المعجم الإلكتروني:
قراءة في بعض التجارب الغربية والعربية

ذ. عمر المهديوي
كلية الآداب، عين الشق الدار البيضاء

توطئـــة

تعتبر المعاجم الإلكترونية أدوات أساسية بالنسبة للتحليل والتوليد الآليـــين في اللغات الطبيعية، وهي أيضا عنصر أساس لمعالجة اللغة البشرية، واللغة المســـتخدمة في الحاسوب. لذا ينبغي أخذهما بعين الحسبان، بالتركيز على اللغة البشرية على مســتوى الاكتساب، وعلى اللغة الاصطناعية على مستوى البناء.

إن بناء معاجم آلية يتوخى خدمة أغراض المدقق الإملائـــي والمـــترجم الآلي، وغيرهما من مجالات وتطبيقات المعالجة الآلية للغات الطبيعية (NLP).

إن أهمية المعجم تكمن في كمية المعلومـــات الـــتي يتوفـــر عليـــها، والأدوات والتقنيات التي يعتمد عليها في الكشف عـــــن معلومـــات المداخـــل المعجميـــة. إن إستراتيجيات تفوق المعجم إذن تحدد من ناحية في المعارف الضرورية التي من المســـلم أن يقدمها عن كل مدخل (مفتاح معجمي)، ومن ناحية أخرى، في طريقة ومنهجيـــة المعالجة والمدة الزمنية التي يستغرقها التفتيش عن المعلومة. فعلى سبيل المثال لا الحصـــر،

عادة ما يبدأ البحث بتقديم معنى المفردات على مستوى التحليل، ثم يأتي بعد ذلك دور التأليف أو التركيب، حيث يتم ضبط مفاهيم وتصورات المفردات.

I- أنواع المعاجم

إن الغنى في المعارف والحاجيات ومتطلبات الصناعة المعجمية الآلية فرض التمييز بين أنماط المعاجم:

- المعاجم أحادية اللغة: وينحصر دورها في وصف اللغة في المعجم، إنها تقوم بدور الإحالة الوصفية بالنسبة لمتكلم اللغة الخاصة، ومنها ما يتخذ طابعا موسوعيا كلسان العرب لابن منظور؛

- المعاجم ثنائية اللغة: يسمى هذا النوع بمعاجم الترجمة، لأنها تقوم بدور الوصف من اللغة المصدر إلى اللغة الهدف؛

- المعاجم متعددة اللغة: وهي التي توظف لغات متعددة، وتنقل المعارف والمعلومات بين متكلمين مختلفين؛

- معجم المترادفات: وهو معجم يبحث في العلاقات الدلالية بين المفردات[1]؛

- المعجم اللغوي الموسوعي؛

- الذخيرة اللغوية؛

- المعجم الآلي أو الإلكتروني؛

- المعجم المحوسب Lexique informatisé : وهو عبارة عن مدونات لغوية صغيرة أو كبيرة الحجم، تخزن في أقراص، باستخدام تقنية المعلومات، يظهر الجانب الحاسوبي فيها على مستويي الإدخال والإخراج، حيث تنجز كل عملية من تينك بسرعة، وبأقل كلفة. وتزخر الشابكة Internet على نسخ إلكترونية من هذا النوع من

(1) مهديوي عمر وسلوى حمادة السيد، **المعالجة الدلالية الآلية للغة العربية**، "مجلة العربية"، ع. 3، 2006.

المعاجم في مجالات شتى، وهي ليست بمعاجم آلية البتة، والجانب الحاسوبي يكمن في بنيتها التركيبية لا أقل ولا أكثر. لهذا لا يمكن أن تقوم بدور قاعدة المعطيات الآلية التي تشكل حجر الأساس في بناء معاجم حاسوبية من جهة، وبالنسبة للتحليل الآلي للنصوص في اللغات الطبيعية من جهة أخرى.

وفي ما يلي جدول يلخص أنواع المعاجم الموجودة في لغات العالم، والتي يمكن تصنيفها حسب زوايا النظر على النحو الآتي:

التصنيف	النوع		النماذج المعجمية
نقطة الانطلاق	معاجم الألفاظ		المصباح المنير للفيومي
	معاجم المعاني		المخصص لابن سيده
طريقة الترتيب	ألفبائي حسب الأوائل	بدون تجريد	معجم الرائد لجبران سعود
		بعد التجريد	أساس البلاغة للزمخشري
	ألفبائي حسب الأواخر	بدون تجريد	التقفية في اللغة لأبي بشر اليمان
		بعد التجريد	القاموس المحيط للفيروز أبادي
	صوتي		العين للخليل بن احمد
	صرفي (علبي الأبنية)		مقدمة الأدب للزمخشري
	موضوعي (معاجم المعاني)		متخير الألفاظ لابن فارس
العموم والخصوص	معاجم عامة		لسان العرب لابن منظور
	معاجم خاصة	اشتقاق	تصريف الأفعال لانطوان الدحداح
		معربات	المعرب للجوا ليقي
		سياقي	معجم التراكيب والعبارات الاصطلاحية لأحمد أبو سعد
		مترادفات	معجم المعاني للمترادف والمتوارد والنقيض لنجيب اسكندر
		شخص أو نص	معجم شعراء المعلقات لندي الشايع

	معجم للنطق لهجة أو لهجات		معجم دانيال في نطلق اللغة الإنجليزية معجم الألفاظ العامية لأنيس فريحة
عدد اللغات	أحادي		المعجم العربي الأساسي (المنظمة العربية للتربية والثقافة والعلوم)
	ثنائي		قاموس سعادة(إنجليزي-عربي) لخليل سعادة.
	متعدد		قاموس(إسباني- فرنسي- عربي) لعلا عبد الحميد سليمان
حجم المعجم	معجم كبير		تاج العروس (للزبيدي)
	معجم وسيط		المعجم الوسيط(لمجمع اللغة العربية بالقاهرة)
	معجم وجيز		المعجم الوجيز(لمجمع اللغة العربية بالقاهرة)
	معجم جيب		المصباح المنير للجيب (عربي- عربي) للفيومي
الفترة الزمنية	معجم معاصر		Longman Dictionary of Contemporary English
	معجم تاريخي		معجم أكسفورد للغة الإنجليزية
الهدف	وصفي		معجم وبستر في طبعتيه الأولى 1882) والثانية (1934)
	معياري		معجم وبستر في طبعته الثالثة (1961)
نوع مستعمل	ابن اللغة		تاج العروس للزبيدي
الآلة / الحاسوب/ الإنسان	حاسوبي محوسب		معاجم صخر معجم الغني

