الفصل التاسع

نحو إطار عام لمدوَّنة لغوية للمعجم التاريخي للُّغة العربية

عبد المحسن بن عبيد الثبيتي

مقدمة

لعل من نافلــة القول الحديث عــن أهمية المدوّنات اللغوية في دراســة اللغة وبناء المعاجم، حيــث تتبدى أهميتها واضحة جليــة في خدمة المعاجم الحديثة في اللغات الغربيــة، خصوصًا اللغة الإنكليزية؛ فـ **قاموس أكســفورد** بشقيه الحديث والتاريخي خير شــاهد على هذا. والحق أن الإشارة إلى أهمية المدوّنات اللغويــة في بناء المعجــم التاريخي للُّغة العربية ليســت خافيةً عن المهتمين بموضوعه، حيث يورد محمد حســن عبد العزيز[1] جهد أعضاء لجنة المعجم التاريخــي، والنقاشــات والاعتراضات التي دارت في شــأن تصميم مدونة لغوية عربية اقترحها علي القاســمي، تتكون من ألف مليون كلمة عربية من العصر الجاهلي إلى العصر الحديث تحوي خمســة عشر مجالًا موضوعيًا موزعة على خمس فترات زمنية.

لا تخلــو بعض الدراســات الخاصــة بالمعجــم العربي من إشــارات واضحة إلى المدوّنات اللغوية الخاصة بالمعجم التاريخي، انظر على ســبيل المثال عبد العلــي الودغيــري[2] وإبراهيم بن مراد[3]. وعلــى الرغم من هذا الاهتمام بالمدوّنات اللغوية في خدمة المعجــم التاريخي للُّغة العربية – وإن كان نظريًا – فإننــا لا نكاد نجد لها ذكرًا في الدراســات والأبحــاث اللغوية

(1) محمد حسن عبد العزيز، **المعجم التاريخي للغة العربية: وثائق ونماذج** (القاهرة: دار السلام، 2008).

(2) عبد العلي الودغيري، «التأريخ لمعجم اللغة العربية: أســئلة وإشــكالات،» **التاريخ العربي**، العدد 54 (2010).

(3) إبراهيم بن مراد، «قضية المصادر في جمع مادة المعجم،» **مجلة مجمع اللغة العربية بدمشق**، السنة 78، العدد 3 (2001).

الأخـــرى؛ ما يجعلنا بحاجة إلـــى التعريف بالمدوّنـــات وأنواعها والمتطلبات اللازمة لتصميمها قبـــل الحديث عن التصور المقترح للمدوَّنة اللغوية الخاصة بالمعجـــم التاريخي للُّغة العربية؛ ليتســـنى للقارئ الوقوف معنا على أســـباب ودواعي التصميم وكذلك انتقاده.

إن المدوّنات اللغوية بمفهومها الحديث لم تبدأ إلا في نهاية الستينيات من القرن الماضي، حيث جُمعت أول مدونة لغوية محوســـبة وهي مدونة براون[4] (Brown Corpus) نسبة إلى جامعة براون الأميركية التي كانت تحوي مليون كلمة من الإنكليزية الأميركية المعاصـــرة في ذلك الوقت. ثـــم تبعتها مدونات عدة بالحجم والتصميم نفســـيهما بعد ذلك مثل مدوّنة LOB[5] التي جُمعت في عام 1980، والتـــي كانت نتيجة تضافر جهد جامعة لانكســـتر في بريطانيا وكل من جامعة أوســـلو والمركز النرويجي لحوسبة العلوم الإنســـانية في النرويج. ومع تزايد قدرات الحاســـب وإمكانية رقمنـــة النصوص توالـــت المدوّنات اللغوية الغربية وبالأخـــص الإنكليزية منها؛ فظهرت مدوّنـــات معروفة عدة مثل مدونة كوبيلد (COBUILD)[6] التي تحوي ما يقرب من ثلاثة مليارات كلمة استخدمت في بناء معاجم **كوبيلد** المتعددة، والمدوّنـــة اللغوية الوطنية البريطانية (BNC)[7] التي تحوي مئة مليون كلمة، ومدونة أكسفورد اللغوية التي تحوي ملياري كلمة وتستخدم خصيصًا في بناء **معاجم أكسفورد** الشهيرة وكذلك OEC[8].

لا شـــك في أن توالي ظهـــور المدوّنات اللغوية وتعدد اســـتخداماتها في بناء المعاجم ودراســـة اللغة كان نتيجة الفوائد والمنافع التي ظهرت نتيجة ذلك

---

(4) W. N. Francis and H. Kucera, «Brown Corpus Manual: Manual of Information to Accompany A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers,» (Brown University, Providence, 1964).

(5) Stig Johansson, Geoffrey Leech and Helen Goodluck, *Manual of Information to Accompany the Lancaster-oslo/Bergen Corpus of British English, for use with Digital Computers* (Oslo: University of Oslo, Department of English, 1978).

(6) David Crystal, *An Encyclopedic Dictionary of Language and Languages*, Blackwell Reference (Oxford, UK; Cambridge, Mass.: Blackwell, 1992), p. 85.

(7) British National Corpus: <http://www.natcorp.ox.ac.uk/>.

(8) <http://oxforddictionaries.com/words/about-the-oxford-english-corpus>.

الاســتخدام. ولندرة الحديث عن المدوّنات اللغوية باللغة العربية، سوف نورد في المبحثين الأول والثاني من هذه الدراســة مقدمــة مختصرة عن المدوّنات اللغوية وأنواعها لأهمية ذلك للقارئ؛ حتى يســتطيع الوقوف معنا على أسباب اختيارنا التصميم المقترح للمدونة اللغويــة الخاصة بالمعجم العربي التاريخي الذي ســنعرضه لاحقًا. ويعرض المبحــث الثالث بعــض التعريفات للمعجم التاريخي للغة العربية التي من خلالها ســوف نضع الغرض الأســاس للمدونة اللغوية العربية الخاصة بالمعجم. ويعرض المبحث الرابع تصميم ومواصفات مدونة المعجم، وفــي المبحث الخامس نضــع تصورنا لــلأدوات والتقنيات اللازمة للتعامل مع المدوّنة.

