

المنظمة العربية للتربية والثقافة والعلوم
مشروع المعجم العربي التفاعلي
اجتماع الخبراء، دمشق 26-28 مايو 2009

مشروع أية -سبل

القاموس العربي للتدقيق الإملائي مفتوح المصدر،

تجربة وآفاق.

طه زروقي¹، محمد كبداني²،

¹ المجلس الأعلى للغة العربية، الجزائر، المدرسة الوطنية العليا للإعلام الآلي بالجزائر

² طاقم مشروع المدقق الإملائي آيسبل، المغرب

البريد الإلكتروني : Taha_zerrouki@gawab.com, med.kebdani@gmail.com

ملخص:

يدخل مشروع المدقق الإملائي المفتوح المصدر ضمن مشروع شامل، هو مشروع آيسبل Ayaspell project ، الذي يهدف توفير أدوات المعالجة الآلية للغة العربية في بيئة البرامج الحرة. وقد دأب المشروع على بلوغ هدفين أساسيين:

الهدف الأول، برمجي: تحسين دعم اللغة العربية في برامج التدقيق الإملائي الحرة المتداولة.

الهدف الثاني، لغوي: بناء قاموس التدقيق الإملائي العربي الحر.

في هذا المقال، نورد تقريرا عما تم التوصل إليه في هذا المشروع.

الكلمات الأساسية : هانسبال، التدقيق الإملائي، القاموس، العربية.

<http://ayaspell.sourceforge.net/>

المواقع

http://sourceforge.net/project/showfiles.php?group_id=205373

التحميل

<http://forge.aaul.net/projects/ayaspell/>

<http://ftp.debian.org/debian/pool/main/a/ayaspell-dic/>

[http://ubuntu.mirror.ac.za/ubuntu-archive/ubuntu/pool/universe/a/ayaspell-](http://ubuntu.mirror.ac.za/ubuntu-archive/ubuntu/pool/universe/a/ayaspell-dic/)

[dic/](http://ubuntu.mirror.ac.za/ubuntu-archive/ubuntu/pool/universe/a/ayaspell-dic/)

<http://groups.google.com/group/ayaspell-dic>

القائمة

البريدية

<http://ayaspell.blogspot.com/>

المدونة

• الجائزة العربية المصادر المفتوحة لفضيلة الشيخ علي جابر العلي السالم الصباح الكويت 2008

الجوائز

• دعم من مدينة الملك عبد العزيز للعلوم والتقنية المملكة العربية السعودية لسنة 2008

• الجائزة الخاصة خلال الملتقى الإفريقي الثالث للبرامج الحرة 2007:

http://rall.logiciels-libres.org/rubrique.php3?id_rubrique=6

• مشروع السنة حسب الجمعية المغربية لتنمية الإعلاميات الحرة 2007:

<http://81.192.48.26/linuxmaroc/modules.php?name=News&file=article&sid=143>

إسماعيل حجير، "من أجل مدقق إملائي مفتوح المصدر" مقدم لنيل درجة الماجستير في علوم اللسان والتبليغ اللغوي . فرع العلاج الآلي للغة، مركز البحوث العلمية والتقنية لترقية اللغة العربية (الجزائر)، تحت إشراف الدكتور عمار بالة.

الدراسات

الأكاديمية

1. مقدمة

إن أهمية التدقيق الإملائي للكتابة تظهر جليا عند كتابة النصوص، رغبة منا في الارتقاء باللغة وسلامتها وقبولها وفهمها من الآخرين، مهما كانت اللغة التي ندون بها، وحين يستعين الكاتب بالحاسوب لإظهار الأخطاء واقتراح الحلول البديلة لها يحتاج برامج تدقيق متخصصة، مما فتح الباب على مصراعيه لمراكز البحوث والشركات المتخصصة لتصميم هذه الأدوات وتسويقها نظرا للحاجة الكبيرة إليها، وهو ما تفسره الأثمان الباهظة لها، إذ ناهز ثمن المدقق الإملائي العربي¹ المليون دولار أمريكي في السوق العالمية سنة 2007.

