



مدينة الملك عبدالعزيز للعلوم والتقنية



المنظمة العربية للتربية والثقافة والعلوم
إدارة العلوم والبحوث العلمي

الاجتماع الثاني لخبراء المعجم الحاسوبي للغة العربية

أبريل 2008

مشروع المعجم الحاسوبي التفاعلي للغة العربية
مقترحات حول إعداد المشروع

أ. مراد لوكام

مشروع المعجم الحاسوبي التفاعلي للغة العربية
مقترحات حول إعداد المشروع

أ.مراد لوكام
جامعة حسيبة بن بوعلي ، الشلف، الجزائر

Loukam@hotmail.com

الفهرس

3	1. مقدمة :
3	2. المحور الأول : عرض بعض أعمالنا في حوسبة اللغة العربية
4	2.1 مشروع 'الفرس' : PHARAS
5	2.2 مشروع 'المحيط'
7	3. المحور الثاني : منهاج إعداد مشروع المعجم الحاسوبي للغة العربية
7	3.1 الجانب اللغوي
8	3.2 الجانب الحاسوبي
12	3.3 الموارد البشرية، المالية و المراحل الزمنية
15	4. خاتمة :
16	المراجع :

1. مقدمة :

الهدف من هذه الوثيقة هو تقديم إسهامنا المتواضع في المجهود النبيل و العمل الضخم الذي تقوم به المنظمة العربية للثقافة و العلوم لإنجاز المعجم الحاسوبي التفاعلي للغة العربية. لا يخفى على أحد أهمية هذا المشروع و الفوائد الكثيرة التي ستجنيها أمتنا بعد إتمامه إن شاء الله.

تتضمن هذه الوثيقة محورين رئيسيين هما :

المحور الأول : سنعرض فيه بعض أعمالنا السابقة و الحالية في مجال حوسبة اللغة العربية ، و ذلك إيماننا بأن بعضها قد يساعد على إنجاز المعجم من الناحية الحاسوبية بل إن بعضها يدخل في صميم عمل المعجم.

المحور الثاني : سنقدم فيه بعض المقترحات العملية لإعداد المشروع معتمدين في ذلك على التوصيات الواردة في وثيقة مشروع المعجم الناتجة عن اجتماع الخبراء المنعقد في الفترة 11-2007/06/13 بدمشق [1].

تنتهي الوثيقة بخاتمة سنركز فيها على بعض الجوانب التي نعتقد بأنها لم تلقى الدراسة المعمقة الكافية في مشروع المعجم و التي قد تتحول إلى عقبات في مرحلة الإنجاز.

2. المحور الأول : عرض بعض أعمالنا في حوسبة اللغة العربية

نشتغل منذ سنوات عديدة على حوسبة اللغة العربية و ذلك بوضع أرضية (platform) لتحليل النصوص العربية في إطار النظرية اللسانية التراكمية ذات التحليل الموجه بالرأس¹ (HPSG : Head Driven Phrase Structure Grammar) . وضعت هذه النظرية في مطلع التسعينيات [2] و تلقى الآن إقبالا كبيرا و اهتماما متزايدا من طرف الباحثين من مختلف بلدان العالم و من مختلف اللغات. و يكفي للدلالة على خصوصيتها أن نعرف أن مؤتمرا عالميا يخصص لها كل عام² تحتضنه بالتداول أكبر مراكز البحث العالمية في مجال الإعلاميات اللسانية. من أهم ميزات هذه النظرية :

1- تعتبر النظرية في حد ذاتها نتاج نظريات لسانية عديدة سابقة (GPSG, TAG,

LFG...) بحيث حاول واضعوها اقتباس أحسن ما يوجد في النظريات الأخرى و

نقادي نقاط ضعفها.

¹ نعتقد أن أحسن ترجمة لكلمة Head هنا هي كلمة 'عامل' كما استعملها النحاة العرب الأولون.
² أنظر الموقع المخصص لهذه النظرية على الإنترنت

- 2- تعتمد هذه النظرية على مبدأ توحيد التمثيل للمادة اللغوية بجميع أقسامها (المدخلات المعجمية، القواعد النحوية، ...الخ)، فكلها ممثلة على شكل مصفوفة تسمى بنية الصفات (features structure). كما تعتمد النظرية على مبدأ دمج كل عناصر التمثيل اللغوي (المعجمي، الصرفي، النحوي و حتي الدلالي) للمدخلات المعجمية (lexical entries) في بنية الصفات.
- 3- مبدأ توحيد التمثيل و دمج المعارف اللغوية سهل كثيرا من عملية التحليل و أعطى لهذه النظرية قابلية كبيرة للحوسبة³.

