الحاسوب وصناعة المعجم

أ.د. محمود فهمي حجازي
عضو المجمع – مصر

أصبح الحاسوب أو الحاسب الآلي من أهم وسائل صناعة المعاجم الحديثة. تتعامل به المؤسسات البحثية اللغوية ودور النشر الكبرى المتخصصة في مراحل كثيرة من صناعة المعجم وتتيحه لملايين المستفيدين وهذا العرض رؤية لغوية تطبيقية.

أولاً: التصنيف الكمي

1. تنوعت المعاجم في القرن العشرين تنوعّاً كبيراً استجابةً لمتطلبات ملايين المستفيدين من فئات المثقفين والطلاب والتلاميذ والأطفال، وزادت أعدادها بانتشار التعليم في العالم كله وبدرجات متفاوتة. ولم يعد المعجم مرجعاً أحادي الهدف لبحث معاني كلمات صعبة أو نادرة بل تعددت أهداف كل معجم وعليه تحقيق هذه المتطلبات بشكل دقيق ومركز ومفيد. ولهذا كله أصبحت المعاجم الحديثة للغة مثل الألمانية أو الفرنسية بالمئات تتعدد أنواعها وتتفاوت أحجامها وتتجدد طبعاتها وتتنوع فئات المستفيدين منها.
2. توصف المعاجم طبقاً لعدد المداخل، نجد ذلك على سبيل المثال على غلاف المعجم أو في مقدمته أو في الكتيبات التعريفية به أو في الإعلان عنه. هناك معاجم للأطفال من عدة مستويات ومعاجم مدرسية للتلاميذ ومعاجم الطلاب وجمهور المثقفين ومعاجم متنوعة ومعاجم شاملة ومعاجم لمراحل زمنية في تاريخ اللغة. كما تتنوع إلى معاجم أحادية وأخرى ثنائية أو متعددة اللغات.
3. زيادة حجم المادة في ضوء إمكانات قاعدة البيانات أصبحت من أهم التحولات، لم يعد الاقتصار على آلاف الكلمات من مئات أو عشرات النصوص هو المتاح، بل أصبحت قاعدة

البيانات تحسب مادتها بملايين الكلمات الواردة في نصوص آلاف الكتب والمطبوعات، ثم يجري الاختيار منها طبقاً لأهداف المعجم وبمعايير واضحة.

4. تعددت مجالات الإفادة من الحاسوب في صناعة المعجم في أكثر مراحله من جمع المادة من النصوص إلى المعاونة في الإعداد إلى الإتاحة للمستفيدين. زاد دور الحاسوب وقل العمل اليدوي بنظام البطاقات.

ثانياً: المداخل المعجمية

1. أصبحت المعاجم الحديثة لا تقتصر على التركيز على معاني الكلمات الصعبة والألفاظ البائدة. والمداخل المعجمية ترتبط بهدف المعجم وحجمه والمستفيدين منه. لم يعد الانطباع الفردي معياراً كافياً لإدخال مدخل ما في المعجم أو عدم إدخاله. لاشك أن الخبرة لها دورها ولكن المعايير العلمية والإفادة من الحاسوب صارت من أسس تحديد المداخل بأنواعها. وتنص مئات المعاجم التي أعدت في أواخر القرن العشرين وبدايات القرن الحادي والعشرين على استعمال الحاسوب.
2. المكونت اللغوية لا تقتصر على بيان الدلالة أو الدلالات بل تتضمن أيضاً معلومات لغوية، ومنها النطق، والتنوعات الإقليمية، والكتابة الصحيحة وصيغ التصريف. ولكن هناك جوانب تتطلب اهتماماً خاصاً وإلا تفقد جدواها، وهي التراكيب السياقية. لا يكفي أن يرشدنا الحاسوب إلى كلمة رغب. بل لابد من ذكر علاقاتها بحرف الجر. (رغب في) لها دلالة مختلفة عن (رغب عن) وعن (رغب إلى).
3. هناك مداخل لها دلالاتها اللغوية ويتطلب التدقيق فيها مراعاة مجالات تخصصها واستعمالها الثقافي. لا يكفي في المادة قهر أن نكتفي بدلالة الفعل، ولابد من ذكر القاهرة. وفي الوقت نفسه هناك كلمات يقتصر استعمالها على المغرب العربي مثل مخبزة وهناك يشار إلى ذلك إذا كان هدف المعجم يتضمن أمثال هذه الكلمات. هذه تُجْمع من مطبوعات مناسبة ومن اللغة المنطوقة ثم تشرح لتتخذ أماكنها في المعجم طبقاً لنظامه وهدفه.