وإن كانت هذه الأثمان معقولة بالنسبة للشركات الكبيرة والمراكز العلمية، إلا أنها تقف عائقا أمام المستعمل العادي من أجل تدقيق نصوصه ورسائله ومكتوباته في البرامج التي لا تحتوي على هذه الخدمة، كما تؤرق الباحثين في اللغة العربية، من أجل تدقيق أبحاثهم حول معالجة اللغة وإمكانية الاستناد على أسس بحثية سابقة، مما تجعل الباحثين يدورون في حلقات مفرغة، ويعيدون اختراع ما سبق اختراعه، ويعوقهم عن إضافة لبنة جديدة في بناء العلم، ما دام عليهم في كل مرة إعادة البناء من الأساس.

ومنذ نشأة البرمجيات الحرة، ظهرت الحاجة الملحة إلى مدققات إملائية للغات المختلفة من أجل تدقيق النصوص في جميع أنواعها سواء مستندات أو رسائل، أو صفحات ويب، مما دعا إلى تصميم برامج تدقيق مفتوحة المصدر، كان أولها برنامج Ispell الذي تناسلت منه العديد من المدققات الأخرى مثل Myspell, Aspell, Hunspell. هذه البرامج زودت العديد من التطبيقات في مجال المكتبية وتحرير النصوص والصفحات والرسائل، ولكونها مفتوحة المصدر، فقد حُوت من أجل تطبيقات ومنصات كثيرة. ولما أشرقت شمس البرامج الحرة على المستخدم العربي، كان أول همه في البداية دعم العربية في التطبيقات من إمكانية كتابة وتعريب الواجهات، وسرعان ما اشتدت الحاجة إلى مدقق إملائي حر لمختلف الاستعمالات. وبما أن الحاجة أم الاختراع فقد ظهرت العديد من المحاولات، اهتمت في البداية بدعم العربية في المدققات الإملائية الحرة أولا، ومن ثم بناء قاموس عربي للتدقيق الإملائي.

كانت أول محاولة لمحمد الزبير من مجتمع عيون العرب² ببرنامج أسماه الدولي³، هو مدقق إملائي خالص، لكن تجربته فشلت بسبب القاموس الذي تم استعماله، والحاجة إلى إدماج هذا المشروع في التطبيقات الأخرى. وقد حاول محمد سمير إعادة برمجته باسم "بغداد"⁴، ولاقت تجربته نفس المصير، لكنه حاول مرة أخرى بناء قاموس إملائي لبرنامج Aspell الذي أضاف دعم اللغات غير اللاتينية. وظهرت محاولة مشابهة لإعداد قاموس لبرنامج Aspell، في مشروع Arabic-Spell⁵.

1- كولتك تنتهي من إبرام صفقة مع ميكروسوفت <http://www.ameinfo.com/ar-69687.html>

2 - <http://www.arabeyes.org>

3 - <http://www.arabeyes.org/project.php?proj=Duali>

4 - <http://home.foolab.org/cgi-bin/viewcvs.cgi/projects/baghdad/>

5 - <http://sourceforge.net/projects/arabic-spell/>

كان سبب فشل التجربتين السابقتين هو الافتقار إلى قاموس عربي حر، ما عدا قاموس تيم بوكولتر⁶ Tim Buckwalter المبني من مكنز مواد صحفية عربية. لكن جهل الباحث تيم بوكولتر بالعربية وعدم دقة المادة الصحفية المرجعية لغوياً، ترتب عنه انعدام دقة القاموس لاحتوائه على مفردات خاطئة رسماً ولغةً مما انجر سلباً على دقة المدققات الإملائية المعتمدة عليه (الدولي، بغداد، arabic-spell)، التي تعطي نتائج رديئة، تدفع المستعمل إلى الاستغناء عنها.

2. مشروع أية-سبل

يدخل المدقق الإملائي هذا، إذن، ضمن مشروع شامل، هو مشروع آيسبل Ayaspell project ، الذي يهدف توفير أدوات المعالجة الآلية للغة العربية في بيئة البرامج الحرة، منها، بالإضافة للمدقق الإملائي، أدوات الترادف المعجمي [Thesaurus](#)، التدقيق النحوي [Grammar-checker](#) وقواميس الأنظمة المحمولة [Embedded systems](#) .

فيما يتعلق بالتدقيق الإملائي، دأب المشروع على بلوغ هدفين:

الهدف الأول، برمجي: تحسين دعم اللغة العربية في برامج التدقيق الإملائي الحرة المتداولة وكان تحت إشراف الأستاذ طه زروقي.

الهدف الثاني، لغوي: بناء قاموس التدقيق الإملائي العربي الحر وكان تحت إشراف الأستاذ محمد كبداني.