2.1 مشروع 'الفرس' : PHARAS

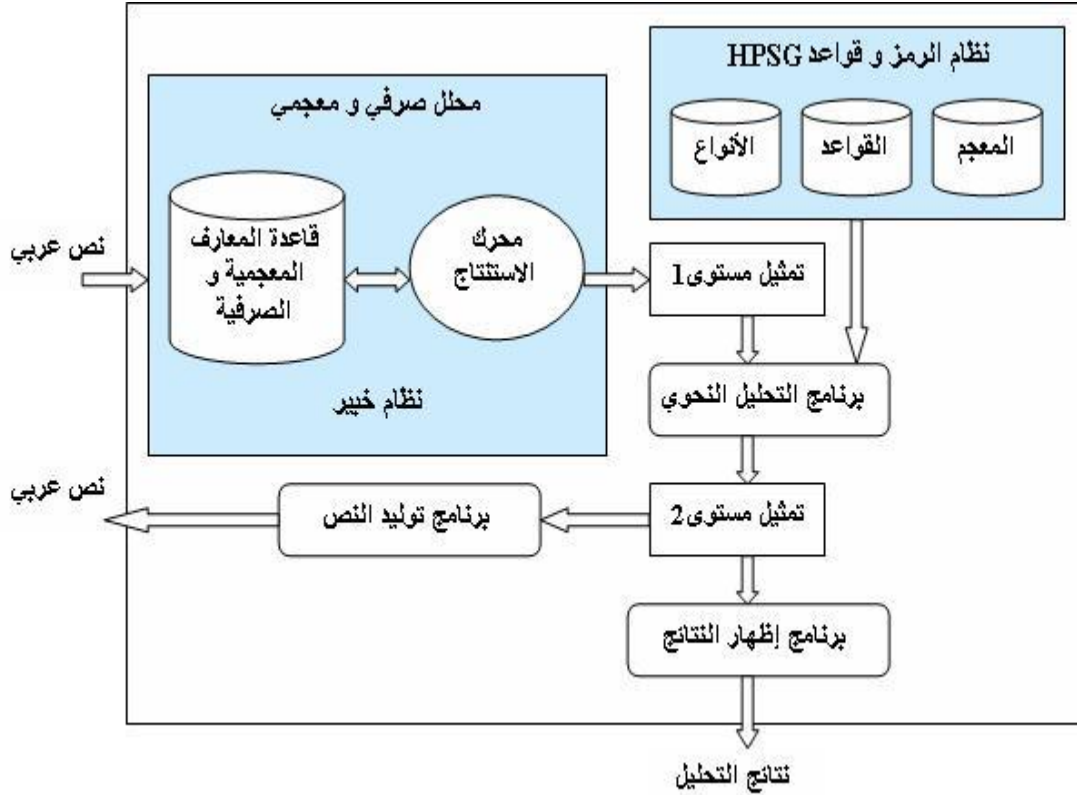
أطلقنا منذ سنوات مشروعا طموحا لوضع أرضية حاسوبية لتحليل النصوص العربية بالاعتماد على نظرية HPSG سمي بمشروع 'الفرس' (PHARAS)⁴. يهدف المشروع [3] إلى وضع نظام متكامل مشكل من عدة برامج حاسوبية مترابطة تعنى بجميع مستويات التحليل اللغوي كما يوضحه الشكل 1.

يمكن أن نشرح المخطط العام لسير الأرضية كالاتي:

1. مرحلة التقطيع : يقطع النص المراد تحليله إلى كلمات.
2. مرحلة التحليل المعجمي و الصرفي : تعرض كل كلمة موجودة في النص إلى محلل معجمي و صرفي صمم على شكل نظام خبير (expert system) يحتوي على قاعدة للمعارف المعجمية و الصرفية و محرك للاستنتاج [4].
3. مرحلة توليد بُنى الخصائص و ذلك باستنتاج كل خصائص الوحدات اللغوية التي عثر عليه المحلل. تخزن هذه الخصائص على شكل مصفوفات (تمثيل رقم 1 على الشكل).
4. مرحلة التحليل النحوي و ذلك انطلاقا من المعلومات و الخصائص الصادرة عن المرحلة السابقة و كذلك بمراعاة نظام الرمز و القواعد الخاصة بنظرية HPSG.
5. مرحلة إظهار نتائج التحليل إما على شكل مصفوفات أو على شكل ملفات من نوع XML.
6. مرحلة اختيارية إضافية يمكن العودة إليها بغرض التحقق من صحة التحليل المنجز و ذلك بتوليد و استرجاع النص الأصلي.

³ و هي بذلك تحاول أن تتفادي ما آلت إليه النظرية التوليدية (generative theory) التي جعلت بعد تطوراتها الكثيرة عملية التحليل و الدراسة للمادة اللغوية شيئا نظريا بحتا و غاية في التعقيد (تجبر الدارس على القيام بمجموعة كبيرة من التحويلات transformations) مما أدى إلى عزوف الباحثين عنها.