أولًا: تعريف المدوّنات اللغوية

تتفاوت التعريفات الخاصة بالمدوّنات اللغوية من كونها أي مجموعة من النصوص إلى كونها مجموعة من البيانات اللغوية المكتوبة أو المنطوقة. ومن أشهر التعريفات ما يراه سو أتكن[9] من أنها نصوص إلكترونية جمعت لغرض معين بنــاء على معايير واضحة. ويــرى جون ســنكلير[10] أن المدوّنة اللغوية مجموعة من نصوص اللغة في صــورة إلكترونية تجمع اعتمــادًا على معايير خارجية؛ لتمثل قدر المســتطاع اللغة أو أحد صورها لتكون مصدرًا للأبحاث اللغوية. ويرى الباحث أن المدوّنات اللغوية ليســت حكــرًا على اللغويين أو لدراســة اللغة من منظور لغوي فحســب، بل إن نمذجة اللغة وحوسبتها أيضًا يُعدان منظورًا آخر للمدونات اللغوية، ولهذا فإن التعريف التالي قد يكون أعم وأشمل من التعريفات السابقة: المدوّنة اللغوية هي نصوص إلكترونية جُمعت لغرض معين بناء على معايير خارجية واضحة؛ لتكون ممثلة لمجال الدراسة.

(9) Sue Atkins, Jeremy Clear and Nicholas Ostler, «Corpus Design Criteria,» *Literary and Linguistic Computing*, vol. 7, no. 1 (1992), pp. 1-16.

(10) J. Sinclair, «Corpus and Text - Basic Principles,» in: Martin Wynne, ed., *Developing Linguistic Corpora: A Guide to Good Practice*, AHDS Guides to Good Practice (Oxford: Oxbow Books, 2005), pp. 1-16.

يحدد التعريف السابق الصيغة التي يجب أن تكون عليها نصوص المدوّنة بأنها إلكترونية، أي إنها بصيغة نصية بسيطة، وقابلة للمعالجة الآلية المباشرة من الحاسب، وهذا يخرج المواد التي لا يمكن معالجتها مباشرة مثل الصور أو الملفات بصيغة PDF. ثم يؤكد التعريف وجود غرض واضح ومحدد من جمع هذه النصوص، قد يكون هذا الغرض محدودًا مثل المدوّنة اللغوية العالمية للعربية[11] (The International Corpus of Arabic) التي تستخدم ضمن مشروع لبناء نظام للتواصل العالمي. وقد يكون الغرض عامًا وشاملًا مثل دراسة اللغة وحوسبتها كما في المدوّنة اللغوية العربية (المدوّنة العربية)[12].

يشدد التعريف أيضًا على وجود معايير واضحة تحدد ماهية النصوص التي سوف تضاف إلى المدوّنة. هذه المعايير يجب أن تكون خارجة عن النص وألفاظه وتراكيبه الداخلية، مثل المنطقة الجغرافية التي تجمع منها النصوص ومصدرها والفترة الزمنية التي ظهرت فيها. كما أن هذه المعايير يجب أن تحقق شرطين رئيسين مترابطين ومتداخلين: التوازن والتمثيل. ويقصد بالتوازن ألا تطغى نصوص ذات طبيعة معينة على باقي نصوص المدوّنة مثل أن تطغى مؤلفات كاتب معين على نصوص المدوّنة، أو يطغى فكر سياسي أو ديني معين ما لم يكن المقصود هو دراسة هذا الكاتب، أو هذا الفكر بالأساس، أو أن هذا هو الواقع خارج المدوّنة. أما التمثيل فهو قدرة المدوّنة، على تمثيل واقع اللغة في مجال الدراسة، أو الغرض الذي من أجله جمعت النصوص. وهذا الشق المهم كان ولا يزال محل دراسة من الباحثين. انظر على سبيل المثال محاولات بايبر لمعرفة تأثير تصميم المدوّنة وطريقة جمع نصوصها في قدرتها على تمثيل اللغة[13]. وكذلك الملاحظات التي أثارها خورشيد أحمد في شأن تمثيل المدوّنة

---

(11) Sameh Alansary, Magdy Nagi and Noha Adly, «Building an International Corpus of Arabic (ICA): Progress of Compilation Stage,» Paper Presented at: The 7th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, 5-6 December 2007

(12) <www.kacstac.org.sa>.

(13) Douglas Biber, «Representativeness in Corpus Design,» *Literary and Linguistic Computing*, vol. 8, no. 4 (1993), pp. 243-257.

اللغوية البريطانية – أشهر المدوّنات اللغوية الإنكليزية – اللغة الإنكليزية[14].

لحساسية موضوع قدرة المدوّنة على تمثيل اللغة محل الدراسة، وعدم الاتفاق في شأنه حتى الآن، يرى بعض الباحثين أن المدوّنة اللغوية لا تمثل اللغة بل تمثل نفسها فحسب، وأن النتائج التي نحصل عليها لا يمكن تعميمها على اللغة ككل بل على المدوّنة اللغوية فحسب. والحقيقة أن محط اهتمامنا في المدوّنات هو الأنماط الشائعة التي تظهر في استخدامات اللغة كما تظهرها المدوّنات اللغوية على المستويات كلها: اللفظية والنحوية والصرفية والقدرة على دراسة اللغة بشكل جديد بفضل الإمكانات التقنية المتاحة حاليًا لمعالجة الأحجام الهائلة من النصوص التي يمكن أن تحويها المدوّنات.