II- المعجمية الحاسوبية

لقد ساعد التطور السريع الذي عرفته تكنولوجيا الحواسيب، في الآونة الأخيرة، على معالجة المعلومات غير العددية بأقل تكلفة وبأسرع ما يمكن من الوقت، مما يتصل بالتخزين والبنينة وإنشاء الإرساليات/الرسائل اللسانية انطلاقا من قواعد مضبوطة وصارمة. ومن نتائج تقنية المعلومات أن المعلوميات، بأدواتها ووسائلها، أحدثت تأثيرا بارزا على مختلف نواحي الحياة اليومية للأفراد،إلى درجة أصبح الكل يستغل نتائج وتطبيقات المعلوميات في إنجاز عمله. هذه الظاهرة أحدثت تغيرات عميقة وجذرية على مختلف الأنشطة التي يقوم بها الفرد، ومن ثمة أصبح الإنسان يعيش على إيقاع الانفجار المعلومياتي الذي ترتبت عنه ثورة وانقلاب على مفاهيم الزمن والمكان، والتنقل، والتواصل والتصدي للمعرفة من خلال شابكة المعلومات أو شابكة الشبكات التي ترتب عنها أيضا، أن كل من البريد الإلكتروني والدردشة والندوات الإلكترونية قلصت مسافة الاتصال بين المتحدثين.إن مفهوم الفضاء، بالمعنى الفيزيائي، قد اندثر و لم يعد له محل من الإعراب في الثقافة الرقمية الحديثة، إذ تهدمت الحدود بين المناطق الجغرافية والقارات إلى درجة أصبح معها العالم أشبه ما يكون بالغرفة الصغيرة، مما يسر سبل تبادل المعلومات بين الشعوب والأمم، كما أن الأقراص المغنطة وأقراص التسجيل الصوتي سهلت تخزين المعلومات وأرشفتها كيفما كان نوعها، ومهما بلغ حجمها.

إن اللقاء بين المعلوميات واللسانيات أدى إلى ميلاد علم جديد يصطلح على تسميته بـ"اللغويات المعلومياتية أو علم اللغة الحاسوبي". وهو ميدان يقع في نقطة الوصل بين العلوم الإنسانية والعلوم الصلبة التي تهتم بدراسة الأنظمة الخاصة بفهم وتوليد اللغة البشرية.

لكن السؤال الذي يطرح نفسه هو: كيف نبني قاعدة بيانات معجمية للمفردات اللغوية موجهة لخدمة المعاجم الإلكترونية للغة العربية؟ ثم ما هي الإمكانات التي توفرها الآلة للمعجمي؟

تتوفر الآلات الضخمة والكبيرة على السرعة الفائقة في إنجاز العمليات الحسابية، حيث تنجز الملايين من العمليات في الثانية، وتتوفر على ذاكرة ضخمة. وقد سبق لشركة (إ.ب.م) منذ سنوات خلت أن عرضت ذاكرة بسعة 3856 وحدة، قادرة على تخزين ما يقارب 470 مليار خاصية. إذ يمكن تخزين مليون كتاب من حجم 200 صفحة، ويمكن الحصول على أية جملة من كل كتاب في ظرف وجيز لا يتجاوز 20 ثانية، بمعنى آخر، يمكن تخزين وأرشفة كتب الخزانة الوطنية الفرنسية أو الخزانة العامة بالمغرب.

وبالمقابل، فإن الآلات المتوسطة أي الأقل ضخامة تتوفر على معالجات تستطيع القيام بعمليات مختلفة، فالحاسبات اليوم تتوفر على معالجات لها من القدرة الاستيعابية ما يمكنها من تنفيذ عمليات حسابية، مثلا معالج INTEL 8086 يستطيع القيام بملايين عملية في الثانية، وتخزين أربع مرات 64000 خاصية.

يمكن للمعجمي أن يستخدم معالجات ذكية سواء في البيت أو المكتب، تساعده على التواصل مع حاسوب آخر أكثر ضخامة، والاستفادة من المعطيات من خلال برنامج معلوماتي معين، وهذا الحاسوب الذكي قادر على القيام بعمليات حسابية سريعة تسرع من عمل المعجمي وتنفذه بأقل تكلفة، وبدقة عالية ومهارة غاية في الإحكام.

يستطيع المعجمي من خلال المعالج الآلي الحصول على المعطيات والمعلومات بطريقة أشبه ما تكون بالمباشرة، ويستمد هذه المعلومات من بنك المعطيات المخزنة في الحاسب على الشابكة، وفي الوقت نفسه يمكنه معالجتها.

إن استخدام الحاسوب بالنسبة إلى المعجمي لا ينحصر فقط في رصد الخصائص المميزة لهذه التقنية، بقدر ما يتعداه إلى الإمكانية التي يتيحها الحاسوب على مستوى التطبيقات الآلية، إذ يساعد على تنظيم بنوك المعطيات الضخمة، وهو ما لا يتيسر له القيام به اعتمادا على الذاكرة البشرية المحدودة.

ومن بين الإمكانيات التي تتيحها الآلة للمعجمية نذكر ما يلي:

- تحرير النص على الآلة؛
- استرجاع المعلومات بكل يسر وفي أسرع وقت ممكن؛
- ضعف كلفة تخزين المعلومات وتكشفيها؛
- استخدام المنهجية الصورية (الخوارزميات) في معالجة النصوص؛
- الاستفادة من تقنية المدقق الإملائي والنحوي في مراجعة النص المعالج؛
- توظيف تقنية القارئ الضوئي؛
- التمكن من تجاوز المراحل التقليدية التي كان يعتقد خطأ أنها إجبارية في معالجة نص من النصوص؛
- إمداد المعجمي بالمعلومات التي تسهل عمله؛
- السرعة في إنجاز العمليات والقيام بالتعديلات الممكنة إما بالحذف أو الزيادة في محتوى الكتاب، مما يسمح بتطوير العمل المعجمي وبلورته في أية لحظة.