وبعد دراسة مستفيضة لتاريخ المدققات الإملائية وإمكانياتها وملاءمتها للغة العربية، تبين لنا الآتي :

- تميز اللغة العربية ببعض الخصائص اللغوية غير المدعومة من قبل المدققات الحالية.
 - اختيار برنامج هانسبال Hunspell لدعمه اليونيكود، وتوفيره بعض المزايا التصريفية التي يمكن استغلالها من أجل العربية، وباعتباره المعتمد من قبل مشروع المكتب المفتوح أوبن أوفيس [OpenOffice.org](#)، مما يسهل انتشاره.
 - اختيار برنامج Aspell لدعمه اليونيكود، وكثرة البرامج المعتمدة عليه.
 - ضرورة بناء قاموس عربي حر جديد اعتماداً على المراجع العربية المعتمدة.
 - تقسيم العمل إلى قسم بناء القاموس حسب الخصائص المتوفرة، في مرحلة أولى، وقسم برمجي لدعم بعض الخصائص المميزة للعربية في مرحلة أخرى.
- كان الرهان العام والتحدي هو تحقيق نتيجة بمستوى لا يقل قيمة عما هو متداول من أدوات التدقيق غير الحرة. لذا تم تأسيس مشروع بمواصفات مهنية حديثة :
- بناء موقع إنترنت⁷ رسمي للمشروع، لعربية والإنجليزية والفرنسية، فيه آخر الأخبار الملفات للتحميل، والوثائق اللغوية.

6- <http://www.qamus.org/>

7- <http://ayaspell.sourceforge.net>

- قائمة بريرية⁸ لمناقشة الاختيارات والاقتراحات وتوضيح الحلول التقنية.
- مدونة⁹ حيث المقالات التي تنظر للمشروع ونصوص تفسر المقاربات وتشرح المنهجيات.

وقد تم تقسيم المشروع إلى مرحلتين :

- المرحلة الأولى بناء القاموس اللغوي، باستعمال الخصائص المتوفرة في برنامج هانسبال، وكذا إضافة خاصية إغفال التشكيل في مرحلة أولى، وقد انتهت هذه المرحلة بإصدار القاموس ونشره على الإنترنت.
- المرحلة الثانية : تحسين آلية برنامج التدقيق وإضافة المزايا العربية، وإعادة هيكلة القاموس ليتلاءم مع التحسينات.

3. المرحلة الأولى من المشروع :

اهتمنا في هذه المرحلة بجمع المادة اللغوية من أجل بناء قاموس يستعمل خصائص البرنامج الحالي، مع إضافة خاصية إغفال التشكيل برمجيا.

أ- تجاهل التشكيل والتطوير

وقد عملنا على إضافة خاصية إغفال التشكيل والتطوير، في المرحلة الأولى، نظرا للحاجة الملحة إليها واستحالة حلها بأي طريقة أخرى. هذه الخاصية تقوم بتجاهل الحركات والتطوير قبل تدقيق النص، وذلك لغياب الحركات عادة عن النص العربي، كما أن مفردات القاموس غير مشكولة. وتم برمجتها بتسيق مع المطور الأصلي لبرنامج هانسبال ونشرت ابتداء من الإصدار 1.1.5، ويتم الإشارة إلى الحروف التي تغفل بتفعيل خاصية IGNORE في ملف الزوائد. هذه الخاصية يمكن أن تستعمل من أجل لغات أخرى غير العربية.

ب- بناء القاموس

استدعى غياب قاموس عربي حر، بناء واحد يلبي شروط التدقيق الإملائي بالاعتماد على المعاجم اللغوية العربية المتداولة التراثية والحديثة. من هذه المعاجم أساساً معجم الأفعال العربية (مجموعة (Bescherelle)، المعجم الوسيط، المعجم الغني، معجم المحيط ولسان العرب. هذه هي إذن الروافد المهيكلة لقاموس آيسبل الذي أصبح ثاني قاموس حر متوفر على الشبكة، حر بمعنى خضوعه للرخصة العمومية الشاملة [GPL].

