

المعجم العربي الإلكتروني:
أهميته وطرق بنائه

الأستاذ الدكتور عبد المجيد بن حمادو
جامعة صفاقس
الجمهورية التونسية

الأربعاء 27 ذو الحجة 1432 هـ - الموافق 23 تشرين الثاني 2011م

1. المقدمة:

يعتبر المعجم من أهم روافد اللغة والمعرفة ويمثل ذاكرة الشعوب. وقد اعتنى اللغويون العرب بالمعجم من حيث المحتوى والهيكل منذ القرن الثاني الهجري ، فازدهرت الصناعة المعجمية العربية وتعددت مدارسها.

ومع انتشار استعمال الحاسوب والشابكة (الإنترنت) ودخول البشرية في ما يسمى "مجتمع المعرفة" ، برزت على الساحة تطبيقات لغوية حاسوبية جديدة ومتنوعة. هذه التطبيقات تعتمد في غالبها على المعجم. فعرف المعجم الإلكتروني (أو الحاسوبي) تطوراً كبيراً على صعيد الهيكله ومحتوى المداخل، وكذلك الخدمات المتطورة التي يمكن أن يسديها إلى المستخدم مستغلاً في ذلك الإمكانيات الهائلة التي يوفرها الحاسوب من طاقة تخزين للمعلومات المعجمية وتحسينها وسرعة البحث عنها واسترجاعها، وقدرة على معالجة البيانات متعددة الوسائط. فتطورت هذه الصناعة بسرعة فائقة وأصبحت تتنافس بصفة جدية مثيلتها التي تنتج المعاجم الورقية.

في هذه الدراسة سنحاول إلقاء الضوء على هذه الجوانب؛ حيث سنبدأ بتقديم مفهوم المعجم الإلكتروني ومقارنته بالنسخة الورقية، ثم نبين أهمية المعجم الإلكتروني العربي في مجتمعنا الذي يتحول بسرعة إلى مجتمع المعرفة ونمر بعد ذلك إلى تفصيل طرق هيكله وبناء المعجم العربي الحديث. ونختم بمناقشة موضوع تقييس المعاجم الإلكترونية والجهود التي تبذل في هذا المجال بالنسبة للغة العربية.

1. المعجم الإلكتروني: تاريخه وأهم مكوناته

المعجم الإلكتروني هو نسخة حاسوبية معدلة من النسخة الورقية. فهو يتكون من عدد كبير من المداخل يحتوي كل واحد منها على المعلومات التي يمكن

تجميعها حوله. تختلف هذه المعلومات من معجم إلى آخر حسب الأهداف التي بني من أجلها وأصناف المستخدمين المستهدفين.

وقد بدأ الاهتمام بالمعاجم الإلكترونية منذ منتصف القرن الماضي ؛ حيث اقتصر في البداية استعمال هذه المعاجم كموارد لغوية للتحليل الآلي للغات الطبيعية على المستوى الصرفي والنحوي والدلالي، فكانت المعاجم بمثابة قواعد بيانات تحتوي على معلومات مشفرة لا يفهمها إلا البرنامج الذي يستغلها. وقد تطورت وتنوعت هذه المعاجم من حيث المحتوى وكذلك من حيث الهيكلية لمواكبة تطور تطبيقات المعالجة الآلية للغات الطبيعية التي تستغلها مثل: تقطيع الجمل، تحليل النصوص واسترجاعها، البحث عن المعلومات، التدقيق الإملائي، التلخيص الآلي للوثائق والترجمة الآلية. فإذا أخذنا على سبيل المثال المعجم المصمم للتدقيق الإملائي نجده يختلف تماما عن المعجم المصمم للتحليل النحوي. فالأول يقتصر على قائمة كلمات اللغة، والثاني يستوجب تمثيل المعلومات الصرفية والنحوية (مثل: قسم الكلم، التعدية واللزوم، أدوات التعدية،...) بالنسبة لكل مدخل. وتجدر الإشارة هنا إلى أن هذه المعاجم التي صممت للألة يصعب استغلالها من طرف الإنسان. ولبلوغ هذه الغاية وجب تطوير قواعد البيانات لتحويل المعلومات المشفرة التي تحتويها إلى نصوص وإضافة تفسيرات لها ليتمكن الإنسان من قراءتها وفهمها، وكذلك تصميم برامج بينية متطورة للبحث عن المعلومات المتوفرة بالقاعدة وعرضها على الشاشة بطريقة مستساغة واستعمال وسائط متعددة (نص، صورة، صوت فيديو وصور متحركة).

التطور الذي حصل على مستوى قواعد البيانات والبرمجيات المصاحبة، بالإضافة إلى تطور الأجهزة الحاسوبية، قد ساعدا في بداية الثمانينات على ظهور أول المعاجم الإلكترونية المتاحة إلى الجمهور العريض على سطح المكتب أو

على الإنترنت (سواء عن طريق الاشتراك أو مفتوحة المصدر) أو على الأقراص المدمجة (CD - ROM) والأقراص الأخرى.

ومن بين القواميس الإلكترونية الغربية العديدة والمتنوعة، نذكر هنا بعض الأمثلة للغة الإنكليزية والفرنسية: ذخيرة اللغة الفرنسية المحوسبة (TLFi)، وقاموس الأكاديمية الفرنسية (2 طبعات)، و"الببليورم لاروس" الفرنسي ومعجم "أوكسفورد" الإنجليزي OED (الطبعة الثانية) والمعجم والمكنز الإلكتروني "كولينز".

أما بالنسبة إلى اللغة العربية فثمة تقصير واضح في إعداد معاجم حاسوبية تلبي حاجيات المستخدم العربي بالرغم من إتاحة أمهات المعاجم العربية على أقراص المدمجة أو على الإنترنت . فهذه الخطوة ليست كافية؛ لأن هذه المعاجم تقتصر لأبسط مقومات المعاجم الحاسوبية بالمعنى الحديث للكلمة. إنها مجرد "نسخ مرقمنة" لتلك المعاجم الورقية (في صيغة " Doc" أو "HTML") لا يمكن الاستفادة منها بالشكل المطلوب (Ait Taleb, 2005). بالإضافة إلى ذلك، فالأدوات المصاحبة للبحث عن المعلومة بسيطة كما يتضح في المشاريع المعروفة مثل "عجيب" (لنظام صخر) و"كلمات" و"لسان العرب" / "القاموس المحيط". هذه السليبات تعود أساساً إلى ضعف الهيكلة الحاسوبية التي بنيت عليها مداخل هذه المعاجم.

٢. تصنيف المعاجم الإلكترونية

إن أبسط تصنيف للمعاجم الإلكترونية الذي يمكن أن يُستخلص مما ذكر ، يُميز بين المعاجم الموجهة إلى الآلة (Machine Readable Dictionary) والمعاجم الموجهة للإنسان (Human Oriented Electronic Dictionaries)، علماً بأن الجهود قد اتجهت السنوات الأخيرة إلى دمج هذين الصنفين، وذلك باقتراح

منهجية بناء معاجم مقيسة تراهن على تلبية حاجة الإنسان والآلة في نفس الوقت
(Francopoulo G. & George M. (2008)).

وفي ما تبقى سنهتم بالمعجم الإلكتروني الموجه للإنسان.

التصنيف الحديث للمعاجم الإلكترونية الموجه للإنسان يأخذ بعين الاعتبار
الجوانب الأساسية التالية: ترتيب المداخل (حسب الجذور، حسب الجذوع،...)،
وظيفة المعجم (للترجمة، للتعليم،...)، نوعية المحتوى (نصي/ متعدد الوسائط، لغوي/
متخصص، أحادي اللغة/ متعدد اللغات،...)، المحمل (سطح المكتب، صفحات
الويب، أقراص مدمجة،...).