ثانيًا: أنواع المدوّنات اللغوية

يذكر سنكلير[15] خمسة أنواع من المدوّنات: المدوّنات المرجعية والمدوّنات الراصدة والمدوّنات المقارنة والمدوّنات المتوازية والمدوّنات المتخصصة. وسوف نلقي عليها الضوء بصورة مختصرة في ما يلي:

– **المدوّنات المرجعية**: هي المدوّنات التي تُصمم بحيث تعطينا معلومات مفصلة بقدر الإمكان عن استخدامات اللغة ويتحقق هذا من خلال احتوائها على عدد كبير من النصوص، بحيث تظهر صور التنوع اللغوي ذي العلاقة والمفردات المميزة له بشكل واضح. ومن سمات هذه المدوّنات تشارك الصور المختلفة للُّغة في عدد كبير من المفردات والقواعد النحوية. فلو كانت لدينا مدونة مرجعية للغة الصحافة العربية في عصرنا الحاضر فيجب أن تكون بحجم كبير قدر المستطاع، وتحتوي على نصوص من صحافة الدول

(14) Khurshid Ahmad, «Being in Text and Text in Being: Notes on Representative Texts,» in: Gunilla Anderman and Margaret Rogers, eds., *Incorporating Corpora: The Linguist and the Translator*, Translating Europe (Clevedon; Buffalo: Multilingual Matters, 2008), pp. 60–94.

(15) John Sinclair, «Corpus Typology – a Framework for Classification,» in: Gunnel Melchers and Beatrice Warren, eds., *Studies in Anglistics*, Acta Universitatis Stockholmiensis. Stockholm Studies in English; 85 (Stockholm: Almqvist and Wiksell, 1995), pp. 17-33.

العربية كلها، وكذلك من أنواع متعددة مـن الأخبار والمقالات والتقارير التي يمكن أن تحويها الصحافة. بذلك تكون المدوّنة قادرة على إبراز صور التنوع اللغوي بين البلدان العربية وموضوعات الصحافة، وسـوف نلاحظ بلا شـك مشـاركة كبيرة بين الألفاظ والقواعـد النحوية المميزة للغـة الصحافة عمومًا لكننا سـنلاحظ بعض الفروقات بين الصحافة السـعودية والصحافة المغربية على سـبيل المثال لا الحصر (لاحظ مثلًا بنوك وأبناك - مليارات وملايير - عمل وشغل).

- **المدوّنات الراصدة**: مدونات مشـابهة للمدونات المرجعية، لكنها تنقح وتحدث بنصوص جديدة بصفة مسـتمرة؛ لتعطي صورة عن التغيرات التي قد تطرأ على استخدام اللغة ومتغيراتها محل الدراسة، مثل ظهور ألفاظ أو تراكيب جديدة لم تكن معروفة من قبل، أو اضمحلال استخدام بعضها (انظر على سبيل المثال مصطلح الربيع العربي).

- **المدوّنات المقارنة**: مجموعات متشـابهة من النصوص من لغات عدة، مثل التعليقات السياسـية الصحفية في الصحافة العربية والإنكليزية والفرنسية. وقد تكون من لغة واحدة، لكن من وسائل اتصال متنوعة مثل المقارنة بين لغة الصحافة ولغة الإذاعة، أو بين المناهج الدراسـية التونسية والمصرية والقطرية. والقصد من هذه المدوّنات هو دراسـة الفروقات اللغوية بشكل أدق بين صور اللغة المختلفة أو بين اللغات.

- **المدوّنـات المتوازية**: مجموعـة من النصـوص وترجماتها إلى لغات أخرى. وتستخدم عادة في أعمال الترجمة.

- **المدوّنات المتخصصـة**: مجموعة من النصوص ذات طابع محدد، مثل المقالات العلمية في مجال الفيزياء، أو النصـوص القانونية. وتُظهِر المدوّنات المتخصصة تشـابهًا لغويًا مع المدوّنات المرجعية، لكنهـا أيضًا تظهر بأنماط خاصة بها تميز لغتها بوضوح عن لغة المدوّنات المرجعية أو الراصدة.

تضيف سوزان هنستون[16] إلى ما سبق نوعين آخرين:

- **المدوّنات التاريخية**: مجموعة من النصوص تجمع من فترات زمنية مختلفة بقصد دراسة تطور اللغة مثل معرفة الألفاظ التي اندثرت أو استجدت أو تغير أساليب الخطاب.

- **مدونات المتعلمين**: مجموعة من نصوص متعلمي اللغة، مثل كتابات طلاب معاهد تعليم اللغة العربية لغير الناطقين بها، أو تلامذة المدارس المتوسطة. ويُستَخدم هذا النوع من المدوّنات لدراسة الأنماط العامة لقدرة متعلمي اللغة على اكتساب اللغة، ومعرفة المجالات التي بحاجة إلى تطوير وتحسين في تعليم اللغة.

يتطلب بناء المدوّنات اللغوية الخطوات التالية:

• تحديد الغرض الذي لأجله سوف تبنى المدوّنة وماذا نتوقع أن نحصل عليه منها.

• تصميم المدوّنة ووضع مواصفاتها، حيث تُحدَّد المعايير الخارجية التي تُجمع نصوص المدوّنة بناء عليها:

- حجم المدوّنة (عدد الكلمات التي تحويها المدوّنة).

- لغة المدوّنة (عربية، إنكليزية، فرنسية...).

- طبيعة النصوص (مكتوبة أو منطوقة).

- حجم العينة (كامل النص، جزء من أوله، أو من أوسطه، أو آخره وكم عدد الكلمات...).

- تاريخ النصوص (سنة بعينها، أو سنوات متفرقة، أو فترة محددة، أو فترات متعددة...).

Susan Hunston, *Corpora in Applied Linguistics*, Cambridge Applied Linguistics Series (16) (Cambridge: Cambridge University Press, 2002).

– المنطقة الجغرافية التي ظهر فيها النص (دولة معينة، أو مجموعة من الدول، أو منطقة بذاتها داخل الدولة...).