يتكون قاموس التدقيق الإملائي من ملفين هم ملف المفردات المصنفة مرفقة بأسماء القواعد التصريفية الممكنة لكل واحدة منها، وملف الزوائد أو قائمة القواعد التصريفية الممكن تطبيقها على المفردات. وقد تطلب إنشاء القاموس بشقيه أكثر من 2000 ساعة عمل على مدى قرابة سنتين من النشاط المتواصل (أبريل

⁸ -<http://groups.google.com/group/ayaspell-dic>

⁹ -<http://ayaspell.blogspot.com/>

2006 - يناير 2008) تم خلالها تحليل آلاف المفردات وتصنيفها وتوليدها حسب القواعد النحوية والصرفية، وتحديد معناها لتمييز تعدي الأفعال. إجمالاً، تمت معالجة أكثر من 50.000 مفردة تتوزع على ما لا يقل عن 10.000 فعل عربي، و40.000 اسم وعشرات الحروف والأدوات النحوية وما استثنى منها.

1. تصريف الأفعال

مثلت معالجة الفعل الشطر الأكبر من المشروع لكثرة التصريفات الممكنة لكل فعل وتعدد الحالات الإملائية فيها من إعلال وضبط للهمزة وتضعيف وإدغام، ويعتمد تدقيق الأفعال على مادة لغوية تحتوي 10.000 فعل عربي أصلي، أضيفت إليها الأشكال الخاصة بالإبدال والإعلال والتضعيف والإدغام وضبط الهمزة، فارتفع عدد المفردات في قاموس آيسبل Ayaspell إلى ما يقارب 15.000 (14523 مفردة بالضبط).

وقد وضعت القواعد التصريفية للأفعال في ملف الزوائد لتغطي كل صيغ التصريف الممكنة في اللغة العربية ما عدا صيغ المؤكد.

ويمكننا في هذه الفقرة أن نقارن القاموس ومع المدقق الإملائي لشركة كولتك المستعمل في برنامج ميكروسوفت أوفيس:

الحالة	مدقق آيسبل	مدقق كولتك	مثال
أفعال القلوب	جزئي	لا	نقول نطننا وتظنينك ولا يجوز قول نضربنا وتضربينك.
التعدي إلى مفعولين	ليس بعد	لا	يعطيكموها
الأفعال النادرة	نعم	لا	ائتثر - أوجى
الشكل/الحركة	ليس بعد	نعم	الكلمة المشكولة (كَتَبَ) صحيحة.
اعتماد كامل للمبني للمجهول	نعم	جزئياً	خوطئ من (خطئ) و شوددت من (شاد).
سابقتين تتضمن همزة الاستفهام	نعم	لا	أوتدري؟ أفتعلم؟
ثلاث سوابق	نعم	لا	أفستكتبها؟
صيغ التوكيد	ليس بعد	لا	ليكتبنا
التعدي النسبي للأفعال اللازمة	نعم	لا	وكم من انتصار انتصرناه بفضل جهاد شعوبنا!

2. معالجة الأسماء والحروف

بعد الانتهاء من معالجة الأفعال لصعوبتها، انكب المشروع على دراسة مشتقات الأفعال بأصنافها، اسم مفعول، اسم فاعل، مبالغة، أسماء التفضيل وصفات مشبهة. ومن ثم تصنيفها وتوليد الهيئات الصرفية

الممكنة منها (المؤنث، المثنى وجمع السالم) حسب القواعد العربية المعروفة. ووضعنا في القاموس مداخل تحتوي على الكلمة في صيغة المفرد المذكر أو جمع التكسير واستثناءً على هيئة المفرد المؤنث أو جمع السالم.

إحصائيات الأسماء والحروف

الفئة	العدد	ملاحظة
اسم-جامد	10328	
مصدر	13372	
اسم-الفاعل	8406	
اسم-مفعول	1807	
مبالغة-اسم-الفاعل	2066	
صفة-مشبهة	1017	
اسم التفضيل	378	
اسم منسوب	862	
مفردات أخرى	4248	تتوزع بين الصفة والنسبة والاسم الجامد
إجمالي	42484	
استثناءات	611	الحروف والأدوات النحوية وما استثنى من هيئات صرفية

من حيث التصريف، نجد في القاموس صيغ الإضافة اللفظية التي أهملتها المدققات الإملائية المغلقة (مثل: المقيمي الصلاة)، وصيغ كثيرة متعلقة بالسوابق مثل أوكاتبُ (أوكاتبون؟) أو وِلكاتبُ (ولكاتبون) التي أهملتها المدققات الإملائية الأخرى لندرة استعمالها في الكتابات الحديثة على ما يبدو.