انطلاقاً من هذه الجوانب يمكن تصنيف المعاجم الإلكترونية كما يلي:

- معاجم لغوية: متكونة من عينة من المفردات. يحتوي كل مدخل على
المعلومات اللغوية الأساسية التالية: تعريف الكلمة، خصائصها الصرفية والنحوية،
طريقة الكتابة (الإملاء)، المعاني المختلفة مع أمثلة وشواهد لمختلف الاستعمالات.
- معاجم متخصصة تحتوي على المفردات المستعملة لعلم ما أو فن: قاموس
الرياضيات، قاموس الطب، قاموس الاقتصاد، قاموس الحاسوب، قاموس أسماء
العلم.
- معاجم متعددة اللغات تعطي ترجمة الكلمات إلى لغة أو لغات أجنبية.
- معاجم بصرية تحتوي على مجموعة من الصور وأشرطة الفيديو مبنوية حسب
المواضيع التي تعالجها (بنايات، حيوانات، ألعاب، ملابس، وسائل نقل،...).

٣. المعجم الإلكتروني مقابل المعجم الورقي

يمكن مقارنة المعجم الإلكتروني مع النسخة الورقية بالاعتماد على الجوانب
التالية: المحتوى، الهيكل، فرص التحديث، طرق البناء ، طرق البحث عن
المعلومات والوقت المستهلك للوصول إليها.

1.3 مزايا المعجم الورقي:

للمعجم الورقي عدة مزايا نذكر منها على وجه الخصوص أنه مألوف وحجمه الخارجي والظاهر يمكن أن يعطيك فكرة عن كمية المعلومات المتوفرة داخله. كذلك المعجم الورقي سهل التصفح وقراءته سلسلة ولا تتعب النظر على عكس القراءة من الشاشة. كما لا يحتاج تصفح المعجم الورقي إلى تشغيل معدات خاصة مثل الحاسوب. هذه الخاصية تجعله مستقلا بذاته ووجوده غير مرتبط بتوفر أشياء أخرى مما يطيل عمره وييسر الحفاظ عليه. في حين أن الوسائط الإلكترونية أو المغناطيسية عمرها قصير نسبيا وسرعان ما يتجاوزها الزمن.

2.3 مزايا المعجم الإلكتروني:

للمعجم الإلكتروني مزايا عديدة متأتية أساسا من التطور التكنولوجي الذي حصل السنوات الأخيرة على مستوى سرعة معالجة البيانات وسعة تخزين المعلومات وكذلك على مستوى البرمجة ومعالجة قواعد البيانات. هذه المزايا يمكن حصرها في النقاط التالية:

- تنوع طرق البحث عن المعلومة: يمكن لمستخدم المعجم الإلكتروني أن يصل إلى المعلومة عبر الجذر أو الجذع (البحث البسيط) أو عبر المعنى (البحث المتقدم). مثلا يمكن البحث عن كلمة "هضبة" باستعمال المعنى الآتي "أرض مرتفعة". كما يمكن البحث عبر الإبحار داخل المعجم باستعمال الروابط النصية (Hyper-text links).

- طاقة التخزين الواسعة وتطور تقنيات قواعد المعطيات تتيح بناء معاجم كبيرة الحجم تجمع بين القديم والمعاصر ومتعددة اللغات والوسائط. هذه المعاجم تمتاز بالدقة والشمولية من حيث أنها توفر لكل كلمة معانيها الأساسية والفرعية وتعطي لذلك أمثلة وشواهد متنوعة.

- إمكانية التوليد الآلي لبعض الكلمات القياسات بدون الحاجة إلى تمثيلها بالمعجم وذلك بالاعتماد على قواعد الاشتقاق. هذه الميزة يصعب (أو يستحيل) توفيرها بالنسبة إلى المعجم الورقي لأن إيراد المشتقات القياسية لجميع الأفعال الممثلة سيضاعف حجم المعجم ويجعله غير قابل للاستعمال.

- احتواء المعجم الإلكتروني على عدة تطبيقات لغوية مهمة يمكن للمستخدم أن يستفيد منها مثل: تصريف الأفعال والأسماء، البحث عن المترادفات، المعالجة على المستوى الصوتي لتحويل المكتوب إلى منطوق، التدقيق الإملائي لتصويب الكلمات المدخلة،... هذه الخدمات غير متوفرة في المعجم الورقي.

- سهولة تعديل المعجم الإلكتروني بإضافة مداخل جديدة أو بتحيين مداخل موجودة. هذه الميزة يجب التعامل معها بكل حذر للحفاظ على مصداقية المعجم وجودته. تعديل محتوى المعجم يجب أن يتم من طرف معجميين وفق ترتيب مقننة تضمن تماسك المحتوى.

- الاعتماد على الوسائل الحاسوبية الحديثة المتعددة الوسائط (Multimedia) من نصوص، وأصوات، وصور ثابتة ومتحركة، وأفلام الفيديو لعرض المعارف. هذه الخاصية لها تأثير إيجابي على استساغة وفهم المعارف المعروضة.

وتجدر الإشارة هنا إلى أن للمعجم الورقي والمعجم الإلكتروني مزايا مشتركة لكن بدرجات متفاوتة نذكر بالخصوص إمكانية الاستعمال في كل مكان. مع تطور الحواسيب المحمولة على مستوى الحجم والاستقلالية عن التزود بالطاقة الكهربائية أصبح استغلال المعجم الإلكتروني متاح في كل مكان لكن بدرجة أقل من المعجم الورقي.

وبالرغم من ما ذكر من مزايا ما زال المعجم الورقي ينعم باستحسان القراء والمستخدمين لكن المستقبل سيكون حتما لفائدة المعجم الإلكتروني.

٤ . أهمية المعجم بالنسبة إلى اللغة العربية

1.4 أهمية المعجم الإلكتروني للمستخدم العربي

المعجم الإلكتروني ضروري للفرد العربي في مجتمع المعلومات والاقتصاد القائم على المعرفة . فقد تطورت وظيفه المعجم السنوات الأخيرة ليصبح وسيلة عمل بالنسبة للطالب وللاستاذ وللباحث وللمترجم ولعدة فئات اجتماعية ومهنية أخرى.

وتتضح هذه الضرورة أكثر إذا تمعنا في واقع صناعة المعاجم في اللغات العالمية، وقارناه بواقع صناعة معاجمنا العربية، خاصة الإلكترونية منها.

أهميته بالنسبة إلى تعليم اللغة تكمن في إقبال التلاميذ أو الطلاب على النسخة الإلكترونية لما توفره من إغراء على مستوى البحث عن المعلومة والعرض باستعمال وسائل متعددة الوسائط.

التحديث السريع يجعل من المعجم الإلكتروني مواكبا لتطور اللغة العربية على جميع مستوياتها.

يمكن أن نضيف أخيرا أن المعجم الإلكتروني يسهل عملية بناء أنتولوجيا للغة العربية (Baccar F. et all. 2011).

2.4 أهمية المعجم للمعالجة الآلية للغة العربية

يعتبر المعجم الإلكتروني أداة ضرورية وأساسية لتطبيقات المعالجة الآلية للغات الطبيعية بصفة عامة. جودة النتائج التي تفرزها هذه التطبيقات مرتبطة كليا بجودة المعجم من حيث عدد مداخله وشمولية المعلومات التي يحتويها كل مدخل. يتضح هذا الترابط إذا تمعنا مثلا في عمل المدقق الإملائي؛ فنجد

أنه يعتبر الكلمات الصحيحة خاطئة إن لم يجدها ممثلة في المعجم الذي يعتمد عليه في عملية التدقيق. كذلك عند التصحيح لا يقترح كلمات غير موجودة بالمعجم من بين المرشحة للتصحيح.

