

الندوة الدولية حول الأداة المعلوماتية في خدمة اللغة العربية في مواجهة تحديات العولمة،
جامعة أبوبكر بلقايد، تلمسان، الجزائر 15-16 نوفمبر 2008.

المدقق الإملائي العربي الحر

لمشروع آيسبل

طه زروقي*، محمد كبداني**

med.kebdani@gmail.com المغرب. ** taha_zerrouki@gawab.com الجزائر، *

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 :

التدقيق الإملائي، القاموس، العربية، هانسبال

البطاقة التقنية

القاموس : محمد كبداني، المغرب البرمجة: طه زروقي، الجزائر	الطاقم
http://ayaspell.sourceforge.net/ http://ayaspell.sourceforge.net/ar.html http://perso.menara.ma/~kebdani/ayaspell-dic/	المواقع
http://sourceforge.net/project/showfiles.php?group_id=205373 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-dic/	التحميل
http://groups.google.com/group/ayaspell-dic	القائمة البريدية
http://ayaspell.blogspot.com/	المدونة
<p>الجائزة العربية المصادر المفتوحة لفضيلة الشيخ علي جابر العلي السالم الصباح الكويت</p> <p>دعم من مدينة الملك عبد العزيز للعلوم والتقنية المملكة العربية السعودية و</p> <p>الجائزة الخاصة خلال الملتقى الإفريقي الثالث للبرامج الحرة:</p> <p>http://rall.logiciels-libres.org/rubrique.php?id_rubrique=6</p> <p>مشروع السنة حسب الجمعية المغربية لتنمية الإعلاميات الحرة:</p> <p>http://81.192.48.26/linuxmaroc/modules.php?name=News&file=article&sid=143</p> <p>إسماعيل حجير، "من أجل مدقق إملائي مفتوح المصدر" مقدم لنيل درجة الماجستير في علوم اللسان و التبليغ اللغوي . فرع العلاج الآلي للغة، مركز البحوث العلمية والتقنية لترقية اللغة العربية (الجزائر)، تحت إشراف الدكتور عمار بالة.</p>	الجوائز
	الدراسات الأكاديمية

إن أهمية التدقيق الإملائي للنص المكتوب لا تخفى على أحد، بحيث أن أول ما يشغل بال المرء أثناء الكتابة، أيًا ما كانت اللغة المستعملة، هو خلو النص المكتوب من الأخطاء الإملائية ولتسهيل هذا العمل، يستعين الكاتب بالحاسوب والبرمجيات المتخصصة لإظهار الأخطاء واقتراح الحلول البديلة لها، مما فتح الباب على مصراعيه لمراكز البحوث والشركات المتخصصة لإنتاج هذه الأدوات والمعينات المعلوماتية - ذات المصدر المغلق - والتسابق فيما بينها من أجل الحصول على ما يسمى بمصححات أو مدققات إملائية بوظائف متطورة أكثر فأكثر تلبى حاجة مستعملي الحاسوب في المجال المكتبي وتغني خدمات موزعو البرامج المعلوماتية التي تدفع الملايين من أجل الحصول عليها.

على صعيد البرامج الحرة وإلى حدود 2006، لم يكن هناك أي مدقق إملائي عربي حر عملي، رغم تعدد المحاولات العربية المرتبطة بطريقة مباشرة أو غير مباشرة بمجموعة عيون العرب Arabeyes أهمها محاولتي الأخوين محمد الزبير ببرنامج "الدوّلي" Duali ومحمد سمير ببرنامج "بغداد" Baghdad . ثم جاءت محاولتنا برنامج التدقيق الإملائي [هانسبل](#) Hunspell المعتمد من قبل مشروع المكتب المفتوح أوبن أوفيس [OpenOffice.org](#) ومن برنامج أسبل Aspell الأكثر شيوعاً في الوسط البرمجي الحر. البرنامجان مطوران أصلاً للغات اللاتينية ولكن بعد إضافة خاصية اليونيكود ودعم ثنائية الاتجاه إليهما أصبحا مؤهلين لدعم اللغات غير اللاتينية من ضمنها اللغة العربية.