تهدف المعلوميات إذن، إلى جمع المعلومات وحفظها في بيانات ومعطيات مبنينة واسترجاعها، وفي هذا الإطار طبقت تقنيات تخزين المعلومات ومعالجة البيانات على بنوك توثيق المعلومات وحفظها، ويمكن الإشارة في هذا السياق إلى بنوك المعلومات المصطلحية المنتشرة في مناطق مختلفة من العالم العربي (بَاسم) بالمملكة العربية السعودية و(مكانز) الشركة العالمية للإلكترونيات بالقاهرة، (المعربي) Lexar، والذي تم بلورته وتطويره فيما بعد بمعهد الدراسات والأبحاث للتعريب بالرباط.

تتوجه المعلومة المعجمية المبنينة المراجعة لخدمة تقنيات التخزين؛ لهذا فالحاسوب يساعد على تخزين الوحدات غير المستمرة وبرمجة الذاكرة على المعلومة النحوية الأكثر أو الأقل تعقيدا مقترنة بوحداتها اللسانية، ويتعرف عليها في معجم اللغة وفق برنامج المؤشرات النحوية وطريقة النطق أو الكلام، وأصول الكلمات (الإيتمولوجيا)، التاريخ والتعريف والأمثلة. إن المعاجم الثنائية اللغة تمدنا بالمقابلات الضرورية للمصطلحات العلمية والتقنية، في حين ينحصر دور المعاجم الأحادية اللغة في تزويد المستعمل بالمصطلحات المعاصرة المرتبطة بكفايته اللغوية.

III- قاعدة المعطيات المعجمية Base de donnée lexicale

ما المقصود بقاعدة المعطيات المعجمية ؟ وكيف يتم تثبت الملفات فيها؟ وما علاقتها بقاعدة المعطيات النصية ؟ وما دورها في بناء المعجم الإلكتروني؟

تحدد المعلوميات قاعدة المعطيات بأنها مجموعة من المعلومات الدقيقة وغير الحشوية والضرورية لسلسلة من التطبيقات الآلية، والتي تنبني على نظام رياضي منطقي يتولى تدبيرها.

إن بناء قاعدة معطيات معينة عملية تنطلق من ملاحظة الحالة المدروسة للوصول إلى نتائج مضبوطة تشكل حجر الأساس لقاعدة المعطيات. وبهذا المعنى، إن قاعدة المعطيات هي كم من المعلومات المجمعة في ميدان معرفي معين بطريقة منظمة وصارمة في الآن نفسه.

وحسب علم الجبر، فإن قاعدة المعطيات تعد بمثابة مجموعة نهائية تحتوي على عناصر مقرونة بخصائصها المشتركة، ووصفها يتم بطريقة محددة، وبناؤها يستند إلى معايير صارمة تشكل أساس التنظيم الصوري الذي يتحكم في محتوى المجالات غير الموحدة والملتبسة قصد تطبيقها على كل عناصر المجموعة المستخدمة في الوصف.

وعليه فإن هذا النوع عليه تقوم معالجة قواعد المعطيات الآلية بتطبيق وإنجاز أنظمة تدبير قاعدة المعطيات[2] المتخصصة، والتي تمكن من القيام بعمليات سريعة على مستوى المراقبة أو المعالجة ،واسترجاع المعلومات إدخالا وإخراجا من وإلى قاعدة المعطيات.

تمثل قواعد المعطيات بالنسبة للمعجمية إحدى الركائز الأساسية التي ينبني عليها العمل المعجمي، وقواعد المعطيات تعد مهمة وأساسية بالنسبة لأي تخصص كيفما كان نوعه. بيد أن مفهوم قاعدة المعطيات (Base dedonnées) يتداخل مع مفهوم بنك المعطيات Banque de données، إذ يقترح لهما برنار كيمادا تعريفين إجرائيين[3]، يرسم من خلالهما الحدود الفاصلة بين المفهومين،حيث يقول:

تحدد قاعدة المعطيات بأنها جهاز أو عتاد حاسوبي، نتاج تجميع مجموعة من المعطيات، أحداثا (وقائعا) كانت أم معلومات، تشترك في خاصية مشتركة ألا وهي أنها تكون مسجلة على الحاسب وفق قواعد التنظيم والترتيب والتنفيذ (الإدخال) المضبوطة وأن برنامج أو برامج المعالجة أو الاستفسار (المساءلة) تخضع بشكل خاص لهذه القواعد. بينما لبنك المعطيات بعد تكميلي أو إضافي، إنه يتناسب مع المؤسسة التوثيقية الآلية التي تدبر أو تشغل جهازا أو أجهزة في بناء المعطيات وموجهة لخدمة جماعة من المستفيدين، ولها أهداف خاصة تتمثل في جمع وحفظ المعلومات المجمعة، والأكثر شمولية (استيعابا) بهدف تأمين الوثائق المستعملة من قبل المستفيدين أو المنفذين. إنها تتوخى حل مشاكل التدبير والتسيير الإداري والقانوني والتي يبقى بعضها بدون حل، ولاسيما ما يتعلق بالجانب المالي، والأمن، والخاصية الافتتاحية والعلمية، إلخ.

---

(2) (**DBMS**) Data Base Management Systems.

(3) Bernard Quemada, *Bases de données informatisées et dictionnaires*, Lexique 2, Presses Universitaires de LILLE, p.105.

إن قاعدة المعطيات فكرة مبتذلة أو تافهة من حيث المبدأ، ولكن ما يهم المعجمي بشكل أساسي هو العتاد المزدوج: فمن جهة هناك النصوص (نصوص أو عناصر نصية عادية، نصوص طبيعية، مواد قاموسية، إلخ). ومن جهة أخرى، هناك رهان خوارزميات المساءلة أو الاستفسار (برامج، قواميس آلة أو آلية، أنظمة توثيقية)، والتي انطلاقا منها تصبح المعالجات أو التحليلات الآلية للنصوص سهلة ومتيسرة للغاية. على أن درجة تعقيدها تحدد نوعية ومستوى تطوير النتائج المحصل عليها.