3. نتائج المرحلة الأولى

في نهاية المرحلة الأولى، صدرت الإصدار النهائي الأول لقاموس التدقيق الإملائي، يوم 10-01-2008 ويمكن تحميله من الإنترنت¹⁰، وقد استقبل هذا القاموس استقبالا حسنا في عالم البرمجيات الحرة، حيث تم تحويله لاستعماله في العديد من البرامج مثل المتصفح فايرفوكس (شكل 1)، وبرنامج البريد الإلكتروني Thunderbird (شكل 3)¹¹، كما يتم استعماله في برنامج المكتب المفتوح (شكل 2)¹². ويمكن استعماله في العديد من البرامج التي تعتمد على برنامجي هانسال وأسبل. كما يمكن استعماله وإعادة تعديله حسب رخص البرامج الحرة GPL/LGPL/MPL. ويعمل على المنصات المختلفة كنظام وندوز (شكل 1) ولينكس

¹⁰ - <http://ayaspell.sourceforge.net/>

¹¹ تحميل القاموس العربي الإملائي برنامج فايرفوكس وثناندر بيرد <https://addons.mozilla.org/ar/firefox/addon/3677>

¹² - تحميل القاموس العربي للمكتب المفتوح

http://wiki.services.openoffice.org/wiki/Dictionaries#Arabic_.28North_Africa_and_Middle_East.29

(شكل 2؛ شكل 3).

كما تحصل المشروع على العديد من الجوائز أهمها الجائزة العربية المصادر المفتوحة لفضيلة الشيخ علي جابر العلي السالم الصباح الكويت لسنة 2008، وتلقى المشروع دعماً من مدينة الملك عبد العزيز للعلوم والتقنية بالمملكة العربية السعودية سنة 2008.

4. المرحلة الثانية من المشروع

من خلال الخبرات المكتسبة في المرحلة الأولى من المشروع، ظهرت الحاجة إلى العديد من المزايا الجديدة والرغبة في تعديل برنامج التدقيق من أجل ملاءمة أحسن لقواعد اللغة العربية، إن النفاص المكتشفة في البرنامج الأصلي لهانسبل، والتي استدعت إحداث تغييرات لدعم كامل للغة العربية بميزتيها الاشتقاقية والالتصاقية، مهدت الطريق للحديث عن المزايا الجديدة المطورة التي ساهمت بشكل جدّ فعّال في حلّ بعض المشاكل التي واجهها هذا البرنامج عند تدقيقه للوثائق المكتوبة باللغة العربية. فمن خلال دراسة برنامج هانسبل وبعد الوقوف على نقاط ضعفه، تبين بوضوح بأن هذا البرنامج في حاجة إلى تطوير وإضافة بعض الوظائف وإدخال بعض التعديلات تساعد على تدقيق النصوص العربية، مثل : مسألة إغفال الشكل والتطوير (عولجت في المرحلة الأولى)، ومسألة الإدغام ومسألة الإعلال والإبدال، ومسألة تراكم الزوائد.

أ- مسألة الإدغام

عند تعامل المدقق مع بعض الكلمات مثل : أنا رددت، هي رددت، يتحتم علينا حالياً معالجة كل حالة على حدى من خلال مدخل إضافي في القاموس مما يؤدي إلى تعقيده وصعوبة صيانتها وحلّ المشكل أضيفت خاصية الإدغام 'GEMENATING' من خلال وضع علم خاص في ملف الزوائد. تعني هذه الخاصية أن العلم الذي سيأتي بعدها سيسعمل لفرض الإدغام في كل قاعدة يكتب فيها، ثم يتم استعماله في قاعدة الزيادة لفرض الإدغام.

ب- مسألة الزوائد المتوسطة

تدخل هذه الخاصية ضمن مسألتني الإبدال والإعلال التي تعتبر واحدة من أهم وأصعب المسائل التي تواجه التدقيق الإملائي المبني على سوابق ولواحق، وخاصة في اللغة العربية، لأنه في هذه الحالة نقوم بدراسة التغييرات الداخلية التي تطرأ على جذع الكلمة.

نعرف الزائدة المتوسطة بأنها زائدة تدرج داخل الجذر أو الساق¹³.