بعد حصول دعم اللغة العربية في هذين البرنامجين، هانسبل وأسبل، توفرت إمكانية العمل على القواميس العربية الخاصة بالتدقيق الإملائي والتي بدونها لن تؤدي هذه البرامج وظيفتها. لم يكن في الساحة إلا قاموساً عربياً واحداً حرّاً هو قاموس [تيم بوكولتر](#) Tim Buckwalter المعتمد في بنائه على مكنز لغوي مكون من مادة صحفية عربية. ولكن للأسف يحتوي القاموس على مفردات خاطئة في نسبة كبيرة منها رسماً ولغة مما أثر سلباً على المدقق الإملائي التي تعتمد عليه وجعلها لا ترقى إلى المستوى المنتظر منها وهذا مثل المدقق [arabic-spell](#) ، منتج شركة [غوغل](#) Google.

أمام هذا الوضع، تم تأسيس مشروع بمواصفات مهنية حديثة: [موقع إنترنت](#) أول مؤقت باللغة العربية ثم آخر رسمي [بالإنجليزية](#) و [العربية](#) ثم [الفرنسية](#) حيث يجد المهتم آخر الأخبار ويستطيع تحميل الملفات ويطّلع على الوثائق اللغوية ويكون على علم بالمنجزات و [قائمة بريدية](#) حيث تناقش الاختيارات وتوضع الاقتراحات وتوضح الحلول التقنية وأخيراً [مدونة](#) حيث المقالات التي تنظر للمشروع ونصوص تقسّر المقاربات وتشرح المنهجيات.

إطار المشروع

يدخل المدقق الإملائي، 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، يهدف المشروع إلى بلوغ هدفين هما:

- تحسين دعم اللغة العربية في برامج التدقيق الإملائي الحرة المتداولة وكان تحت إشراف الأستاذ طه زروقي.
- بناء قاموس التدقيق الإملائي العربي الحر وكان تحت إشراف الأستاذ محمد كبداني.

الجانب البرمجي من المشروع

إن النقائص المكتشفة في البرنامج الأصلي لهانسبل، و التي استدعت إحداث تغييرات لدعم كامل للغة العربية بميزتيها الاشتقاقية والاتصافية، مهّدت 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 و التطويل و مسألة الإدغام و مسألة الإعلال و الإبدال.

أولاً، تجاهل التشكيل والتطويل

بتنسيق مع المطور الأساسي لبرنامج هانسبل، تمت برمجة خاصية مهمة لها علاقة بغياب التشكيل وعلامات التطويل من القاموس المرجعي وذلك ببرمجة خاصية التجاهل نحوها عند التدقيق ويتم ذلك بتفعيل خاصية IGNORE في ملف الزوائد حيث توضع الحروف والحركات التي سيتم إغفالها أمامها. هذه الخاصية يمكن أن تستعمل من أجل لغات أخرى غير العربية مثل اللغة العبرية التي تحمل بعض حروفها علامات النقط.

• ثانياً، مسألة الإدغام

عند تعامل المدقق مع بعض الكلمات مثل : أنا رددت، هي ردت، و التي كانت تحتم زيادة مدخل آخر للفعل في القاموس مما يؤدي إلى تعقيد هذا الأخير و كبر حجمه ولحلّ المشكل أصيقت خاصية الإدغام 'GEMENATING' من خلال وضع علم خاص في ملف الزوائد، دون اللجوء إلى إضافة مدخل جديد في

القاموس. في ملف الزوائد تعني هذه الخاصية أن العلم الذي سيأتي بعدها سيستعمل لفرض الإدغام في كل قاعدة يكتب فيها، ثم يتم استعماله في قاعدة الزيادة لفرض الإدغام.

ثالثاً، الزوائد المتوسطة

تدخل هذه الخاصية ضمن مسألتي الإبدال والإعلال التي تعتبر واحدة من أهم وأصعب المسائل التي تواجه التدقيق الإملائي المبني على سوابق و لواحق، وخاصة في اللغة العربية، لأنه في هذه الحالة نقوم بدراسة التغيرات الداخلية التي تطرأ على جذع الكلمة مقلاً قال، يقول، قلت.

