

لسانيات المدونة الحاسوبية وصناعة المعجم العربي

بقلم: الدكتور علي القاسمي

المدونة، في اللغة والاصطلاح:

" المدونة "، في اللغة، اسم مفعول مشتق من الفعل "دَوَّنَ يدوِّن تدويناً" بمعنى كتب. والفعل " دَوَّنَ " مشتق، بدوره، من كلمة فارسية مُعَرَّبَةٌ هي " ديوان " التي استعملها العرب لتدلّ على الدفتر الذي تُكتب فيه أسماء العمال والجند وأهل العطية، وكذلك على المكان الذي تُحفظ فيه هذه الدفاتر. و" دَوَّنَ الكُتُبَ والصحف " : جمعها ورثتها. ويُقال إنّ الخليفة عمر بن الخطاب (ت 23هـ/644م) أوّل من " دَوَّنَ الدواوين " في الدولة الإسلامية، أي أنشأها ونظّمها (1).

وتولّى أسد بن الفرات (ت 213هـ/828م)، قاضي القيروان وأحد تلاميذ الإمام مالك (ت 179هـ/795م) السابقين، جَمَعَ صُحُفٍ تَضَمَّ دروساً فقهية في كتاب أسماه " المدونة " يُنسب إلى الإمام مالك. ثم جاء قاضي القيروان عبد السلام بن سعيد التنوخي (ت 240 هـ/854م)، الملقّب بـ " سحنون " واضطلع بتوسيع هذا الكتاب، كما يبدو، ونشره باسم " المدونة الكبرى " التي هي عماد الدرس الفقهي عند المالكية (2).

وفي النصف الأول من القرن الميلاديّ العشرين وما بعده، استعمل العرب لفظَ " مدونة " لتدلّ على مجموعة أحكام قانونية أو فقهية، مثل " مدونة القانون المدني " و " مدونة الأحوال الشخصية " (3)

وعندما شاع استعمال كلمة (Corpus) في الدراسات اللسانية الحديثة باللغتين الإنجليزية والفرنسية في النصف الثاني من القرن الميلادي العشرين، ترجمها بعض اللغويين العرب بكلمة "مدونة"، وشاع استعمالها حتى أثبتها "المعجم الموحد لمصطلحات اللسانيات" على الوجه التالي:
"Corpus (Eng.), Corpus (Fr.)
ما يشكّل الرصيد اللغوي أو مجموع المعطيات اللغوية التي يُخضعها الباحث للتحليل والدرس." (4)
وكلمة (Corpus)، في اللغتين الإنجليزية والفرنسية ذات أصل لاتيني: (Corpus) ومعناه: جسد.

وتُجمع هذه الكلمة باللغة الإنجليزية على (Corpora) وعلى (Corpuses)، وفي اللغة الفرنسية يبقى جمعها على لفظ (Corpus)، ولها معنيان هما:
1) مجموعة نصوص، خاصةً إذا كانت مكتملة وقائمة بذاتها، كما في عبارة "مدونة الشعر الأنجلوسكسوني".
2) في علم اللغة وصناعة المعجم: كتلة من نصوص، مكتوبة أو منطوقة، تمثّل نماذج من اللغة؛ وتكون، عادةً، مخزّنة في قاعدة بيانات إلكترونية (5).

فالمُدونة هي مجموعة من النصوص تمثّل اللغة في عصر من عصورها أو في مجال موضوعي من مجالات استعمالها، أو في منطقة جغرافية معيّنة، أو في مستوى من مستوياتها، أو في جميع عصورها ومجالاتها ومناطقها ومستوياتها. والمدونة إما أن تُجمع يدوياً وتُقرأ ورقياً، وإما . كما هو الشائع حالياً . تُخزّن في الحاسوب وتُعالج وتُقرأ إلكترونياً. ونستعمل المدونة من أجل دراسة اللغة وتحليلها ومعرفتها من خلال نماذج منها وردت بصورة طبيعية، مستخدمين قدراتنا على الملاحظة التي هي أمّ المعرفة.

تاريخ استخدام المدونات في البحث اللغوي:

إنّ استخدام اللغويين لمجموعات من النصوص اللغوية في دراساتهم وأبحاثهم وتصنيف معاجمهم ليس بالأمر الجديد. إذ يُبشّرنا تاريخ اللسانيات أنّ اللغويين العرب، في القرن الثاني الهجري

/ القرن الثامن الميلاديّ، كانوا يرحلون من البصرة والكوفة في اتجاه بوادي جزيرة العرب، يجوبونها، وهم محمّلون بدفاترهم ومحابرهم، ويشافهون الأعراب ويستنطقونهم، ويسجّلون كلامهم وما ينطقون به من جُمَل وتعبيرات ومفردات، ويعودون بعدها إلى البصرة والكوفة ليدرّسوا ما جمعوه من نماذج لغويّة، وتحليلها، واستنباط القواعد النحويّة منها، أو استخدامها في تصنيف المعجمات بصورة علميّة (6).

ومن هؤلاء اللغويين الخليل بن أحمد الفراهيدي (ت 170هـ/786م) الذي سُئِل ذات مرّة : " من أين أخذتَ عِلْمَكَ هذا؟ فقال: من بوادي الحجاز ونجد وتهامة" (7). ثم صنّف لنا الخليلُ أوّل معجم متكامل باللغة العربيّة هو " كتاب العين " .

ومن هؤلاء اللغويين الذي استخدموا النصوص التي جمعوها من البادية في درّسهم النحوي علي بن حمزة المعروف بالكسائي (ت 180هـ/796م)، أحد الأئمة القراء من أهل الكوفة، الذي خرج ذات مرّة إلى البادية ورجع " وقد أنفذ خمس عشرة قنينة حبر في الكتابة عن العرب سوى ما حَفِظ. " (8)

ومن هؤلاء اللغويين النضر بن شُمَيْل (ت 204هـ/820م)، الذي نشأ في البصرة وأخذ عن الخليل بن أحمد، ورحل إلى البادية لمشاهدة الأعراب وتدوين كلامهم، ثم عاد إلى البصرة، وعاود الترحال إلى البادية، وهكذا دواليك حتّى بلغ ما أمضاه في البادية قرابة أربعين عاماً (9).

واشتهر عبد الملك بن قريب الأصمعي (216 هـ/831م) اللغوي البصري، برحلته العديدة إلى البادية: المرید، فالقصيم، فبادية بني أسد، فبلاد بني عامر، وهو " يترك العراق رويداً رويداً متّجهاً إلى قلب البادية يسمع، ويعي، ويرى، ويدوّن كلّ ما يعينه على استيعاب البادية قولاً، وحياتاً، وعادةً" (10) ليصنّف لنا عدداً من المعجمات الموضوعيّة القيّمة تتناول أهل البادية، ونباتاتها، ومياهها، وحيواناتها، مثل كتب: الصّفات، خلق الإنسان، خلق الفرس، الشاة، الدلو، النحلة، الخيل، الأخبية والبيوت، الرّحل، النبات والشجر، مياه العرب، الأنواء، الميسر والقّداح، الإبل، الوحوش،

الأمثال، السُّلَّاح، جزيرة العرب، السَّرَج واللِّجَام، أسماء الخَمَر. وتنبني معاجمه الموضوعية هذه على ما دوّنه من نماذج لغوية حيّة.

واستمرّ هذا التقليد اللغويّ في تصنيف المعاجم بالاستناد إلى مدوّنات منطوقة أو مكتوبة حتّى القرنين الرابع والخامس الهجريّين؛ فأبو منصور محمد بن أحمد الأزهري (ت 370هـ/980م)، الذي " رحل وطاف في أرض العرب في طلب اللغة " كما يقول ابن خلكان (11)، يذكر لنا في مقدّمة معجمه ذي المجلدات العشرة " تهذيب اللغة" أسماء القبائل التي قصدتها ودوّن كلامها، ويعدّ ذلك من خصائص معجمه المميّزة، فيقول:

" ... منها (أي من خصائص معجمه) تقييدي نكتاً حفظتها ووعيتها من أفواه العرب الذين شاهدتهم وطالت أيام مقامي معهم، إذ كان ما أثبتته كثير من أئمة أهل اللغة في الكتب التي ألفوها والنوادر التي جمعوها لا ينوب مناب المشاهدة، ولا يقوم مقام الدربة والعادة."

وهذا ما فعله أبو نصر إسماعيل بن حمّاد الجوهري (ت 393 هـ/1003م) قبل تصنيف معجمه الشهير بـ "الصّحاح": "تاج اللغة وصحاح العربية"، الذي قال في مقدمته:
" ...فإني قد أودعتُ في هذا الكتاب ما صحّ عندي من هذه اللغة، ...بعد تحصيلها بالعراق رواية، وإتقانها دراية، ومشافهتي بها العرب العاربة في ديارهم بالبادية..." (12)

وعندما يروم محمود بن عمر الزمخشري (ت 538 هـ / 1144م) تصنيف معجمه " أساس البلاغة " الذي يُعنى فيه بالاستعمالات المجازية للألفاظ، فإنّه يجمع مدوّنة كبيرة من النصوص اللغوية، المنطوقة والمكتوبة، ليستخلص مادته منها، فيقول عن الإعداد لتصنيف هذا المعجم، في المقدمة:

"...فُلِيَتْ له العربية وما فَصَح من لغاتها، ومُلِح من بلاغاتها، وما سُمِعَ من الأعراب في بواديتها، ومن خطباء الحِلَل في نواديتها، ومن قراضية نجدٍ في أكلائها ومراتعها، ومن سماسرة تهامة في أسواقها ومجامعها؛ وما تراجرت به السُّقاة على أفواه قُلُبها، وتساجعت به الرُّعاة على شفاه

عُلبها؛ ما تقارضته شعراء قيسٍ وتميمٍ في ساعاتِ المماتنة، وما تزاملتْ به سفراءُ ثقيفٍ وهذيلٍ في أيامِ المفاتنة؛ وما طُولِعَ في بطونِ الكُتبِ ومُتُونِ الدفاترِ من روائعِ ألفاظِ مُفَتَّنَةٍ، وجوامعِ كَلِمٍ في أحشائها مُجَنَّنَةٍ. " (13)

ويبدو أنّ هذا التقليد الصحيح في إرساء الدراسات اللغوية وتصنيف المعاجم على تحليل نصوصٍ تمثّل اللغة في سياقاتها الطبيعية، قد تضاءل واضمحَلّ في عصور الانحطاط في البلاد العربية. ولكننا نجد أن هذا التقليد يتزعزع في أوروبا خلال العصور الوسطى حين يأخذ اللغويون الأوربيون في وضع قوائم المفردات الواردة في نصٍّ من النصوص مع سياقاتها المختلفة، أي الكلمة، ثم الجمل التي ترد فيها تلك الكلمة للوقوف على معناها واستعمالها، وهذا ما نسمّيه اليوم بـ "المطابقة" Condordancing. كما استعمل اللغويون الأوربيون النصوص في تعليم اللغة، والنحو، وعلم الدلالة، وعلم اللغة المقارن، وغيرها من المجالات اللسانية. وعلى الرغم من أنّ مصطلح "لسانيات المدوّنة" لم يكن مُستعملاً آنذاك، فإنّ تقنيات البحث التي كان يجريها اللغويون على مجموعات نصوصهم، هي مشابهة للبحث اللغوي المعاصر القائم على المدوّنة، مع فارق أساسي واحد هو استخدام الحاسوب في تخزين النصوص ومعالجتها وتحليلها في البحث اللغوي المعاصر (14).

وفي نهاية القرن الميلاديّ التاسع عشر، جمع كادنج J.Kading مدوّنة من نصوص لغوية ألمانية تتألف من مليون كلمة ليقارن توزيع الحروف الألمانية وشيوعها (15)

وفي أوائل القرن الميلاديّ العشرين درس الإنثروبولوجي اللغويّ الأمريكيّ بواس F. Boas بعض لغات الهنود الحمر في الولايات المتّحدة الأمريكيّة، عن طريق الرحيل إلى محمياتهم وتجمعاتهم، ومشافهتهم، وتدوين كلامهم، ثم تحليله ومَعجمته وتقعيده، أي تماماً كما فعل الأصمعي قبله بأكثر من ألف عام. ونشر بواس أبحاثه في كتاب (16).