⁴ PHARAS : une plate-forme d'analyse basée sur le formalisme HPSG pour l'Arabe standard.



الشكل 1. المخطط العام لمشروع الفرس PHARAS

نعمل حالياً على إدماج برامج إضافية للأرضية الغرض منها جعلها أكثر انفتاحاً و قابلية للاستعمال في تطبيقات اللغة العربية.

يمكن طبعا استخدام أجزاء عديدة من الأرضية ، مع بعض التعديلات، في مشروع معجم اللغة العربية. نخص بالذكر المحلل الصرفي الذي لا بد منه لتشغيل المعجم. و هنا تجدر الإشارة إلى تطابق وجهات النظر بين اختيارنا لفكرة النظام الخبير (expert system) و بين ما ورد في وثيقة إعداد المعجم [1] الذي جنح إلى نفس الفكرة.

كما نلفت الانتباه إلى إمكانية اقتباس طريقة تخزين المعارف المعجمية الموجودة في أرضية الفرس و استعمالها في المعجم. فعلا ، من الضروري أن نولي أهمية بالغة لمبدأ تقييس المدخلات المعجمية، لأن ذلك شرط أساسي لتحقيق استعمال المعجم في تطبيقات أخرى للغة العربية.

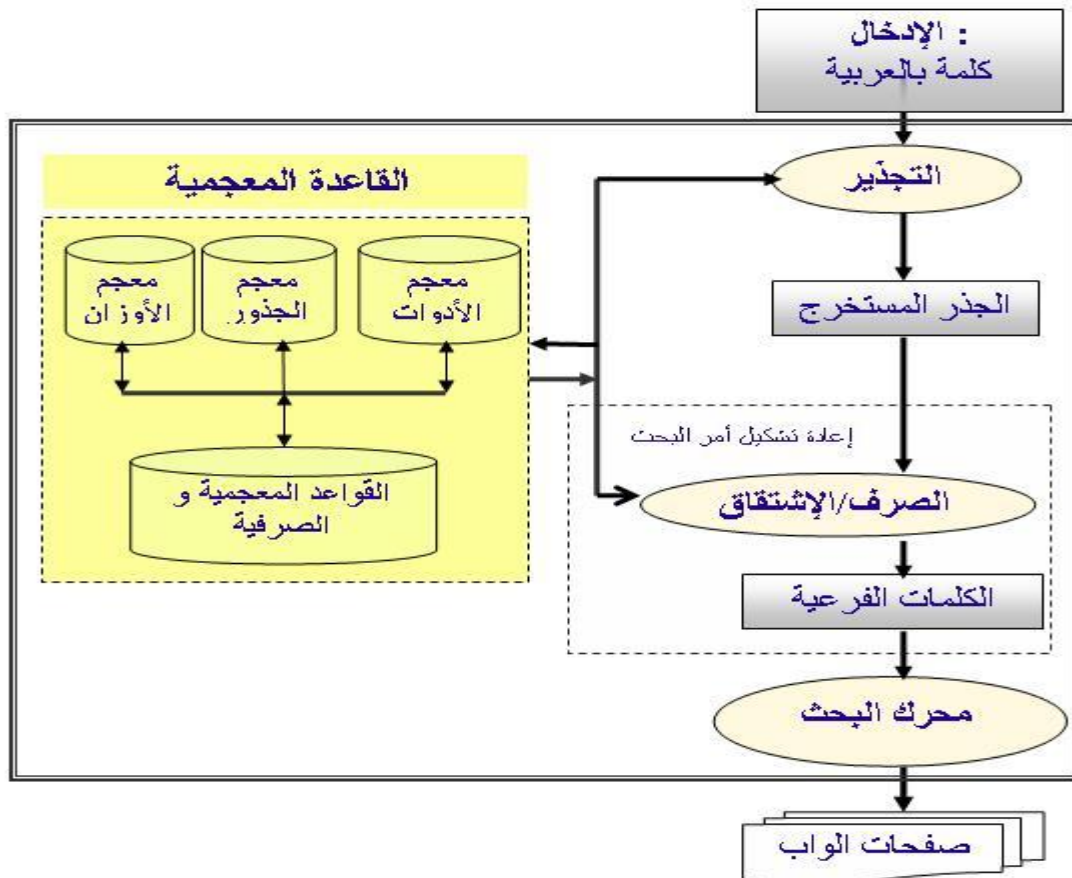
2.2 مشروع 'المحيط'

