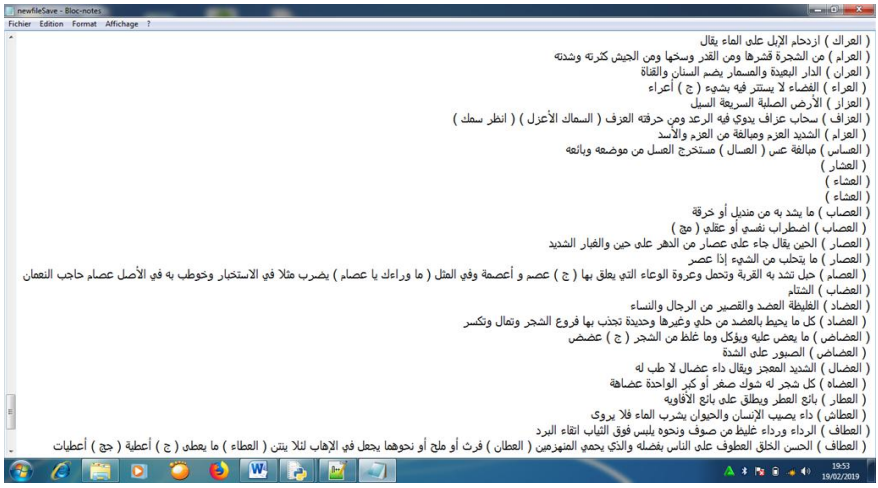
إنّ القراءة الآليّة أو استرجاع الصّيغ الصّرفيّة المشتقة من الفعل الثلاثي المجرد من المعجم الوسيط الرّقمي انتهت بإشكالات تتعلق بعضها بخصائص اللغة العربيّة، وبعضها الآخر تتعلق بطريقة رقمنة المعجم الوسيط، وبعضها مردّها إلى واضعي المعجم؛ ما تطلب منّا إعادة تصفيّة النّتائج المحصّل عليها في كلّ قراءة لصيغة صرفيّة معيّنة. ومن هذه الإشكالات نذكر:

1– إشكالات رقمنة المعجم الوسيط: الإشكالات المتعلقة برقمنة المعجم الوسيط تتمثّل فيما يلي:

– جاءت جميع المداخل في المعجم بين قوسين، وبما أنّ البرنامج يقرأ بالمطابقة؛ بمعنى أنه يقرأ تماما ما طُلب منه في التّعليمة، فإذا ورد مثلا في التّعليمة قراءة المدخل الذي بين قوسين، وهذين الأخيرين يكتبان بفراغ عن المدخل، ولكن في قاعدة البيانات أي مدوّنة المعجم الوسيط الرّقميّة ورد القوس ملتصقا بالمدخل، فإنه لا يتم قراءته ولا يظهر في النّتيجة؛ لأنّه يعتبر حرفا مضافا إلى المدخل حسب البرنامج.

– يأتي بعد المدخل المعجمي مباشرة التّعريف المعجمي على نفس السّطر، ولكن في بعض المداخل نجد أنّ التّعريف يأتي في السّطر الموالي، فلا يتم مواصلة قراءة التّعريف المعجمي، كما هو وارد في التّعليمة، ونمثّل

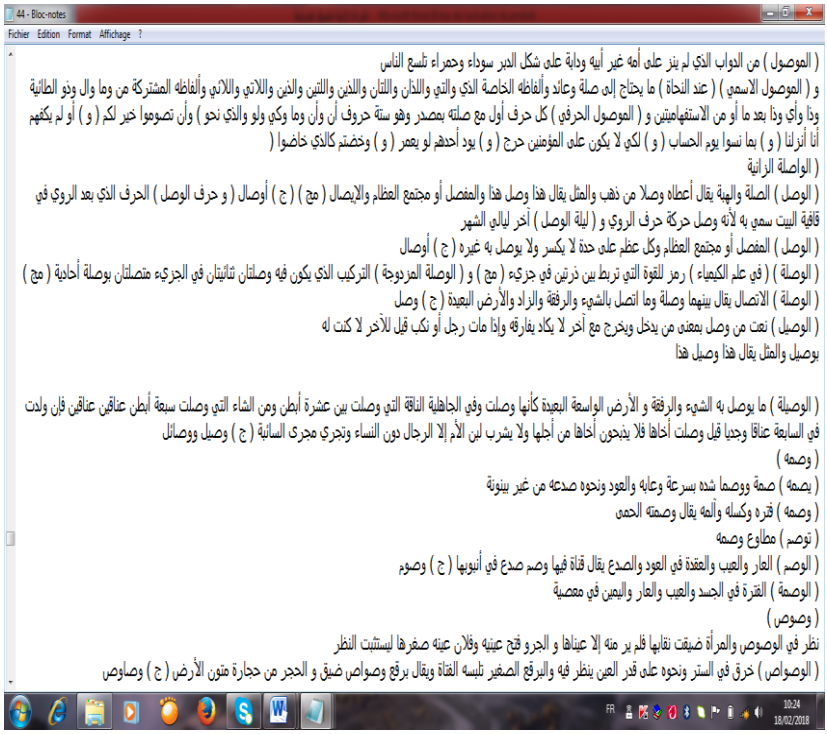
لذلك في مدوّنة الصّيغ الصّرفيّة فِعال مدخل العِشاء الذي ظهر في النّتائج دون تعريف معجمي كما في هذه الصّورة:



ونفس الحال في الصّيغة الصّرفيّة فعْل إذ نجد مدخل الكوث العشر والنّيص ظهرت في نتائج البحث دون تعريف معجمي أيضا.

– إنّ البرنامج يقوم بالقراءة إلى غايّة نهايّة السّطر، يعني عند العودة إلى السّطر فهو مدخل آخر، ولكن في بعض التّعاريف المعجميّة نجدها تحوي أبياتا شعريّة كشواهد، وطبعا كتابة هذه الأبيات تستدعي الرّجوع إلى السّطر عند كتابة بيت شعري آخر، وعندها لا يواصل البرنامج قراءة الأبيات التي تلي البيت الأوّل. والحال نفسه عند ورود جزء من النّص المعجمي في صفحة مواليّة؛ يعني عند نهايّة الصّفحة يُنهي القراءة. مثل مدخل: الوصيل في صيغة فعيل، كما هو موضّح في الصّورة:

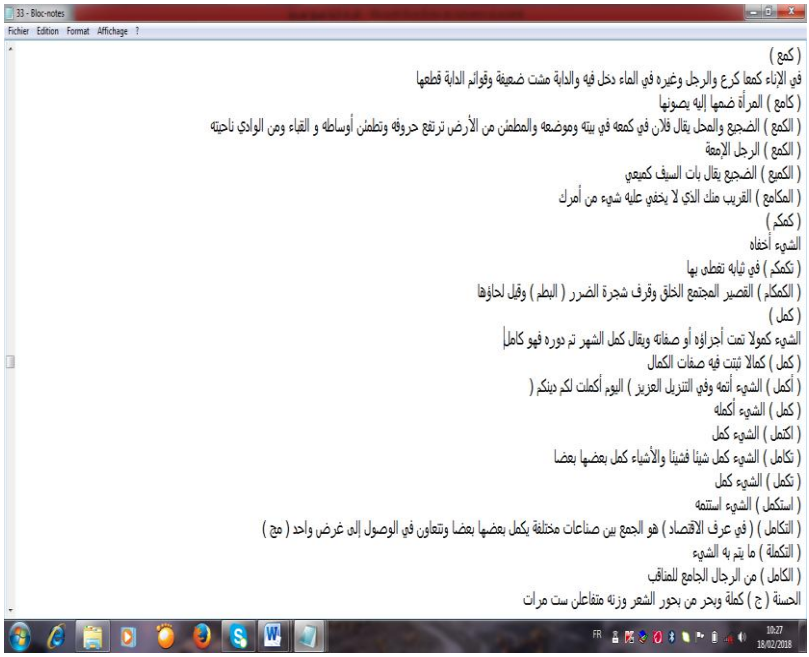
الاسترجاع الآلي للصيغ الصّرفيّة من المعجم الوسيط الإلكتروني ...



وأيضا المدخل المعجمي الرّتم والوقف في صيغة فعْل، ومدخل السّيوبة في صيغة فعولة.

– ووجدنا بعض المفردات بين قوسين داخل التّعريف المعجمي، والتي يقوم البرنامج بقراءتها مثلما تُقرأ المداخل المعجميّة الواردة بين قوسين، فمثلا في قراءة الصّيغة الصّرفيّة (فعْل) ظهرت في النّتائج كلمة (النّقد) التي وردت في التّعريف المعجمي لمدخل التّأثريّة، التي وردت بين معقوفين في المدونة الورقيّة ولكن كتبت في المدونة الرّقميّة بين قوسين. وكذلك

ظهرت في النّتائج كلمة (البطم) التي وردت في التّعريف المعجمي للمدخل كمكم، كما توضّح لنا هذه الصّورة.