ومن هنا نستخلص أن تطوير تطبيقات حاسوبية للغة العربية تلبية حاجة المستخدم العربي وتجعل اللغة العربية مواكبة لتطور المجتمع رهين وجود معاجم إلكترونية عربية مناسبة وذات جودة عالية وتخضع إلى مقاييس عالمية على مستوى المحتوى والهيكلية. وعدم توافر هذه المعاجم بالجودة المطلوبة له انعكاس سلبي على استعمال اللغة العربية في الوطن العربي وانتشارها عالمياً، لأنه أصبح من المؤكد أن اللغة التي لا يقع الاهتمام بها من حيث الحوسبة يتناقص دورها تدريجياً كلغة ناقلة (vehicular language) وتصبح عرضة للاضمحلال. والأمثلة تكاثرت في العشريات الأخيرة التي شاع فيها استعمال التكنولوجيات الحديثة بين الناس.

والمسؤولية هنا مشتركة بين اللغويين (المعجميين) والمعلوماتيين المتخصصين في اللغويات الحاسوبية، وكذلك صناع القرار لتوفير الدعم المادي والأدبي، لأن بناء وتطوير المعاجم الإلكترونية عمل متعدد الاختصاصات ويتطلب اعتمادات تفوق غالباً إمكانيات البلد الواحد. وفي هذا الإطار يمكن ذكر بعض المشاريع المعروفة بالنسبة للغات الأجنبية مثل مشروع EUREKA الممول من طرف إيطاليا وإسبانيا لتطوير معجم مقيس، وكذلك المشاريع العديدة الممولة من طرف الإتحاد الأوروبي في مجال تطوير موارد معجمية متعددة اللغات والترجمة الآلية مثل مشروع EUROTRA.

5. طرق بناء المعجم الإلكتروني

بغض النظر عن الطريقة المعتمدة، فإن بناء معجم إلكتروني ليس بالعمل الهين، فهو يتطلب مجهودا جبارا يقوم به فريق يتكون من معجميين ومعلوماتيين. يهتم المعجميون بتجميع المادة اللغوية من مدونات ومعاجم ورقية وانتقاء المداخل وتحديد المعلومات الملحقة بكل مدخل.

ويهتم المعلوماتيون أو الحاسوبيون بالتصميم والهيكل لضبط البنية الفوقية (Macro-structure) التي تعنى بترتيب مداخل المعجم والبنية المصغرة (Micro-structure) التي تهتم بترتيب مكونات المدخل الواحد. كذلك يهتم الحاسوبيون بتصميم البرامج الضرورية لإدخال المعلومات المعجمية وتحسينها والبحث عنها وعرضها ولتوفير مختلف الخدمات الملحقة بالمعجم من تدقيق إملائي وتصريف الأسماء والأفعال وغيرها. ولبلوغ درجة عالية من الدقة والجودة يجب أن يعمل أعضاء هذا الفريق بطريقة متعاونة ومتكاملة وكذلك متوازية ربحا للوقت.

وفي ما يلي أهم طرق بناء المعاجم الإلكترونية.

1.5 الطريقة اليدوية

تعتبر الطريقة اليدوية من أسهل طرق بناء المعاجم لكنها مكلفة من حيث الجهد والوقت والمال.

يتم بناء المعجم باتتباع المراحل التالية:

١. الاتفاق على هيكل المعجم (ترتيب المداخل ومحتوى كل مدخل) انطلاقا

مما يسمى بكراس الشروط الذي يحدد الهدف من بناء المعجم، المستخدم المستهدف (مبتدئ، خبير،...) طريقة استغلال المعجم (على الشبكة، على سطح المكتب أو على قرص مضغوط،...).

٢. تحديد المصادر اللغوية (مدونات، معاجم ورقية،...) التي سيقع اعتمادها كمادة أولية.
٣. انتقاء وتجميع المادة المعجمية من المصادر المذكورة في جذاذات ورقية.
٤. تصميم وبناء قاعدة البيانات التي ستحتوي المعجم.
٥. بناء برمجية لإدخال المعلومات المعجمية يدويا من الجذاذات الورقية.
٦. إدخال المعلومات يدويا من الجذاذات الورقية.
٧. تجريب المعجم للتأكد من تحقيقه كافة الوظائف المطلوبة منه.
٨. تصويب الأخطاء اللغوية والبرمجية.

المرحلة الأولى يقوم بها المعجميون بتنسيق مع الحاسوبيين، المرحلة الثانية والثالثة يقوم بها اللغويون مع إمكانية الاستعانة بالحاسوبيين لمدهم ببعض الأدوات التي تساعدهم على تحليل وجرد المدونات أو القيام ببعض العمليات الإحصائية. المرحلة الرابعة والخامسة يقوم بها الحاسوبيون. المرحلة السادسة يقوم بها كتابة المرحلة السابعة يقوم بها لغويون والمرحلة الأخيرة يقوم بها المعجميون والحاسوبيون.

2.5 انطلاقا من معجم ورقي مرقم

هذه الطريقة ممتازة جدا لأنها تقتصر المسافات وتقتصد المال والجهد. الفكرة المركزية هي تحويل معجم ورقي مرقم إلى معجم إلكتروني يخضع إلى مواصفات دقيقة وحديثة. هذا التحويل يتم عن طريق برنامج حاسوبي يقع تصميمه للغرض. النتيجة المتحصل عليها قاعدة بيانات معجمية. ومن مزايا هذه الطريقة إمكانية الاعتماد على أكثر من معجم ثم يقع دمج قواعد البيانات المتحصل عليها للحصول على قاعدة ثرية على مستوى عدد المداخل وكذلك على مستوى محتوى

المدخل. وقد قام فريق من مخبر ميراكل بجامعة صفاقس باقتراح وتجربة هذه الطريقة على معجم "الغني" (A Khemakhem 2009) وكذلك على معجم الوسيط بالتعاون مع مدينة الملك عبد العزيز للعلوم والتقنية.

مراحل بناء قاعدة البيانات المعجمية هي الآتية:

١. اختيار المعجم (أو المعاجم) الذي سيعتمد عليه في صيغته المرقمة.
٢. تصميم قاعدة البيانات المعجمية التي ستحتوي مداخل المعجم.
٣. تصميم برنامج تعبئة قاعدة البيانات المعجمية انطلاقاً من النسخة أو النسخ المرقمة.
٤. تجريب القاعدة.
٥. تصحيح الأخطاء اللغوية والحاسوبية.
٦. إثرائها يدوياً إذا اقتضت الحاجة.

3.5 انطلاقاً من مدونة نصية

هذه الطريقة تعتبر من أصعب الطرق وأقلها دقة. مزاياها تكمن في أنها غير مكلفة على المستوى البشري والمادي وكذلك تمكن من الحصول على معجم مواكب.

بناء معجم بهذه الطريقة يتم باتباع المراحل التالية:

١. تحديد المدونة أو بناؤها.
٢. إنجاز برمجية لاستخراج المداخل والعلاقات بينها والمعلومات المعجمية الخاصة بكل مدخل بصفة آلية من المدونة.
٣. تدقيق المادة المعجمية المتحصل عليها من طرف المعجميين.

٤. إدخال المادة المعجمية المدققة في قاعدة البيانات التي ستحتوي المعجم.

4.5 الطريقة التعاونية (ويكي)

المقاربة التعاونية (الويكي) تمكن من بناء معاجم إلكترونية بتكلفة زهيدة وذلك بتكاتف جهود المتطوعين المتخصصين في مجال المعجمية. هذه المقاربة برزت ونجحت مع ظهور الموسوعة "ويكيبيديا". فهي تسمح لكل متطوع يرى في نفسه القدرة على المساهمة البناءة بالعبور إلى المعجم وتغيير محتواه سواء بإضافة مدخل جديد أو بإثراء محتوى مدخل موجود.