تعتمد معظم محركات البحث على الإنترنت على مفاتيح الكلمات (key words) . فهي تقوم ببساطة بالبحث في صفحات فهارس الشبكة عن سلسلة الحروف المشكلة للكلمة المدخلة. فمثلا إذا بحثنا عن

كلمة 'الجزائريات' فإن المحرك سيستخرج كل الصفحات التي تحتوي على هذه الكلمة بالضبط و سيهمل كل الصفحات التي تحتوي على كلمات مشابهة تنتمي لنفس الأصل مثل : الجزائر، الجزائر، الجزائريون، ... الخ بالرغم من أنها قد تكون مهمة للباحث. انطلاقاً من هذه الملاحظة ، أطلقنا مشروع 'المحيط' [5] الذي كان الغرض منه إضافة هذه الإمكانيات الجديدة في البحث و التي تسمح بالبحث عن الصفحات التي تحتوي على كلمة ما و في نفس الوقت عن كل الكلمات التي لها نفس الأصل.

يحتوي نظام المحيط على ثلاثة أقسام كما يوضحه الشكل 2 :

1. نظام التجذير الذي يستخرج أصل الكلمة بالاعتماد على نظام خبير قائم على أساس القواعد المعجمية و الصرفية للغة العربية.
2. نظام الإشتقاق الذي يستخرج كل الكلمات الفرعية التي لها نفس الأصل.
3. محرك البحث الذي يطلق البحث بإضافة الكلمات الفرعية .



الشكل 2. المخطط العام لمشروع 'المحيط'

جُرب هذا النظام في البحث على الإنترنت [5] و يجري تطويره حالياً.

نعتقد أن المبدأ الذي يقوم عليه نظام المحيط (التجذير/الإشتقاق) من المفيد جداً، بل من الضروري، أن يستعمل في معجم اللغة العربية ، ذلك لأن الحالة هي نفسها. فمستعمل المعجم سيكون بكل تأكيد مهتماً عند البحث عن كلمة ما، بالبحث أيضاً عن الكلمات التي تنتمي لنفس الأصل.

3. المحور الثاني : منهاج إعداد مشروع المعجم الحاسوبي للغة العربية

نقدم في هذا المحور بعض المقترحات العملية لإنجاز مشروع المعجم الحاسوبي للغة العربية.

3.1 الجانب اللغوي

يجب تشكيل فرقة كاملة من المختصين اللغويين من أجل الحسم في الأسئلة و الخيارات التالية :

- مصادر المادة اللغوية للمعجم (لكل من :المدخل، والتعاريف، والأمثلة، والمترادفات، والتراكيب اللغوية، والفروق اللغوية، والمصطلحات ...)
- كيفية استقاء المعلومات من هذه المصادر
- كمية المعلومات المستقاة من هذه المصادر
- أسلوب العمل في حسم أوجه الخلاف بين مصادر المعلومات
- الوسط الذي تخزن فيه هذه المعلومات، وآلية العمل به
- آلية العمل في مراجعة هذه المعلومات، والتثبت من صحتها
- مقدمة المعجم، والملاحق

غير أننا يمكن أن نذكر بالمقترح الوارد في وثيقة المشروع [1] و الإضافة عليه ما يأتي:

أولاً : تكليف مجموعات عمل ، قد تكون مستقلة عن بعضها البعض، متشكلة من متخصصين لغويين توكل لكل واحدة منها مهمة التحضير و التقيح لكل نوع من أنواع المادة المعجمية التالية :

- المدخل والتعاريف: و مصدرها الأساسي (المعجم الوسيط).
- المدخل والتعاريف: و مصادرهما من المعاجم الحديثة نسبياً كالمدرسي، والأساسي، والهادي، والمحيط... وثانيهما: الكتب والمقالات والصحف والمجلات في مواقع الإنترنت.
- في الأمثلة والشواهد: من مصادرهما المختلفة، و منها (معجم أمهات الأفعال - معانيها وأوجه استعمالها) ، المكتبة الشاملة وموسوعة الشعر العربي .

- في المجالات الدلالية والمترادفات والأضداد: و مصادرها المكنز الكبير، وكنز المترادفات.
- في العبارات الاصطلاحية والتراكيب اللغوية: و مصادرها : معجم التراكيب والعبارات الاصطلاحية: القديم منها والمولد، المعجم العربي الأساسي، المكنز الكبير.
- في الأخطاء الشائعة: و مصادرها معجم أخطاء الكتاب، نحو إتقان الكتابة العلمية باللغة العربية، دليل الأخطاء الشائعة في الكتابة والنطق.
- في المصطلحات: من مصادرها المختلفة.
- في الفوائد اللغوية والنحوية والصرفية والبلاغية: و مصادرها معجم محيط المحيط، معجم الكليات، المعجم المدرسي.
- في الكلمات المصاحبة: و مصادرها المكتبة الشاملة وموسوعة الشعر العربي .
- في الفروق اللغوية: و مصادرها الأساسي (الفروق في اللغة).