وفي عام 1921م جمع المرابي وعالم النفس الأمريكي أدورد ثورندايك E. Thorndike (1874 . 1949) مدوّنة مؤلّفة من 4.5 مليون كلمة في نصوص مقتبسة من 41 مصدراً مختلفاً من

أجل تكوين قائمة بالكلمات الشائعة. وفي العشرينيات من القرن العشرين أجرى بالمر وهورني Palmer et Hornby دراسات على نظم (نحو) اللغة الإنجليزية المنطوقة من خلال نصوص كثيرة جمعها، وذلك للاستفادة منها في تعليم اللغة الإنجليزية لغير الناطقين بها لناطقين بها. وقد كانت تلك الدراسات أساساً لتألف المعجم الشهير Learner's Dictionary of Current English الذي شارك في تصنيفه هورني، وغاتني، وويكفيلد Hornby, Gatenby, Wakefield ونشرته مطبعة جامعة أكسفورد عام 1952. وتطور هذا المعجم في طبعاته اللاحقة ليصبح Advanced Learner's Dictionary of Current English الذي نشر سنة 1963م. كما أن المستشرق الألماني هانز فير Hans Wehr بني معجمه العربي . الألماني Arabisches Worterbuch fur die Schriftsprache der Gegenwart (1952) على قراءات واسعة في الصحف والكتب العربية المعاصرة له واتبع المستعرب الأمريكي كوان J.Milton Cowan في توسيع هذا المعجم عند إصدار الطبعة العربية للإنجليزية منه عام 1961 بعنوان: A Dictionary of Modern Written Arabic.

واستخدم اللغويون البنيويون الأمريكيون في النصف الأول من القرن العشرين، وعلى رأسهم بلومفيلد Bloomfield وأستاذنا آر تشبولد هيل A.A.Hill طريقة بواس في استخراج القواعد اللغوية، وخاصة الصوتية، من نصوص منطوقة ومكتوبة يجمعونها ويحللونها. وفي تلك الأثناء، استخدم اللغوي الأمريكي تشارلز فريز C. Fries مدونة نصوص مجموعة يدوياً للوقوف على شيوخ مفردات اللغة الإنجليزية للاستفادة منها في تدريس الإنجليزية للناطقين باللغات الأخرى (17)

واستخدم عالم النفس السويسري جان بياجيه Jean Piaget (1896-1980م)، مدونات من كلام الأطفال لتحليلها ومعرفة تطور لغة الطفل. كما كان عدد من الآباء الأمريكيين يحتفظون بمدونات (مذكرات) تشتمل على الكلمات والعبارات التي يتفوه بها أطفالهم لمعرفة تطور لغة الطفل.

ويطلق اللغويون الأمريكيون اسم "لسانيات المدونة المبكرة" Early Corpus Linguistics وبالفرنسية "Corpus primitive" على البحوث اللسانية الأولى التي استخدمت المدونات اليدوية، مثل بحوث بواس المذكورة. أما مصطلح لسانيات المدونة فيحتفظون به لنعث دراسة المدونات الإلكترونية.

المدونة الحاسوبية:

المدونة الحاسوبية هي عبارة عن مجموعة مهيكلة من النصوص اللغوية الكاملة المكتوبة (أو المنطوقة) التي تُقرأ إلكترونياً. وكثيراً ما تكون هذه النصوص مصحوبة بالشارات الشارحة لمكوناتها اللغوية. وتمدنا المدونة بالأدلة والأمثلة على كيفية استعمال اللغة في سياقات طبيعية، بحيث يستطيع اللغوي إجراء بحوثه عليها، ويستطيع المعجمي أن يختار مداخل معجمه ويكتب موادّه بصورة دقيقة وعلمية. وما لم تستجب المدونة إلى احتياجات محدّدة وتجب على أسئلة معينة، فإنها ستسمي مجرد حقبة قديمة من الكلمات التي لا قيمة لها.

كانت أول مدونة حديثة مقروءة إلكترونياً هي مدونة جامعة براون الأمريكية **Brown University**. وقد أنشئت هذه المدونة سنة 1961م، وتتألف من مليون كلمة من نصوص اللغة الإنجليزية الأمريكية، التي استُقيت من 15 نوعاً من النصوص مثل: الصحافة (تقارير، افتتاحيات، مقابلات، إلخ.)، والهوايات، والأديان، والنصوص العلمية، والروايات، إلخ.).

وصارت مدونة جامعة براون نموذجاً لعدد من المدونات التي أنشئت بعد ذلك، مثل مدونة لانكستر. أوسلو. برغن للغة الإنجليزية البريطانية، ومدونة اللغة الإنجليزية الهندية. وكلّ منهما يتألف من مليون كلمة اختيرت من 15 نوعاً من النصوص، تماماً مثل مدونة جامعة براون.

وعلى الرغم من أنّ هذه المدونات الصغيرة الحجم مفيدةٌ لكثير من الباحثين، فإنّ بعض اللغويين الآخرين، كالمعجميين، شعروا بالحاجة إلى مدونات أكبر تستطيع أن تمدّهم بالكلمات الحوشية، والتراكيب النادرة، والألفاظ والمعاني والاستعمالات الجديدة التي تدخل اللغة باستمرار. ومن هنا بدأت مدونات خاصة بالصناعة المعجمية مثل مدونة معجم التراث الأمريكي، ومدونة معاجم كوليز كويلد، اللتين سنتطرق إليهما بعد قليل.

وبعد أن تُجمع نصوص المدوّنة الحاسوبية، يركّب عليها (مُحرّك بحث). ومُحرّك البحث عبارة عن برنامج حاسوبيّ صُمم لمساعدة مُستعمله على البحث في مدوّنة حاسوبية.

ويميّز اللغويّ الفرنسيّ فرونسوا راستيه F. Rastier بين أربعة أنواع من مجموعات النصوص المقروءة آلياً، هي:

- 1) الأرشيف (المحفوظات)، التي يضمّ مجموعة من الوثائق التي يمكن الاطلاع عليها. ولكنّ الأرشيف لا يشكّل مدوّنة، لأنّه يعوزه الشرح (الشارات التعريفية بمكوّنات النصّ).
- 2) مدوّنة المراجع، التي تتألّف من مجموعة النصوص التي سيبحث فيها اللغويّ ويستخرج منها مدوّنة البحث.
- 3) مدونة البحث، وهي النصوص المختارة من مدوّنة المراجع في ضوء احتياجات الباحث.
- 4) مدوّنة العمل الفرعية، التي يتطلّبها العمل، وتختلف من بحث لآخر، ومن مرحلة في البحث إلى أخرى. وقد تضمّ فقط الفقرات أو السياقات ذات العلاقة بالبحث.

فمثلاً، في التحليل الموضوعيّ للمعطيات أو البيانات النصّية المتعلقة بالمشاعر والأحاسيس، فإنّ الأرشيف هو بنك النصوص الفرنسيّ Frantext؛ ومدوّنة المراجع التي تُستقى من الأرشيف مكوّنة من 350 رواية نُشرت بين سنتي 1830 و1970م؛ ومدوّنة البحث تتكوّن من تلك المقاطع التي تشتمل على أسماء المشاعر والأحاسيس؛ أما مدوّنة العمل الفرعية (أو المدونة المختارة)، فهي مجموعة السياقات المتعلقة بهذا الإحساس أو ذاك (18)

كيفية البحث في المدوّنة الحاسوبية:

هنالك نوعان من استخدام المدوّنة في البحث اللغوي:

- أ. تحقيق مبنيّ على المدوّنة Corpus-Based Investigation، فالباحث يبحث في المدوّنة للعثور على الأدلّة التي تدعم وصفه اللغويّ أو لتأنيث نظريّة قائمة بالأمثلة.

ب . تحقيق توجهه المدونة **Corpus-Driven Investigation** ، فالباحث هنا لا يملك تصوراً مسبقاً للوصف اللغوي، ولا إيماناً بنظرية معينة. وإنما يستخلص نتائجه من ملاحظته على الاستعمال الفعلي كما تقدّمه المدونة (19)

ومهما كان نوع الاستخدام، فإنّ طريقة البحث في المدونة تتبع الخطوات التالية:

أ . يختار الباحث المدونة التي يريد البحث فيها.

ب . يدخل الباحث الكلمة أو التعبير الذي يريد البحث عنه في المستطيل المُخصّص في صفحة البحث.

ج . يستعمل الباحث الفأرة للضغط على كلمة " ابحث " الموجودة بجانب المستطيل.

د . يقوم (مُحرّك البحث) بفحص المدونة للعثور على الكلمة أو التعبير الموجود في المستطيل بصفحة البحث، ثم يعرض نتائج البحث على شاشة الحاسوب. ولا يستغرق البحث أكثر من ثوانٍ معدودة.

لسانيات المدونة:

لسانيات المدونة **Corpus linguistics/ Linguistique de Corpus** هو العلم الذي يدرس الظاهرة اللغوية من خلال مدونة أو مجموعة كبيرة من النصوص التي يمكن قراءتها آلياً. فلسانيات المدونة ترمي إلى دراسة اللغة كما هي ماثلة في سياقات تزوّدنا بها المدونة أي في نصوص مقتبسة من العالم الحقيقي.

وعلى الرغم من أن استعمال الأمثلة الحقيقية المستقاة من النصوص، في البحث اللغوي ليس بالأمر الجديد، فإنّ لسانيات المدونة قد تطوّرت كثيراً في العقود الأخيرة من السنوات، بفضل الإمكانيات الهائلة التي تقدّمها المعالجة الآلية للغة الطبيعية. فوجود الحاسوب وتوفّر النصوص التي تُقرأ آلياً جعلاً من الممكن الحصول على البيانات بسرعة ويسر وبصورة ملائمة للتحليل. ولهذا فإنّ مصطلح " لسانيات المدونة " يرتبط اليوم ارتباطاً وثيقاً باستخدام الحاسوب، حتّى أنّ كثيراً من

اللغويين يعرفون هذا العلم بأنه " استعمال مجموعة من النصوص المحوسبة التي يمكن قراءتها إلكترونياً من أجل البحث اللغوي".

ولكنّ لسانيّات المدوّنة ليس مجرد الحصول على البيانات والأمثلة الحيّة بواسطة الحاسوب، وإنّما هي، في حقيقتها، دراسة وتحليل تلك المعلومات والنصوص بعد الحصول عليها من المدوّنة. فمهمّة المختصّ بلسانيّات المدوّنة هي ليس الحصول على البيانات وإنّما تحليل تلك البيانات. وما الحاسوب إلا أداة مفيدة، بل لا غنى عنها أحياناً، في عمليّة التحليل.

وتتداخل لسانيّات المدوّنة مع اللسانيّات الحاسوبية في بعض ميادين البحث، خاصة عندما تستفيد لسانيّات المدوّنة من المعالجة الآليّة للغة. فالمعالجة الآليّة للغة المكتوبة أو المنطوقة، التي هي جزء من اللسانيّات الحاسوبية، تساعدنا في البحث في المادة اللغوية وإعداد برنامج محرك البحث. فمثلاً، قد يستعمل الباحث المعالجة الآليّة في تعيين صيغ معينة من الأفعال في المدوّنة واستخراجها في سياقاتها لغرض الدرس والتحليل.

حققت لسانيّات المدوّنة نجاحاً باهراً في الولايات المتّحدة الأمريكيّة عام 1967م، عندما نشر هنري كوشيرا ونلسن فرانسز دراستهما عن التحليل الحاسوبيّ للغة الإنجليزيّة الأمريكيّة المعاصرة (20). وقد بنيا عملهما على مدوّنة براون Brown Corpus التي تضمّ نصوصاً مختارة بعناية ومقتبسة من مصادر متنوّعة ويبلغ عدد كلماتها المليون. وتضمنت دراسة كوشيرا وفرانسز جوانب متعدّدة من التحليل اللسانيّ تفيد الباحثين في اللسانيّات، وتعليم اللغة، وعلم النفس، والإحصاء، وعلم الاجتماع.