– نجد أحيانا ورود مدخلين أو مفردتين داخل القوس الواحد بدل المدخل الواحد، وطبعا لا تتحقق القراءة؛ لكونه لا يطابق ما هو وارد في التّعليمة، ونمثّل لذلك ب (الكشح، الكُشاح) داء يصيب... والمدخل (القفن والقفنّ)

– جاءت جميع الأسماء والمصادر الواردة في المعجم معرّفة بالألف واللام، إلاّ أنّ هناك أفعالا حرفها الأوّل ألف بهمزة الوصل ثم لام مثل (التّجأ، التّبَن، التّقَط، التّبَس...) والبرنامج يجعل هذا النّوع من الأفعال أسماء تبتدئ بالألف واللام.

– وردت بعض المداخل على صيغة الجمع وليس على صيغة المفرد، فمثلا عند قراءة الصّيغة الصّرفيّة (فَعَل) ظهر في النّتائج مدخل الرّقن واحدته رقنة، وفي صيغة (فُعول) ظهر سعوف الواحد سعف، وقُحوف مفرده قِحف، وفي صيغة (مفاعل) ظهر مدخل مغارف الذي مفرده مغرفة. وفي صيغة (فعْل) ورد الزّهر الذي واحدته زهرة. فنكون عندها مجبرين على حذفها من قائمة الصّيغ الواردة في النّتائج عند تصفيتها.

– نجد بعض المداخل المعجميّة خاليّة من التّعريف المعجمي؛ أي يحيلك إلى مدخل آخر فيكتب أنظر إلى مدخل كذا، وعندها يجب علينا ملء الفراغ بالعودة إلى المدوّنة والمدخل الذي أُشير إليه، ونمثّل لذلك في صيغة فعْل بالمدخل (الغيْط) أنظر غوط.

– هناك بعض الصّيغ الصّرفيّة لا وجود لها كمداخل، وإنما نجدها في التّعريف المعجمي، فلا تظهر في النّتائج رغم انتمائها إلى قائمة تلك الصّيغ، من مثل (صعوبة في صيغة فُعولة، صعود في الصّيغة الصّرفيّة فُعول) وهذا مأخذ يؤخذ على واضعي المعجم، وليس على الذين قاموا برقمنة المعجم.

2– إشكالات متعلقة بخصائص اللغة العربيّة: إنّ من خصائص اللغة العربيّة أنّ بعض الصّيغ الصّرفيّة المشتقة لها نفس البنيّة أو الشّكل مع الصّيغ الجامدة وعند القراءة الآليّة لهذه الصّيغ المشتقة تظهر في النّتائج

الصّيغ المشتقة وكذلك الجامدة، فنكون مجبرين عندها على تصفيّة النّتائج؛ بالإبقاء على المشتقة منها فقط، ونمثّل لذلك بالصّيغ التّاليّة:

– ظهرت في النّتائج الصّيغ الجامدة: الوطر، الونع، اليلق، الينم...عند استرجاع الصّيغة الصّرفيّة (فعَل)

– ظهر في النّتائج مدخل: الإجاص، عند استرجاع الصّيغة الصّرفيّة (فِعال).

– ظهرت في النّتائج الصّيغة المقترضة: الكريك، عند استرجاع الصّيغة الصّرفيّة (فعيل).

لقد تمّت رقمنة المعجم الوسيط دون تدوين الحركات؛ أي دون تشكيل، ونحن نعلم أنّ الحركات في اللغة العربيّة تمكّننا من تمييز صيغة عن أخرى، وهذا يؤدي إلى ظهور صيغ تحمل نفس الشّكل لكن بحركات مختلفة، ونمثّل لذلك بما يلي:

– ظهرت في نتائج استرجاع الصّيغة الصّرفيّة (فِعالة) المفردات: زراعة صناعة حلاقة بكسر الحرف الأوّل، ولكن تظهر معها نُفايّة قُمامة حُثالة أي بضم الحرف الأوّل، وأيضا زَمالة نَقاهة وَقاحة بفتح الحرف الأوّل، ومسّاكة سمّاعة وزمّارة بتضعيف الحرف الثاني. والحال نفسه مثلا في صيغة (فُعال) مثل سعال ثغاء نواح يظهر في النّتائج أيضا المداخل التي صيغتها الصّرفيّة (فِعال) أو صيغ مشابهة لها شكلا دون حركات مثل:

نِزال نِكال نَشاط وفاة نساء بلاد دماغ نهار... ما جعلنا نقوم بتصفيّة النّتائج بالعودة إلى المعجم الوسيط الورقي، أو الذي تمّ مسحه ضوئيا (مجمع اللغة العربيّة بالقاهرة، 2004).