ثانيا : من الأهمية البالغة أن يفضي عمل الفرق اللغوية السالفة الذكر إلى تفتيح جيد للمادة المعجمية يسمح بإدخالها ببسر و وفق المعايير المطلوبة. يمكن أن يتم ذلك بأحد الأمرين :

- يوضع لكل كلمة بطاقة ورقية تحتوي على جميع المعلومات المعجمية المتصلة بها و المستقاة من المصادر. تستعمل هذه البطاقات بعد ذلك كأدوات (supports) تسمح بإدخال المادة المعجمية في المعجم من طرف الأشخاص المكلفين بذلك.
- توضع للفرق اللغوية واجهات حاسوبية (interfaces) تسمح لهم بإدخال المعلومات مباشرة في ملفات المعجم.

3.2 الجانب الحاسوبي

1. لغة البرمجة المستعملة في المعجم :

نقترح أن تكون لغة البرمجة الأساسية المستعملة لإنجاز المشروع هي لغة جافا (java) و ذلك لمزاياها المتعددة، و نذكر منها :

- تسمح بجعل البرنامج مفتوح المصدر (open source).
 - تسمح باستعمال المعجم على نظم تشغيل مختلفة مثل windows و linux.
- غير أننا نقترح أن تستعمل أيضا ، لإنجاز بعض أجزاء المعجم ،لغة Php و ذلك لنفس الأسباب مع الإضافة بأنها أكثر أهلية من java فيما يتعلق بإنجاز البرامج على الشبكة (internet programming).

2. الوسط البرمجي لقواعد معطيات المعجم :

يتضمن المعجم معطيات لغوية يحتاج إليها في عمله تتوزع على ملفات عديدة (الجزور، الأوزان، قواعد اشتقاق، المدخلات المعجمية، ...الخ). يمكن أن تخزن هذه الملفات باستعمال نمط XML الذي يضمن تخزين المعطيات بطريقة نصية لكن وفق هيكلية محددة مما يسمح استخدامها وتعديلها بيسر. لكن النمط النصي (text format) هذا قد يتحول إلى عائق بالنسبة للملفات كبيرة الحجم التي تحتوي على الآلاف من المدخلات، كما هو الحال بالنسبة للمعجم، لأن البحث فيها من قبل البرمجيات لا يتم إلا بطريقة الوصول المتسلسل (sequential access) مما سينجر عنه بطء في التنفيذ و الإستخراج. لهذا السبب نقترح أن يلجأ عند الضرورة إلى استخدام مُسَيِّر لقاعدة المعطيات (Data Base Management System DBMS) من نوع MySQL⁵ إذا تعلق الأمر بالملفات الثقيلة التي يضمها المعجم⁶.

3. البرمجيات المساندة لعمل المعجم :

من المفيد جدا استعمال برمجيات مساعدة لعمل الحاسوب كالتالي عرضناها في أعمالنا السابقة مثل : المحلل الصرفي المعتمد على النظام الخبير، برنامج التجدير، برنامج الصرف. غير أننا يجب أن ننبّه بأن هذه الأعمال المنجزة يجب أن تُدخل عليها بعض التعديلات كي تتوافق تماما مع خطة عمل المعجم و لا يجب أن ينظر إليها على أنها في أتم الجاهزية. يجب الأخذ بالحسبان فترة التعديلات هذه أثناء إعداد المشروع.

4. مصادر الوسائط المتعددة (الصور، الرسوم، الأصوات، مقاطع الفيديو...) :

إدماج الوسائط المتعددة (الصور، الرسوم، الأصوات، مقاطع الفيديو...) في المعجم يعتبر بكل تأكيد من الأعمال التي تتطلب جهدا كبيرا ، بل نعتبره بحق مشروعا في حد ذاته. الصعوبة لا تكمن بالطبع في كيفية إدخال المادة (صوت و صورة) ، فذلك أمر بسيط نسبيا، بل الصعوبة تكمن في مصادر اقتناء هذه المادة. صحيح يمكن اللجوء لاستعمال الصور و الأصوات الموجودة على الإنترنت ، لكن يجب أن نذكر بأن معظم هذه المادة تعتبر ملكية فكرية ، تحفظها قوانين حماية التأليف، و بالتالي لا يمكن التصرف فيها بحرية. هناك عدة أسئلة و عقبات يمكن أن تطرح نفسها من هذا الجانب، مثلا :

• ماذا سنفعل إذا أردنا أن ندمج ضمن تعريف كلمة 'أسطرلاب' في المعجم صورة لهذه الآلة ؟

⁵ مع العلم أن MySQL يوافق أيضا مبدأ المصدر المفتوح.