– الوعاء الذي ظهر فيه النص (الصحف، الكتب، النشرات الإخبارية، مواقع الإنترنت...).

– المجال الموضوعي (أخبار، مقالات صحفية، شعر، رواية، علوم طبيعية...).

– المعلومات التي يجب إضافتها للتعريف بالنصوص (اسم الكاتب، المنطقة، دار النشر، تاريخ النشر، وعاء النشر، المجال الموضوعي...).

– تحديد الأجهزة والبرامج التي تحتاجها المدوّنة بناء على حجمها والغرض منها.

– جمع النصوص وحفظها.

– استخدام المدوّنة وأدواتها.

– تنقيح تصميم المدوّنة ومحتوياتها بحسب نتائج الاستخدام.

ثالثًا: تعريف المعجم التاريخي

يتطلب وضع تصميم ملائم للمدونة اللغوية للمعجم التاريخي للُّغة العربية معرفة الغرض الأساس للمعجم وما هي حدود متطلباته، وما الذي يراد منه، وما سوف يضمه بين دفتيه. ولبيان هذا المطلب نعرض في ما يلي ثلاثة تعريفات للمعجم التاريخي للُّغة العربية أوردها بعض أهل الاختصاص، ثم نستخلص منها بعض النقاط المهمة التي تساعدنا في وضع التصور العام للمدونة اللغوية للمعجم التاريخي للُّغة العربية.

يعرّف علي القاسمي المعجم التاريخي بأنه «نوع من المعاجم يرمي إلى تزويد القارئ بمعلومات عن أصل الألفاظ وتاريخها ومعانيها من خلال تتبع

تطورها منذ أقدم ظهور مسجّل لها حتى يومنا هذا. وذلك يعني أمرين:

الأول، أن يضمّ المعجم التاريخي كل لفظ استُعمِل في اللغة، سواء يُستعمل في الوقت الحاضر أم لا.

الثاني، أن يوثّق المعجمُ تاريخَ كل لفظ في شكله ومعناه واستعماله ممثِّلًا لهذا اللفظ بعدد من الشواهد»[17].

يعرف إحسان النص المعجم التاريخي بقوله: «والمعجم التاريخي للغة العربية هو المعجم اللغوي العام الشامل الذي يجمع أشتات الوحدات المعجمية العربية – ما دُوّن منها في المعاجم وما لم يدوَّن – وأن يؤرخ لظهورها في الاستعمال وما طرأ على دلالاتها من التطور بحسب ما توفره النصوص. فإن النصوص هي مصادر التأريخ، لأن التأريخ لوحدات المعجم ليس تأريخًا لأول ظهورٍ لها في اللغة عامة، بل هو تأريخ لأول ظهور لها في نص مكتوبٍ، قد يكون نقيشة وقد يكون صحيفة، وقد تكون الصحيفة مطبوعة وقد تكون مخطوطة، وتعتمد النصوص المتوفرة بمختلف أجناسها ومختلف المعارِف التي تمثّلها، ومختلف العصور والأمصار التي كتبت فيها»[18].

يرى محمد حسن عبد العزيز أن المعجم التاريخي للغة العربية ديوان «يضم بين دفتيه مفرداتها وأساليبها: مبانيها ومعانيها، ما استخدم منها وما أميت أو هجر، ما حدث لها من تغير عبر الأزمان والأصقاع، بل سيكون كذلك ديوانًا لتاريخ العرب والمسلمين..»[19].

من التعريفات السابقة نستخلص أن المعجم التاريخي للُّغة العربية هو «التوثيق الجغرافي والتاريخي لمباني ومعاني الألفاظ العربية طوال فترة

---

(17) علي القاسمي، «المعجم التاريخيّ للغة العربية هل نستطيع أن ننجزه بعد مائة عام؟،» (عرب 48 (موقع إلكتروني)، 19/ 4/ 2006): <http://www.arabs48.com/?mod=articles&ID=36212>.

(18) إحسان النص، «مشروع المعجم التاريخي للغة العربية: مسيرة وتاريخ،» **مجلة مجمع اللغة العربية بدمشق**، السنة 82، العدد 1 (2007).

(19) عبد العزيز، **المعجم التاريخي للغة العربية**.

استخدامها مدعما بالشواهد». وباختصار فالمعجم التاريخي هو «السيرة الذاتية للألفاظ»[20].

حتى يتحقق لنا الغرض من المدوّنة اللغوية في بناء المعجم التاريخي للُّغة العربية؛ نحن بحاجة إلى جمع كل ما يمكن جمعه من نصوص كتبت بالعربية، وتزداد الحاجة إلى هذا كلما كان تاريخ النص أقدم. وفي القسم التالي نضع الإطار العام للمدونة اعتمادًا على ما سبق ذكره في نهاية المبحث الثالث.

رابعًا: تصميم مقترح للمدونة العربية

سبق الإيضاح أن تصميم المدوّنة اللغوية يعتمد على تعريفنا للمعجم التاريخي الذي ذُكر في المبحث السابق. وفي ما يلي نتبع الخطوات التي ذكرناها سابقًا في متطلبات بناء المدوّنات في المبحث الثالث من هذه الدراسة:

- **الغرض من بناء المدوّنة**: المساعدة في بناء معجم تاريخي للُّغة العربية يوثقُ جغرافيًا وتاريخيًا مباني ومعاني الألفاظ العربية طوال فترة استخدامها مدعمًا بالشواهد آخذين في الاعتبار اختلاف الدلالات باختلاف المجالات والسياقات التي قد تظهر فيها الألفاظ.

- **معايير جمع النصوص**:

- حجم المدوّنة: 2,000,000,000 (ألفا مليون) كلمة.
- لغة المدوّنة: اللغة العربية الفصحى.
- طبيعة النصوص: النصوص المكتوبة مشكولة ما أمكن.
- حجم العينة: كامل النص متى ما توافر.
- تاريخ النصوص: من العصر الجاهلي حتى العصر الحالي.