وبالنسبة للأعمال المعجمية، إن التحليلات تنقسم إلى نوعين: التحليلات البسيطة من النوع التسلسلي/التعاقبي، تمكن من كشف أو استكشاف المتتاليات أو تلازمية الكلمات حسب ترتيبها في الخطاب (توزيعات، وتواردات)، والنوع الآخر من التحليلات يمكن من الحصول على العناصر التي تنتمي إلى هذه الأصناف، والمقولات أو البراديغمات المعجمية، إلخ.

وانطلاقا مما سبق، يعتبر المعجم الإلكتروني بمثابة قاعدة معطيات لاعتبارات عدة نحصرها فيما يلي:

- أنه يحيل على مجال معرفي معين له صلة مباشرة باللغة، وبشكل محدد المعجم؛
- أن بنيته مبنية على شروح وتفسيرات خاصة بالمظاهر الاستعمالية للمعجم؛
- أن العناصر التي يتضمنها هي وحدات معجمية تشترك في خصائص مشتركة؛
- أنه قابل للاستخدام في الدعامات والركائز الإشعاعية، وفي أنظمة التدبير المنطقي، وبمعنى أوضح برامج التحليل اللساني.

يعتبر المعجم الإلكتروني إذن بمثابة قاعدة معطيات معجمية قابلة للاستخدام الوظيفي كمحرك بحث بالنسبة لبرامج اللسانيات على الشابكة، كما يعد أيضا أداة معلومياتية، إذ أنه يفيد في تحقيق تطبيقات عدة،نذكر على سبيل الذكر أن حوسبة

الكتابة تقدم خدمات إيجابية تتمثل في سرعة الأداء، والدقة في الإنجاز، وضعف في الكلفة.

VI– أصناف قواعد المعطيات المعجمية الآلية

تنقسم قواعد المعطيات الآلية إلى أصناف وأنماط، كل صنف له خصائصه الذاتية التي يتقاسمها مع الصنف الآخر، نذكر منها ما يلي:

أ– **بنك المعطيات المعجمية**: تحدد بنوك المعطيات بأنها مجموعة من الأدوات والتقنيات التي تقبل التطبيق بسهولة على الإنجاز المعجمي. وتتسم بالسمات التالية[4]:

- **الإجبارية**: يتضمن المعجم الأحادي أو الثنائي أو المتعدد اللغة، اللغوي أو الموسوعي، العام أو المتخصص (جذور، عبارات، حكم وأمثال...)، معلومات نحوية صرفية (فعل، نوع،..، طريقة النطق، إملاء، إيقاع)، ودلالية: (ترادف، اشتراك، تضاد،...). وهي مجبرة للاستفادة من تقنية المعلومات وما تتيحه من إمكانيات وافرة فيما يتصل ببنوك المعطيات.
- **البنية**: تتسم بنوك المعطيات ببنيتها العلاقية أو الترابطية، وهذا يسمح بتمثيل المعلومات بشكل أوضح وأدق.

وتعد بنوك المعطيات المعجمية في الأساس، بنوكا نصية مختصة موجهة لخدمة التحليلات المعجمية، وهي كما يظهر من تسميتها تعنى بالعتاد الحاسوبي الذي يستجيب لمطالب الملفوظات بنوعيها المكتوب والمسجل ذات الخصائص خارج لسانية التي تؤشر وتحدد ضوابط الإنتاج، والغايات، والنقل، ومختلف التأليفات الممكن قيامها بين هذه العناصر برمتها.

---

(4) René Moreau et Isabelle Warnesson ; *Ordinateur et lexicographie*, Lexique 2, Presses Universitaires de LILLE ; 1983, p.126.

إن مجموع النصوص الموجهة للمعالجة الآلية تتطلب تهيئة وإعدادا خاصا، قبل مساءلتها بواسطة البرامج المعدة لهذا الغرض.

ب- بنك النصوص: les banques de textes تمثل تجميعا لأكبر عدد من النصوص المتضمنة لجميع أصناف الأشكال والصيغ والوثائق المكتوبة أو المنسوخة والمضبوطة وفق المعايير التي تحقق مطالب الجماعة المستعملة(5).

تهدف "المكتبات الإلكترونية" إلى تغذية وإثراء قواعد المعطيات النصية بالمعلومات الأولية، حيث تضطلع الوثائق النصية بدور تجميع الوثائق الآلية وحفظها.

إن بنوك النصوص تختص بالنصوص العلمية والتقنية والمنهجية، ونأخذ مثالا لذلك أرشيفات ووثائق جامعة أكسفورد المخزنة في بنوك نصية، وبنك النصوص الفرنسية لمعهد اللغة الفرنسية القديمة والحديثة والمعاصرة .

إن وظائف بنوك النصوص لا تنحصر فقط في المهام والأغراض التقليدية للمكتبات، وإنما ترتبط بمهام أخرى لها صلة بالتشغيل الحاسوبي في حالة النصوص على الآلة.إن مشكلة بنك النصوص إذن تتجاوز مرحلة القدرات والاهتمامات الخاصة بالمعجميين والقاموسيين.

ج- قواعد المعطيات النصية les bases de données textuelles : تعتبر بنوك المعطيات النصية مكونا مهما وأساسيا في المعالجة الآلية للغات الطبيعية، ولاسيما الحوسبة المعجمية، إنها تستجيب لمطالب العتاد الحاسوبي.

V- المعجم الإلكتروني: تحديدات أولية

ما المقصود بالمعجم الإلكتروني، وما هي مجالات استخدامه ؟ وما هي أنواعه، وما الفرق بينه والمعجم الورقي العادي ؟.

---

(5) Bernard Quemada *Ibid*, p.108.

للجواب على هذه الأسئلة، ننطلق مما أنجز في المختبرات العالمية للبحث الآلي، ومما أعده بعض الأفراد في تطبيق نظرية المعجم التركيبي، وسنكتفي هنا بالوقوف عند أعمال موريس غروس سابقا، وزلين هاريس، وغيرهما ممن أسهم في بناء معاجم آلية للفرنسية في مختبر المعالجة الآلية والتوثيق اللساني.