وبعد مدة قصيرة طلب الناشر هوتن . مفلن Houghton-Mifflin في مدينة بوسطن من كوشيرا وفرانسز أن يوفّرا له قاعدة شواهد تتكوّن من مليون كلمة على أن يتألّف كلُّ شاهد من ثلاثة أسطر، ليستفيد منها المعجميون الذي كانوا سيصنّفون معجماً أمريكياً ينشره الناشر المذكور، هو معجم التراث الأمريكيّ American Heritage Dictionary. وهكذا أصبح هذا المعجم أوّل معجم

يستفيد من لسانيات المدونة. وقد جمع هذا المعجم بين المقاربتين: المقاربة المعيارية (كيف ينبغي أن تُستعمل اللغة؟) والمقاربة الوصفية (كيف تُستعمل اللغة فعلاً؟).

وسار ناشرون آخرون على نهج الناشر هوتن . مفلن، فقام ناشر بريطاني يتولى نشر معاجم كولنز كوبلد Collins' COBUILD المخصّصة لتعليمي اللغة الإنجليزية، بالاستفادة من بنك الإنجليزية Bank of English في إنشاء مدونة CBUILD عام 1980 (21). وقد بلغت طاقة هذه المدونة 20 مليون كلمة عام 1987م عندما استخدمها مصنّفو معاجم كولنز كوبلد للغة الإنجليزية Collins' COBUILD English Language Dictionary . ثم بلغت طاقتها عام 1996 م حوالي 320 مليون كلمة من النصوص المكتوبة والمنطوقة.

وفي الوطن العربي، أجرى الدكتور علي حلمي موسى، في مطلع السبعينيات من القرن الميلادي العشرين، دراسات حاسوبية إحصائية على جذور معاجم " لسان العرب " و " الصحاح " و " تاج العروس " (22). ولكن هذه الخطوة بقيت يتيمة ولم تُستثمر في صناعة المعجم أو تعليم اللغة العربية.

لقد تطورت الدراسات في ميدان لسانيات المدونة تطوراً هائلاً في النصف الثاني من القرن الميلادي العشرين في الولايات المتحدة الأمريكية، بفضل ثورة تكنولوجيا المعلومات، والنمو الهائل في اللسانيات الحاسوبية والهندسة اللغوية والترجمة الآلية وغيرها من العلوم اللغوية التي تستعين بالحاسوب. ويقدم لنا جونسون إحصائية بنمو أعداد الدراسات المتعلقة بلسانيات المدونة، نجملها في الجدول التالي(23):

عدد الدراسات	المدة
10	إلى 1965
20	1966 . 1970
30	1971 . 1975

80	1976 . 1980
160	1981 . 1985
320	1986 . 1990

(جدول يبين تطوّر عدد الدراسات في مجال لسانيات المدوّنة في النصف الثاني من القرن العشرين)

تشومسكي ولسانيات المدوّنة:

معروف أنّ نعوم تشومسكي N. Chomsky في أطروحته المتعلقة بالنحو التوليديّ . التحويليّ، انتقل باللسانيات من التجريبيّة المحسوسة إلى العقلانيّة. ففرق بين القدرة اللغويّة (Competence) الكامنة في الفرد وفي الجماعة الناطقة باللغة K وبين التعبير أو الإنجاز (Performance). فالقدرة هي مجموعة القواعد التي تشكّل معرفة الناطقين اللغويّة. وهذه القدرة قدرتان: الأولى، قدرة كليّة مكوّنة من القواعد الفطريّة التي تحكم اللغات عامّة؛ والثانية، قدرة خاصّة مكوّنة من القواعد المتعلقة بلغة معيّنة. أما التعبير أو الإنجاز فهو التحقّق الفعليّ من خلال الكلام للقدرة اللغويّة المستنبطة عند المتكلمين.

ويرى تشومسكي أنّ مهمّة اللغويّ هي استنباط نموذج اللغة من القدرة اللغويّة لدى المتكلمين، وليس من التعبير أو الإنجاز الذي هو بمثابة مرآة غير مصقولة للقدرة اللغويّة. فهذه القدرة اللغويّة هي التي تمثّل معرفتنا الذاتية باللغة، وينبغي استخلاص القواعد منها.

ولهذا كله، فإن تشومسكي لم يحدّد إنشاء مدوّنات من أجل تحليل اللغة كما تُستعمل في الحياة الحقيقيّة؛ لأنّ المدوّنة، في نظره، تشتمل على أخطاء كثيرة نتيجة التعبير أو الإنجاز. وكان يرى أن خير مقارنة هي تحليل جيّد لعدد صغير من الأمثلة التي تمثّل القدرة اللغويّة في شروط مختبريّة نتحكم فيها. فجوهر اللغة يكمن في القدرة لا التعبير الذي يتعرض للخطأ والتشويه. وهو بذلك أقرب إلى مدرسة الكوفة منه إلى مدرسة البصرة؛ لأنّ المدرسة الأولى كانت تعتدّ بأمثلة صحيحة

معدودة ثم تتوسع في القياس عليها لاشتقاق القواعد، على حين أنّ مدرسة البصرة كانت لا تشتقّ قواعدها اللغوية إلا من كمية وفيرة كافية من التعبيرات الحقيقية التي وردت في سياقات طبيعية.

ولكن أنصار لسانيات المدونة لم يأبهوا بنظرية تشومسكي تلك وتركوا جانباً ثنائية القدرة اللغوية مقابل التعبير أو الإنجاز اللغوي، كما تركوا جانباً ثنائية البنية العميقة للغة مقابل البنية السطحية؛ وقالوا إنّ التحليل اللغويّ الوحيد الذي يمكن الركون إليه هو ذلك التحليل الذي يقوم على أمثلة كثيرة مستقاة من الحياة، ووردت في سياقات طبيعية، وبأقل درجة ممكنة من التدخلات التجريبية (24)، وكأنهم بذلك الرأي يشايعون مدرسة البصرة النحوية.

واليوم تقوم كثير من الجامعات حول العالم باستخدام المدونات في التعليم والبحث أو ضمن برامج جامعية في لسانيات المدونة حيث يُدرّب الطلاب على طرائق وأدوات إنشاء المدونات المختلفة، وتجميع نصوصها، وتحليل بياناتها المكتوبة والمنطوقة. وهناك جامعات تنظّم دورات تدريبيّة صيفيّة قصيرة تتعلق بنوع محدد من المدونات مثل المدرسة الصيفية التي أقامتها جامعة نانت Nantes الفرنسية صيف 2006م باسم المدرسة اللسانية للمدونات المنطوقة L'Ecole . Corpus Oraux Linguistique de

خصائص المدونة الحاسوبية:

- تمتاز المدونة الحاسوبية على المدونة اليدوية بثلاث ميزات:
- 1) السهولة، التي تتمثل بسهولة الوصول إلى الوحدات والتراكيب اللغوية المطلوبة.
 - 2) السرعة، إذ يمكن البحث في نصوص تتألف من ملايين الكلمات وتحليلها بسرعة أكبر من استعمال العين المجردة.
 - 3) الدقة، فمعالجة النصوص أكثر دقة من معالجتها بالعين المجردة، فإذا كنا نبحث في مدونة كبيرة جداً عن كلمة واحدة، مثلاً، يستطيع الحاسوب أن يعثر على جميع السياقات التي ترد فيها تلك الكلمة دون أن يفوته سياق واحد.

مجالات استخدام المدونة وأهدافها:

تُستخدم المدونة في بحوث تنتمي إلى مجالات لغوية متعدّدة أهمّها:

1 علم اللغة:

تُستخدم المدونة لدراسة القدرة اللغوية والتعبير (الإنجاز) اللغويّ، كما تكشف عنهما البيانات التي جُمعت من سياقات طبيعيّة. والبحث في هذا الميدان يتطلّب استعمال مدونة ذات نصوص مشروحة، أي مؤشّرة ومُعَلِّم على مكوّناتها من مفردات وتراكيب.

2 علم اللغة التاريخي:

يُعنى علم اللغة التاريخي بدراسة اللغة باعتماد البعد التاريخي في تفسير التحوّلات والتغيّرات الطارئة عليها. وهذا النوع من الدرس يحتاج إلى نصوص تنتمي إلى عصور اللغة المتعاقبة. ويتّصل بهذا، استخدام المدونة لتصنيف معجم تاريخي.

3 علم اللغة الحاسوبي:

يرمي علم اللغة الحاسوبي إلى وضع نظام آلي لمعالجة اللغة الطبيعية. ويحتاج إلى مدونة لتطبيق أو اختبار النظام الذي يضعه على نماذج ممثّلة لجميع أنواع النصوص التي من المتوقع أن يعالجها ذلك النظام، وكذلك لبناء معجم مختصّ في مجال مُعيّن.

4 علم اللغة التطبيقي:

كثيراً ما تُستخدم المدونة في بحوث علم اللغة التطبيقي، مثل البحوث المتعلّقة باكتساب اللغة الأم أو اللغة الثانية، والبحاث المتعلّقة بتعليم اللغة وتعلّمها، إذ نستعمل المدونة لإحصاء شيوع المفردات والتراكيب. وتستخدم المدونة كذلك في البحوث في مجال استعمال اللغة لإغراض خاصة، وفي هذه الحالة الأخيرة تكون المدونات المتخصصة مثل مدونة لغة الصحافة، أو مدونة اللغة الطبية، ذات فائدة عظيمة للوقوف على المصطلحات والتراكيب المستعملة في ذلك التخصص (25).

وإذا نظرنا إلى مدوِّنة عربيّة، مثل مدوِّنة عدنان عيدان المسماة بـ " بنك اللغة العربية " التي تتألّف من نصوص يبلغ تعدادها مئات الملايين من الكلمات، نجد أنّ صاحبها يسطّر الأهداف المباشرة، كما يلي:

- 1) تيسير الإملاء العربيّ، وكشف الأخطاء الإملائيّة الشائعة.
- 2) الوقوف على الذخيرة اللغويّة العربيّة، وتحديد المفردات الشائعة.
- 3) تطوير تعليم اللغة العربيّة في المدارس من خلال معرفة مفردات اللغة الفصيحة وتراكيبها الأكثر استعمالاً.
- 4) تصنيف المعاجم العامّة والمتخصّصة والتاريخيّة بالاعتماد على هذه المدوِّنة الشاملة.
- 5) القيام بدراسات تتناول مفردات اللغة وبنياتها ونحوها.
- 6) المعالجة الآليّة للغة العربيّة، تحليلاً وتوليداً.
- 7) الترجمة الآليّة القائمة لا على معرفة مفردات اللغة وتعابيرها الاصطلاحية ومقابلاتها في اللغة الأخرى فحسب، وإنما على نحوها المقارن كذلك.
- 8) تزويد الباحثين بالمصادر اللازمة لدراساتهم العلميّة المختلفة (26).

وبصورة عامّة، يمكن القول إن المدوِّنة التي يتمّ تحديثها بانتظام وباستمرار، تساعدنا في الوقوف على التغيّر الذي يطرأ على اللغة عبر الزمن: كلمات تصبح أكثر شيوعاً أو تختفي من الاستعمال، ظواهر لغوية تنتشر من منطقة إلى أخرى، معانٍ جديدة تظهر إلى جانب المعاني القائمة للكلمات أو تحلّ محلّها، استعمالات جديدة للمفردات القائمة، كلمات أجنبيّة تدخل إلى اللغة وتصبح جزءاً منها، إلخ.

التصنيف النوعي للمدوّنات:

التصنيف هو تقسيم الأشياء إلى أنواع أو أصناف أو ضروب طبقاً لخصائص مشتركة بينها، أو طبقاً للعلاقات القائمة بينها كعلاقة الجنس بالنوع أو علاقة الكلّ بالجزء. ويفيد التصنيفُ الباحث في اختيار المدوِّنة المناسبة التي تستجيب إلى احتياجاته.

ويمكن أن تُصنّف المدوّنات طبقاً لمعايير مختلفة، مثل:

أ) اللغة:

تصنّف المدوّنات من حيث لغة النصوص إلى:

. مدوّنات أحادية اللغة مقابل مدوّنات متعدّدة اللغات، وقد تضمّ المدوّنات من النوع الثاني النصوص ذاتها في عدد من اللغات مثل المدونة المسماة " اتحاد البيانات اللغوية " The Linguistic Data Consortium التي تشتمل على مجموعة من المحادثات الهاتفية في عدد من اللغات. وتفيدنا المدوّنات من الصنف الأخير، مثلاً، في بحوث علم اللغة المقارن، وفي الترجمة، والترجمة الآلية، وخاصة الترجمة الآلية الإحصائية (27)

. مدوّنات أصلية اللغة مقابل مدوّنات مترجمة، وتفيدنا المدوّنات من الصنف الثاني في تعليم تقنيات الترجمة، مثلاً.