4. المداخل المعجمية ليست مجرد تراكم كلمات وتراكيب ومعلومات، بل ترتبط في جمعها بمادة لغوية محددة وتناسب هدف المعجم وتسدّ حاجة المستفيدين منه. وتقوم على جهد بشري للغويين ومتخصصين إلى جانب الإفادة في التخزين والترتيب والإتاحة من الحاسوب.
5. المداخل الموسوعية من أهم مكونات المعجم ولا يجوز الاعتماد فيها على الانطباع، ومنها أسماء الدول والعواصم والمدن الكبرى والحوادث المهمة والشخصيات التي لها دور في التاريخ أو الثقافة أو العلم. لابد من تَعَرُّفِ هذه المداخل طبقاً لهدف المعجم ثم تشرح وتعرّف لتدخل الحاسوب تمهيداً للإفادة منها.

ثالثاً: قاعدة البيانات اللغوية

1. زاد حجم المدونة لقواعد البيانات اللغوية من سنة 1960م حتى نهاية القرن العشرين من مليون كلمة إلى مئتي مليون كلمة او أكثر.
2. مفهوم المدونة Corpus في هذا السياق: مجموع النصوص اللغوية التي تختار لتدخل ذاكرة الحاسوب لأغراض البحث اللغوي من حيث الصيغة أو التركيب أو الدلالة. ويكون الاختيار اعتماداً على المستوى اللغوي والحدود المكانية والزمنية وحجم المادة المنشودة.
3. من أمثلة المدونات:

آ. مدونة جامعة براون للإنكليزية والأمريكية

ب. مدونة بنك اللغة الإنكليزية

ج. مدونة Cobuid

د. مدونة لانكستر أوسلو/برجين للغة الإنكليزية

هـ. مدونة لونغمان، وصدرت عنها معاجم كثيرة من أنواع شتى.

وهناك مدونات في معهد اللغة الألمانية في مانهايم في ألمانيا ومراكز البحوث اللغوية في دول أوربية كثيرة.

4. المعايير اللغوية لخطة المدونة:

آ. حجم المدونة

ب. المستوى اللغوي/المستويات اللغوية

ج. مصادر المادة

د. المجالات الدلالية العامة

5. هناك جهود كبيرة بذلت لعمل برامج حاسوبية لتيسير التحليل الصرفي والسياقي والدلالي. وبهذا كله أُنْتجت معاجم للغات أوربية، تتفاوت من حيث مدى الاعتماد على العمل اليدوي الإنساني ومدى الإفادة من الحاسوب.

رابعاً: مرصد البيانات والتراكيب السياقية والدلالية

1. كان الحاسوب في بدايات عمله يقدم الكلمات مفردة، ولكن التطور جعل مراصد البيانات تتيح الكلمات الواحدة في سياقات استعمالها. ومن هنا يكون بحث دلالتها أو دلالاتها ممكناً على نحو أدق. وفي معاجم لونغمان للإنكليزية ومنها المعجم المدرسي المتداول في مصر نجد على سبيل المثال كلمة go تشغل ثلاثة أعمدة (260 + 261 + 262) وذلك لأن دلالاتها الأساسية الكثيرة مع عبارات أو جمل استعمالها يأتي بعدها عشرات التراكيب السياقية، مثل: go after، go along، go back و go by وعددها يتجاوز العشرين ولكل منها دلالة تتضح من سياق الجملة ويشرحها المعجم. والعربية فيها تراكيب سياقية كثيرة مع الكلمات الأساسية. وبذلك تتعدد الدلالات.
2. المادة الكبيرة في مرصد البيانات اللغوية تتيح لِمُعِدِّ المعجم تَعرُّفَ علاقات دلالية، ومن المفيد أن توضع في المعجم. وفي معجم لونغمان المدرسي نجد مع بعض الكلمات علاقة التضاد (opposite) في آخر المداخل، كما نجد إحالات المقارنة مع كلمات أخرى (compare) وإحالات أخرى بعد (see).
3. تتيح قاعدة البيانات الكلمة في سياق أطول، أي في تتابع من عدة كلمات قبلها وبعدها. وهذا ما يتيح تعرُّف جوانب متعددة من استعمال الكلمة.