إن البحث في المعجم يقودنا إلى فهم القواعد التي تتحكم في تأليف المفردات، والحصول على معرفة واضحة بخصوص المظاهر التركيبية لها. ولقد ظهر مفهوم المعجم الإلكتروني في النصف الثاني من القرن العشرين(6)، مما أتاح الفرصة لتوفر مفرداته على الحاسب، وعلى الرقائق الإلكترونية، بحيث أصبح في الإمكان الحصول على أية مفردة بكل يسر، بل وتوظيفها في تطبيقات أكثر تعقيدا كمعالجة النصوص والتدقيق الإملائي، والترجمة الآلي، وغيرها، ومن ثمة ظهرت معاجم عامة، وأخرى متخصصة في الطب واللسانيات والهندسة والفلك.

نظريا، يمكن النظر إلى المعجم الإلكتروني من زاويتين: زاوية اللغويات المعلومياتية، وزاوية المعجمية أو المعجماتية التي تهتم بكيفية صناعة المعاجم الورقية، وتحديد مبادئها وآلياتها.

يشترك المعجم الإلكتروني مع المعجم الورقي في مستوى المقولة، والوصف المورفوتركيبي للمقولة المعجمية. ومن وجهة نظر معلوماتية ولسانية، يشكل المعجم الإلكتروني مادة خام لتقديم وصف شامل للمفردات صرفيا ونحويا. وباختصار، إن المعجم الإلكتروني عبارة عن قاعدة معطيات مشفرة.

يعتبر المعجم نموذجا للكفاية اللغوية لدى المتكلم (في دماغه البشري). وهو عبارة عن منظومة من الوظائف المسجلة في الكفاية المعجمية على شكل متواليات

(6) محمد منصور الغامدي وآخرون، الترجمة عبر الشبكة العالمية: **نظام مقترح للترجمة من وإلى العربية**، مجلة جامعة الملك سعود، "اللغات والترجمة"، 2006، ص. 8.

لغوية يصطلح عليها في علم المعجم بالمداخل المعجمية. إنه لائحة طويلة من المفـــردات مقترنة بمعطيات، ومرتبة ترتيبا خاصا. وتدعى هذه اللائحة من المفردات بالمدونـــة أو المتن. والمعجم عبارة عن زوج: المفتاحية والمعطيات، يقصـــد المفتاحيـــة المفـــردة في المعجم، وبالمعطيات المعلومات الصرفية والنحوية المقدمة عن كل مفردة، وقد تكـــون مسننة أي مشفرة، أو عبارة عن قاعدة، كما هو الأمر في المعجم الاستعمالي. كمـــا يمكن أن نقترح تعريفات وتحديدات أخرى بناء على مناهج وطـــرق التصديـــق Les méthodes d'accés.

يمثل المتغير ت حجم القاموس أو المعجم على شكل نص مكون من حـــروف تسمى بـــ les octets، أو مفردات الذاكرة (Mots-mémoire)، كما قد يمثل علـــى شكل مفاتيح أو معلومات في المعجم أو مفاتيح فقط.

يمثل الرابط T رابط الضغط القائم بين حجم المعجـــم "ت" في صـــورة نـــص والحجم التمثيل'T الممثل لها بالمعادلة التالية: R= F [T'T;]

ولكي يتحقق هذا الضغط، لابد أن يكون الرابط أقل من واحد. ويمثل المتغير n عدد المداخل المعجمية في القاموس أو المعجم، و يرمز إلى وظيفة الخوارزميـــات بـــــ Log bc التي تستخدم باعتبارها وظيفة لضبط تعقيدات الخوارزميات، ومن أجل ذلـك نوظف الترميز الآتي: o(f(n))

إن الدالة g(n) تنتمي إلى التعقيد o(f(n)) شريطة:

إذا وجدت c توجد No، بشرط أن تكون n > Nog(n) <= c.f(n)

حيث إن: c و N ثابتين/ ساكنين.

إن المتغير(n) لابد أن يكون منسجما مع الخوارزميات (المعطيات)، ذلـــك أن كتابة /تحرير المعجم على شكل نص عملية خطية تتوقف على حجم المعجم الـــذي يتخذ التركيبة التالية: o(I).

وبمجرد تطبيق هذا الإجراء، إن العملية الأساس لن تكون خطية. وفيما يتعلق بوحدة التخزين نستعمل الميغا أوكتي المرموز لها ب:Mo.

IV- الفرق بين المعجم الورقي والمعجم الآلي

يتجلى الفرق بين المعجم الإلكتروني والمعجم العادي في أن الأول مبني وموجه أساسا لاستخدام البرامج الحاسوبية التي لا تربطها أية صلة بالمعرفة اللسانية. بينما المعاجم العادية مثل المعاجم المتخصصة كقاموس التجارة يبنى أساسا لخدمة أغراض تجارية، حيث يرمي إلى توسيع مدارك الأشخاص في ميدان المعرفة التجارية والاقتصادية والمبادلات. في حين أن المعاجم اليدوية أو العادية يمكن أن تختزن على شكل مذكرات منسوخة أو في أقراص مدمجة، وتقبل الخضوع للبرمجة الحاسوبية أو الحوسبة، وهذا لا يمنع من أن توظف برامج التحليل اللساني، يسوغ ذلك الاعتبارات التالية:

- ✓ القاموس الإلكتروني ويشترط فيه أن يكون شاملا وعاما، لأن البرنامج اللساني المعد للمعالجة الآلية، لا ينبغي أن يفشل في العثور على أية معلومة كيفما كان نوعها وكيفما اتفق، لأن أي خطأ في المعلومات المدخلة من شأنه أن يتسرب إلى باقي مفردات الجملة أو قل النص برمته، ومن ثمة يعرقل عملية اشتغال البرنامج المعلومياتي.
- ✓ يشترط في المعلومات أن تكون في المعجم الإلكتروني واضحة، بينما في المعجم العادي غالبا ما يؤشر على المعلومات الأكثر شهرة أو تداولا بكلمة معروف أو بديهي، في حين لا شيء بديهي في الحاسوب، فكل المعلومات أساسية وضرورية، وينبغي أن تكون واضحة ومضبوطة.
- ✓ ترتبط المعاجم الإلكترونية ببرامج المعالجة الآلية؛ ذلك أن المعلومات المتضمنة فيها ينبغي أن تتلاءم والمداخل المعجمية.