– ظهر في النّتائج استرجاع صيغة (فُعول) كلمات من مثل: الحشوة الحلوى الخروف الزّبون الطيور القهوة الكسوة... يعني أنّ الواو لا يكون حرف مد، وإنما أصلي في الكلمة.

– ظهرت في صيغة (فعلان) مثلا المفردات الانسان البرهان الطغيان الفنجان النّسيان الفرقان... يعني أنّ الشّكل نفسه لكن الحركات مختلفة.

– ظهرت في صيغة (فعيل) مفردات تشبهها شكلا مثل: التّنين الثريا الجزيء الحنيّة المدير... فنكون مجبرين عندها من إعادة تصفيّة النّتائج حسب الصّيغة التي نودّ اشتقاقها.

ولكون هذه الإشكالات تتعلق بخصائص اللغة العربيّة، فلا يمكننا تفاديها، بتغيير أشكال وبنى وحركات كلماتها.

الخاتمة: خلصنا من خلال تعاملنا آليا مع المدوّنة الرّقميّة (المعجم الوسيط) إلى وجوب إعادة النّظر في رقمنة المعجم الوسيط، ونفس الحال بالنّسبة لكل المدوّنات الأخرى التي يشتغلون على رقمنتها حاليا وذلك باعتماد هذه المقترحات:

– التّشكيل والتّضعيف وبخاصّة الكلمات التي يكون فيها لبس؛ لأنّ البرنامج يُظهر في نتائج البحث تماما ما يطابق التّعليمات التي قدّمت له، وفي غياب الحركات نكون مجبرين على تصفيّة النّتائج.

– الكتابة بطريقة دقيقة ومنظمة؛ مثل وضع الأقواس، والفراغ بين الكلمات، والانتقال إلى السّطر فقط عند نهايّة التّعريف المعجمي؛ لأنّ البرنامج يتعرف على الكلمات من خلال الفراغات الموجودة بينها.

– كتابة التّعريف المعجمي بعد كلّ مدخل معجمي، وليس الإحالة إلى مدخل آخر؛ لأنّ ذلك يجعلنا نبحث عن المدخل الآخر، لملء الفراغ في التّعريف المعجمي.

– الاكتفاء بكتابة مفردة واحدة كمدخل معجمي، لأنّ القراءة الآليّة تكون دقيقة وبالمطابقة.

– كتابة مفردات المداخل المعجميّة في صيغة الإفراد، لأنّ صيغة الجمع تشبه شكل أو بنيّة بعض الصّيغ الصّرفيّة

– كتابة الصّيغ الصّرفيّة المشتقة كمداخل معجميّة، وليس ضمن التّعريف المعجمي؛ ليسهل قراءتها آليا.

والأخذ بهذه المقترحات في رقمنة المدوّنات هو تيسير للتعامل معها آليا، وتحقيق لمشروع الذخيرة اللّغويّة، وكذا مشروع المعجم التّاريخي كما

هو مخطط له. ونقدّم نموذجا عن المشاريع الجاهزة للتعامل معها آليا وهو مدوّنة (Brown) الانجليزيّة التي تمثّل مكنزا، ومدوّنة ضخمة تمّ رقمنتها بطريقة صحيحة ومنظمة وجاهزة للتعامل معها آليا بلغة البرمجة البايثون (python).

قائمة المصادر والمراجع:

– رمضان عبد الله، الصّيغ الصّرفيّة في العربيّة في ضوء علم اللغة المعاصر، مكتبة بستان المعرفة ط1، طبرق،2006.

– طارق المالكي، أنطولوجيا حاسوبيّة للنحو العربي، نحو توصيف منطقي ولساني حديث، دار النّابغة للنشر والتّوزيع ، ط1، المغرب، 2015.

– عبد الرّحمن محمد طعمة، البناء العصبي للغة دراسة بيولوجيّة تطوريّة في إطار اللسانيات العرفانيّة العصبيّة، دار كنوز المعرفة، ط1، الأردن، 2017.

– مجمع اللغة العربيّة، المعجم الوسيط، مكتبة الشّروق الدوليّة، ط4، القاهرة، 2004.

– Gérard swinnen, apprendre à programmer avec python, , Editions Eyrolles 61, bd saint-germain 75240 Paris cedex 05 . 2009

Tarek ziadé, programmation python conception et – optimisation, Editions Eyrolles 61, bd saint-germain 75240 Paris cedex 05. 2e édition, 2009