. مدوّنات الناطقين باللغة مقابل مدوّنات متعلّمي اللغة، وتفيدنا المدوّنات من الصنف الأخير في تحليل الأخطاء الشائعة بين المتعلّمين، واتخاذ الاحتياطات اللازمة لمكافحتها في المناهج وطرق التعليم.

. مدوّنات النصوص المكتوبة مقابل مدوّنات الكلام المنطوق.

ب) الوسيلة:

تُصنّف المدوّنات من حيث الوسيلة المستخدمة في تجميعها واستعمالها إلى مدوّنات مطبوعة يدوية، أو إلكترونية تُخزّن وتُعالج وتُسترجع بالحاسوب، أو مدوّنات منطوقة مسجلة رقمياً أو بالفيديو، أو مدوّنات مختلطة الوسيلة.

ت) التخصص:

تُصنّف المدوّنات من حيث نوعيّة النصوص التي تشتمل عليها إلى مدوّنات متخصصة تنتمي نصوصها إلى مجال علميّ معيّن، مثل الطبّ أو الهندسة، ومن أمثلة هذه المدوّنات المتخصصة "مدوّنة الرقابة الجوية" Air Traffic Control Corpus في ولاية أوريغون في

الولايات المتحدة الأمريكية التي تُستخدم للتعرف على الكلام المستعمل في وسائل النقل الجوية، وقد جُمعت نصوصها المنطوقة من أفراد طُلب إليهم التحدّث في الهاتف وذكر أسمائهم أو نطق عبارات وأرقام وحروف؛ مقابل مدوّنات عامّة تنتمي نصوصها إلى جميع المجالات العلمية والأدبية والفنية.

ث) الفترة التاريخية:

تُصنّف المدوّنات من حيث الفترة التاريخية التي تنتمي إليها النصوص إلى مدوّنات تزامنية، تختص بفترة زمنية واحدة، مثل "مدونة لامبتر" *TheLampeter Corpus* التي استُمدت نصوصها من اللغة الإنجليزية في الفترة الواقعة بين سنتي 1640م و1740م، ومثل "مدوّنة أكسفورد الإنجليزية" *The Oxford English Corpus* التي تنتمي نصوصها إلى الفترة من سنة 2000م إلى سنة 2006؛ مقابل مدوّنات تاريخية تنتمي نصوصها إلى جميع عصور اللغة المختلفة، مثل "مدونة هلسنكي" للغة الإنجليزية *Helsinki Corpus* التي تنتمي نصوصها إلى فترات اللغة الإنجليزية المتعاقبة: الإنجليزية القديمة، والإنجليزية الوسيطة، والإنجليزية الحديثة.

ج) التوثيق:

تُصنّف المدوّنات من حيث توثيق نصوصها إلى مدونات موثّقة، يُنسب كلّ نصّ فيها إلى مؤلّفه، وعنوان الكتاب، ومكان النشر وتاريخه؛ مقابل مدوّنات غير موثّقة. والمدوّنات التي تُستخدم في تصنيف المعاجم التاريخية يجب أن تكون موثّقة.

ح) الشرح:

المقصود بالشرح *annotation* هو وضع شارات أو علامات على المفردات والتعبيرات والتراكيب الواردة في نصوص المدوّنة لمعرفة قسم الكلام الذي تنتمي إليه المفردات، وأنواع التراكيب النحوية، وغير ذلك من المعلومات النحوية والبلاغية. ومن هذا المنظور تُصنّف المدوّنات إلى مدوّنات مشروحة مقابل مدوّنات غير مشروحة.

(خ) طريقة التصميم:

تُصنّف المدوّنات من حيث طريقة تجميعها وتصميمها ونسب نوعية النصوص المكوّنة لها، إلى مدوّنات متوازنة، ومدوّنات هرميّة، ومدوّنات اعتباطية. وسنتناول هذه النقطة بالتفصيل لاحقاً.

(د) الحجم:

تُصنّف المدوّنات من حيث حجمها الذي يُقاس بعدد الكلمات في النصوص التي تضمّها إلى مدوّنات محدودة تتألّف من مليون كلمة، ومعظم المدوّنات من هذا النوع، مقابل مدوّنات شاملة تبلغ عدد كلمات نصوصها عدّة مئات من الملايين ومن أمثلتها "مدوّنة أكسفورد الإنجليزيّة" التي تضمّ بليون كلمة (أي ألف مليون كلمة).

ويمكن تصنيف المدوّنات إلى أصناف أخرى طبقاً للحيثيّة التي ينظر إليها المُصنّف.

إنشاء المدونة:

لا توجد قواعد متفق عليها تتعلّق بإنشاء المدوّنة والإجراءات الواجب اتّباعها لتصميمها. ولعلّ السبب يعود إلى وجود أنواع كثيرة للمدوّنات، وكلُّ نوع يتطلّب إجراءات تختلف عمّا يتطلّبه النوع الآخر. فنوع المدونة التي نريد إنشاءها، أو حتّى اختيارها من بين المدونات القائمة فعلاً، يعتمد على نوع النتائج التي نتوخّى الحصول عليها منها.

السؤال الرئيس الذي ينبغي أن يجيب عليه المسؤول عن المدوّنة قبل أن يُقدّم على تجميع النصوص، هو: ما هو الغرض من إنشاء المدوّنة؟ هل ستكون قاعدة بيانات لتصنيف معجم تاريخي؟ أو لمجرد إعداد قائمة شيوع المفردات في اللغة المعاصرة؟ أو لتكون مصدراً تعليمياً لطلاب اللغة من غير الناطقين بها؟ أو للوقوف على المصطلحات الطيّبة العربية؟ إلخ. فالغرض من المدوّنة هو الذي يحدّد طبيعة نصوصها، وتركيبها، وحجمها، وشرحها، إلخ.

والسؤال الثاني الذي ينبغي أن يجيب عليه المسؤول عن إنشاء المدونة قبل تجميع النصوص، هو: من هم مستعملو المدونة المراد إنشاؤها؟ هل هي للاستعمال الشخصي، أم للاستعمال اللغويين الباحثين، أم للاستعمال التجاري العام؟ (28)

والسؤال الثالث الذي يُثار عند التخطيط لإنشاء المدونة هو: ما هي الموارد المتاحة لتجميع النصوص، من مال ووقت ومعدّات وخبرات بشرية؟ ولإعطاء فكرة عن التكاليف المادية، نذكر أن مدونة أكسفورد الإنجليزية التي تستخدم طريقة ذكية غير مكلفة في تجميع النصوص من مواقع الشبكة الإلكترونية (الإنترنت)، تبلغ ميزانيتها 35 مليون جنيه إسترليني.

وبصورة عامة، هنالك ثلاثة عوامل هامة تتحكّم في نوع المدونة التي ينبغي إنشاؤها أو استعمالها، وهي:

- 1) نوع المادة اللغوية المطلوبة،
- 2) كمية البيانات المطلوبة،
- 3) المدونات المتوفرة.

1) نوع المادة اللغوية المطلوبة:

يتوقّف نوع المادة اللغوية على نوع البحث الذي نريد إجراءه بمساعدة المدونة. ففي عملية تجميع المدونة، يستطيع المسؤول أن يؤثّر في نوعية النصوص التي تتألّف منها المدونة وتركيبتها، وذلك بتفضيل نوع من النصوص على بقية الأنواع، من حيث:

- . نوعية النصوص: أدبية (سردية، غير سردية، كتب أطفال، إلخ.)، علمية، فنية، رياضية، إلخ.؟
- . مستوى النصوص اللغوي: اللغة المكتوبة، اللغة المنطوقة، كلتاها، إلخ.
- . مواصفات كُتاب النصوص: من حيث العمر، الجنس، الثقافة، إلخ.؟
- . فترة كتابة النصوص: العصر الجاهليّ، العصر الإسلاميّ، العصر الوسيط، العصر الحديث، جميع عصور اللغة؟

. مكان كتابة النصوص: الجزيرة العربية، العراق، الشام، مصر، المغرب، جميع البلاد التي تُستعمل فيها اللغة العربية، إلخ.؟

وهكذا، فإنّ تركيبة المدوّنة تعكس اختيارات المسؤول عن تجميع المدوّنة. فقبل أن يقوم بتجميع نصوص المدونة عليه أن يقرّر نسبة كلّ نوع أو صنف من هذه الأصناف في تركيبة المدونة. هل يريد لها مدوّنة متوازنة أم مدوّنة هرميّة. فإذا افترض هذا المسؤول، مثلاً، أنّ لغة الإعلام هي المؤثر الرئيس في لغة الناس، أعطى نسبةً أكبر في المدوّنة لنصوص الصحفّية والإذاعيّة والتلفزيّة، وجعلها قاعدة الهرم.

لنأخذ مثلاً على ذلك تركيبة مدوّنة أكسفورد الإنجليزيّة *The Oxford English Corpus* التي تضمّ نصوصاً تعدادها بليون كلمة مقتبسة من مواقع شبكة المعلومات الدوليّة (الإنترنت) وبعض المصادر المطبوعة ورقياً مثل المجلات العلميّة لتكون نصوصاً متمّمة في بعض المجالات العلميّة التي يندر وجود نصوصها في الإنترنت. وتنتمي النصوص إلى الفترة من سنة 2000 إلى سنة 2006م، وهدفها دعم صناعة المعاجم في أكسفورد في القرن الواحد والعشرين، بحيث يستطيع المعجميّ أن يرى الكلمات في سياقاتها، ويقف على ظهور الألفاظ والمعاني والتراكيب الجديدة، وإدراك التوجهات الجديدة في استعمال الكلمات وتهجيتها.

تتألّف هذه المدوّنة من نصوص اللغة الإنجليزيّة، البريطانيّة والأمريكيّة، بنسبة 80% (أي حوالي 800 مليون كلمة)، والـ 20% الباقية (أي حوالي 200 مليون كلمة) فتتألّف من أنواع من الإنجليزيّة المستعملة في أستراليا، وجنوب أفريقيا، وكندا، والكاربي، إلخ. وتضمّ المدوّنة نصوصاً من مناطق تستعمل الإنجليزيّة لغةً ثانية، مثل الهند، وسنغافورة، وهونغ كونغ. فتركيبة هذه المدوّنة هرميّة من حيث اختيار اللهجات.

وتقسّم هذه المدوّنة إلى عشرين مجالاً موضوعياً، ونصوص كل مجال من هذه المجالات تُسمى مدونة فرعية *Subcorpus*. وكلّ مدوّنة فرعيّة مقسّمة هي الأخرى إلى مجموعة من المجالات الموضوعيّة المتخصّصة. فإذا أخذنا المدوّنة الفرعيّة الخاصّة بمجال (الرياضة)، نجد أنّها مقسّمة إلى 40 مجالاً متخصّصاً مثل: كرة القدم، كرة السلة، ألعاب الميدان، السباحة، إلخ. وهكذا يمكن القول إنّ تركيبة هذه المدوّنة متوازنة موضوعياً.

ويبين الجدول التالي تركيبة مدونة أكسفورد الإنجليزية من حيث مجالاتها الموضوعية، وعدد كلمات كل مجال، ونسبته من مجموع المدونة:

المجال الموضوعي	عدد كلماته بالمليون	% من المجموع
السرد (رواية، قصة ...)	220	22
الأخبار	190	19
مواقع الشبكة	150	15
هوايات أوقات الفراغ	75	7.5
الفنون	66	6.6
الرياضة	60	6
المجتمع	58	5.8
القانون	52	5.2
العلوم	44	4.4
الحاسوب	36	3.6
الإنسانيات	35	3.5
إدارة الأعمال	29	2.9
الدين	27	2.7
الطب	24	2.4
مجالات أخرى	50	5

(جدول يبين تركيبة نصوص مدونة أكسفورد الإنجليزية) (29)

إذا أردنا أن تكون المدونة صادقة في تمثيلها للغة المعاصرة في أنواعها المختلفة ومستوياتها المتباينة، وجب علينا أن نتأكد من أن النصوص التي نجمّعها مستقاة من المصادر المتنوعة مثل:

البحوث الأكاديمية، الأدلة التقنية، الدوريات، التقارير الصحفية، مواقع الشبكة الإلكترونية، الروايات والقصص القصيرة، كتابات الهواة، كتابات الحركات السرية، الخ. وتعطي الصحف والمجلات صورة صادقة عن الاستعمال اللغوي السائد، لأنّ موادّها لا تخضع لكثير من التنقيح والتصحيح والتحرير، كما هو الحال في بعض الكتب الأدبية التي تصدر عن دور نشر مشهورة. وهكذا يستطيع الباحث أن يقف من خلال الموادّ الصحفية على كثير من " الأخطاء " الشائعة في الإملاء والنحو والدلالة على الرغم من وجود القواعد " الصحيحة".