⁶ أذكر هنا أن أغلب ما كان يعاب على النظام الحاسوبي ل ذخيرة اللغة الفرنسية (Trésor de la Langue Française) عند عرض إصداراته الأولى سنة 2003 كان بطئه الشديد لأنه يعتمد كلية على نمط XML.

· ماذا سنفعل إذا أردنا أن ندمج ضمن تعريف كلمة 'مركبة' و استعمالاتها في اللغة العربية صورة لمركبة فضائية ؟ هل يجب الإتصال بمصدر الصورة (وكالة الأنباء أو غيرها) للحصول على موافقة استعمال الصورة ؟.

· ما العمل إذا أردنا أن ندرج ضمن تعريف كلمة 'القلب' شكلاً يبرز مقطعاً للقلب مع شرايينه ؟

و الواقع أن هذه الأسئلة و أخرى يجب أن تحظى بدراسة معمقة للفصل فيها، ذلك لأن لها انعكاسات مباشرة على مدة و تكاليف إنجاز المعجم . لكن هذا لا يمنعنا من أن نطرح الأفكار و المقترحات التالية :

· تكليف فريق عمل توكل إليه مهمة رصد و إحصاء كل المواد المتعددة الوسائط (الصور، الرسوم، الأصوات، مقاطع الفيديو...) على الشبكة التي يمكن أن تصنف ضمن الإستعمال العام (public domain) و التي يمكن أن تفيده في المعجم و أن تستعمل بكل حرية دون أي إشكال .

· إمكانية التعاقد مع هيئة متخصصة في مجال الإعلام توكل إليها مهمة تكوين مدونة متعددة الوسائط (الصور، الرسوم، الأصوات، مقاطع الفيديو) وفق قائمة معدة من طرف الهيئة المشرفة على المعجم.

5. خيارات الدخل والخرج :

من المفيد أن نذكر هنا بأهم الخصائص التي يجب أن يتوفر عليها المعجم كما وردت في وثيقة المشروع [1].

- أولاً : في الدخل

ينبغي أن يكون الدخل على مستويين:

المستوى الأول هو مستوى الجذر. فإذا أدخل المستعمل جذراً ما، حصل على كل ما يتعلق بمدخل هذا الجذر. ولهذا المستوى جملة تسهيلات تتعلق بـ: عدد حروف الجذر المدخل، وانتمائها إلى حروف المعجم... الخ.

والمستوى الثاني هو مستوى الكلمة. فإذا أدخل المستعمل كلمة ما (فعلًا أو اسماً أو صفةً أو مصدرًا أو أداة)، انتقل العرض مباشرة إلى المكان الذي ذُكرت فيه تلك الكلمة، ضمن الجذر المنتمية إليه.

ومن التسهيلات المتاحة في هذا المستوى قبولُ الكلمة المدخلة مشكولةً كلياً أو جزئياً أو غير مشكولةٍ البتة.

- ثانياً: في الخرج

من المؤكّد أن عرض معلومات المعجم الحاسوبي ينبغي أن يتمتع بالسهولة، والوضوح، والمرونة [1]. إذا كانت المادة المعجمية غنيةً بمداخلها، فينبغي عدم عرض كلّ تفاصيل هذه المداخل جملةً واحدة، لئلا يطول العرض كثيراً، فيرتبك المستعمل في الحصول على مبتغاه. بل تُعرض عناوين هذه المداخل في قائمة نختار منها أحدَ مداخلها ليقوم المعجم بعرضه.

6. أنواع البحث التي يشتمل عليها المعجم :

يشتمل المعجم على عدة أنواع من البحث يمكن أن تقسم إلى ثلاثة مستويات وفق غاية المُستعمل :

- النوع الأول : يسمح ببساطة تصفح المعجم نصاً بنص مع إمكانية التركيز على جانب معين من المادة اللغوية ، كالتعريفات مثلاً.
- النوع الثاني : يسمح بالبحث بطريقة أفقية في المادة اللغوية ، وفق معايير و تخصيصات يختارها المُستعمل، مثلاً : استخراج كل المواد المعجمية (المصطلحات) التي تخص ميدان الفيزياء أو الهندسة المدنية.
- النوع الثالث : يسمح بالبحث وفق أمر مركب يشتمل على تخصيصات عديدة ، مثلاً : استخراج كل الكلمات التي هي على شاكلة 'لو؟؟' أي التي تبدأ بحرفي الام و الواو و تكون في قائمة أسماء الآلة مع استظهار مصدرها و صورها إن وجدت.

طبعا في كل حالات البحث يجب أن يسمح المعجم بالانتقال بسهولة إلى كل المواد المعجمية ذات العلاقة بالمضمون المعروف و ذلك عن طريق النصوص الممنهلة (hypertexts).