رابعاً، مسألة الزوائد المزدوجة

إن تركيب السوابق واللواحق لكلمة معينة (جذر في القاموس) من أجل الحصول على كلمات أخرى يخضع في الغالب لقيود تنظم هذا التركيب و تحترم خصوصيات اللغة المراد تدقيقها. فمن خلال الدراسات التي تناولت اللغة العربية، تبين بأن الكلمة العربية تقبل سابقتين ولاحقتين. الزائدة السياقية أو الزيادة المزدوجة أو المتلازمة circumfix هي زيادة توضع على طرفي الجذر أو الساق. أي أنها مكونة من جزأين متلازمين سابقة تسبق الكلمة ولاحقة تلحق بها، ويكثر هذا النوع في الأفعال المضارعة كتب ي-كتب-ون.

إدماج المزاي البرمجية الجديدة:

لقد تم إدماج ميزة إغفال التشكيل والتطويل في الإصدارات الرسمية للبرنامج بينما لاتزال المزاي الأخرى قيد التجربة على برنامج هانسبال وكذلك برنامج Aspell.

الجانب القاموسي من المشروع

استدعى غياب قاموس عربي حر، العمل على بناء قاموس عربي 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 المهيكلية لقاموس آيسبل الذي أصبح أول قاموس عربي إملائي حر متوفر على الشبكة، حر بمعنى خضوعه للرخصة العمومية الشاملة [GPL].

يتكون القاموس من ملفين أولهما ملف المفردات الذي يحتوي على المفردات مرفقة بعلامات تدل على القواعد التي يمكن تطبيقها عليها، بينما نجد القواعد في ملف الزوائد. وقد تطلب إنشاء القاموس بشقيه (أكثر من 2000 ساعة عمل على مدى قرابة سنتين من النشاط المتواصل (أبريل 2006 - يناير 2008) وتحليل آلاف المفردات من فعل واسم وأداة وحرف وتصنيفها وتوليدها حسب القواعد اللغة العربية النحوية والصرفية، ثم تحديد معناها لتمييز الفعل اللازم والمتعدي لعقل أو غير عقل والصفة العائدة على عقل أو غير عقل ومعرفة الشاذ منها والعادي. إجمالاً، تمت معالجة أكثر من 50.000 مفردة تتوزع على ما لا يقل عن 10.000 فعل عربي، 40.000 اسم وعشرات الحروف والأدوات النحوية وما استنتهي من هذا أو ذاك.

المُكَوِّنُ الفَعْلِي

مثلت معالجة الفعل العربي الشطر الأول من المشروع الجانب الأكثر استهلاكاً للوقت واستدعت الوقوف على العشرات من المراجع اللغوية، ويعتمد التدقيق الإملائي للفعل العربي للمشروع على مادة لغوية تحتوي على ما يفوق 10.000 فعل عربي وبعد إضافة الأشكال الخاصة بالإبدال/الإعلال والتضعيف/الإدغام وما يجري على الهمزة من تحولات، ارتفع عدد المفردات في قاموس آيسبل Ayaspell إلى ما يقارب 15.000.

بالنسبة للهياكل المتولدة بواسطة ملف الزيادات فإنها تغطي كل صيغ التصريف الممكنة في اللغة العربية ما عدا صيغ المؤكد وتتركب هذه الأفعال مع كل الزيادات السابقة الممكنة (سوابق Prefixes) وبأغلب الزيادات اللاحقة (لواحق Suffixes) باستثناء تلك المتعلقة بالتعدي لمفعولين. من مميزات مدقق هانسبل Hunspell معتمداً على قاموس آيسبل Ayaspell في الجزء الخاص بالأفعال بالنسبة لمدقق MsOffice:

اعتماد تصريف أفعال القلوب: جزئياً [مدقق MsOffice: لا] فمثلاً نقول نظننا وتظنينك ولا يجوز قول نضربنا وتضربينك.

اعتماد التعدي إلى مفعولين: ليس بعد [مدقق MsOffice: لا] نحو يعطيكموها.

اعتماد الأفعال النادرة: نعم [مدقق MsOffice: لا] نحو ائثر - أوجى.

اعتماد الشكل/الحركة: ليس بعد [مدقق MsOffice: نعم] نحو اعتبار الكلمة المشكولة (كُتِبَ) صحيحة.

اعتماد كامل للمبني للمجهول: نعم [مدقق MsOffice: جزئياً] مثل خوطئ من (خطئ) و شوددت من (شاد).