وقد يواجه المسؤول عن تجميع المدوّنة صعوبات قانونية وغيرها في الحصول على أنواع معينة من النصوص، مثل الرسائل الشخصية، أو النصوص المحمية بحقوق المؤلف، أو تلك النصوص التي يتطلّب تضمينها في المدوّنة إذناً من الناشر.

2) كمية البيانات المطلوبة وحجم المدوّنة:

إنّ كميّة البيانات المطلوبة تعتمد على نوع الدراسة التي نستخدم المدوّنة من أجلها. فإذا كنا نريد أن نضع قواعد شاملة للغة برمتها أو نقوم بإحصاءات لغوية، أو أن نصنّف معجماً تاريخيّ، أصبح من الضروريّ أن نخزّن كميات كبيرة من النصوص تمثّل اللغة في عصورها المختلفة ومناطقها المتباينة ومجالاتها المتنوعة. أما إذا كان هدفنا أن نعثر على بضعة أمثلة عن كيفية استعمال كلمة أو تعبير ما، فلسنا بحاجة إلى كميات كبيرة من النصوص.

ليست هنالك حدود معينة لكميّة النصوص في المدوّنة، ولا حدود لعدد الأمثلة التي يجب دراستها؛ وإنّما يمكن القول، بصورة عامّة، إنّ من المهمّ أن تكون لدينا نصوص " كافية". أمّا حدود الكفاية فتعتمد على نوع البحث الذي نقوم به (30). علينا أن نجيب على عدّة أسئلة منها: هل نريد أن تكون المدونة نموذجاً ممثلاً للغة أو أن تكون مستوعبة للغة؟ فمثلاً إذا كنا نعتزم إنشاء مدوّنة للشعر الجاهليّ، فيمكن أن تقتصر المدوّنة على نماذج من الشعر الجاهليّ، كالمعلّقات مثلاً، أو أن تستوعب جميع قصائد الشعراء الجاهليّين.

قد يتبادر إلى الذهن أنّ مدوّنة كبيرة أفضل من مدوّنة صغيرة. بيد أنّنا يجب أن نتذكّر أنّه حتّى المدوّنة الكبيرة قد لا تتوفّر على جميع البيانات اللغويّة إذا لم تكن نصوصها قد اختيرت بطريقة جيّدة. ولهذا يمكن القول بوجود عاملين يتحكّمان في مدوّنة نريد إنشاءها لتأليف معجم تاريخي، مثلاً: كمّيّة النصوص، وكيفيّة اختيار النصوص، بحيث تكون متوازنة موضوعيّاً وزمانيّاً ومكانيّاً.

في بداية ظهور المدوّنات الحاسوبيّة، كانت المدوّنة تتألّف، عادةً، من مليون كلمة. وبعد مدة انتقل حجم المدوّنات إلى مائة مليون كلمة، أما اليوم فنجد مدونة أكسفورد الإنجليزيّة تتكون من بليون كلمة. ومع ذلك، فإن معظم المدوّنات الموجودة تتألّف من مليون كلمة. ولكي نتصوّر مقدار المليون كلمة، علينا أن نعرف أنّصفحة واحدة من جريدة كبيرة تشتمل على حوالي ألف كلمة. فالمليون كلمة تساوي ألف صفحة من جريدة كبيرة (31)

3) المدوّنات المتوفرة:

في التخطيط لإنشاء مدوّنة، يجب دراسة إمكان الاستفادة من المدوّنات الموجودة فعلاً. فمثلاً، ثمة مدونات عديدة متوفّرة في البلاد العربيّة، يمكن الاستفادة من بعضها في إنشاء المدوّنة الخاصّة بتأليف معجم عربيّ تاريخيّ. ولكن هنالك عوامل تتحكم في الاستفادة من هذه المدوّنات. ومن أهم هذه العوامل ما يأتي:

أ. هل المدوّنة عمومية متاحة للاستعمال المجانيّ من طرف المؤسّسات والأفراد.

ب. أم أن المدونة خاصة لأغراض الاستغلال التجاريّ؟

ج. وإذا كان بالإمكان الاستفادة منها، فهل هي متوفّرة على الشبكة بحيث يمكن استنساخها، أو هي مسجّلة على أقراص يمكن اقتناؤها ونقلها؟ وفي كلتا الحالتين، هل يمكن نقلها بصيغة Plain Text كي نستطيع أن نركّب عليها محرّك بحث، فتستجيب لأسئلتنا واحتياجاتنا؟

وحثّى عندما نستطيع الاستفادة من المدوّنات المتوفرة، لا بدّ لنا من استكمالها بنصوص لتكون المدوّنة النهائيّة متوازنة طبقاً لأهدافنا.

صعوبات إنشاء المدوّنة باللغة العربيّة:

قرر اتحاد المجامع اللغوية العلمية العربية سنة 2004 إنشاء هيئة تتولّى تصنيف معجم تاريخي للغة العربية ينبنى على مدوّننة تساعد المعجميين على تتبّع تطوّر الألفاظ والتراكيب العربية في شكلها ومعناها واستعمالاتها. ولكنّ الاتحاد يواجه صعوبات ماليّة وبشريّة وتقنيّة، نجملها فيما يلي:

1) عدم توفّر الاتحاد على المال اللازم لتصنيف المعجم المطلوب، فهو ينتظر التبرعات من المُحسنين في زمن ندر فيه هذا النوع من الناس في بلادنا العربية. وندرك حجم الصعوبة التي يواجهها الاتحاد إذا ما علمنا أن الأكاديمية الإسبانية الملكيّة قد أعلنت في مطلع هذه السنة 2006م أنّها ستشرع هذا العام، بالتعاون مع أكاديميّات بلدان أمريكا اللاتينيّة، في إعداد معجم تاريخي للغة الإسبانية يشتمل على 120 ألف مدخل، ويقع في حوالي 12 مجلداً، ويستغرق تصنيفه بضع سنوات. وسيستفيد محررو هذا المعجم من قاعدة معلومات الأكاديمية، خاصّة مدوّنتها التاريخيّة للغة الإسبانية التي تضمّ 50 مليون معلومة لغويّة، وكذلك من طبعات معجمها للغة الإسبانية. ويضيف إعلان الأكاديمية أنّ الحكومة الإسبانية قرّرت تخصيص مبلغ إضافي للأكاديمية مقداره 1,5 مليون دولار في كلّ سنة من سنوات تأليف هذا المعجم (32). ولا شكّ في أنّ استجابة الحكومة الإسبانية السريعة نابع من إدراكها أنّ إيجاد مجتمع المعرفة الذي يحقّق التنمية البشريّة يتطلّب تنمية اللغة الوطنيّة أولاً وأنّ الأداة الرئيسيّة في تنمية اللغة هي المعجم، على حين أنّ جُلّ حكوماتنا العربية لا تدرك ذلك، بل على العكس تصرّ على استعمال لغات أجنبية بدل العربيّة في التعليم والإدارة.

2) عدم توفّرنا في البلاد العربيّة على عناصر بشريّة كافية لها دراية بإنشاء المدوّنات وبصناعة المعجم المبنية على البحث في المدونة. ناهيك بندرة اللغويين المختصين باللغات العربيّة (الساميّة) وغيرها من اللغات التي تفاعلت معها اللغّة العربيّة خلال عمرها المديد. وعندما عقد اتحاد المجامع اللغوية العلميّة العربيّة، بالتعاون مع مجمع اللغة العربيّة الأردني، ندوةً في عمّان في المدة من 19.16 سبتمبر/أيلول 2002م حول " قضايا اللغة العربيّة في عصر الحوسبة والعولمة" لم يُقدّم في الندوة أيُّ بحث يتعلّق بلسانيات المدونة، ما يشير إلى ندرة هذا النوع من البحوث في بلادنا العربية (33).

3) عدم وجود مدونات، بالمعنى الصحيح، يُمكن استثمارها في صناعة المعجم التاريخي. فالمدونات العربية الموجودة التي تعدّ على أصابع اليد الواحدة، لا تتوفر على الشروط المطلوبة؛ فهي جُمعت لأغراض تجارية، وهي ليست متوازنة من حيث نوعية نصوصها، ولا موثقة بصورة كاملة، ومكوناتها اللغوية ليست مشروحة ولا مؤشّر عليها.

4) إنّ تجميع نصوص مدونة جديدة وتخزينها بالحاسوب يتطلبان جهوداً جبارة، خاصة لعدم توصل الحاسوبيين إلى نظام جيد للنسخ الضوئي للنصوص العربية. ولهذا يتوجب طباعة النصوص يدوياً على لوحة مفاتيح المطراف. فهناك طريقتان لإدخال النصوص في الحاسوب: (أ) الطباعة على المطراف، و(ب) المسح الضوئي. ويتحكم في الاختيار بين الطريقتين تكلفة الطباعة على المطراف، والنوعية، إذ يؤدي المسح الضوئي إلى أخطاء كثيرة. وتحتاج كلتا الطريقتين إلى مراجعة النصوص وتصحيحها بعد إدخالها في الحاسوب.

5) عدم توفر محركات بحث بالعربية في الأسواق التجارية، ولا بُدّ من تطوير محرك بحث للمدونة المزمع إنشاؤها. وهذا يتطلب كثيراً من البحث في المعالجة الآلية للغة العربية.

توثيق المدونة وشرحها:

لنفترض أنّنا نضطلع بتصنيف معجم تاريخي للغة العربية، واستخدمنا مدونة تضم نصوصاً عربية تنتمي إلى عصور اللغة المختلفة، وطلبنا من المدونة تزويدنا بالسياقات (الجملة) التي ترد فيها كلمة "مدونة"، فأخرج لنا الحاسوب مطبوعة تشتمل على ألف سياق وردت فيها كلمة مدونة، دون أن يخبرنا عن التاريخ الذي كُتب فيه كل سياق، ومن هو الكاتب، وفي أي مجال علمي: فقهي، أدبي، علمي، إلخ..، فإن هذه المدونة لا قيمة لها في عملنا.

ولهذا فإننا بحاجة إلى معلومات توثيقية ولغوية وبلاغية عن كل نص من نصوص المدونة. فإذا حملت المدونة هذه المعلومات في صورة شارات وعلامات ورموز توضع بعد كل نص وبعد كل مكون من مكوناته، وُصفت بأنها مدونة مشروحة Annotated Corpus.

ويتفق الباحثون على أنّ المعلومات التوثيقية الضرورية عن كل نص، لصناعة المعجم التاريخي، تتناول ما يأتي:

- . عنوان النص، أو عنوان الكتاب الذي اقتطف النص منه.
- . اسم الكاتب، وجنسه (الذي يُعرف من اسمه باللغة العربية)
- . تاريخ كتابة النص أو تاريخ ميلاد الكاتب وتاريخ وفاته، أو على الأقل العصر الذي عاش فيه الكاتب.
- . جنسية الكاتب، البلاد التي ينتمي إليها، أو المكان الذي عاش فيه.
- . موضوع النص.

إن هذه المعلومات التوثيقية تُذكر عادة مع كل نص من نصوص المدونة عند إدخاله في الحاسوب، وتتولى البرمجة بعد ذلك مهمة ذكر هذه المعلومات، إذا طُلبت، بعد كل جملة أو سياق يستخرجه الباحث.

وبالإضافة إلى ذلك، فإنّ هناك حاجة، خاصة لدى الباحثين في قواعد اللغة، إلى معلومات تساعدهم في تحليل اللغة وتقعيدها. وتستجيب المدونات لهذه الاحتياجات بوضع شارات بعد المفردات تدلّ على أقسام الكلام التي تنتمي إليها تلك المفردات، وشارات نحوية بعد العبارات النحوية مثل عبارتي المُسنَد والمُسنَد إليه، وشارات بلاغية تبيّن نوعية اللغة (شعرية، علمية، دينية، إلخ...)، وهكذا.