IIV- قراءة في بعض التجارب المعجمية الآلية

إن بناء معاجم إلكترونية للغة العربية يتطلب الإلمام بالنظريات اللسانية الحديثة والنظرية الحاسوبية، كما يتطلب أيضا الاستفادة من التجارب العلمية للأمم والشعوب الأخرى وأخص بالذكر هنا تجربة مختبر LADL باعتباره مركزا علميا رائدا في ميدان معالجة اللغات الطبيعية، وقد أنجز معاجم إلكترونية للغة الفرنسية.

ما هي أهم الإنجازات الغربية والعربية فيما يتصل بالمعاجم الآلية والمَحوسبة؟

* **التجربة الغربية الفرنسية:** مخبر المعالجة الآلية والتوثيق اللساني LADL

يعتبر مخبر المعالجة الآلية والتوثيق اللساني LADL نموذجا من المراكز البحثية الرائدة في العالم على مستوى المعالجة الآلية للغات الطبيعية، إذ تمكن من بلورة معاجم إلكترونية للغة الفرنسية استنادا إلى نظرية المعجم التركيبي Lexique Grammaire .

يتألف نظام المعاجم الإلكترونية للغة الفرنسية من مستويات :

- DELAS؛ أي المعجم الإلكتروني للمفردات البسيطة، وهو يتضمن أكثر من 80000 مدخل معجمي، ويسند لكل مدخل رقم هو عبارة عن سنن، يعتبر جزءا من الخطاب، كما يسند إليه سنن رقمي يحد د الإعراب (flexion) مثل: (جناح، س21)، حيث يشير 21 إلى أن الاسم مذكر، وتضاف إليه لاحقة الجمع (ات). ويوظف DELAS حوالي 500 رقم، ويتراوح السنن الرقمي ما بين (1- 500)، كما يستعمل 2Mo لتخزين ملفات النصوص؛
- DELAC: تمكن الأرقام في من توليد حوالي 60000 شكل لغوي مأثور: صيغ تصريف الأفعال، الجمع والإفراد بالنسبة للأسماء، المذكر والمؤنث، المفرد والجمع بالنسبة للصفات، ويسمى المعجم الإلكتروني للصيغ المأثورة DELAF (B.Courtois 1984).

- DELAF: تتألف مداخله من مفردة وسنن رقمي، ويوظف أقل من 8500 سنن رقمي، يتراوح ما بين (1– 8500)، ويشغل مساحة 12Mo مخصصة لملفات النصوص، يشغل منها DELAF فقط 6,7 Mo .
- DELAC و DELAF مداخل الأول هي المفردات المركبة [Si–90] [Gr]. وتعرف المفردة المركبة بأنها متوالية من المفردات البسيطة، ذات خصائص تركيبية غير صالحة للمفردات البسيطة نفسها.

ومن الناحية الصورية، فإن مداخل DELAC هي سلسلة من المفردات البسيطة المقترنة بمعلومات ترتبط بالخطاب والإعراب. وعلى العموم فإن الأرقام الموظفة في DELAC هي نفسها الموظفة في DELAF.

بالإضافة إلى عناصر مجموعة المعاجم الإلكترونية المبلورة من قبل (LADL) هناك معاجم أخرى لا تقل أهمية عما سبق ومنها: DELAP وDELAPF.

DELAP أي المعجم الإلكتروني الفونيمي، مداخله تصل إلى 80000، مقترنة بأوصاف صوتية. المفتاحية فيه هي المفردة، والمعلومة المقدمة هي الفونيمية (Phonémisation) التابعة للمعلومات في DELAS.

إن ما يميز هذا النوع من المعاجم هو كونه مركب أو معقد بمعنى آخر، كل مفتاح أي مدخل تسند إليه معلومات مغايرة.

– المعجم التركيبي Grammaire Lexique وهو آخر عنصر من مجموعة المعاجم الإلكترونية، ويتعلق أساسا بقائمة الأرقام المسندة للجمل، والممثلة انطلاقا من فعلها، وتشكل مجموع هذه الأرقام الصيغ النحوية للمفردة.

إن تشغيل معاجم LADL يطرح مشاكل كثيرة على مستوى قواعد المعطيات، لأن مفاتيح التفوق أو النجاح تتسم بالتغير مما يجعل من المساحة المخصصة للتمثيل

المعلومياتي، والزمن المستغرق لنجاح المعطيات لا يسران على قدم المساواة، ومن ثمة ذلك سينعكس لا محالة على النتائج.

*** التجربة العربية: نماذج للحصيلة، ونماذج للتقويم**

1- شركة صخر للمعلوميات والإلكترونيات

تعتبر شركة صخر[7]، الشركة العربية الوحيدة التي لديها حالياً أكبر طيف من التقنيات والمنتجات، الموجهة لمعالجة اللغة العربية. وتتوفر حالياً على مدقق إملائي ونحوي، ومشكل آليٌّ، تسميه "المصحح الآلي الثنائي اللغة" يمكن أن يدمج في برنامج مايكروسوفت وورد، أو يستخدم مع بعض البرامج الأخرى، ناهيك عن التقنيات الأساسية التي طورتها في مجال المعالجة الآلية للغة العربية، مما يتصل ب "قاعدة بيانات المعلومات اللغوية"، والمعالج الصرفي متعدد الأوجه Multi-mode Morphological Processor، والمحلل النحوي متعدد الأوجه Multi- mode Syntactic Processor ، والمشكل الآلي Automatic Diacritizer ، بينما تتضمن قاعدة بيانات المعلومات اللغوية: "قاعدة بيانات المعاجم العربية"، و"قاعدة بيانات النحو العربي" و"الذخيرة اللغوية العربية".

2- تجربة الدكتور الحناش : المعاجم الآلية للغة العربية

يتألف المعجم الإلكتروني للغة العربية من ثلاثة مستويات متشابكة ومتكاملة فيما بينها، كل مستوى منها يجسد مستوى لسانيا بدءا بالمستوى الصوتي، ومرورا بالمستوى الصرفي، وانتهاء بالمستوى التركيبي.

- المعجم الإلكتروني العربي للجذور DEAR وهو عبارة عن قاعدة معطيات من المعلومات النحوية والصرفية التي تتحكم في استخراج الجذر من الكلمة في اللغة

(7) www. sakhr. com.