ولتقليل الأخطاء يمكن لكل مستخدم أن يتعرف إلى التغييرات الحديثة التي أدخلت على المعجم والتثبت فيها مع إمكانية تصويب الخطأ.

من أهم مزايا هذه الطريقة أنها:

- تضمن المواكبة السريعة للتطورات التي تحدث على المعجم من بروز مفردات جديدة وتوسع مجال مفردات موجودة.

- تعتبر أن بناء المعرفة ليس مجرد مسألة شخصية. بل هو مقاربة جماعية من أجل الوصول إلى توافق في الآراء.

- لا تقصي أحداً من عملية بناء المعجم. فهي تضع على قدم المساواة المبتدئ (الطالب مثلاً) والخبير (المعجمي).

بالنسبة للعربية يمكن ذكر مشروع "ويكاموس" الذي انطلق في شهر مايو 2004 والذي تقدم أشواطاً مهمة لكنه لا يزال يشكو من ثغرات عديدة في المحتوى لأن بناءه لم يكن ممنهجاً، بل موكول إلى إرادة المساهمين في عملية الإثراء.

٦. تقييس المعاجم الإلكترونية

تقييس الموارد المعجمية هو عمل جماعي تقوم به مجموعة من الخبراء لوضع تصور موحد لهذه الموارد يشمل جميع اللغات. هذا التصور يصبح مقياسا إذا وافقت عليه جهة رسمية وطنية أو عالمية مثل منظمة إيزو " ISO " أو اتحاد شبكة الويب العالمية (W3C). وكمثال نذكر مقياس " LMF " (Lexical Markup Framework) لمنظمة "إيزو" والمعرف بـ ISO 24613.

لتقييس الموارد المعجمية مزايا عديدة نذكر أهمها:

- المواصفات المقترحة صالحة لبناء معجم إلكتروني للاستعمال البشري وأيضا للمعالجة الآلية.
- تيسير عملية تبادل المعاجم بين الأشخاص والمؤسسات قصد الإثراء المشترك وذلك باستغلال البنية الموحدة.
- إمكانية دمج معاجم ثنائية اللغة بقصد توليد معجم متعدد اللغات للترجمة الآلية مثلا.
- بناء معاجم تستجيب لحاجيات المستخدم.
- تبادل الأدوات الحاسوبية مثل:
 - البحث عن المعلومات داخل قواعد البيانات والمعاجم.
 - التوقي الإملائي.
 - التحليل الصرفي.
 - الشائلي الآلي.
 - التصريف الآلي.

○ احتساب البيانات الإحصائية.

اعتمادا على مقياس LMF وقع بناء العديد من المعاجم الإلكترونية لأغلب اللغات (الإنجليزية، الفرنسية، الإسبانية، الإيطالية، البنغالية، وعدة لغات آسيوية،...).

بالنسبة إلى اللغة العربية يمكن القول إن عملية تقييس المعاجم لم تلقَ الاهتمام الكبير من الجهات المختصة. وفي ما يلي سنركز على الأبحاث التي أنجزت بمخبر "ميراكل" بجامعة صفاقس بالتعاون مع منظمة "إيزو" للتقييس تركز على مقياس LMF وأفضت إلى بناء قاعدة بيانات معجمية للغة العربية (Baccar F. et al. (2008)). من هذه القاعدة يمكن توليد معاجم متعددة حسب الحاجة كما يمكن توليد شبكة دلالية عربية "أنتولوجيا".

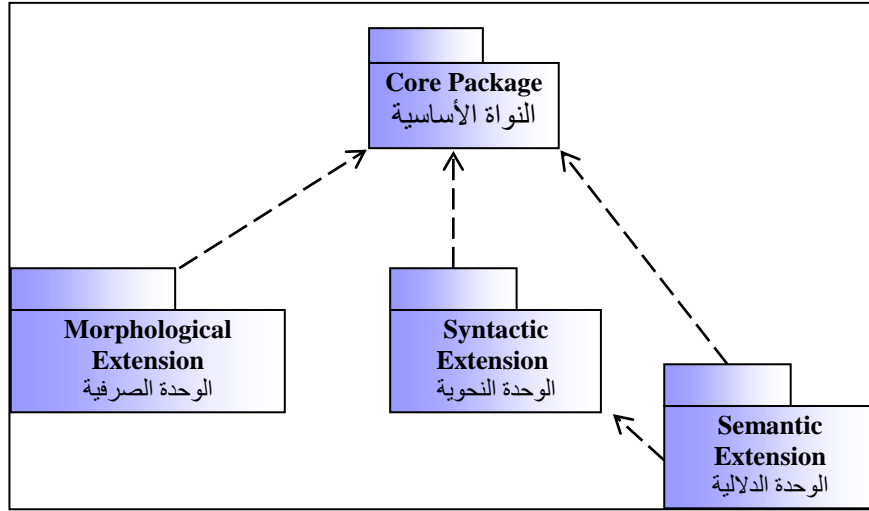
يمتاز مقياس LMF بمرونته وبإمكانية تغطيته مختلف مستويات اللغة (الصرفي، النحوي، الدلالي،...). يعتمد LMF على تصميم وحدات (modular) يمكن من:

أ - اختزال حيز تمثيل المعلومات المعجمية بعدم تمثيل المعلومات الخاصة بالصيغ المصرفية القياسية والاقتصار على تمثيل القواعد التي تمكن من توليدها.

ب - الفصل بين المستويات اللغوية لمدخل المعجم في شكل وحدات متخصصة تعنى كل وحدة بمستوى معين مع إمكانية الربط بين هذه الوحدات عند الحاجة إلى ذلك. هذا الفصل يمكن من إضافة وحدات جديدة دون المساس بالوحدات الموجودة.

وبناء على ما ذكر يتكون المعجم من نواة وثلاث وحدات اختيارية (انظر الرسم البياني عدد1):

- "النواة الأساسية" (CorePackage) تجمع المعلومات العامة للمدخل مثل أصل الكلمة، تعريفها، استعمالاتها (شواهد وأمثلة).
- وحدة صرفية تهتم بالمستوى الصرفي
- وحدة نحوية تهتم بتمثيل السمات النحوية للكلمة
- وحدة دلالية تهتم بالعلاقات الدلالية بين الكلمات.



الرسم البياني عدد 1: النواة الأساسية والوحدات المختصة

النواة الأساسية

تتكوّن هذه النواة من ستة أصناف (classes) من المعلومات أهمها (انظر الرسم البياني عدد2):

- المعلومات العامّة (Global Information): تحتوي على معلومات تخص نسخة المعجم مثل الجهة أو الجهات الممولة للمشروع، تاريخ الإصدار، رمز النسخة، المؤلف/ المؤلفون، العلامات الاصطلاحية، ملاحظات مختلفة،....

- المدخل المعجمي (Lexical Entry): يشمل معلومات خاصة بالمدخل

الذي يمكن أن يكون الجذع (Lemma) أو الجذر (Root) والتي تصنف كما يلي:

- النوع (type): هذه المعلومة تمكن من تحديد نوعية المدخل إن كان جذرا أم جذعا.
- قسم الكلام (Pos): اسم أو فعل أو حرف،... هذه المعلومة تخص الجذع.
- الوزن الصرفي (scheme): خاصّة بالأسماء المشتقة والأفعال المشتقة.
- معدلات الاستخدام (frequency): معلومة إحصائية حول مدى شيوع الكلمة (كثيرة الاستعمال/ متوسطة الاستعمال/ قليلة الاستعمال)، و يمكن تحديد هذه المعلومة بالرجوع إلى المدونة.
- التاريخ (dating): لتحديد الحقبة الزمنية التي ظهرت فيها الكلمة (قديمة أم حديثة،...).
- أصل الكلمة (etymology): لتحديد أصل الكلمة (فارسية، يونانية، تركية،...).
- الجذع (Lemma): يمثل الكلمة في صيغتها المجردة (بدون تصريف).
يحتوي هذا الصنف على المعلومات التالية:

○ طريقة كتابة الجذع.