خامساً: بنوك المصطلحات والمعاجم المتخصصة

1. تكونت فكرة بنوك المصطلحات في موقعين نحو عام 1970م، وذلك في إطار التقدم المطرد لتطبيقات الحاسوب وتقنياته آنذاك. أولهما في أوربا مع تقدم مراحل التعاون والتكامل بين دولها ذات اللغات الكثيرة، كانت الحاجة واضحة إلى الدقة في المقابلات باللغات الأوربية المتعددة حرصاً على التعامل بكفاءة ووضوح وسرعة. وظهرت الفكرة أيضاً في إقليم كوبيك الفرنسي اللغة في كندا، حرصاً على الدقة والكفاءة والسرعة في التعامل بين الفرنسية والإنكليزية هناك وعلى استمرار العلاقة اللغوية مع فرنسا.
2. التصنيف الموضوعي للمجالات هو أساس وضع كل مصطلح في مكانه. الترتيب ليس هجائياً والعمل يراعي المجالات الثقافية والمعرفية والإدارية والعلمية. يضم كل مجال كبير عدة مجالات فرعية وفي كل مجال فرعي منظومة من المفاهيم تنتظم بداخلها مصطلحاتها. وبذلك لا يحدث غموض دلالي بأن يستعمل المصطلح لعدة مفاهيم علمية مختلفة. وهذه الفكرة لها أصولها في التراث العربي، على نحو ما نجد في المخصص لابن سيده، وفي كتب لغوية أخرى. ولها بداياتها في معجم روجيه الإنكليزي. في بنك المصطلحات العمل الكبير المنشود يتضمن جميع المصطلحات وتصنيفها ومقابلاتها.
3. لغات بُنوك المصطلحات ترتبط بالهدف منه، منها ما يكون بلغتين ومنها ما يكون بأكثر من لغتين. والهدف هنا تيسير الترجمة والتدقيق فيها وعدم التداخل، حرصاً على التعبير الدقيق في الصياغة أو الترجمة. وبعض الشركات الكبرى في أوربا فيها بنوك مصطلحات تساعد على صياغة العقود والكتيبات الإرشادية والترجمة.
4. بنوك المصطلحات وصناعة المعاجم المتخصصة جانبان متكاملان، بنك المصطلحات هو في الواقع مجموعة معاجم متخصصة وضعت في منظومة. وما أسهل استدعاء مصطلحات تخصص دقيق محدد بلغتين أو أكثر ليكون أمامنا معجم متخصص.
5. الإتاحة من بنوك المصطلحات في داخل كل مؤسسة كبرى أو كل وزارة لها عدة أشكال، منها الطباعة الورقية ومنها الأقراص المدمجة ومنها الطرفيات المرتبطة ببنك المصطلحات

الرئيسي ومنها الاستدعاء الانتقائي لمصطلحات معينة مع مقابلاتها بلغة أخرى.

سادساً: الحاسوب وتصنيف المداخل

1. المعجم عمل لغوي مرجعي لا تقتصر أهميته على شرح دلالة أو دلالات الكلمة الصعبة أو النادرة، بل هو مرجع للقارئ والباحث لبيان نطق الكلمة وتصنيفها الصرفي وعلاقاتها السياقية ودلالتها ومستوى استعمالها ومنطقة استعمالها وكونها مستعملة أو بائدة.
2. تعدد صيغ التدوين مما يلاحظ في عدد من اللغات العالمية التي لها انتشار في مناطق جغرافية كبيرة. بعض الكلمات الإنكليزية له تدوين بريطاني وآخر أمريكي، وفي العربية كلمات تدون في المشرق العربي تدويناً يختلف عنه المغرب العربي.
3. تدوين النطق عرف مستقر في تقاليد المعاجم في بعض اللغات، ومنها اللغة الإنكليزية. ويتطلب ذلك تحديد المستوى اللغوي المقبول أو المنشود مثل RP في الإنكليزية ولغة المسرح والإعلام في الألمانية، ثم عمل رموز دالة على نطق كل صوت وعلى النبر والتنغيم، وعلى التنوع الجغرافي العام للنطق.
4. التصنيفات الصرفية مهمة في المعجم، ومنها أنواع الكلم والجمع واسم الجمع واسم الجنس والتعدي واللزوم والتصغير والنسب وغير ذلك.
5. المعلومات النحوية منها ارتباط الفعل بحرف جر وعلاقات الإضافة والموصوف والصفة والعمل الإعرابي وحركة البناء وغير ذلك.
6. المستوى اللغوي متنوع في اللغات، منه مستوى النصوص الدينية، مستوى فصحى التراث، مستوى الفصحى المعاصرة، مستوى لغة الإعلام، عامية المثقفين، عامية محلية... إلخ. وتحديد هذه المستويات وعمل رموز لها مهم عند التخطيط للمعجم.
7. التوزيع المكاني للعربية واضح بين المشرق والمغرب، وهناك استعمالات إدارية خاصة بقطر من الأقطار. وفي الإنكليزية نجد استعمالاً أمريكياً أو استرالياً أو هندياً إلى جانب بريطانيا.
8. زمن الاستعمال يشار إليه في بعض المعاجم بذكر رمزٍ دالٍّ على كون الكلمة بائدة أو كون

الكلمة جديدة في لغة الشباب، وقد تذكر دلالتها في قرن محدد.