الزيادة المتوسطة هي شائعة في اللغة الإندونيسية، واللغات الأفيانوسية والأسترالية الآسيوية. وتوجد في العربية وحتى في حالات شاذة في الإنجليزية مثل sing, sang, sung, song، وفي استعمالات

¹³ - <http://en.wikipedia.org/wiki/Infix>

لغوية معينة. كما يمكن اعتبار الاشتقاق بزيادة حروف وسط الجذر نوعاً من الزيادة المتوسطة. ونستعملها في تصريف الأفعال الجوفاء والمهموزة، مثل قام، يقوم، قم، قيم؛ سأل، سئل. فيصبح تصريف الفعل قام في المضارع بإضافة السابقة ي.و والتي تعني أن الحرف الياء زائد في البداية، والنقطة تعبر عن أول حرف في الكلمة، والواو يزداد بعد أول حرف.

قم+ ي.و (سابقة) = يقوم

قم+ ا.وا (لاحقة) = قاموا

وقد تمت تجربتها في برنامج هانسبال وكذا أسبل، وقد أبدى مطور البرنامج الأصلي رغبة في استغلالها في اللغات الأخرى.

ج- مسألة الزوائد المزدوجة

من خلال تصريف الأفعال وخصوصاً في المضارع، يتبين لنا وجود تلازم بين السوابق واللواحق في حالات عديدة، مثل كتب+ ي-ون = يكتبون، ونلجأ حالياً إلى تركيبها من خلال الفصل بين السابقة واللاحقة وتركيبها من خلال خاصية الزيادة المدورة circumfix مما يعقد قواعد الزيادة ويصعب صيانتها. لذا نرغب في اختزال تلازم السابقة واللاحقة من خلال كتابتها في قاعدة واحدة، نسميها الزائدة المزدوجة. وقد تم تجربتها على برنامج هانسبال وكذا برنامج أسبل وأعطت نتائج جيدة. ومن الممكن أن تحتوي أيضاً على زائدة متوسطة.

د- مسألة تراكب الزوائد

فمن خلال الدراسات التي تناولت اللغة العربية، تبين بأن الكلمة العربية تقبل سابقتين ولاحقتين. وهو غير ممكن حالياً في برنامج هانسبال الذي يدعم سابقة واحدة ولاحقتين، لذا نختزل السوابق في مجموعة مركبة من السوابق، فبدلاً من أن نعتبر السابقة الأولى (و) والسابقة الثانية (ال)، نعتبر المجموعة (وال) سابقة. لذا نفكر في زيادة قدرة البرنامج على معالجة تعدد السوابق واللواحق.

5. آفاق المشروع:

يعاني القاموس من بعض مظاهر الضعف وهي عموماً :

- تداخل المستويات بين المستوى الإملائي والمستوى النحوي والمستوى المعنوي، مثل إدراج تعدي الفعل في تدقيق الكلمة، بينما يجب أن يكون في المستوى النحوي.
- تعقيد قائمة الكلمات وقائمة قواعد تطبيق الزوائد، مما يجعل إضافة مفردة جديدة للقاموس، عملاً صعباً.
- صعوبة استخراج المعلومات من القاموس من أجل إعادة استعماله في تطبيقات أخرى.
- صعوبة تكييف بعض الحالات مع قواعد البرنامج، مثل التغييرات الحاصلة في وسط الكلمة.

- القاموس موجه أساسا لبرنامج هانسبال، ويمكن استعماله في برنامج أسبل، في الإصدارات الجديدة.
 - بطء في اقتراح البديل وضعف الدقة.
 - القاموس الحالي، لا يدعم التشكيل.
- لذا نرغب في مستقبلا في تلبية الحاجيات ومعالجة المسائل السابقة من خلال :
- التفريق بين المستويات المختلفة الإملائي والنحوي، تمثيل التعديية في مستوى آخر.
 - تبسيط قواعد الإلحاق، ووضع طبقة مجردة، للانتقال من قاموس عام، إلى قاموس خاص ببرنامج التدقيق، مما يسهل العمل على القاموس وطريقة عرضه وصيانتته وإضافة كلمات جديدة، وتمكين إعادة استعمال القاموس في تطبيقات أخرى.
 - إضافة المزايا الخصوصية للغة العربية في البرنامج ومن ثم تطويع البرنامج مفتوح المصدر وليس تطويع قاعد اللغة.
 - تجريد القاموس، وتوسيع مجالات استعمالته في برامج تدقيق أخرى ما عدا هانسبال.
 - تحسين فعالية المدقق الإملائي من حيث السرعة والنجاعة
 - تطوير خاصية اعتبار التشكيل في الكلمات.
 - إعادة هيكلة القاموس بطريقة تتماشى مع التدقيق النحوي مستقبلا.
 - إعادة بناء القاموس بالاقتنصار على المستعمل من المفردات والتركيز على التصريف المتداول فعلا في الكتابات الحديثة.
 - هيكلة القاموس في عدة أصناف ومستويات للغة، من القاموس الخفيف المتداول، إلى القاموس الشامل.