○ وطريقة نطقه (كتابة صوتية أو تسجيل صوتي).

- المعنى (Sense): يحتوي هذا الصنف على المعلومات التالية:

○ المعرف الوحيد (Sense-Id).

○ معدل استخدام (frequency): كثير الاستعمال/ متوسط الاستعمال/ قليل الاستعمال . تستغل هذه المعلومة الإحصائية لترتيب المعاني حسب شيوعها وشهرتها عند عرضها على الشاشة.

تتكرر هذه المعلومات بحسب تعدد معاني المدخل . كل معنى مرتبط بتعريف أو

بأكثر.

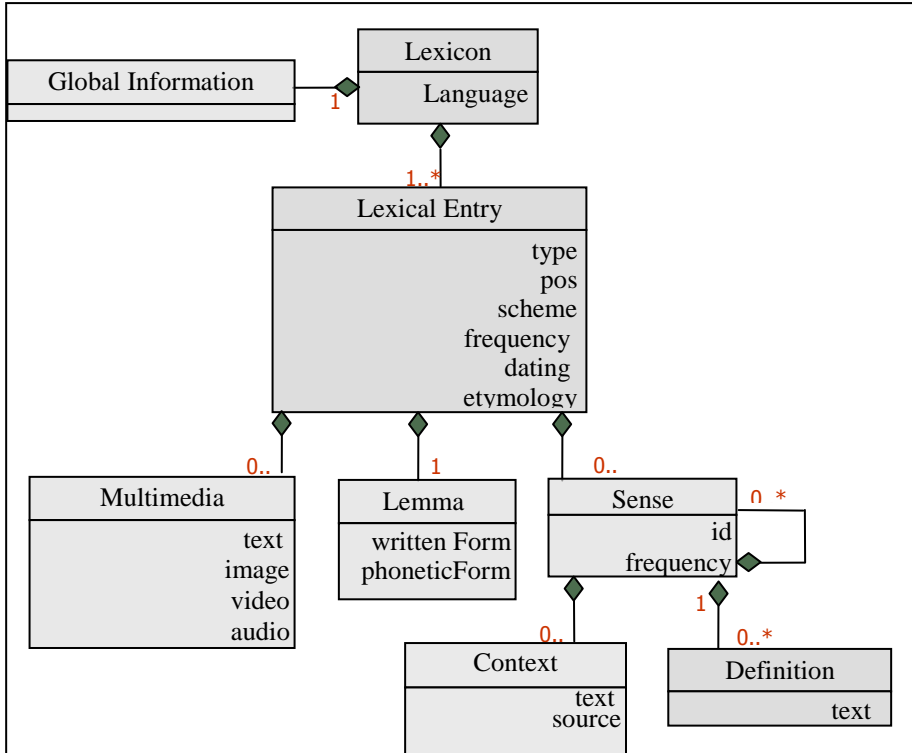
- **التعريف (Definition)**: يشتمل على نص لتقديم تفسير المعنى الواحد وتدقيقه.

- **الشواهد (Context)**: يشتمل هذا الصنف على:

○ الشواهد (أو الأمثلة)

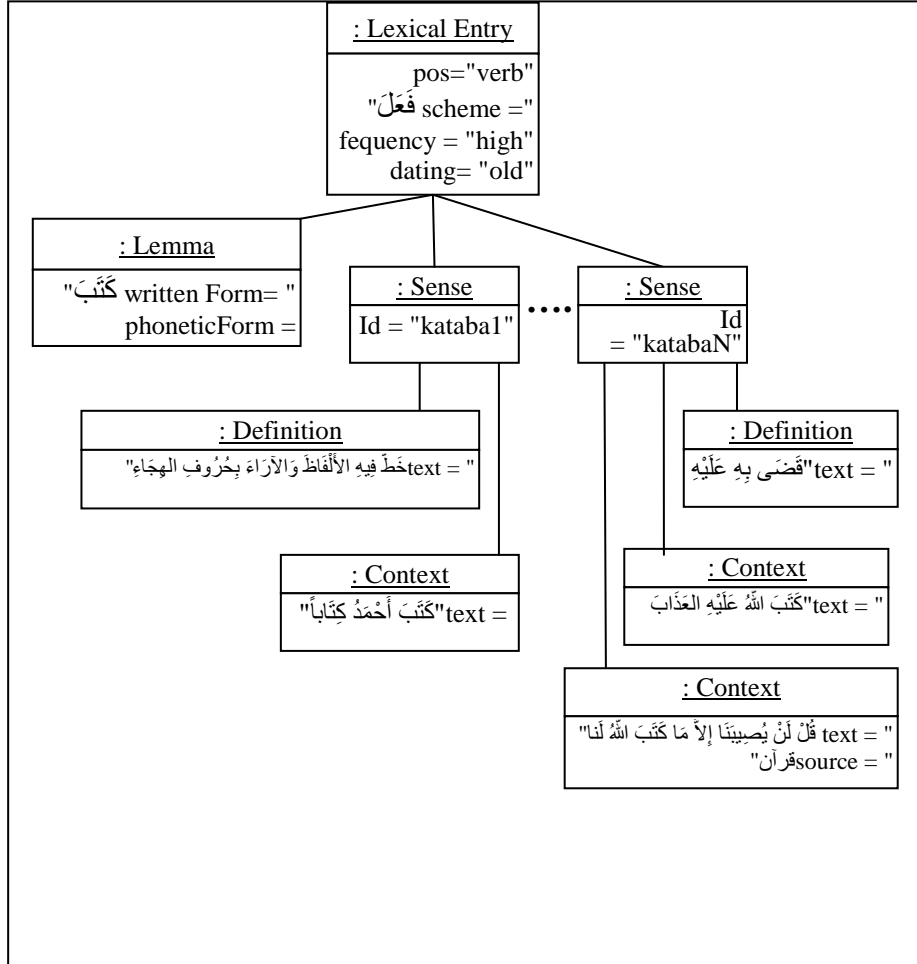
○ مصادرها (source): معاجم قديمة وحديثة، كتب الأدب ، القرآن الكريم، الحديث النبوي.

- **الوسائط المتعددة (multimedia information)**: يحتوي هذا الصنف على أصوات، صور ثابتة ومتحركة، أفلام الفيديو وتفسيرات عامة حولها.



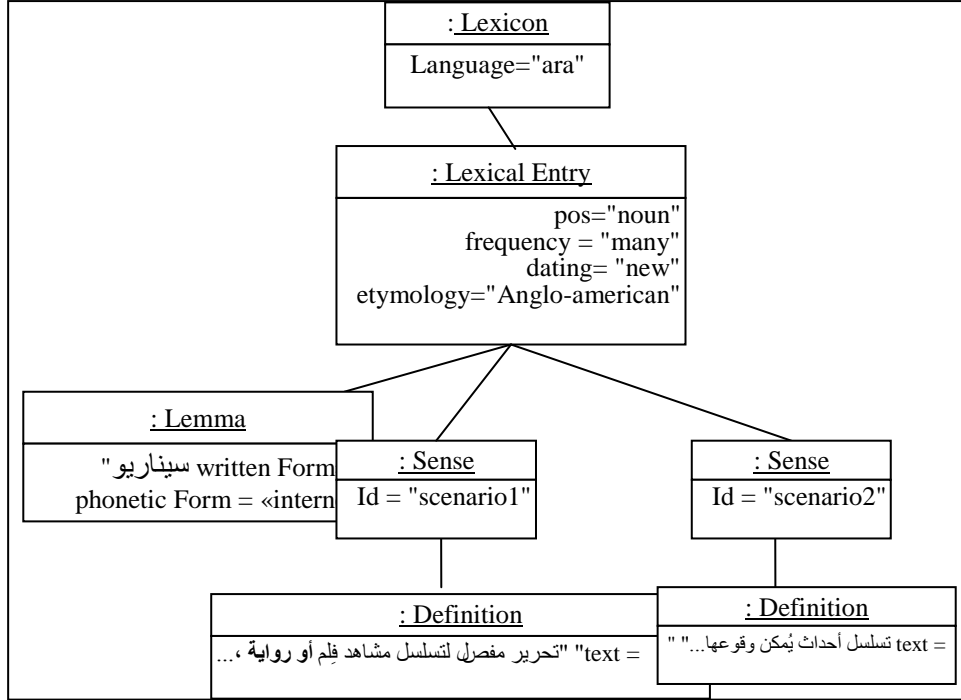
الرسم البياني عدد 2: مخطط النواة الأساسية

وفيما يلي أمثلة لتوضيح التصميم المقترح للنواة الأساسية تخص الفعل "كتب" والاسم "سيناريو" .