اعتماد الهياكل المتغيرة في صيغة الأمر للأفعال المهموزة والمضاعفة: نعم [مدقق MsOffice: لا] مثل "بديا" و"بدئا" من أدب و"دو" و"دديا" من ود.

اعتماد سابقتين تتضمن همزة الاستفهام: نعم [مدقق MsOffice: لا] مثل أوتدري؟ أفتعلم؟

اعتماد ثلاث سوابق: نعم [مدقق MsOffice: لا] نحو أفسنتكتبها؟

اعتماد صيغ التوكيد: ليس بعد [مدقق MsOffice: لا] نحو ليكتبنان.

اعتماد التعدي النسبي للأفعال اللازمة: نعم [مدقق MsOffice: لا] نحو: وكم من انتصار انتصرناه بفضل جهاد شعوبنا...!!!؛-)

تجدر الإشارة إلى كون هذه النتيجة، حصلنا عليها باستثمار خصائص هانسبل Hunspell الأصلية فقط.

المُكَوِّنُ الاسمي والحرفي

بالإضافة لأسماء الجامدة والمصادر وصيغ النسبة المرتبطة بهما، انكب المشروع على دراسة مشتقات

الأفعال بأصنافها، اسم مفعول، اسم فاعل، مبالغة، أسماء التفضيل وصفات مشبهة. بعد نسخها من المعاجم المرجعية، تم تصنيفها وتوليد الهيئات الصرفية الممكنة منها (المؤنث، المثنى وجمع السالم) حسب القواعد النحوية للغة العربية المعروفة. مداخل القاموس تحتوي إذن على الكلمة في صيغة المفرد المذكر أو جمع التكسير واستثناءً على هيئة المفرد المؤنث أو جمع السالم.

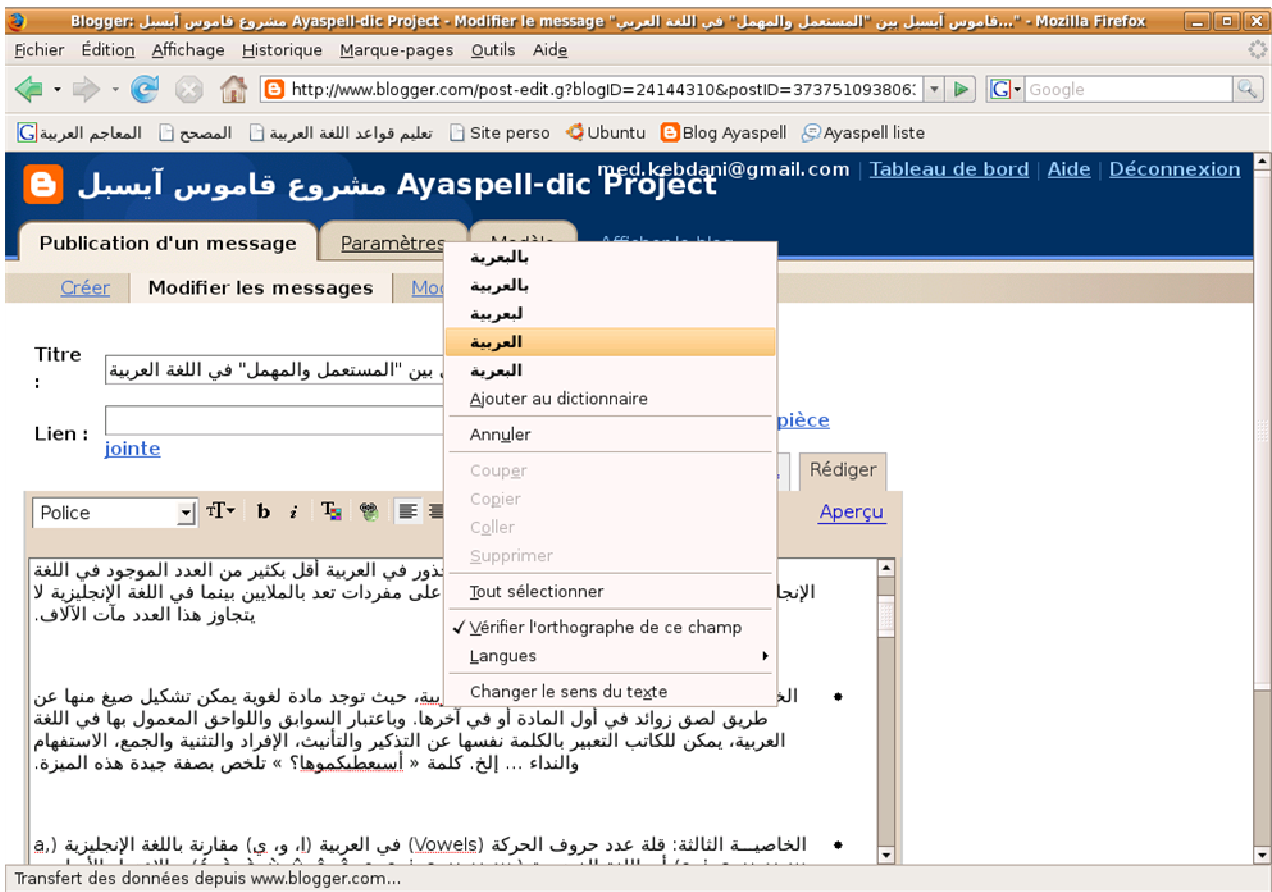
إجمالاً، نجد في القاموس : 10328 اسم-جامد، 13372 مصدر، 8406 اسم-الفاعل، 1807 اسم-مفعول، 2066 مبالغة-اسم-الفاعل، 1017 صفة-مشبهة، 378 اسم التفضيل، 862 اسم منسوب، بالإضافة إلى مفردات أخرى بعدد 4248 تتوزع بين الصفة والنسبة والاسم الجامد نسخت من معاجم مختلفة ومكانز ونصوص متنوعة. الحصيلة هي إذن: 42484 مفردة أما عدد الحروف والأدوات النحوية وما استثنى من هيئات صرفية في قاموس آيسبل وصل إلى 611.