وفيما يتعلّق بنوعية اللغة المُستعملة في النصّ، يرى اللغوي الفرنسي راستيه ضرورة أن تبيّن شارات المدونة المشروحة ثلاثة مستويات: نوع الخطاب (علمي، أدبي، قانوني، ...)، والحقل النوعي، ففي الخطاب الأدبيّ، مثلاً، يوجد (شعريّ، سرديّ، مسرحيّ، ...)، والنوع، ففي حقل السرد

توجد أنواع (رواية، رواية بوليسية، قصة، حكاية، أدب رحلات، ...)، وحتى نوع الرواية يمكن أن يكون جنساً تنضوي تحته أنواع عديدة (رويات الرسائل، رواية المذكرات، ...).

ويرى راستير ضرورة شرح المدونة بشارات تبين المستويات الثلاثة الأولى من الأنواع اللغوية، على الأقل. ولكنّ الشارات لا تغني عن تحليل السياقات المطلوبة تحليلاً دقيقاً؛ لأنّ النص العلمي قد يشتمل على لغة دينية وبالعكس. ولا يوجد هناك نوع لغوي خالص. فاللغة تمثّل الحياة، والحياة زاخرة بأصناف النشاط الإنساني المتنوعة.

قد لا نحتاج إلى وضع شارات صرفية على كلّ فعل أو اسم أو ضمير في المدونة العربية، لأنّ معالجة اللغة العربية آلياً حققت تطوّراً على مستوى الحرف والكلمة والجملة. وتوجد حالياً عدّة مُعالجات صرفية تستطيع تحليل الكلمة العربية إلى عناصرها الاشتقاقية والصرفية. فمثلاً، يستطيع المُعالج الصرفي تحليل الكلمة وتفكيكها إلى نواتها الأساسية، والعناصر اللاصقة قبلها مثل أداة التعريف وواو العطف، والعناصر اللاصقة بعدها مثل الضمائر المتصلة والعلامات الإعرابية. فكلمة " واستفساراتهم" يفكّكها المُعالج الصرفي آلياً إلى (و + استفسار + ات + هم)، ثم يُمكنه استخراج الجذر (ف س ر)، والفعل (استفسر)، وصيغة الفعل (استفعل) وهكذا(34).

ويذكر عبد الفتاح إبراهيم وسالم غزالي العاملان في قاعدة البيانات المعجمية العربية (أو مشروع معجم العربية الآلي) في المعهد الإقليمي لعلوم الإعلاميّة والاتّصال عن بُعد في تونس، أنّ هدف المشروع هو تحقيق المعالجة الآلية الكاملة للغة العربية، تحليلاً وتوليداً، وأنّ المشروع يتكون من قاعدة البيانات الفعلية (حوالي 20700 فعل)، وقاعدة البيانات الأسمية (حوالي 40000 وحدة مفردة)، وأدوات المعالجة الآلية المتكونة من محلل صرفي آلي للكلمات، ومولّد الكلمات الممكنة بحركاتها. وقد جُرّبت أدوات المعالجة الآلية هذه على مدونة حاسوبية فتعرّفت على 90 % من مفرداتها (35).

نخلص من هذا كله إلى أنّ المدوّنة الحاسوبية للغة العربية ليست بحاجة إلى وضع شارات صرفية بعد كلّ مفردة من مفرداتها، فالمعالجة الآلية للنصوص تغنيها عن إضافة تلك الشارات إلى المفردات في المدوّنة. فإذا طلب الباحث ضميراً من الضمائر المتصلة مثلاً، قام المحلّل الصرفي بتحديد ذلك الضمير في جميع السياقات التي يرد فيها.

المدوّنة وأزمة المعجم العربي المعاصر:

يعاني المعجم العربي المعاصر أزمة جعلته عاجزاً عن أداء وظيفته، ما صرف الناس عنه. وتتجلّى هذه الأزمة في ظاهرتين هما:

(1) انقسام المعجم العربي عن واقعه، أي أنه لا يصف اللغة العربية الفصيحة كما يستعملها أهلها اليوم.. والسبب في ذلك أنّ المعجميين العرب منذ عصر الانحطاط حتّى اليوم تخلّوا عن عمليّة جمع النصوص ودراستها وتحليلها لاختيار مداخلهم وكتابة موادهم، كما كان يفعل رواد المعجمية العربية كالخليل والأصمعي. لقد اكتفى هؤلاء المعجميون منذ عصر الانحطاط بالتقل من المعاجم السابقة واستنساخ التعريفات منها، فلم تُعدّ المعاجم المعاصرة قادرة على مساعدتنا في معرفة المعاني الجديدة التي اكتسبتها المفردات بعد ذلك. فالمعاجم القديمة التي نُقلت منها لا تمثّل اللغة كما هي عليه اليوم.

(2) الاكتفاء بتسجيل معاني المفردات والخلط في ترتيبها، فهذه المعاجم لا تُعنى بتسجيل استعمالات المفردات في متلازمات لفظية من تعابير اصطلاحية وسياقية وغيرها (36)، كما أنّها تُدرج المعاني كما اتفق دون تفريق بين الممات والحيّ، والحسيّ والعقلي، والحقيقيّ والمجازي، والعامّ والخاصّ. وحتى عندما تسوق بعض الشواهد على تلك المعاني، فإنّها لا ترتبها ترتيباً زمنياً، فيأتي الشاهد من العصر العباسيّ مثلاً قبل الشاهد من العصر الجاهليّ، وهكذا (37). وهذا ناتج عن عدم توقّر المعجميين على بيانات كافية تساعدهم في ترتيب المعاني والوقوف على استعمالات الألفاظ.

ويضرب لنا سالم غزالي وعبد الفتاح إبراهيم مثلاً على النقص في معاجمنا المعاصرة بالفعل (أخذ)؛ إذ قارنا المعاني التي أوردها " المعجم الوسيط" لهذا الفعل بمعانيه التي توصلوا إليها بعد

تحليل سياقاته في مدونة صغيرة لا يتجاوز حجمها مليون ونصف مليون كلمة، جُمعت نصوصها من الصحف والمجلات والكتب المدرسية وبعض الكتب الأدبية مثل " الأيام " لطفه حسين وبعض كتب أبي حيان التوحيدي(38). وطبعاً، لا تُعدّ هذه المدونة كافية لتصنيف معجم عربيّ.

وجد الباحثان أن "المعجم الوسيط" يعطي المعاني التالية للفعل (أخذ):
. أخذ (الشيء): حازه وحصله؛ تناوله؛ قبّله.
. أخذ (فلاناً) : حبسه ؛ عاقبه ؛ قتله ؛ أسره ؛ غلبه ؛ أمسك به ...

أما المدونة الصغيرة فقد أعطت 1440 سياقاً ورد فيها الفعل (أخذ). وفي 35% من هذه السياقات (505 سياقات)، كان معنى الفعل (أخذ): تناول أو أمسك بـ أو حاز، وهذا المعنى، كما يبدو، هو المعنى الحقيقي أو المركزي.

والمعنى الثاني للفعل (أخذ). كما ظهر في 325 سياقاً من سياقات هذه المدونة، أي 23% من مجموع السياقات. لا علاقة له بالحيازة والتناول. وهذا المعنى يظهر عندما يتقدّم الفعل (أخذ) فعلاً آخر في صيغة المضارع، فيصير معنى (أخذ): بدأ، كما في : أخذ يقفز، وأخذ الولد يقفز، أخذ في القفز. وسنستخلص من سياقات (أخذ + فعل مضارع)، أنّ الفعل (أخذ) يكون دائماً في صيغة الماضي، ولا يرد أبداً في صيغة المضارع؛ وعلى المعجمي أن ينبّه القارئ إلى ذلك. قد يُطرح السؤال: وهل يمكن أن نقول: " سيأخذ الولد في القفز." أو " يأخذ الولد في القفز كلما انتهى من عشائه" ؟ يكمن الجواب في المدونة إذا كانت وافية بالغرض، لأنّ مداخل المعجم ومعانيها واستعمالاتها يجب أن تنطلق من المدونة، وجميع قواعد اللغة التي نضمّنها في المعجم ينبغي أن تبني على المدونة كذلك.

لم تشتمل المدونة الصغيرة على أيّ سياق يرد فيه الفعل (أخذ) بمعنى: قتل، أو حبس، أو قبّل؛ وهي المعاني التي أوردها "المعجم الوسيط". ولكن المدونة أشارت إلى معانٍ واستعمالاتٍ للفعل (أخذ) تبلغ نسبتها 53% من مجموع السياقات التي يرد فيها الفعل (أخذ)، لم يذكرها "المعجم الوسيط" (بمعنى أنه "فاته نصف اللغة أو أكثر" على حدّ تعبير مجد الدين محمد بن يعقوب

الفيروزآبادي (ت 817هـ / 1414م) صاحب القاموس في معرض نقده لمعجم الصحاح للجوهري (39). ومن أمثلة استعمالات الفعل (أخذ) التي لم يذكرها المعجم الوسيط ما يأتي:

. أخذ في النظر، أخذ بالنظر، أخذ في الاعتبار، أخذ بالاعتبار، أخذ في عين الاعتبار أخذ بعين الاعتبار،
. أخذ على عاتقه،
. أخذ الاحتياطات اللازمة/الكافية...
. أخذ مبادرة...
. إلخ. (40)

من هذه الأمثلة البسيطة المستقاة من مدونة محدودة الحجم جداً وغير متوازنة، تظهر لنا بوضوح ضرورة إرساء تصنيف المعجم العربي على مدونة كافية، إذا أردنا لهذا المعجم أن يكون ممثلاً صادقاً للغة.

المدونة وصناعة المعجم المختص:

لم يعد علم المصطلح مجرد دراسة قوائم المصطلحات ومقابلاتها بلغة أخرى. فتطور الدراسات النظرية في علم المصطلح وتطبيقاتها العملية أدى إلى ظهور ما نسميه بعلم المصطلح النصي *Textual Terminology/ Terminologie textuelle*. وعلى رأس اهتمامات هذا العلم الانطلاق من مدونة مزدانة بنصوص علمية يتولى المصطلحي فحصها بعناية، واستخلاص المصطلحات في سياقاتها منها.

ففي علم المصطلح النصي يتبع البحث خطوات منهجية يمكن إجمالها في ما يلي:

- . التحديدات النظرية
- . تعريف دور المدونة
- . تحديد الاحتياجات
- . تحديد الطرائق والأدوات: برامج المعالجة الآلية للغة العربية

. وصف النماذج المُنتجة أو المستعملة

. تقييم النتائج (41)

لننظر الآن في الفوائد التي نجنيها من استخدام مدونة عربية كبيرة الحجم تضم نصوصاً من جميع البلاد العربية، عبر عصور اللغة المختلفة، وتغطي جميع المجالات العلمية والأدبية والفنية، بما في ذلك مدونات فرعية في علوم الكيمياء والفيزياء والرياضيات، إلخ.

ولنفترض أننا نبحث في مصطلحات الكيمياء من أجل إعداد مشروع معجم مختصّ بمصطلحات الكيمياء، يشارك في تنفيذه مجموعة من الكيميائيين والمصطلحيين واللغويين، والغاية منه تنسيق المصطلحات الكيميائية العربية، وتقييمها، وتوحيدها على مستوى الوطن العربي.

نستطيع أن نحصل من هذه المدونة على معلومات وإجابات على أسئلة كثيرة تفيدنا في مشروعنا هذا وفي أمور عديدة تتعلق في العلوم وتاريخها وتطور الثقافة العربية عموماً، وتعريب التعليم العالي، وطرائق تدريس الكيمياء، وغيرها من الأمور. ومن هذه المعلومات ما يأتي:

1) معلومات إحصائية تجيب على الأسئلة التالية:

أ. ما نسبة المصطلح الكيميائي في النصوص العامة غير المتخصصة؛ في اللغة العربية عموماً، وفي كل عصر من عصورها، وفي كل قطر من أقطارها؟

ب. ما نسبة المصطلحات العلمية (الكيميائية) في نصوص علم الكيمياء؟

ج. ما نسبة المصطلح الكيميائي في مجموع المصطلحات العلمية؟

2) معلومات تتعلق بتاريخ العلوم، مثل:

أ. تطوّر مفاهيم علم الكيمياء، وذلك بفحص المصطلح الكيميائي الواحد في مختلف العصور ومراقبة خصائص المفهوم الذي يعبر عنه في كل عصر.