الرسم البياني عدد 3: تطبيق النواة الأساسية على فعل كتب

نلاحظ في هذا الرسم البياني أنّ فعل كتب ليس بحاجة لتحديد أصله (etymology) لذلك وقع الاستغناء على هذه المعلومة بالنسبة لهذا المدخل.



الرسم البياني عدد 4: النواة الأساسية لكلمة حديثة ودخيلة "سيناريو"

نلاحظ أن بالنسبة لكلمة سيناريو وقع الاستغناء على تمثيل الوزن (scheme) الذي كان ضروريا بالنسبة لكلمة كَتَبَ. بالمقابل يصبح أصل الكلمة (etymology) معلومة أساسية.

٢.٦ الوحدة الصرفية

تتم هذه الوحدة بتمثيل الخصائص الصرفية للمدخل والعلاقات الصرفية بين المداخل.

تتكوّن الوحدة الصرفية من صنفين من المعلومات (انظر الرسم البياني عدد5):
كلمة مُصرفة (word Form): هذا الصنف يمكّن من تمثيل المعلومات التالية:

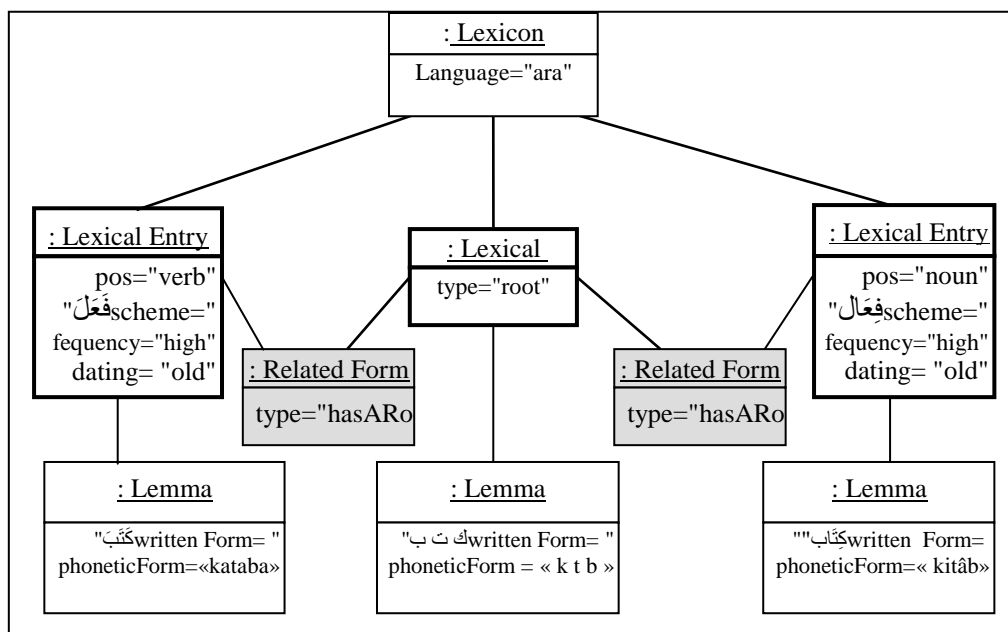
- العدد (Grammatical Number): مفرد، مثنى، جمع.
- الجنس (Grammatical Gender): مذكّر، مؤنث.
- الضمير (Person): المتكلم، الغائب، المخاطب.
- الصيغة (Verb Form Aspect): الماضي، المضارع، الأمر.
- البناء (Voice): المعلوم، المجهول.
- الإعراب (Casual Flexion): المرفوع، المجزوم، المنصوب.
- قابلية الجمع (Countability): نعم، لا.
- قابلية التعريف (Definiteness): نعم، لا.
- ...

- الروابط بين المداخل (Related Form): هذا الصنف يمكننا من ربط الكلمة بجذرها والعكس:

- الرّابط: hasARoot يمكن من الانتقال من الكلمة المشتق إلى جذرها. مثلا من فعل كَتَبَ إلى جذره "ك، ت، ب".
- الرّابط hasAStem يمكن من الانتقال من الجذع المشتق إلى الجذع الأصلي مثل زاد و استزاد.

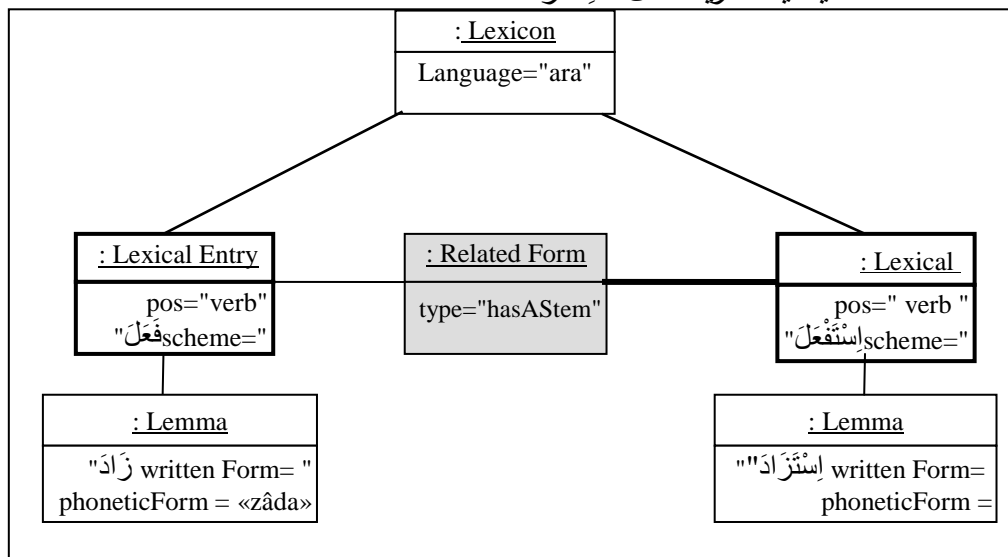
تمكّن هذه الرّوابط من تيسير التعرف (أو البحث) إلى جميع مشتقات المدخل الواحد (جذر أو جذع).

الرسم البياني عدد 5 يقدم توضيحا للمخطط المقترح للوحدة الصرفية بالنسبة إلى المدخل الخاص بالجذر "ك ت ب". كما يُبيّن العلاقة بين هذا المدخل والمدخلين المشتقين "كِتَاب" و "كَتَبَ".



الرسم البياني عدد 5: مثال للرباط "has A Root"

الرسم البياني عدد 6 يقدم توضيحا لكيفية الرّبط بين فعل ثلاثي مجرد مثل "زَادَ" و أحد المشتقات القياسية المزينة مثل "إِسْتَزَادَ".