(20) «للألفاظ سير ذاتية كالناس، ولها حظوظ متفاوتة في الحياة، فهي تولد وتنمو وتبلغ مرحلة النضج، ومنها ما يعمّر، ومنها ما يصيبه الهرم فيموت، وربما تبعث بعد موتها». انظر: عبد الرزاق الصاعدي، في: **المدينة** (ملحق الرسالة)، 14/ 9/ 2012.

• المنطقة الجغرافية التي ظهر فيها النص: جميع البلدان والمناطق حتى إن لم تكن عربية.

• الوعاء الــذي ظهر فيه النص: تســعة أوعيــة مقترحــة: المخطوطات المحققة والكتب والصحف والمجلات والدوريات العلمية والرسائل الجامعية والإصدارات الرســمية ومواقع الإنترنت والتواصل الاجتماعي (ما كان سليم اللغة مخصصًا للإنترنت فحســب). بطبيعة الحال فإن هذه القائمة محل تعديل ونقاش وهي تتدرج من أسمى استخدامات اللغة كما في المخطوطات المحققة والكتب إلى ما هو أدنى من ذلك لكنه يكفل الســلامة اللغوية في حدها الأدنى مثل الإنترنت ومواقع التواصل الاجتماعي.

• المجال الموضوعي: لــكل وعاء من الأوعية الســابقة المجالات التي توافقه، ويندرج تحت كل مجال منها موضوعات متخصصة تناسبه. فعلى سبيل المثال فإن مجال الفقه الإســلامي هو أحد مجــالات المخطوطات المحققة، ويمكن أن يندرج تحته موضوعات عدة مثل: الفقه المالكي والشافعي والحنفي والحنبلي والإثني عشــري والزيدي. ويجب أن نراعي هنــا أننا مهتمون باللغة والألفاظ وليــس الأفكار والمعتقدات. وأن التباين والتنــوع في ما بين الأفكار والموضوعات يثري بلا شك حصيلة المدوّنة اللغوية. الجدول (9–1) يوضح الأوعية والمجالات الموافقة له تحت كل وعاء.

• المعلومات التــي تجب إضافتها إلى التعريــف بالنصوص: المعلومات التالية هي الحد الأدنى المطلوب للتعريف بالنص:

- عنوان النص: ما وضعه المؤلف كعنوان للنص أو ما اشتهر به.
- اسم المؤلف كاملًا قدر الإمكان.
- جنسيته وبلده الأصلي بالنسبة إلى القدماء من المؤلفين.
- جنسه ذكر أم أنثى.
- الوعاء الذي ينتمي إليه النص من أحد الأوعية المذكورة سابقًا.

- المجال الذي ينتمي إليه النص.

- موضوع النص المتخصص.

- تاريخ صدور النص الذي يمكن تحديده بدقة للنصوص الحديثة، أما بالنسبة إلى نصوص المخطوطات القديمة المحققة فيمكن استخدام متوسط زمن حياة المؤلف تاريخًا لصدور النص[21].

- مكان الإصدار والنشر وهذا موجود في النصوص الحديثة لكنه صعب بالنسبة إلى المخطوطات؛ لذا يكتفى بذكر البلد التي يعتقد أن المخطوط كُتب فيه.

- مصدر النص مثل اسم الصحيفة أو المجلة أو عنوان الموقع أو اسم الناشر.

قبل الانتهاء من التصميم المقترح والانتقال إلى الأدوات والوظائف الحاسوبية للمدونة تبقى الإجابة عن الأسئلة التالية: لماذا اختير حجم المدوّنة بهذا العدد من الكلمات؟ هل ستكون المدوّنة مفتوحة أم مغلقة؟ كيف سيوزَّع عدد الكلمات على الفترات الزمنية وعلى الأوعية والموضوعات؟

للحق فإنه من الصعب جدًا تحديد عدد الكلمات الملائم والكافي لبناء المعجم التاريخي، لكن القاعدة المهمة في أي مدونة لغوية هي أنه كلما ازداد حجم المدوّنة وتنوعّت مصادرها وموضوعاتها كان ذلك أفضل. والمعجم التاريخي للُّغة العربية يسعى إلى جمع ما يمكن جمعه من ألفاظ العربية وكل دلالة أو إشارة لمبنى أو معنى اللفظة سيكون ذا فائدة. ومهما كان الأمر فإن تحديد حجم المدوّنة مطلوب وضروري جدًا لنعرف أين سنقف، وما يجب جمعه من كل عصر ومن كل مجال. ومن دون معرفة ذلك لن نستطيع أن نخطط لجمع النصوص المطلوبة. إن جمع ألف مليون كلمة أصبح سهلًا وميسورًا

(21) في إحصائية أجريتها بنفسي على مدونة تاريخية عربية من عصر النبوَّة إلى القرن الثامن عشر الميلادي كان متوسط عمر المؤلف سبعين عامًا وبذلك يكون منتصف عمره هو 35 عامًا ويكون تاريخ التأليف هو عام ولادته زائدًا 35.

جدًا، وبالتالي فإن جمع الألفي مليون كلمــة قد يكون فيه بعض التحدي، لكنه ممكن التحقيق في فترة وجيزة. وبعد جمع العــدد المطلوب يعاد النظر في ما أمكن وما لــم يمكن، وهل نحن بحاجــة إلى الزيادة أم لا؟ ومــا هي الفترات والمجالات التي نحن بحاجة إلى التركيز عليهــا وإضافة نصوص أخرى إليها متى توافرت؟ وما الفتــرات أو المجالات التي يكتفى بمــا جُمع من نصوص فيها؟ إن عملية المراجعة هذه عملية مهمة، ويجب أن تجري باســتمرار بحيث تشمل كامل التصميم، ولا يتأتى ذلك من دون العمل على النصوص والأدوات المتوافرة، ومعرفة مــا ينقص التصميم أو ما جُمع مــن نصوص. وهذا يتطلب أن تكون المدوّنة مغلقة في البداية، ثــم تكون مفتوحة قابلة للإضافة بعد إنجاز النسخة الأولى من المعجم.