العربية. وعلى الرغم من تنوع مدونة الجذور العربية وتعددها،فإنه قد تم بنـــاء هـــذا المعجم، وذلك بإرفاق كل جذر لغوي بجميع الاحتمالات الاشتقاقية الممكنة،مما طبعـه بالشمولية والسهولة.وقد بني هذا المعجم على الشكل التالي:

ض.ر.ب : (ف= 0+)، (مص = 1+)، (فو = 2+)، (س= 0 +) إلخ.

حيث أن:

+ = الحروف الزائدة عن الجذر

– = الحروف الناقصة في الجذر

0 = ليست هناك زيادة في بنية الجذر

1 = أو أي عدد آخر يفيد عدد الحروف الزائدة عن الجذر لتوليـــد الصيغـة المشار إليها بالحرف قبل علامة =، وعندما يتعلق الأمر بنقص في عدد حـــروف الجذر يشار إلها بالرمز (ن) التي تعني نقصان.

إن صعوبة هذا المعجم تكمن في كيفية تشفير الحروف والأدوات التي لا تعـرف جذورا مطردة وعادية، كما هو الحال مع أقسام الكلم، فقد تم النظر إليها من خـــلال صورتها الإستعمالية، لا صورتها الأصلية غير الموجودة أصلا

– **المعجم الإلكتروني للمفردات البسيطة DEAMS** وفيه تثبت كل مفـردة في قاعدة المعطيات بناء على المعلومات النحوية والصرفية المرتبطة بكل واحدة منها. كمـا يسند لكل مدخل معجمي رقما هو عبارة عن سنن أو شفرة، يعتبر جزءا في الخطـاب اللساني، وسنن رقمي يمثل الإعراب.

لقد صنفت معظم المعاجم العربية المفردات اللغوية العربية بناء على خصائصـــها النحوية، وذلك بتمييزها بين الأفعال اللازمة والأفعال المتعدية بحرف واحد أو أكـــثر، كما صنفتها إلى أفعال عادية وأفعال قلوب وأفعال ناقصة. في حين انبنى التصنيـــف في

هذا المعجم على أساس أقسام الكلم مع أخذ بعين الاعتبار الخاصية النحوية للمدخل المعجمي (المفردة)، وعلاقتها بالمستوى المورفولوجي المناسب لها (الصيغة والاشتقاق):

ف: الفعل في الزمن الماضي للمفرد الغائب المذكر.

س : للمفرد المذكر والمؤنث.

ص : الصفة وسائر المشتقات بجميع أنواعها.

ظ : الظرف بجميع أنواعه.

أما المصدر فلم يخصص برمز معين، إذ تم إدماجه في معجم المفردات البسيطة المعربة.

- **المعجم الإلكتروني للمفردات البسيطة المعربة DEAMSF**، وفيه تسند المعلومات الأساسية لكل مفردة معربة (تصريف الفعل ونسبته إلى الضمائر وعلاماته الإعرابية مع كل واحد منها إلخ) بمعنى آخر، أن كل شكل من أشكل هذا المعجم يجب أن تعطى عنه جميع الاحتمالات التصريفية والصرفية التي يوجد عليها في اللغة العربية، وهذا ما يشكل القسم المتعلق بقاعدة معطيات المفردات البسيطة المعربة:

الشكل أ = جذر ب، صرف ج

حيث يراد بالشكل المدخل المعجمي (المفردة البسيطة) وبالجذر الصيغة التي يظهر بها في المعجم نفسه، وبالصرف مختلف الأوجه التصريفية والصرفية التي تلحق المدخل المعجمي البسيط وهو ما يسمى بالإعراب الصرفي في المعجم الآلي.

- **المعجم الإلكتروني للمفردات المركبة DEAMC**: إن أي برنامج معد للمعاجلة الآلية للمفردات في نص معين لابد وأن يتعرف على جميع مفردات النص/الهدف لكي تكون المعالجة شاملة وخالية من أي نقص وذلك بأخذ بعين الحسبان المفردات المركبة التي تشغل حيزا مهما في الذاكرة اللغوية مثل: علامة المرور، وأحد عشر،... وهي مداخل معجمية لا تقبل التجزيء ولا التقطيع، لأنها تشكل

مداخل مسكوكة، ومن ثمة تستدعي وصفا صرفيا وآليا من نوع خاص، أي الجمع بين نوعين من الوصف: العناية بوضعها المورفولوجي للعناصر التي يتألف من الاسم المركب من جهة، ومن جهة أخرى، وصف وضعها في التأليفية التي أنشأها منها الاسم المركب. وينبني هذا الصنف من المعاجم على أساس قاعدة معطيات المفردات المركبة، والتي تفرع بدورها إلى مفردات مركبة، تضم العادي منها والمسكوك، ومفردات مركبة معربة.

3– تجربة الدكتور أبو العزم: من الغني إلى الغني/ الشامل[8]

يعتبر الأستاذ عبد الغني أبو العزم عميد الدراسات المعجمية المعجماتية بالمغرب بامتياز، لما أنتجه من معاجم ودراسات قيمة أغنت الخزانة المغربية والعربية، وسدت ثغرة في باب البحث المعجمي تأليفا ونشرا. ويعد معجم الغني[9] أشهر معجم عربي محوسب، حظي بالسمعة والتداول ما لم يحظ به غيره من معاجم العربية المؤلفة حديثا، فهو يتميز بخصائص معجمية ومعجماتية قلما توجد في معجم آخر، فقد عب من معين التقنيات اللسانية عامة، وتقنيات الصناعة المعجماتية والآلية خاصة، حيث تجاوز الكثير من المعاجم الحديثة سواء على مستوى الشكل أو المضمون، مما يجعله بحق مصدرا لا غنى عنه في صناعة معاجم إلكترونية للغة العربية.

كانت مداخل الغني في البداية لا تتعدى 20.000 مدخل، لكن سرعان ما أصبحت 30.000 مدخل بفعل الإضافات والتعديلات التي أدخلها عليه المعجمي أبو العزم، مما أدى إلى استخلاص رصيد لغوي يضم طائفة موسعة من المفردات غير الموجودة أصلا في المعجم، ومعان جديدة، ومصطلحات أدبية وعلمية، ناهيك عن

(8) أي النسخة المعدلة والمزيدة من **الغني**.