الرسم البياني عدد 6: مثال للرباط "hasAStem"

الوحدة النحوية

تهتم الوحدة النحوية بتمثيل السمات النحوية للمداخل (Syntactic Behaviour). فهي تمكن من تعداد التراكيب الممكنة للكلمة (خاصة بالنسبة للفعل والحرف) وبالتالي يمكن من تحديد تعدي الفعل ولزومه، وكذلك تقسيم الأفعال المتعدية إلى ثلاثة أنواع : متعد لمفعول بنفسه، متعد لمفعول بأداة "حرف جر" أو "ظرف"، متعد لمفعولين.

تتكون هذه الوحدة من أربعة أصناف من المعلومات وهي (انظر الرسم البياني عدد8):

- السمة النحوية (Syntactic Behaviour): يجمع هذا الصنف التراكيب النحوية المختلفة التي ترد فيها الكلمة. ويمكن تحديد المعنى الذي تؤديه الكلمة إذا ذكرت في هذا السياق.

يقع تدقيق مكونات كل تركيبية بوضعها في إطارها النحوي الخاص بها.

- الإطار النحوي (Subcategorisation Frame): يقوم هذا الصنف بتفصيل السمة النحوية للكلمة. فهو يهتم بتركيبية ما من جملة التراكيب الممكنة للكلمة. ويشار إلى ذلك باستعمال نمط تركيبية يضع هذه الكلمة في سياقها. هذا النمط يمكن أن يكون من بين الأنماط التالية:

- فعل/فاعل (VS): إذا كان الفعل لازما.
- فعل/فاعل/مفعول (VSO): إذا كان الفعل متعد لمفعول واحد بنفسه،
- فعل/فاعل/أداة/مفعول (VSPO): إذا كان الفعل متعديا لمفعول بأداة.
- فعل/فاعل/أداة/مفعول 1/مفعول 2 (VSPOO): إذا كان الفعل متعديا لمفعولين.
- ...

وتجدر الملاحظة أن استعمال هذه الأطر وتدقيقها لا يقتصر على تحديد

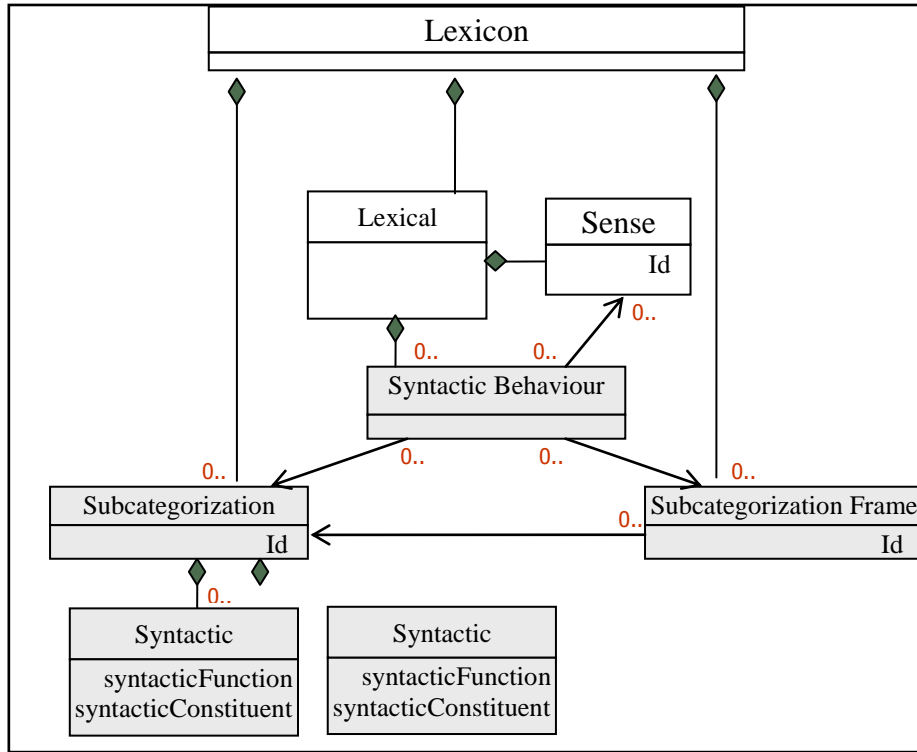
التعدي واللزوم فحسب، بل يمكن من التعرف إلى الاستعمالات الممكنة للكلمة في الجملة.

- الدالة النحوية (Syntactic Argument): يستعمل هذا الصنف لتدقيق

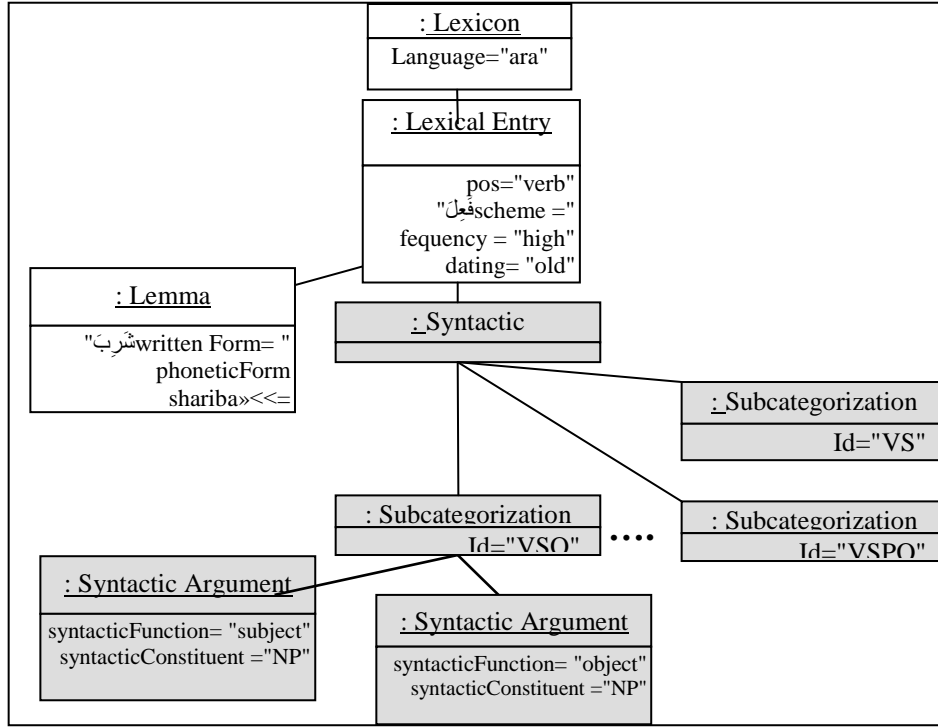
أحد خصائص الإطار النحوي. وذلك بتحديد المعلومات التالية:

○ الوظيفة النحوية (Syntactic Function): فعل subject، فاعل object.

○ المركب النحوي (Syntactic Constituent): مركب اسمي (NP)،
...،PP



الرسم البياني عدد 7: الوحدة النحوية



الرسم البياني عدد 8: الوحدة النحوية لفعل "شَرِبَ"

يوضح هذا الرسم البياني كيف يقع ربط فعل "شَرِبَ" مع الإطارات النحوية

الممكنة التي تخصه وهي:

- VS: فعل/ فاعل (شَرِبَ الْوَلَدُ)
- VSO: فعل/ فاعل/ مفعول به (شَرِبَ الْوَلَدُ الْمَاءَ)
- VSPO: فعل/ فاعل/ حرف جر/ مفعول به (شَرِبَ الْوَلَدُ بِسُرْعَةٍ)
- ...