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

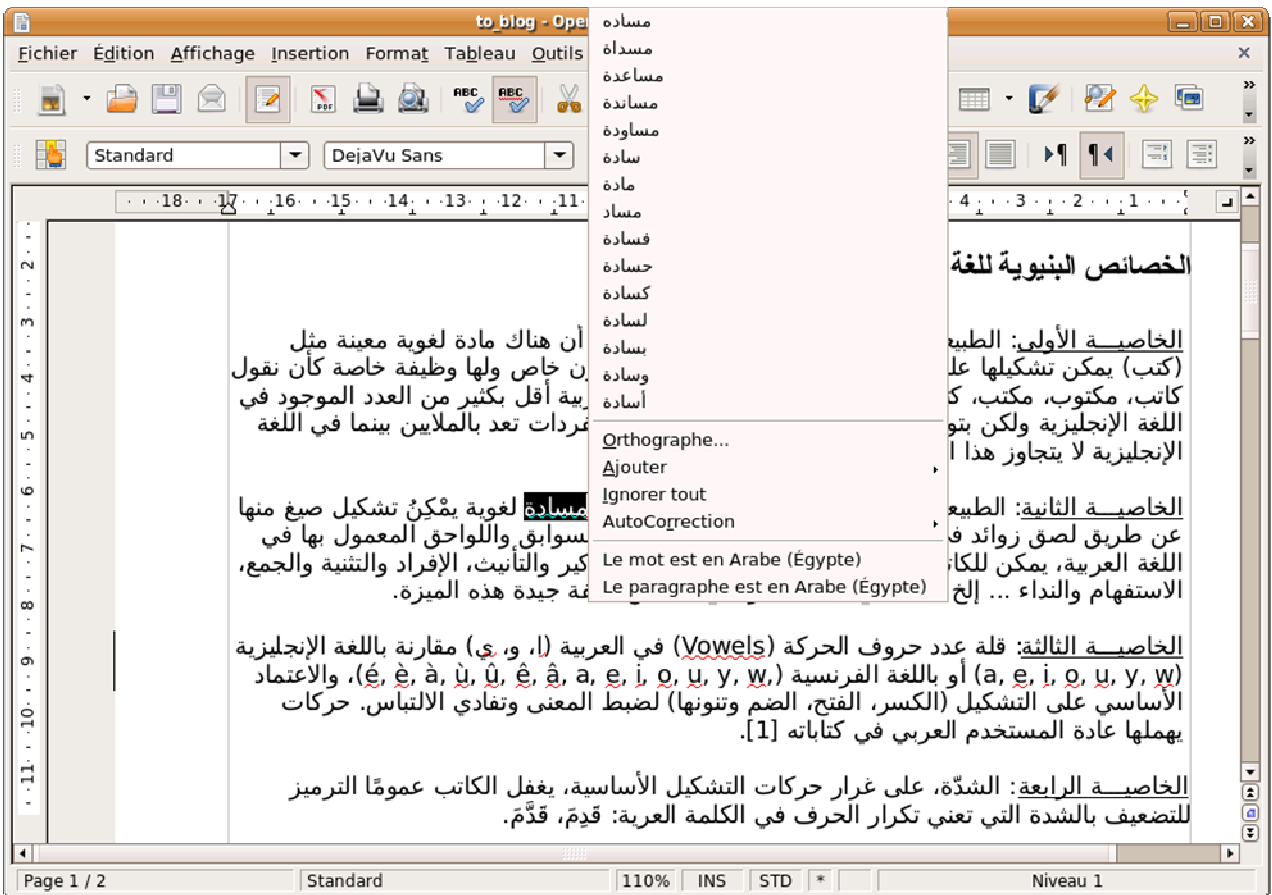
تميز آخر في قاموس آيسبل هو تعيين تنوين النصب "صراحة" لتفادي الأخطاء المتعلقة بالمنوع من الصرف وتكيفاً مع تعود الكاتب العربي على رسم هذه الحركة مع إغفال حركات التشكيل الأخرى عادةً.