(9) نسخة إلكترونية من إصدار شركة صخر.

عدد مهم من الاستشهادات القرآنية والنبوية، والأقوال المأثورة، والأمثال والحكم، والمتلازمات.

إن أي معجمي يود صناعة معجم عربي حديث (ورقي أو إلكتروني)، لا غنى له من الاعتماد على معجم الغني باعتباره ذخيرة لغوية مفتوحة المصدر، لها من المتطلبات المعجمية المعجماتية ما يسهل عمله، ويجعله يتخطى ثغرات المعاجم العربية القديمة والحديثة.

4- تجربة الدكتور نبيل علي

يعتبر نبيل علي أول مهندس لغوي عربي، قام بأبحاث رائدة في العالم العربي فيما يتصل بالمعالجة الحاسوبية للمنظومة اللغوية العربية، وقد أغنى الخزانة العربية بكتب قيمة في اللسانيات الحاسوبية وثقافة المعلومات، وفي مقدمتها كتابيه "اللغة العربية والحاسوب"، و"الثقافة العربية في عصر المعلومات"، بالإضافة إلى العديد من المقالات والبحوث المهمة التي قدمها في مؤتمرات وندوات علمية عربيا ودوليا.

إن ميكنة المعجم في نظر نبيل علي لابد أن تخضع لثنائية التحليل والتركيب[10]، شأنها شأن المكونات اللغوية الأخرى. والمعجم الآلي بالنسبة إليه يشمل المعطيات الصرفية والنحوية والدلالية. كما أنه يعتبر مصدرا أساسا لا غنى عنه في دراسة الإنتاجية الصرفية، وعلاقة الترابط للعناصر المعجمية المختلفة. وهكذا فقد ارتأى أن تنظم ميكنة المعجم في أربع مستويات مترابطة ومتكاملة في ما بينها:

- ❖ قائمة مفردات اللغة.
- ❖ مصفوفات علاقات نحوية وصرفية ودلالية.
- ❖ قاعدة بيانات.

([10]) نبيل علي، **اللغة العربية والحاسوب**، تعريب الكويت، 1988 .

❖ قاعدة معرفة متكاملة.

5 – تجربة الحاج صالح أو ما يسمى بمشروع الذخيرة اللغوية العربية

يعتبر الحاج صالح من الباحثين الرواد الذين أسهموا ببحوث ودراسات هامة، كان لها الأثر البالغ الأهمية في تأصيل المعالجة الآلية للغة العربية.

الذخيرة اللغوية العربية عبارة عن بنك آلي أو محوسب من النصوص القديمة والحديثة، يشتمل على التراث العربي والعالمي، وعلى الإنتاج العلمي العربي والعالمي.

و تهدف الذخيرة إلى تحقيق الأهداف التالية:

✓ أن تكون قاعدة معطيات نصية (أي مدونة لغوية) يحرر على أساسه عدد من المعاجم وخاصة "المعجم الجامع للغة العربية".

✓ أن تكون مصدرا موثوقا من المعلومات العلمية والتكنولوجية والأدبية واللغوية والاجتماعية والثقافية عامة، موجهة للباحثين والمهتمين على حد سواء.

✓ أن تستخدم كأداة تعليمية وتربوية في تعليم اللغة العربية خاصة، واللغات الأجنبية عامة.

أما فوائدها فتتحدد في النواحي الآتية:

– الناحية اللغوية والمعجمية، فإذا كانت أكثر المعاجم العربية المؤلفة اليوم تركز أساسا على المعاجم القديمة، فإن الذخيرة باعتبارها مجموعة نصوص، تهتم باللغة المستعملة من أقدم العصور إلى اليوم.

– الناحية التربوية والثقافية: تعتبر الذخيرة مصدرا موثوقا للمعلومات التي تستجيب لحاجيات وطلبات التلميذ والطالب الجامعي والباحث بصفة عامة. إنها تمكن من تعلم مهارات في اللغة كتعلم العربية أو الفرنسية أو الإنجليزية.

خـــاتمة

عالجت هذه الدراسة مفهوم المعجم الإلكتروني في اللغات الطبيعية، وعلاقتـــه بقاعدة المعطيات الآلية، وأنواعها، والتطبيقات الحاسوبية المنجزة في بعض الأقطـــار، مركزة على بعض التجارب الغربية والعربية، وأخص بالذكر أعمال الأستاذ الدكتـــور عبد الغني أبو العزم من خلال معجم الغني، وأعمال الدكتور الحناش محمد، وأعمـــال الدكتور نبيل علي، وأعمال الدكتور الحاج صالح، كما اقتصرت على تجربة غربيـــة فرنسية في مختبر المعالجة الآلية والتوثيق اللساني (LADL) باعتبارها تجربـــة رائـــدة في ميدان المعالجة الحاسوبية للغات الطبيعية، تمكنت من إنجاز معـــاجم إلكترونيـــة للغـــة الفرنسية.

المراجـــع

أولاً: بالعربية

- عبد الغني أبو العزم، **معجم الغني**، نسخة إلكترونية.
- الغامدي محمد منصور وآخرون، الترجمة عبر الشبكة العالمية: نظام مقترح للترجمـــة من وإلى العربية، مجلة جامعة الملك سعود، "م. اللغات والترجمة"، 2.
- مهديوي عمر وسلوى حمادة السيد، **المعالجة الدلالية الآلية للغة العربيـــة**، "مجلـــة العربية"، ع. 3، 2006.
- نبيل علي، **اللغة العربية والحاسوب**، تعريب الكويت، 1988.

ثانياً: بالأنجليزية والفرنسية

- Mahmoud Mostafa Mohamed, *The Development of A Computer-based Arabic Lexicon*, M.Sc. Thesis, Helwan University, Jan, 2004.
- Bernard Quemada, *Bases de données informatisées et dictionnaires* ; Lexique 2, Lexique 2, Presses Universitaires de LILLE, 1983.
- René Moreau et Isabelle Warnesson; *Ordinateur et lexicographie*, Lexique 2, Presses Universitaires de LILLE; 1983.