٤.٦ الوحدة الدلالية

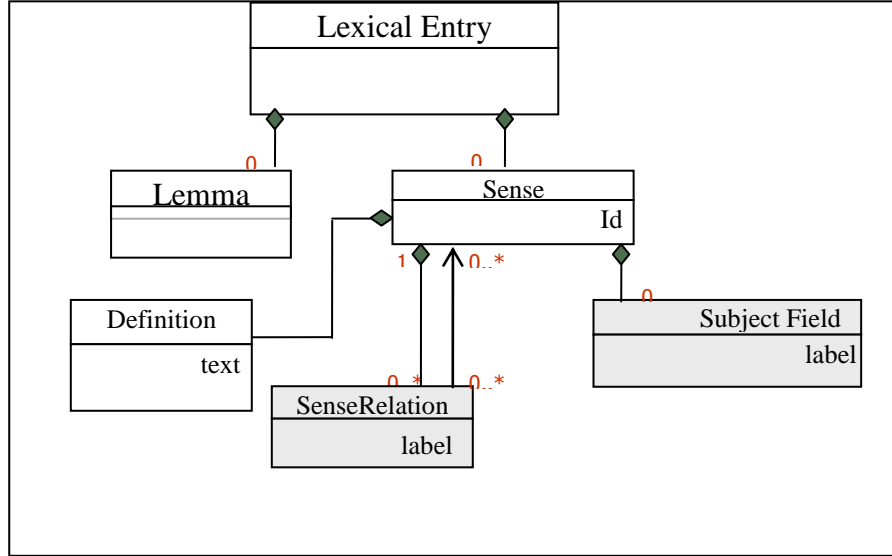
تهتم الوحدة الدلالية بتدقيق معاني المداخل وذلك بتمثيل العلاقات الدلالية بينها (الترادف والتضاد) وكذلك بتحديد مجالاتها.

تحتوي الوحدة الدلالية على صنفين من المعلومات (انظر الرسم البياني عدد9):

- العلاقة الدلالية (Sense Relation): تربط هذه العلاقة بين معنيين مختلفين.

يمكن تصنيف هذه العلاقة الدلالية إلى عدة أنواع: الترادف، التضاد، التعميم (طالب - إنسان)، التخصيص (إنسان - طالب)، جزء من (اليد - جسم الإنسان)، التسلسل الزمني (الخميس - الجمعة)،.... هذا الصنف من المعلومات يلعب دورا مهما في بناء شبكة دلالية "أنتولوجيا" انطلاقا من المعجم.

- مجال المعرفة (Subject Field): يعني هذا الصنف بمجال معنى الكلمة. الكلمة الواحدة يمكن أن تتعدد مجالات استعمالها: رياضيات، علم النبات، لغة، طب،...



الرسم البياني عدد 9 : مخطط الوحدة الدلالية الذي تعنى بالعلاقة الدلالية

7. الخاتمة

في هذه الدراسة تطرقنا إلى جوانب عديدة تميز المعجم الإلكتروني عن المعجم الورقي من حيث الهيكلية والمحتوى وطرق الاستعمال. بينا بالخصوص أهمية المعجم الإلكتروني العربي بالنسبة إلى مجتمعنا الذي يتحول بسرعة إلى مجتمع المعرفة؛ حيث تلعب فيه التكنولوجيا دوراً محورياً. لقد أصبح المعجم الإلكتروني أداة عمل لا يمكن الاستغناء عنه بالنسبة إلى الكثير من الفئات الاجتماعية والمهنية، وهو كذلك حجر الزاوية للعديد، إن لم نقل لجل، تطبيقات المعالجة الآلية للغة. هذه التطبيقات تساهم بصفة جدية في إشعاع اللغة العربية واستمرارها كلغة ناقلية. كذلك تطرقنا إلى مختلف طرق بناء المعجم العربي وختماً بموضوع تقييم المعاجم الإلكترونية لجعلها تخضع إلى مواصفات المعاجم الحديثة من خلال تجربة مخبر ميراكل. إتاحة مثل هذه المعاجم للغة العربية سيدعم حتماً دورها في مجتمعنا الذي يعيش منافسة شرسة من لغات أجنبية أهمها الإنجليزية والفرنسية. فمسؤوليتنا اليوم أعظم مما كانت عليه بالأمس لأن اللغة أبعاداً ثقافية، واقتصادية واستراتيجية.

المراجع

- Baccar F., Gargouri B., Ben Hamadou A. "Methodology for Generating Core Domain Ontologies from LMF Standardized Dictionaries ". In International Journal of Applied Ontology. Mark Musen and Nicola Guarino (Eds.), IOS Press, Amsterdam, 2011.
- Khemakem A., Gargouri B., Ben Hamadou A. "An approach for generating personalized views from normalized electronic dictionaries: A practical *experiment on Arabic language*", CLA'2010
- Khemakhem A., Elleuch I., Gargouri B., Ben Hamadou A., (2009). Towards an automatic conversion approach of *editorial Arabic dictionaries into LMF- ISO 24613 standardized mode. MEDAR 2009, Cairo, Egypt.*
- Baccar F., khemakem A., Gargouri B., haddar K., Ben Hamadou A. (2008). "LMF standardized model for the editorial electronic dictionaries of Arabic", NLPCS 2008, 12-13 june, Barcelone.
- Francopoulo G. & George M. (2008). ISO/TC 37/SC 4 N453 (N330 Rev.16). Language resource management- Lexical markup framework (LMF).
- Ait Taleb S., (2005). Dictionnaires électroniques arabes: le modèle des dictionnaires de Sakhr, revue de l'Association Marocaine des Etudes Lexicographiques, Numéro 3-4, 15-31.

التعليقات والمناقشات

- د. بشير الخضرا/ جامعة الإسراء

أشار إلى تقصير مجامع اللغة العربية بتأخر تجربة المعجم الإلكتروني العربي، وتساءل: أين وصلت هذه التجربة، وهل من جهودٍ بُذلت في هذا الإطار لنلحق بركب العلم الإلكتروني في مجال المعاجم؟

- د. حامد قنبيبي/ جامعة الإسراء

قال إن المعجم العربي الإلكتروني يحتاج إلى دراسة عميقة على شتى المستويات حيث إنه كتاب يجمع بين دفتيه ألفاظ اللغة ومفرداتها وتراكيبها ومعالمها الحضارية بغية شرحها وإيضاحها شريطة أن يرتب ترتيباً معيناً. وأهم ما يجب مراعاته في المعجم الإلكتروني طبيعة اللغة الاشتقاقية التي تقوم على أساس استكمال الناقص من واقع فهم اللغة العربية وطبيعتها الاشتقاقية والقياسية.

كما أشار إلى أهمية مواءمة "معجم ألفاظ الحياة العامة" إلكترونياً ليصل إلى المتكلمين باللغة العربية لما له من أهمية بالغة في حضارة الأمة.

- أ. د. سرى سبع العيش

في هذا السياق أشارت إلى تجربة شخصية في الترجمة مستفيدة فيها من المعجم الطبي العربي الإلكتروني الموحد، حيث تقوم حالياً بترجمة المعجم الطبي الدلالي، وأشادت بمادة المعجم الطبي العربي الإلكتروني إلا إنه ورغم احتوائه على مئة وخمسين ألف كلمة مدونة تقريباً ينقصه الكثير من المفردات التي - وللأسف- يلبيها المنجد أحياناً.

- رد أ. د. عبد المجيد حمادو

أكد أنه قدّم في بحثه طريقةً لبناء معجم عربي إلكتروني.

كما أشار إلى واقع العمل في هذا المعجم حيث انتهوا - في جامعة صفاقس التونسية- من إعداد قاعدة البيانات المعجمية والنظام الذي يمكّن المعجمي من إدخال كافة متطلبات المعجم، وهذا عملٌ مضمّنٌ يحتاج لإدخال بشريّ ضخم، والدعوة مفتوحة لكل من لديه مادة معجمية صحيحة متكاملة شاملة مرتبة مدروسة وواضحة المعالم لإدخالها وفق قاعدة البيانات المعدّة.