الجدول (9-1)
التوزيع المقترح للمجالات ومواضيعها
على أوعية المدوّنة اللغوية للمعجم التاريخي للُّغة العربية

المخطوطات المحققة	
المجال	**الموضوع**
أصول الفقه	المالكية والشافعية والأحناف والحنابلة والزيدية والاثنا عشرية والإباضية والظاهرية
الفقه الإسلامي	المالكية والشافعية والأحناف والحنابلة والزيدية والاثنا عشرية والإباضية والظاهرية
المعتقدات الإسلامية	السُنّة والشيعة والزيدية والإباضية
المعتقدات غير الإسلامية	المسيحية واليهودية والصابئة
الحديث	السُنّة والشيعة والزيدية والإباضية
علوم الحديث	السُنّة والشيعة والزيدية والإباضية
التفسير	السُنّة والشيعة والزيدية والإباضية

يتبع

تابع

علوم القران	السُنّة والشيعة والزيدية والإباضية
السياسة والقضاء	السياسة الشرعية والقضاء والسياسة
الأخلاق	الزهد والرقائق
السِيَـر والتراجم	السيرة النبوية والتراجم
علوم اللغة العربية	النحو والصرف والبلاغة والمعاجم والشعر والإنشاء
العلوم الاجتماعية	التاريخ والجغرافيا والأنساب
العلوم الطبيعية	الفيزياء والكيمياء والرياضيات والفلك والطب
الثقافة والآداب	الشعر والرحلات والأدب
الفلسفة	الفلسفة الإسلامية والفلسفة اليونانية
الكتب والرسائل الجامعية والدوريات المحكمة	
المجال	**الموضوع**
أصول الفقه الإسلامي	المالكية والشافعية والأحناف والحنابلة والزيدية والاثنا عشرية والإباضية والظاهرية
الفقه الإسلامي	المالكية والشافعية والأحناف والحنابلة والزيدية والاثنا عشرية والإباضية والظاهرية
المعتقدات الإسلامية	السُنّة والشيعة والزيدية والإباضية
المعتقدات غير الإسلامية	المسيحية واليهودية والصابئة
الحديث	السُنّة والشيعة والزيدية والإباضية
علوم الحديث	السُنّة والشيعة والزيدية والإباضية
التفسير	السُنّة والشيعة والزيدية والإباضية
علوم القران	السُنّة والشيعة والزيدية والإباضية
السِيَـر والتراجم	السيرة النبوية والتراجم والسير الذاتية

يتبع

تابع

علوم اللغة العربية	النحو والصرف والبلاغة والمعاجم والشعر والإنشاء واللسانيات الحديثة
العلوم الاجتماعية	التاريخ والجغرافيا والأنساب والاقتصاد وعلم النفس وعلم الاجتماع والإعلام والإدارة والسياسة والقانون والتعليم
العلوم الطبيعية	الفيزياء والكيمياء والرياضيات والأحياء وعلم الأرض والفلك والطب والبيئة
العلوم التطبيقية	الهندسة المدنية والهندسة الميكانيكية والهندسة الكهربائية والهندسة الكيميائية وهندسة الطيران والهندسة الصناعية والهندسة النووية والهندسة الطبية وتقنية المعلومات وتقنية الاتصالات
الثقافة والآدب	الشعر والنقد الأدبي والقصة والقصة القصيرة والرواية والمسرح والسينما والفنون التشكيلية
الفلسفة	الفلسفة الإسلامية وفلسفة العلوم وفلسفة العقل ونظرية المعرفة وفلسفة اللغة
القوانين والأنظمة	السياسية الاقتصادية والاجتماعية والأمنية والتعليمية والتجارية

الجدول (9-2)

التوزيع المقترح للمجالات ومواضيعها على أوعية المدوّنة اللغوية للمعجم التاريخي للُّغة العربية

الصحف	
المجال	**الموضوع**
الأخبار	الاجتماعية والرياضية والاقتصادية والتقنية والثقافية والسياسية والعامة والعلمية
المقالات	الرأي والافتتاحيات والاجتماعية والرياضية والاقتصادية والثقافية والسياسية والعلمية والتقنية والعامة والهوايات والدينية
التحقيقات	الاجتماعية والرياضية والاقتصادية والثقافية والسياسية والعلمية والتقنية والعامة والهوايات والدينية

يتبع

تابع

المجلات	
المجال	**الموضوع**
الاخبار	الاجتماعية والرياضية والاقتصادية والتقنية والثقافية والسياسية والعامة والعلمية
المقالات	الرأي والافتتاحيات والاجتماعية والرياضية والاقتصادية والثقافية والسياسية والعلمية والتقنية والعامة والهوايات والدينية
التحقيقات	الاجتماعية والرياضية والاقتصادية والثقافية والسياسية والعلمية والتقنية والعامة والهوايات والدينية والأمنية
الدراسات	علوم اللغة والعلوم الاجتماعية والدينية والأدب والثقافة والاقتصادية والعلوم الطبيعية والعلوم التطبيقية والقانونية والفلسفية الأمنية والعسكرية
الإصدارات الرسمية	
المجال	**الموضوع**
القوانين والأنظمة	السياسية والاقتصادية والاجتماعية والإدارية والصحية والأمنية والتعليمية والتجارية
التقارير	السياسية والاقتصادية والاجتماعية والإدارية والصحية والأمنية والتعليمية والتجارية
الإنترنت والتواصل الاجتماعي	
المجال	**الموضوع**
مواقع الإنترنت	الحكومية والاجتماعية والرياضية والاقتصادية والتقنية والثقافية والسياسية والعلمية والطبية والدينية
تويتر وفيس بوك	الحسابات الشخصية لأصحاب الفكر والرأي