تنفيذ المشروع

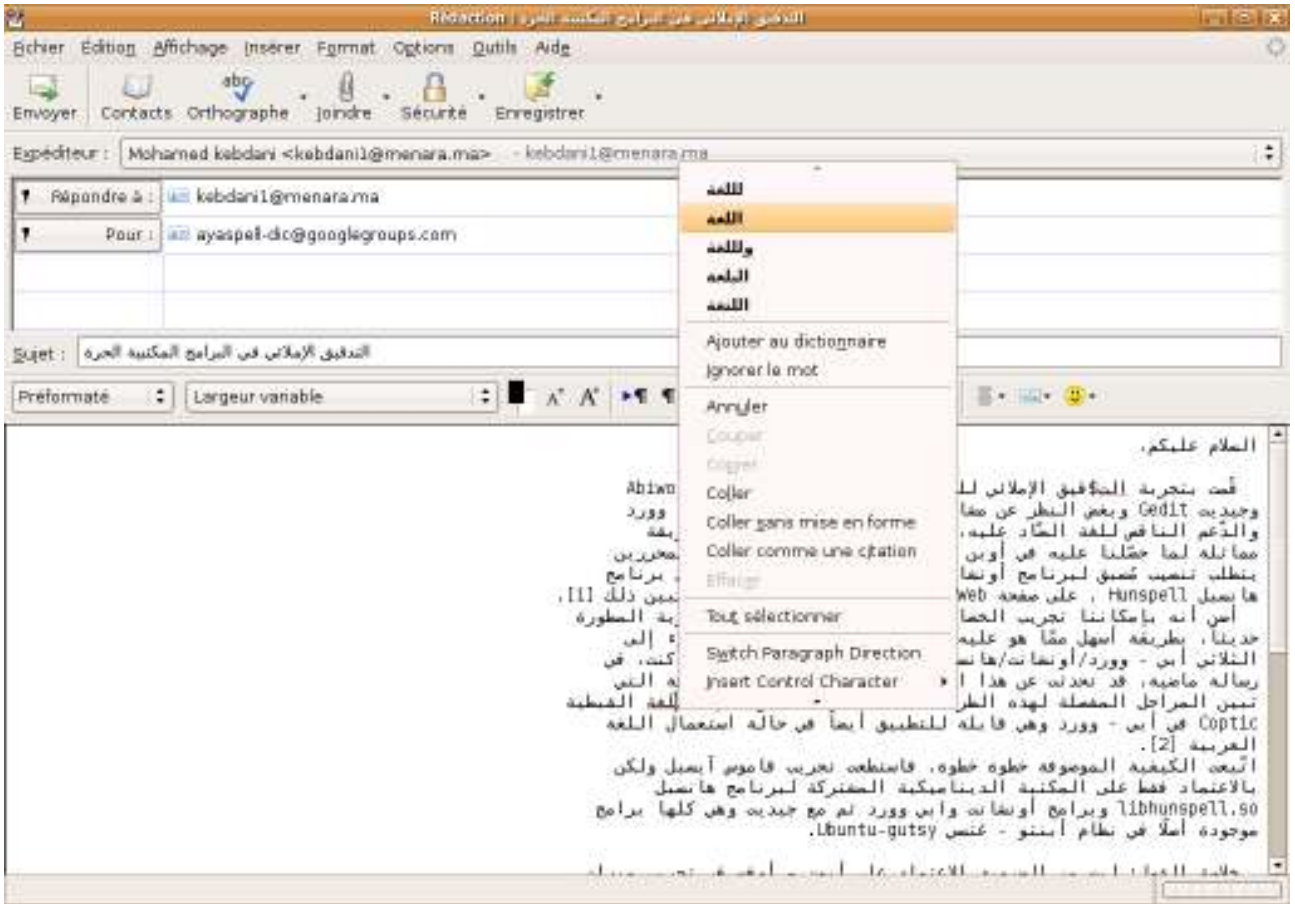
ظهرت الإصدار النهائية الأولى للقاموس في شهر يناير 2008، وقد لاقت قبولا حسنا في ميدان البرامج الحرة وهي تستعمل حاليا في التدقيق على مجموعة أوبن أوفس OpenOffice.org ، متصفح فابرفوكس Firefox، مراسل تاندربيرد Thunderbird ، أبيوورد Abiword ، جيديت... Gedit الخ. وهو متوفر للتحميل على الموقع <http://ayaspell.sourceforge.net> كما يمكن استعماله وإعادة تعديله حسب رخص البرامج الحرة GPL/LGPL/MPL. ويعمل على المنصات المختلفة كنظام windows و لينكس.