ب. تطوّر مصطلحات علم الكيمياء، وذلك بفحص مفهوم واحد والمصطلحات المختلفة التي تعبر عنه في جميع العصور.

3) معلومات ضرورية لكشف الازدواجية المصطلحية والتخلص منها، مثل:

- أ. ما هي المصطلحات المختلفة التي تعبّر عن المفهوم الواحد في القديم والحديث؟
ب. ما هي المصطلحات المختلفة التي تعبّر عن المفهوم الواحد في الأقطار العربية المختلفة حالياً؟

4) معلومات ضرورية لكشف الاشتراك اللفظي المصطلحي والتخلص منه، مثل:

. ما هي المصطلحات المستعملة في أكثر من مجال علمي واحد؟

5) معلومات ضرورية لكشف الترادف المصطلحي والتخلص منه، مثل:

. ما هي المصطلحات التي تعبّر عن مفهوم علمي واحد؟

6) معلومات تتعلق بشكل المصطلح، مثل:

. هل يُكتب المصطلح بأكثر من شكل إملائي واحد؟ وأيها أشيع؟ أوكسجين أو أكسجين؟
إلخ.

إنّ جميع هذه الأسئلة ذات أهميّة كبيرة لمن يعمل في الميدان المصطلحي، ولا يمكن الإجابة عليها بصورة علميّة دقيقة ما لم تتوفر على مدوّنة حاسوبية تضمّ نصوصاً كافية يُعَوَّل عليها. وبطبيعة الحال، فإنّ المدوّنة لا تقوم بتنسيق المصطلحات، ولا بوضعها، ولا بتقييسها وتوحيدها، وإنّما تزودنا بالبيانات الضرورية للقيام بذلك العمل بوجه فعّال.

استنتاجات وتوصيات:

يتضح مما قدّمناه أنّ تصنيف معجم جيّد للغة العربية، سواء أكان معجماً عاماً أو مختصّاً، ينبغي أن يبني على دراسة وتحليل أمثلة حيّة وافية في سياقاتها الطبيعيّة تبيّن لنا استعمال اللغة العربية في الحياة الحقيقيّة، وأن لا يكون المعجم الحديث نقلاً عن معجم قديمة انقطعت صلتها بالتغيّرات والتحوّلات التي عرفتتها اللغة العربية بعدها في ألفاظها واستعمالاتها وتراكيبها.

ولما كانت اللغة العربية لغة معمرة مرتبطة بثقافة عريقة مستمرة منذ آلاف السنين، فإنّ المعجم العام ينبغي أن يكون معجماً تاريخياً، يرتب معاني المداخل ترتيباً زمنياً (من القديم إلى الحديث أو بالعكس، حسب الجمهور الذي يرمي إلى خدمته، فمعجم الطلاب، مثلاً، قد يُرتب من الحديث إلى القديم)، وأن يميّز الترتيب بين الاستعمالات الحقيقية والمجازية، العامة والخاصة، الشائعة والنادرة. ولا يمكننا إنجاز ذلك ما لم تتوفر لنا معلومات كافية توضح تطوّر الألفاظ واستعمالاتها ومعانيها. ولما كانت اللغة العربية لها مستويان فصيح وعامي، وأنّ معجمنا يتناول العربية الفصيحة، فلا بُدّ أن يزواج بين المقاربتين الوصفية والمعيارية لتمكين الطالب أو الباحث الذي يستعين به من التعبير بلغة عربية فصيحة سليمة.

يبدّ أن تجميع نصوص المدوّنة، وإنشائها، واستخدامها بشكل فاعل، يحتاج إلى باحثين متمرسين في اللسانيات الحاسوبية. وهذا النوع من الباحثين غير موجود في بلادنا العربية في الوقت الحاضر، لأنّ أقسام اللغة العربية وشعبها في جامعاتنا ما زالت تستخدم مناهج قديمة تُدرّس بطرائق قديمة. ولهذا ينبغي تغيير المناهج والطرائق التدريسية وتحديثها، وبعبارة أخرى، من الضروريّ تمخّور أقسام اللغة العربية على الهندسة اللغوية. فالكتاب في شكله الحاليّ قد يندثر قريباً ويحلّ محله كتاب إلكترونيّ بحجم الكتاب العادي، نستطيع أن نستخدمه لمطالعة الكتب، وسماع الأغاني، ومشاهدة الأفلام، والاتّصال مع الآخرين شفويّاً وكتابياً، والاطّلاع على الأخبار، وإجراء المعاملات التجارية، وتوقيع العقود، وإجراء الحجوزات الفندقية والجوية والطرقية اللازمة لسفرتنا، وجمع الأنشطة الحياتية التي تخطر على بالنا، كما تنبأ بذلك بيل جيتس في كتابه "المعلوماتية بعد الإنترنت" (42).

- (1) ابن منظور، لسان العرب (بيروت: دار صادر، ب ت) مادة (دون)، وكذلك :
علي القاسمي (المنسّق) وآخرون، المعجم العربي الأساسي (باريس: الألكسو/لاروس،
1989) مادة دون
- (2) الإمام مالك بن أنس، كتاب الموطأ (بيروت: دار الآفاق الجديدة، ط 3، 1985) مقدمة
فاروق سعد، ص 8.
- (3) علي القاسمي، مرجع سابق، مادة (دون).
- (4) مكتب تنسيق التعريب، المعجم الموحد لمصطلحات اللسانيات (الرباط: مكتب تنسيق
التعريب، 2002).
- وقد جرى الخلط في هذا التعريف بين مصطلح (Lexicon) [وقد ورد في موضع آخر في
المعجم محرّفًا هكذا: Lexikon] الذي يدلّ على الرصيد اللفظي العام للغة، وبين مصطلح
(Corpus) الذي يدلّ على نصوص لغوية مجموعة لغرض تحليلها.
- (5) McArthur & McArthur (ed.). The Oxford Companion to the English
Language, as cited in:
www.georgetown.edu/faculty/ballc/corpora/tutorial2.html
- (6) عبد الحميد الشلقاني، الأعراب الرواة (طرابلس: المنشأة العامة للنشر والتوزيع والإعلان،
1975) 153 . 173.
- (7) جمال الدين القفطي، إنباه الرواة على أنباه النحاة ، تحقيق : محمد أبو الفضل إبراهيم
(القاهرة: دار الفكر العربي، 1986) ج 2، ص 258.
- (8) المرجع السابق.
- (9) عبد الحميد الشلقاني، المرجع السابق، ص 156.
- (10) المرجع السابق، ص 158.
- (11) أحمد بن محمد بن خلّكان، وفيات الأعيان وأنباء أبناء الزمان، تحقيق: إحسان
عباس (بيوت: دار الثقافة، ب ت) ج 4، ص 334.
- (12) إسماعيل بن حماد الجوهري، الصحاح ، تحقيق: أحمد عبد الغفور عطار (بيروت:
دار العلم للملايين، ط 3: 1984) ج 1 ص 33.

(13) محمود بن عمر الزمخشري، أساس البلاغة، تحقيق: عبد الرحيم محمود (بيروت: دار المعرفة، 1979) صفحة ك من المقدمة.

(14) “ Corpus Linguistics” in:

www.essex.ac.uk/linguistics/clmt/w3c/corpus-ling/content/history.html

(15) J. Kading (1879). Hanfigkeitwörterbuch der Deutschen Sprache. Steglitz.

(16) F. Boas (1940). Race, Language and Culture. New York: Macmillan.

(17) C. Fries & A. Traver (1940). English Word List. Washington, D.C. American Council of Education.

(18) François Rastier, “ Enjeux épistémologiques de la linguistique de corpus” in : www.revue-texto.net/inédits/rastier_enjeux.html

(19) Salem Ghazali & Abdelfattah Braham,” Dictionary Definitions and Copus-Based in Modern Standard Arabic”. A printed paper.

(20) Henry Kucera & Nelson Francis, Computational Analysis of Present-Day American English, 1967.

(21) www.en.wikipedia.org/wiki/corpus_linguistics

(22) انظر : علي حلمي موسى، إحصائيات جذور معجم لسان العرب باستخدام

الكمبيوتر (الكويت: مطبوعات جامعة الكويت، رقم 19 : 1972)، و علي حلمي موسى،

دراسة إحصائية لجذور معجم الصحاح باستخدام الكمبيوتر (الكويت: مطبوعات جامعة

الكويت رقم 33، 1973)، و علي حلمي موسى وعبد الصبور شاهين، دراسة إحصائية

لجذور معجم تاج العروس باستخدام الكمبيوتر (الكويت: مطبوعات جامعة الكويت، رقم

32، 1973).

(23) S. Jonson (1991) “ Times change and so do corpora” in: Aijmer&Altenburg (ed.) English Corpus Linguistics. London. Longman. Pp. 305-314.

(24) www.en.wikipedia.org/wiki/corpus_linguistics

(25) www.georgetown.edu/faculty/ballc/corpora/tutorial2.html

(26) عدنان عيدان، " نظام بنك اللغة العربية" ورقة تعريفية بالنظام.

(27) للاطلاع على هذا النوع من الترجمة الآلية انظر: "الترجمة الآلية ما زالت هدفاً بعيد

المنال" بقلم ستكس، ترجمة عمر البزري، مراجعة عدنان الحموي، في مجلة العلوم (الترجمة

العربية لمجلة ساينتفك أمريكان)، الكويت، المجلد 22، العددان 7/6 (يونيو/يوليو

2006) ص 21.18.

(28) www.georgetown.edu/faculty/ballc/corpora/tutorial2.html

(29) www.askoxford.com/oec/mainpage/?view=print

(30) www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/choosing.html

[ml](#)

- (31) www.georgetown.edu/faculty/ballc/cprpora/tutorial2.html
- (32) من موقع الأكاديمية الإسبانية الملكية www.rae.sp
- (33) انظر بحوث الندوة في: اتحاد المجامع اللغوية العلمية العربية، ندوة قضايا اللغة العربية في عصر الحوسبة والعولمة (القاهرة: اتحاد المجامع، 2002)
- (34) نبيل علي ونادية حجازي، الفجوة الرقمية (الكويت: عالم المعرفة، رقم 318، 2005) ص 357 . 371.
- (35) عبد الفتاح إبراهيم وسالم غزالي، " قاعدة البيانات المعجمية العربية أو مشروع معجم العربية الآلي " معالي دينار " في المجلة العربية للعلوم، الألكسو ، تونس ، عدد 32 (ديسمبر 1998) ص 15 . 23.
- (36) علي القاسمي، "التعابير الاصطلاحية والسياقية ومعجم عربي لها" في مجلة اللسان العربي، المجلد 17، ج 1 (1979) ص 3417.
- (37) شوقي ضيف، " صعوبات الاستشهاد الشعري في المعجم العربي التاريخي " في ندوة المعجم العربي التاريخي، تونس، نوفمبر 1989.
- (38) Salem Ghazali & Abdelfattah Braham. Op.cit.
- (39) محمد بن يعقوب الفيروزآبادي، القاموس المحيط (بيروت: مؤسسة الرسالة، 1986م) مقدمة المؤلف، ص 34.
- (40) Salem Ghazali & Abdelfattah Braham. Op.cit.
- (41) M. Bouveret, " Linguistique de corpus et terminologie" en : www.irit.fr/asstccot/
- (42) بيل جيتس، المعلوماتية بعد الإنترنت (طريق المستقبل)، ترجمة عبد السلام رضوان (الكويت: عالم المعرفة رقم 231، 1998) ص 13 . 63

ملحق: مراجع ورقية وإلكترونية في لسانيات المدونة

أ) مراجع باللغة الإنجليزية

Corpus Linguistics

I

Bibliography

Please excuse us while we are working on this page to improve the structure and make the information more easily accessible.