خاتمة

عملنا من خلال هذا المشروع على توفير بديل برمجي مفتوح المصدر للتدقيق اللغوي يمكن استغلاله في العديد من البرامج الحرة، ويميز هذا المشروع أنه جاء من حاجة ملحة، لذا ولد حرا طليقا بين المستخدمين المهتمين دون أي قيد وشرط.

وقد عملنا عليه من جانبين هما بناء القاموس حسب المزايا الموجودة وتوفيره في أقرب وقت، ومن ثم تطوير الجوانب البرمجية لتحسين أداء البرنامج وإعادة هيكلة القاموس في المستقبل القريب.

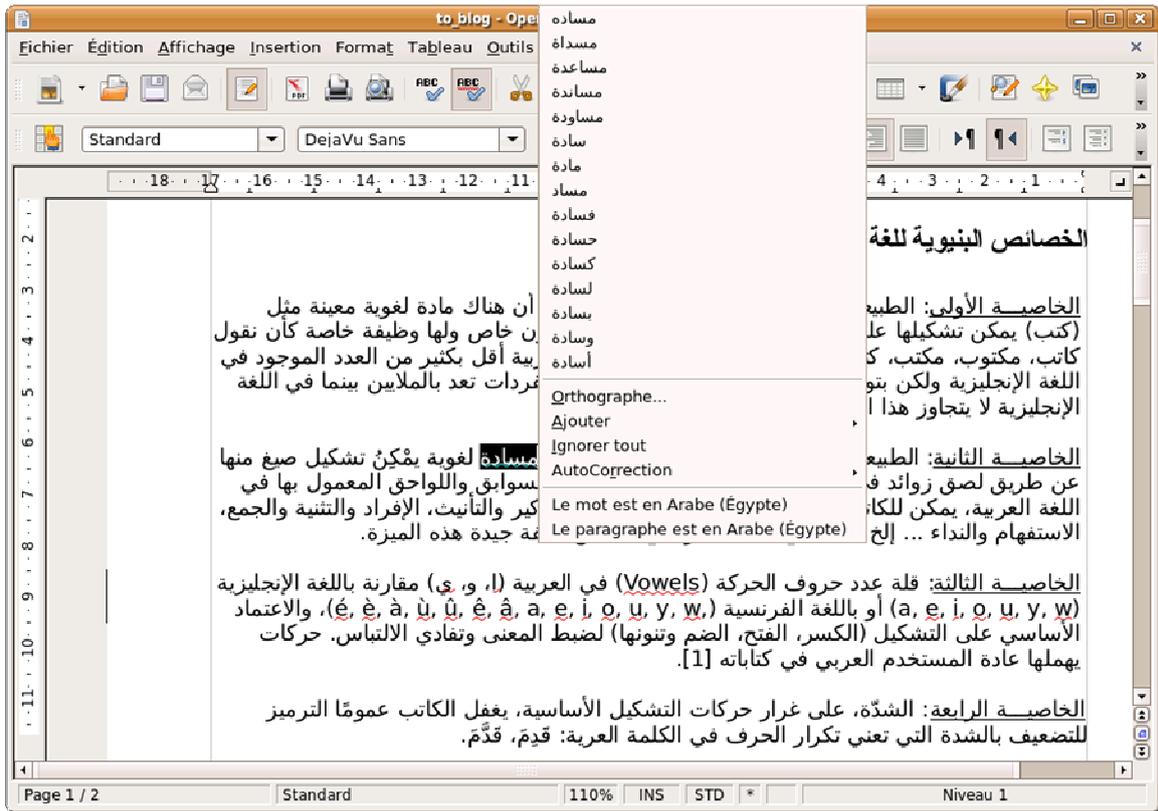
6. المراجع

1. طه زروقي، "الجوانب البرمجية لدعم العربية في المدقق الإملائي مفتوح المصدر هانسبال"، الأيام الدراسية حول المعالجة الآلية للغة العربية، المركز الجامعي بشار، ماي 2007.
2. طه زروقي، "الزوائد المتوسطة والمزدوجة في المدقق الإملائي مفتوحة المصدر"، ندوة " البرمجيات التطبيقية باللغة العربية: خطوات نحو الإدارة الإلكترونية"، المجلس الأعلى للغة العربية، الجزائر، 09-10 ديسمبر 2007.