رسم توضيحي 1 استعمال القاموس العربي في برنامج Firefox



رسم توضيحي 2 استعمال القاموس العربي في برنامج openOffice



رسم توضيحي 3 استعمال القاموس العربي في برنامج thunderbird

في من المظاهر السلبية في عمل المدقق الإملائي في مجال اللغة العربية خصوصاً، البطء في اقتراح البديل الصحيح وضعف الدقة والحل المنتظر تجسيده في الإصدارات القادمة إن شاء الله، قصد تحسين فعالية المدقق الإملائي من حيث السرعة والنجاعة، هو إعادة بناء القاموس بالاقتران على المستعمل من المفردات والتركيز على التصريف المتداول فعلاً في الكتابات الحديثة حسب المقاربة المبينة في الورقة الخاصة بها في مدونة المشروع والمعنونة [بقاموس آيسبل بين "المستعمل والمهمل" في اللغة العربية](#). هذا من جهة، ومن جهة أخرى العمل على هيكلة القاموس بطريقة تتماشى مع [المدقق النحوي العربي](#) الذي بدأ الإخوة في عربآيز التفكير في برمجته ليعمل ضمن المجموعة المكتبية أوبن أوفيس.

الخاتمة

عملنا في هذا المشروع على إنجاز القاموس العربي للندقق الإملائي بشقيه البرمجي الذي عني بدعم العربية وخصائصها في برامج التدقيق الحرة، وكذلك إنشاء القاموس بمفرداته وقواعد الزيادات المختلفة، وللعلم فإن البرنامج مستعمل حالياً في التدقيق على العديد من البرامج ومتوفر للتحميل مجاناً، ويمكن تعديله في إطار البرامج الحرة.

وقد تحصل البرنامج على العديد من الجوائز والدعم من قبل مجموعات البرامج الحرة والمؤسسات الداعمة للتطوير بهذه المنهجية. ونطمح في المستقبل إلى المزيد من الدعم لخصوصيات اللغة العربية وتبسيط

القاموس وقواعده، وإضافة المفردات الجديدة إليه.

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 و التبليغ 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 كولتاك <http://www.coltec.net/ArabicProofingTool.aspx>
10. مذكرة رمزي عباس لنيل شهادة الدكتوراه من المعهد الوطني للعلوم التطبيقية في ليون - فرنسا، بعنوان : ' La conception et la réalisation d'un concordancier ' électronique pour l'arabe ، 2004.