7. أنواع الإحصائيات اللغوية التي يمكن أن يقدمها المعجم :

يجب أن يوفر المعجم جملة من الإحصائيات المفيدة. لا بأس أن نذكر بالقائمة التي وردت في مشروع المعجم [1] .

ففي الأفعال يمكن أن نحصل على إحصائيات من مثل:

- الأفعال الثلاثية المجردة التي تتصرف من باب معيّن.
- الأفعال الثلاثية المزيدة على وزن (استفعل)،

- الأفعال التي تبدأ بحرف النون،
- الأفعال التي تبدأ بحرف الكاف وتنتهي بحرف الباء،
- أفعال اللفيف المقرون،
- الأفعال المهموزة الآخر،
- الأفعال المزيدة التي لا مجرد لها،
- الأفعال التي تتعدى بحرف الباء،
- الأفعال المضعفة من المثال اليائي، وغيرها...
- وفي الأسماء يمكن أن نحصل على إحصائيات من مثل:
 - أسماء الآلة على وزن (فاعول)،
 - الأسماء الجامدة على وزن (فعل)،
 - جموع التكسير على وزن (فُعول)،
- وفي المصادر يمكن أن نحصل على إحصائيات من مثل:
 - مصادر الأفعال التي تتصرف من الباب الخامس،
 - مصادر الأفعال الجوف الواوية،
 - المصادر على وزن (فَعِيلِي)، وغيرها...
- وفي الصفات يمكن أن نحصل على إحصائيات من مثل:
 - الصفات المشبهة على وزن (فعلان)،
 - الصفات التي على وزن (أفعل) وليس مؤنثها (فعلاء)،
 - الصفات على وزن (فعليل) بمعنى مفعول، وغيرها...

8. أسلوب العمل في تحديث المعجم :

من الضروري أن يتبع إتمام مشروع المعجم الحاسوبي بمجهود كبير آخر وهو تحديثه دورياً. عملياً يتم ذلك بوضع إصدارات التحديث على موقع المعجم بالشابكة ، التي يمكن للمستعملين تحميلها و تنصيبها على حواسيبهم الشخصية.

3.3 الموارد البشرية، المالية و المراحل الزمنية

قبل الإدلاء بمقترحاتنا حول الموارد الواجب تسخيرها لإنجاز مشروع المعجم الحاسوبي للغة العربية من المفيد أن نذكر بتجربة إنجاز معجم ذخيرة اللغة الفرنسية [6] و [7] (TLF : Trésor de la Langue Française).

تحتوي ذخيرة اللغة الفرنسية على :

- 16 جزءا في شكلها الورقي.
- 100.000 كلمة.
- 270.000 تعريف.
- 430.000 مثال.

تم إنجاز ذخيرة اللغة الفرنسية :

- في شكلها الورقي بين 1960 و 1990.
- في شكلها الحاسوبي بين 1993 و 2001.
- شارك في إنجازها قرابة 100 باحث من ميادين مختلفة.

الغرض من التطرق لهذه التجربة هي التذكير بأن إنجاز عمل ضخم كالمعجم يتطلب جهدا جبارا و تضافرا كبيرا للجهود.

أما فيما يخص المعجم الحاسوبي للغة العربية المراد إنجازها ، فمقترحاتنا تكون كالآتي :

1. عدد اللغويين ومهمة كل منهم (بحث، مراجعة وتدقيق ..).

نقترح، كما سبق، أن يشكل فريق مشكل من عضوين على الأكثر لكل نوع من العناصر المعجمية التالية :

- المداخل والتعاريف: و مصدرها الأساسي (المعجم الوسيط).
- المداخل والتعاريف: و مصادرها من المعاجم الحديثة .
- الأمثلة والشواهد.
- المجالات الدلالية والمترادفات والأضداد.
- العبارات الاصطلاحية والتراكيب اللغوية
- الأخطاء الشائعة.
- المصطلحات: من مصادرها المختلفة.
- الفوائد اللغوية والنحوية والصرفية والبلاغية.

- الكلمات المصاحبة.
- الفروق اللغوية.

2. عدد المبرمجين ومهمة كلٍّ منهم (برنامج المعجم، قواعد المعطيات، برامج مساعدة) و عددهم أربعة :

- برنامج المعجم : 02
- قواعد المعطيات : 01
- برامج مساعدة : 01

3. عدد الفنيين (إخراج، صوت، صور، مقاطع فيديو ...) .
نقترح، كما سبق، أن توكل هذه المهمة لهيئة متخصصة.

4. عدد المنضدين ومهمة كلٍّ منهم (إدخال نصوص واقتباسات عن طريق لوحة الملامس، والكتب الإلكترونية، والمساحة ...) عددهم ستة (06).