3. إسماعيل حجير، "من أجل مدقق إملائي مفتوح المصدر" مقدم لنيل درجة الماجستير في علوم اللسان والتبليغ اللغوي . فرع العلاج الآلي للغة، مركز البحوث العلمية والتقنية لترقية اللغة العربية 2008.
4. مشروع القاموس العربي للمدقق الإملائي هانسبال <http://ayaspell.sourceforge.net>
5. موقع برنامج هانسبال <http://hunspell.sourceforge.net>
6. Ispell المدقق الإملائي <http://www.gnu.org/software/ispell/ispell.html>
7. Aspell المدقق الإملائي <http://aspell.net/>
8. <http://www.sakhr.com/Technology/Correction> موقع مدقق صخر
9. مدقق كولتك <http://www.coltec.net/ArabicProofingTool.aspx>
10. مشروع الكلمات المستبعدة arabic stopwords <http://arabicstopwords.sf.net>
11. مشروع قطرب لتصريف الأفعال <http://qutrub.arabeyes.org>
- 12.

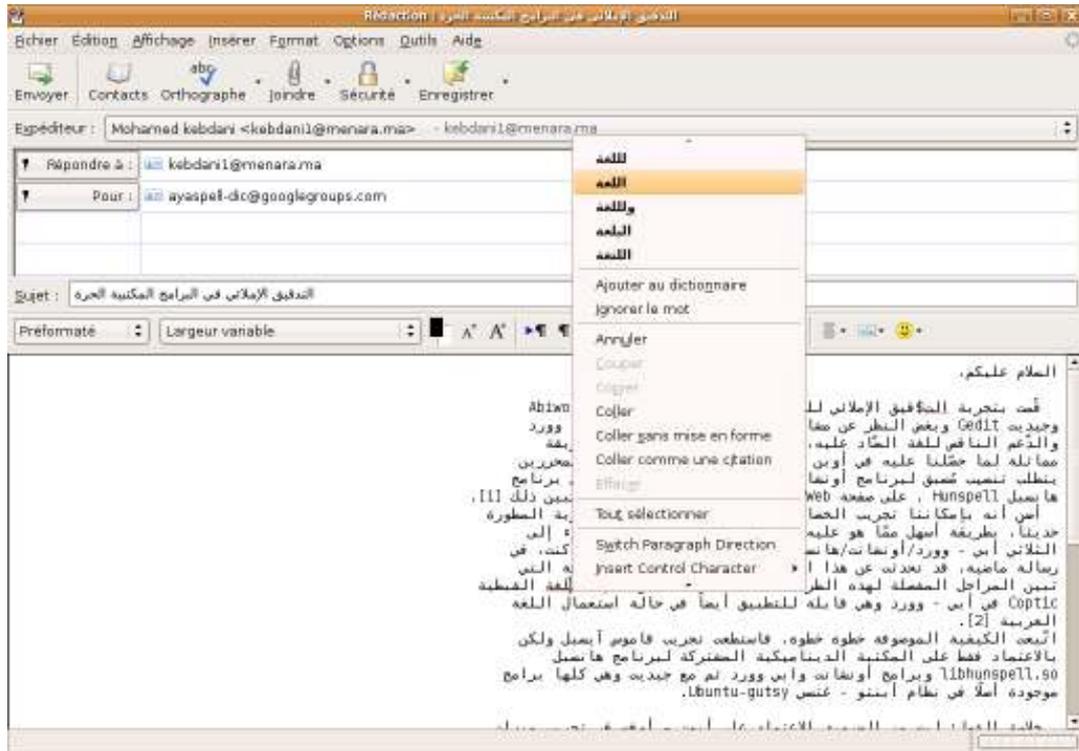
7. الأشكال



شكل 1 استعمال القاموس العربي في برنامج Firefox على منصة وندوز



شكل 2 استعمال القاموس العربي في برنامج OpenOffice على لينكس



شكل 3 استعمال القاموس العربي في برنامج Thunderbird على لينكس