إن حجم المدوّنة المقترح يجب أن يُـوزَّع على الفترات الزمنية أولًا بما يلائم النشـاط الفكري والنشـر في كل زمن. والحق أن المتاح من النصوص القديمة بصيغـة قابلة للمعالجة الآلية مباشـرة نادر مقارنة بمـا هو متاح في العصـر الحديـث، وإن المتوافر من ذلك أيضًـا تقل فيه النصـوص العلمية التطبيقية. لذلـك فإن أي نص يمكن جمعه وإضافته إلـى المدونة من العصر الجاهلـي حتى نهايـة القرن الثامن عشـر الميلادي مرحب بـه. ومن تجربة للباحث في جمع نصوص عربية لهذه الفترة استطاع جمع 350 مليون كلمة، أكثر مـن 80 مليونًا منها كانت مشـكولة. ومن المتوقع جـدًا قدرة المعجم التاريخي – كونه مؤسسة لا فردًا – على جمع 500 مليون كلمة للفترة نفسها بكل سـهولة. والجدول (9–2) يوضـح التوزيع المقتـرح لكلمات المدوّنة على الفتـرات الزمنية. وهو تقدير لا يسـتند إلا إلى حقيقـة واحدة وهي أن المعرفة الإنسانية وتوافر المعلومة يتطور بشـكل متسارع، ويأخذ هذا التطور والازدياد منحنى أُسـيًا مع مرور الوقت. يضاف إلى ذلك خبرات الباحث في العمل على المدوّنات.

نالت قضية توزيـع الفترات الزمنية لمصادر ومدونـة المعجم التاريخي للُّغة العربية نقاشًـا مسـتفيضًا تباينت فيـه وجهات النظـر، وكان رأي لجنة المعجـم التاريخي[22] تقسـيم الفتـرات الزمنية إلى خمس فتـرات: العصر الجاهلـي والعصر الإسـلامي والعصر العباسـي والعصر الوسـيط والعصر الحديث. وما أراه ملائمًا هو أن تقسـم الفترات السـابقة إلـى فترات أقصر طول كل فترة منها مئة عام تبدأ من عام 600 ميلادية، وأن تكون التقسـيمات الخمس المقترحة من لجنة المعجم التاريخي أو غيرها من التقسـيمات مظلةً أكبر لهذه التقسـيمات مع مراعاة التالي: أن يكون مـا قبل 600 ميلادية فترة واحدة، وأن يقسم القرنان التاسع عشر والعشرين إلى أربعة فترات، طول كل فترة منها 50 عامًا؛ والسـبب في ذلك أنها فترات حراك ثقافي وسياسي كبير من المهم رصده بدقة أكبر.

(22) عبد العزيز، **المعجم التاريخي للغة العربية**.

الجدول (9-3)
توزيع محتوى المدوّنة على الفترات الزمنية والأوعية

الفترة الزمنية	الأوعية	عدد الكلمات
من عصر ما قبل الإسلام وحتى نهاية القرن الثامن عشر تقسم إلى فترات طول كل منها مئة عام	المخطوطات المحققة	500 مليون
القرن التاسع عشر	الكتب، والصحف، والمجلات	100 مليون
القرن العشرون	الكتب، والصحف، والمجلات، والرسائل الجامعية والدوريات المحكمة، والإصدارات الرسمية	400 مليون
القرن الحادي والعشرون	الكتب، والصحف، والمجلات، والرسائل الجامعية، والدوريات المحكمة، والإصدارات الرسمية، والإنترنت ومواقع التواصل الاجتماعي.	1000 مليون

بعد الاتفاق على تصميم المدوّنة وجمع النصوص هي بحاجة إلى أجهزة وبرامج حاسوبية للتعامل معها، والغوص في الكم الهائل من المعلومات المتوافرة، وما يمكن أن يستخرج منها من معارف.

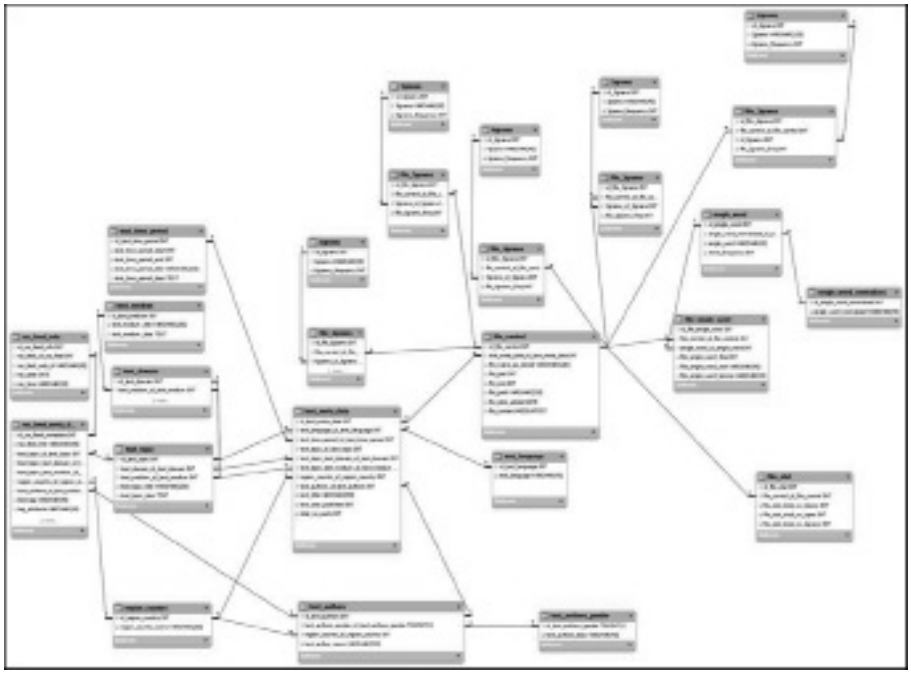
خامسًا: الوظائف الحاسوبية

تحتاج أي مدونة لغوية إلى برامج خاصة للتعامل مع النصوص، واستخراج ما يتطلبه مجال الدراسة من معلومات ومعارف. وأول ما تحتاج إليه المدوّنة اللغوية للمعجم التاريخي للُّغة العربية هو جهاز خادم ملائم لها، ومن الصعب تحديد هذا بالدقة المطلوبة، لكنه خادم حاسوبي يلائم الأعمال المتوسطة بحسب العرف الحاسوبي مع قدرات عالية في التخزين. وقد يكون توفير الأجهزة أسهل الأمور وأيسرها.