- عدد المشرفين ودور كلٍّ منهم (إدارة، سكرتارية ...) .
نقترح أن يتم الإشراف من طرف :
 - مدير للمشروع ككل.
 - منسق الجانب اللغوي
 - منسق الجانب الحاسوبي

5. القيمة التقديرية لتكلفة المعجم
يمكن أن نأخذ بنفس المقترح الوارد في وثيقة مشروع المعجم [1].

6. المراحل الزمنية لإنجاز المشروع

بالنسبة للجانب الحاسوبي نعتقد أن تهيئة قاعدة المعطيات و وضع البرمجيات الرئيسية و المرافقة يمكن أن تتجز في ظرف سنتين و نصف.
أما بالنسبة للجانب اللغوي ، فبالنظر إلى حجم المادة المعجمية الواجب تحضيرها مع صعوبة تشكيل المدونة المتعددة الوسائط (الأصوات، الصور، الأشرطة) فإن المدة المقترحة في وثيقة مشروع المعجم [1] تبدو لنا غير ممكنة لإنجاز المعجم ككل. و مقترحنا هو أن نختار بين أحد الأمرين :

- إما أن تمتد مدة تحضير المادة اللغوية ، إذا أردنا أن ننجز المعجم كاملاً، بثلاث سنوات أخرى لتصبح خمس سنوات كاملة.
- و إما أن نبقى على المدة المقترحة ، أي سنتين، لكن مع خفض أنواع المواد المعجمية المرجو دمجها في المعجم، ثم التفكير في إصدارات أخرى لاحقة يستكمل فيها عمل المعجم.

4. خاتمة :

عرضنا في هذه الوثيقة بعض أعمالنا السابقة و الحالية في مجال حوسبة اللغة العربية، التي يمكن الاستفادة منها لإنجاز مشروع معجم اللغة العربية. نذكر منها على الخصوص : المحلل الصرفي و نظام التجدير/الإشفاق.

من بين الأهداف المرجوة من وضع المعجم هو استعماله في تطبيقات مختلفة للغة العربية، و هذا ما يجعل من تقييس المادة المعجمية شيئاً بالغ الأهمية. في هذا الإطار اقترحنا أن نقتبس من النظرية اللسانية HPSG طريقتها في تمثيل المعارف المعجمية.

- من جهة أخرى ، لفتنا الانتباه إلى بعض الجوانب التي قد تكون لها أثارا سلبية على خطة إعداد المشروع و التي يجب أن تلقى دراسة إضافية و مستفيضة. يمكن حصرها في نقطتين :
- عوائق تشكيل مدونة متعددة الوسائط ملحقه بالمعجم.
 - عدم مناسبة المدة المقترحة لإنجاز المشروع ، خاصة ما تعلق منها بالجانب اللغوي.

المراجع :

- 1- محمد مرياتي، نحو وثيقة لمشروع المعجم الحاسوبي التفاعلي للغة العربية، اجتماع خبراء المعجم الحاسوبي للغة العربية، دمشق 11-13/06/2007.
- 2- Carl Pollard, Ivan A. Sag. « Head-driven Phrase Structure Grammar ». The Ohio State University and Stanford University, 1994.
- 3- Loukam M., « PHARAS : Une plateforme d'analyse basée sur le formalisme HPSG pour l'arabe standard », Actes du premier séminaire sur le langage naturel et l'intelligence artificielle LANIA'2007, 20/21 Novembre 2007, Chlef/Algérie, p 31-40.
- 4- Loukam M., Abbache A., Laskri M.T., « Un analyseur morpho-lexical à base de système expert en vue d'une analyse en HPSG », Actes de la conférence Internationale sur le traitement automatique de la langue arabe CITALA'07, 18/19 Juin 2007, Rabat/Maroc, p 159-166.
- 5- Loukam M., « Al Mouhit : Un analyseur morpho-lexical pour l'arabe standard dans l'optique d'une recherche par racine sur Internet », soumis au International Conference on Web and Information Technology CIWIT'08, 29/30 juin 2008, Université Sidi Bel Abbès, Algérie.
- 6- Pascale Bernard, Jacques Dendien, Josette Lecomte, Jean-Marie Pierrel, Les ressources de l'ATLF pour l'analyse lexicale et textuelle : TLFi, FRANTEXT, Stella, JADT 2002 : 6es Journées internationales d'Analyse de Données Textuelles.
- 7- E. Jacquy et J-M. Pierrel, , Exploitation automatique d'informations à partir d'un dictionnaire (le TLFi) pour une représentation homogène et hiérarchisée des informations, INSTITUT DE LINGUISTIQUE FRANÇAISE (ILF), 2003.