CONTENT

- Books on **Corpus Linguistics**/use of corpora
 - [By title](#)
 - [By author](#)
 - [By topic](#)
 - **Corpus Linguistics in [general](#)**
 - [Special issues](#) within **Corpus Linguistics**
 - [Particular corpora](#)
 - [Other areas](#) where corpora are

used

- **On-line bibliographies and lists of publications**
 - [General sites](#)
 - [Specific topic areas](#)
 - [Other material](#) available on-line
 - [Journals](#) available on-line
-

Books on **Corpus** Linguistics/use of corpora

(* = Links to information given by the publisher)

BY TITLE

- **The BNC Handbook**(*) by Guy Aston and Lou Burnard (Edinburgh University Press) 1998
Extensive information about how to use the SARA search engine. Step-by-step exercises. The first two chapters provide good general information about **corpus** linguistics and the BNC.
- **Concordances in the Classroom: a resource book for teachers** by Chris Tribble and Glyn Jones (Longman Group UK Limited) 1990
An introduction to concordancing with illustrations of how it can be used in the classroom.
- **Computer **Corpus** Lexicography**(*) by Vincent B.Y. Ooi (Edinburgh University Press) 1998
- **Corpus Annotation** (*) by Garside, Leech, McEnery (Addison Wesley Longman) 1997
A collection of papers on various aspects of

corpus annotation, from grammatical tagging and semantic annotation to software development and the emergence of standards for the evaluation of **corpus** annotation. ([external link](#))

- **Corpus-Based Methods in Language and Speech Processing** (*) by Steve Young (ed). (Kluwer Academic Publishers) 1997
- **Corpus Linguistics**(*) by Biber, Conrad & Reppen (Cambridge University Press) April 1998
- **Corpus Linguistics** (*) by Tony McEnery and Andrew Wilson (Edinburgh University Press) 1996

An easy to read, comprehensive coursebook on **corpus** linguistics with web-site [supplement](#).(Abstracted and reviewed in [ICAME Journal 21, 1997](#).([external link](#)).

- **Corpus Processing for Lexical Acquisition** (*) by Branimir Bougraev & James Pustejovsky (eds.) (MIT Press) 1996
- **Current Issues in Computational Linguistics: In Honour of Don Walker**(*) by A. Zampolli, N. Calzolari, M. Palmer (Kluwer Academic Publishers) 1994
- **Data-Intensive Linguistics** by [Chris Brew](#) and [Marc Moens](#) ([work in progress](#))
- **An Empirical Grammar of the English Verb: Modal Verbs** by Dieter Mindt (Cornelsen Verlag) 1995
Descriptive grammar fully based on studies of corpora. ([external link](#)) [review abstract](#)
- **English Corpus Linguistics**(*) by Karin Aijmer & Bengt Altenberg (eds.) (Addison

Wesley Longman) 1991

- **Exploring Textual Data (*)** by L. Lebart, A. Salem, L. Berry. (Kluwer Academic Publishers) 1997
- **An Introduction to Corpus Linguistics(*)** by Graeme Kennedy (Addison Wesley Longman) 1998
- **An Introduction to Spoken Interaction(*)** by Anna-Brita Stenstrom (Addison Wesley Longman) 1994
- **Language and Computers(*)** by Geoff Barnbrook (Edinburgh University Press) 1996
Basic introduction to corpus-based language research and computer analysis techniques for those with no or little experience in the field.
- **Learner English on Computer (*)** by Sylviane Granger (ed). (Addison Wesley Longman) 1998
- **Projects in Linguistics A Practical Guide to Researching Language (*)** by Alison Wray, Kate Trott, Aileen Bloomer. (Arnold) 1998
Introduction to themes, tools and techniques of linguistic research with a small section on use of computers and corpora.
- **Spoken English on Computer (*)** By Geoffery Leech, Greg Myers, and Jenny Thomas. (Addison Wesley Longman) 1995
A collection of papers on different aspects of spoken corpora.; creation, transcription, coding, use. ([external link](#)).
- **Statistics for Corpus Linguistics(*)** by Michael Oakes (Edinburgh University Press)

- **Survey of the State of the Art in Human Language Technology** by Ron Cole et al. (1996)
A collection of papers by 97 contributing authors. Available [on-line](#) or as [post-script files](#).
- **Teaching and Language Corpora(*)** by Anne Wichmann, Steven Fligelstone, Tony McEnery, and Gerry Knowles (eds). (Addison Wesley Longman) 1997
- **Text and Corpus Analysis(*)** by Michael Stubbs (Blackwell Publishers) 1996
An introduction to **corpus** linguistics in terms of linguistic tradition, with examples of analyses of texts and corpora.
- **Using Computers in Linguistics: A Practical Guide (*)** by John M. Lawler & Helen Aristar Dry (eds) (Routledge) 1998
"provides a non-technical introduction to recent developments in linguistic computing and offers specific guidance to the linguist or language professional who wishes to take advantage of them" [web-page](#) with regularly updated appendices of resources
- **Using Corpora for Language Research** by Jenny Thomas and Mick Short (eds). (Addison Wesley Longman) 1996
A collection of papers on different aspects of **corpus** linguistics (dedicated to Geoffrey Leech). (Abstract in [ICAME Journal 21:103, 1997](#)).

Selected on-line bibliographies and lists of

publications

General sites

- **ICAME Bibliography**, compiled by Bengt Altenberg.
A comprehensive list of works "based on or related to computerized English text corpora".
[Part 2](#) (-1989) [Part 3](#) (1990-96). Part 1 published in [ICAME News](#) 10 (1986).
- Michael Barlow's [Corpus Linguistics Bibliography](#)
A long list of various written sources.
- [Annotated Bibliography](#) by Patrik Svensson.
- [Addison Wesley Longman Corpus Linguistics](#) catalogue.
- [Edinburgh University Press](#) Edinburgh Textbooks in Empirical Linguistics
- [John Benjamins Publishing](#) Computational & [corpus](#) linguistics Catalogue

Specific topic areas

- [FrameNet Bibliography](#)
"list of references relevant to computational lexicography and related fields".
 - Articles about [Project Gutenberg](#)
 - [Classroom Concordancing/Data-driven learning Bibliography](#) "references to the direct use of data from linguistic corpora for language teaching and language learning".
 - [List of publications from UCL](#) (United Kingdom)
- Much on learner corpora.

- [UCREL Publications Bookshelf](#)
Publications produced by members of [UCREL](#), Lancaster University.
- [Publications of the Department of Language and Speech](#), University of Nijmegen.
[TOSCA Publications](#)
Publications by the TOSCA members, Nijmegen.
- [BNC-related publications and presentations](#)
Papers, articles, presentations concerning the BNC or research using the **corpus**.
- [Elsnet publications](#) (European Network in Language and Speech).
- [Cobuild](#)
Select bibliography of papers by Cobuild staff members. Focus on lexicography.
- [CSLU publication](#)
Publications from the Center for Spoken Language Understanding.

Other material available on-line

- [General Description of Parsers](#) Georgetown University
- [Systematic Dictionary of Corpus Linguistics](#)
"an attempt to group, systemize, define and explain the basic English terms in **Corpus** linguistics and relative fields"

Journals (with issues available on-line)

- [Computers and the Humanities](#)

- [ICAME Journal](#)
- [International Journal of Corpus Linguistics](#)
- [Journal of Quantitative Linguistics](#)
- [Machine Translation](#)

البيولوجرافية مقتبسة من برنامج جامعة أسكس

www.essex.ac.uk/ البريطانية

Références bibliographiques

AHMAD K. et ROGERS M., 1997, « Corpus-related applications », in *Handbook of terminology management*, vol. 2 , S. E. Wright et G. Budin (dir.), John Benjamins Publishing Company, 1997, pp. 725-760.

BEJOINT H. et THOIRON P. (dir.), 2000b, *Le sens en terminologie*, Presses universitaires de Lyon, 281 p.

BOURIGAULT D. et SLODZIAN M., 1999, « Pour une terminologie textuelle », *Terminologies nouvelles* n°19, pp. 29-32.

BOUTIN-QUESNEL R. et alii, 1985, *Vocabulaire systématique de la terminologie*, Office de la langue française, Québec, 39 p.

CABRE M. T., 1998, *La terminologie. Théorie, méthode et applications*, Presses de l'Université d'Ottawa/Armand Colin, Ottawa/Paris, 322 p.

CANDEL D., 1992, "Rapport sur les besoins exprimés par les utilisateurs virtuels de corpus linguistiques français", Working paper, INALF Paris, *NERC-WP2-31*.

CONDAMINES A. (dir.), 2000, *Cahiers de grammaire* « Sémantique et corpus », Equipe de syntaxe et sémantique, Université de Toulouse Le Mirail.

COWIE A.P., 2001 (1998), *Phraseology. Theory, analysis and applications*, Oxford University Press.

- DAILLE B. et ROMARY L., 2001 (dir.), *Linguistique de corpus, Revue Traitement automatique des langues* vol. 42- n°2.**
- DROZD L., 1981, « Science terminologique : objet et méthode », dans Rondeau et Felber 1981, pp. 115-131.**
- DUBUC R., 1992 (3ème édition), *Manuel pratique de terminologie*, Linguatex/CILF, Montréal/Paris, 102 p.**
- ELRA (2001), *Catalogue*, version 6.2. en 19 p. (exclusion faite du corpus oral).**
- HABERT B., NAZARENKO A. et SALEM A., (1997), *Les linguistiques de corpus*, Armand Colin, Paris, 240 p.**
- International journal of corpus linguistics*, vol.1, 1, 1996, John Benjamins Publishing company, editorial iii**
- KENNEDY G., 1998, *An introduction to Corpus Linguistics*, Longman.**
- KOCOUREK R., 1982, *La langue française de la technique et de la science*, Brandstetter Verlag, 259 p.**
- KOCOUREK R., 1994/1995, « Les textes spécialisés et la terminologie en tant qu'objet de l'analyse linguistique », *Terminologie et linguistique de spécialité, ALFA* vol.7/8, pp. 9-16.**
- LIEBERT W. - A., 1992 (a), "Textual reference corpora : user needs, A report on the Relevant Literature in the Years 1985 to 1992", Working paper, IdS Mannheim, *NERC-WP-4, draft version*.**
- LIEBERT W. - A., 1992 (b), "User needs : survey on IdS corpus users", Working paper, IdS Mannheim, *NERC-WP-5*.**
- McENERY T. et WILSON A., 2001 (1996), *Corpus linguistics, an introduction* , Edinburgh University Press.**
- MEYER I. et MACKINTOSCH K., 1996, The Corpus from a terminographer's viewpoint, *International journal of corpus linguistics*, vol.1, 2, John Benjamins Publishing Company, pp. 257-286.**
- NIEF, C., 1995, *Bases textuelles francophones de spécialité : méthodes d'analyse pour le traitement des corpus textuels de***

spécialité, rapport de stage (AUPELF-UREF, Equipe EST de l'ex-INaLF, Nancy), 149 p. + 22 p. + annexes.

REY A., 1979, *La terminologie : noms et notions*, Presses universitaires de France (Que sais-je ? n°1780), Paris, 127 p.

RONDEAU G. et FELBER H. (dir.), 1981, *Textes choisis de terminologie. Vol.1. Fondements théoriques de la terminologie*, Girstem-Université Laval, Québec, 334 p.

RONDEAU G., 1984 (1981), *Introduction à la terminologie*, Gaëtan Morin éditeur, Chicoutimi, Québec, 238 p.

SAGER J. C., 1990, *A practical course in terminology processing*, John Benjamins Publishing Company, 258 p.

SCHAETZEN (DE) C., 1996, « Corpus et terminologie : constitution de corpus spécialisés pour la confection de dictionnaires », *Revue française de linguistique appliquée* vol.I n°2, Editions De Werelt, Amsterdam, pp. 57-76.

SINCLAIR, J. et J. BALL, juillet 1995, *Eagles Text Typology*, 32 p. typogr. + annexes.

SINGLETON D., 2000, *Language and the Lexicon*, Arnold-Oxford University Press

SLODZIAN M., 2000, « L'émergence d'une terminologie textuelle et le retour du sens », dans Béjoint et Thoiron, pp. 61-85.

ZAMPOLLI, A. (coord.), 1995, N. Calzolari, M. Baker, J. G. Kruyt (ed.), *Towards a Network of European Reference Corpora*, Report of the NERC Consortium feasibility study, Giardini Editori, E Stampatori, Pisa, 234 p. *

*البليوغرافية مقتبسة من

www.irit.fr/assticcot/Groupe.LING/prepLing11.06.02.doc