تأتي بعد ذلك كيفية حفظ النصوص والتعامل مع المعلومات المستخلصة منها. ومن الأساليب الشائعة استخدام واصفات XML، وأن تكون الرسوم ومعلومات النص محفوظة في داخل النص. وقد تكون هذه الطريقة مقبولة في المدوّنات الصغيرة أو التي لا تتعدى المئة مليون كلمة، لكن في حالة المدوّنات الكبرى مثل مدونة المعجم التاريخي فنحن بحاجة إلى تقنية أخرى. وأثبتت تقنيات قواعد البيانات العلائقية كفاءة أعلى في التعامل مع المدوّنات الكبيرة. بحيث تُحفظ النصوص ومعلوماتها وما يستخرج منها من معلومات في جداول منفصلة لكنها مترابطة، ويمكن الوصول من خلالها إلى المعلومات المطلوبة بسرعه عالية. الشكل (9-1) يوضح التصميم المقترح لقاعدة بيانات المدوّنة اللغوية للمعجم التاريخي للُّغة العربية.

الشكل (9-1)

تصميم لقاعدة بيانات علائقية لمدونة المعجم التاريخي للُّغة العربية



تحتاج المدوّنة إلى الأدوات الرئيسة التالية للمساعدة في بناء المعجم:

• سـرد الألفاظ في المدوّنة وتكرارها وتوزيعها الإحصائي على المناطق الجغرافية وعلى الفترات الزمنية والأوعية والمجالات والموضوعات.

• إمكانية البحث بلفظة أو بجزء منها، وسـرد الألفاظ كلها التي ورد فيها الجزء المراد البحث عنه فيها وإمكانية الوصول إلى جميع المعلومات المتعلقة بها.

• إيراد جميع السـياقات التي وردت فيها لفظة معينـة مع إيضاح مصدر السياق وتاريخه ومكانه.

• إيضاح الألفاظ الأقوى ارتباطًا باللفظة محل الدراسة.

• ربط الألفاظ بالجذر والساق آليًا ويدويًا.

• ومـا ينطبق أعلاه على الألفـاظ المفردة فهو أيضًا مطلـوب للمتابعات اللفظية، وما يرتبط بها من معلومات.

• أن يتمكن العاملون على المعجم مـن الوصول إلى المدوّنة ومصادرها عن طريق الإنترنت.

إن كل ما ذكر أعلاه من أدوات سـهل وميسـور جدًا، ويمكن إنجازه في فترة قصيرة ما عدا قضيتي الاكتشاف الآلي للجذع والجذر بدقة عالية، والوسم النحـوي الآلي للألفاظ العربية، فـلا يزال أمام ذلك وقـت طويل، وقد يكون المعجم التاريخي فاتحة خير لمثل هذا الجهد.

خاتمة

إن العمل على المدوّنات اللغوية يتطلب إعادة النظر فيها باسـتمرار حتى يسـتقيم بناؤها ومحتواها مع الغرض الذي وضعت لأجله. ومن خبرات سابقة لي ولغيري ممن عمل في هذا المجـال فإن ما يتبدى بعد الانتهاء من العمل أو في أثنائه يجعل إعادة النظر في التصميم أو بعض مكوناته أمرًا محتملًا. فما تبينه حقائق الواقع قد يختلف بعضه أو كله عما يظنه القائم بالعمل قبل مباشرته. وما

أظنه ملائمًا بالنسبة إلى مدونة المعجم التاريخي للُّغة العربية هو المضي قُدُمًا في أي تصميم يراه الأغلب ملائمًا، ثم يُطوَّر ويُعدَّل بعد العمل عليه والنظر إلى مخرجاته.

مهما اجتهد المجتهدون في هذا قبل العمل الفعلي فسيكون عملهم عرضة للانتقاد؛ بسبب اختلاف وجهات النظر في شأن المدوّنات، والمعجم التاريخي نفسه وما يجب أن يحويه. وبالإمكان التدرب ووضع الخطط لبناء المعجم التاريخي على مدونة صغيرة نسبيًا قد تتكون من مئة مليون كلمة. وعند جمع النصوص يجب التفكير أولًا في جمع ما توافر في الإنترنت وهو كثير، ثم الانتقال إلى مخاطبة دور النشر، والجهات المهتمة بالنشر، مثل الهيئات الثقافية، والجامعات ثم استخدام برامج التعرف الضوئي.

من المهم جدًا أن يكون المعجم منذ بداياته متاحًا على الإنترنت حتى وإن كان العمل قاصرًا لم يكتمل، وأن ندع الفرصة لكل محب للغة أن يبدي رأيه في ما يتم التوصل إليه. وسيكون من المفيد جدًا إتاحة المعجم في صورة رقمية على الإنترنت أو على أسطوانات ضوئية، ولا أعني بذلك أن نضع صورته الورقية بل أن يكون معجمًا بمستويات عدة بحسب اهتمامات الباحث، وأن يكون ثريًا بالوسائط المتعددة، وهذا ممكن ومتاح وليس صعبًا.