سابعاً: خبرات عربية

1. مشروع أسماء الأعلام جرى بالتعاون بين سلطنة عُمان وجامعة القاهرة (1985م). أفاد من الحاسوب في عمليات متعددة بعضها خاص بالمادة اللغوية. روافد متن المعجم كانت متعددة: التدوين الصوتي، التحليل الصرفي، المعلومات الجغرافية عن انتشار اسم العلم، استعمال الاسم في عدة أجيال، وأشهر من سُمِّي بالاسم. وكل هذه الروافد دخلت في أماكنها في الحاسوب أمام كل اسم. وأخذت الصيغة النهائية للعمل كله بطريقة حاسوبية إلى المطبعة.
2. تضم الدول العربية عدة مجامع ومراكز للبحث اللغوي، وفي أكثرها مشروعات للإفادة من الحاسوب، ومنها في مجمع اللغة العربية بالقاهرة مشروع معجم لغة الشعر، وثمة عمل تسجيلي يقوم به الحاسوب في المعجم الكبير ومعاجم المصطلحات العلمية. وفي الجزائر والمملكة العربية السعودية مشروعات لغوية حاسوبية، منها الذخيرة اللغوية العربية في الجزائر ومشروعات مركز الملك عبد الله بن عبد العزيز لخدمة اللغة العربية. أما مشروع المعجم التاريخي فينفذه اتحاد المجامع اللغوية العربية بمشاركة عدد من المجامع.
3. اللغة العربية ذات تاريخ طويل ومكانة حضارية وأهمية معاصرة والدول العربية كثيرة وهنا دور نشر متعددة. ولهذا كله لابد من التفكير في مشروعات لغوية حاسوبية ضخمة، وذلك بتعرف المشروعات الحالية والخبرات العالمية في استعمال الحاسوب في صناعة المعاجم والبحوث اللغوية.

ثامناً: الإتاحة

1. بدايات الإتاحة الورقية للمعاجم كانت في صورة المخطوطات، وكان عددها من كل معجم قليلاً جداً، كما يتضح لنا من فهارس المخطوطات وحدث التحول الكبير منذ خمسمئة عام مع ظهور المطابع وأصبح المعجم الواحد متاحاً في مئات النسخ. ومع انتشار التعليم زادت

الحاجة إلى المعاجم وأصبحت المعاجم تطبع منها آلاف النسخ الورقية.

2. الإتاحة من خلال شبكات المعلومات اتجاه جديد، بدأ بتخزين المعاجم الورقية. وبذلك أصبح الحصول على مادة المعجم سهلاً دون الارتباط بوجود المستفيد في مكان محدد، والمستقبل واعد لهذا النوع من الإتاحة، وعلى وجه الخصوص بعد ارتفاع أسعار الورق وضخامة أحجام بعض المعاجم.
3. الإتاحة من خلال الأجهزة الإلكترونية بدأ بإدخال معجم كامل أو عدة معاجم متكاملة في الحاسوب، مثل معجم للغة الإنكليزية مشروح بها ومعجم لها مع اللغة الوطنية ومعجم ثالث للغة الوطنية مع الإنكليزية. ويكون وجود المعجم الأحادي للغة والمعجمين المزدوجي اللغة تيسيراً للطالب ولغيره.
4. المنافسة بين المعجم الورقي والمعجم الحاسوبي قائمة حالياً في إطار ارتفاع أسعار المعاجم الورقية ورخص الإتاحة عن طريق الحاسوب. وعلى الرغم من تحفظ مجمع اللغة العربية بالقاهرة فإن عدة دور نشر وضعت بعض معاجمه في الحاسوب وأتاحتها من خلال شبكات المعلومات.
5. المادة المحوسبة يمكن إدخالها إلى المطابع المتقدمة بشكل إلكتروني مباشر. ويمكن أخذ هذه المادة المحوسبة إلى قواعد البيانات اللغوية، وبذلك يرتفع رصيدها إلى نصوص بملايين الكلمات.

تاسعاً: اللغويون والحاسوبيون في صناعة المعاجم

1. تختلف صناعة المعجم في عصر الحاسوب عن تأليف المعاجم على مدى عصور التاريخ:

آ. المعجم كان عملاً فردياً يقوم به لغوي معجمي وأصبح عملاً مؤسسياً.

ب. في عصر الاستعمال الورقي كانت البطاقات وسيلة تدوين المادة، وأصبح الحاسوب وقواعد البيانات مراصد المصطلحات وسائل العمل في المادة اللغوية.

ج. كانت اعتماد كثير من المعاجم القديمة على معاجم أقدم تختار منها المداخل والشروح،

وأصبح الاعتماد على قواعد البيانات.

د. الشواهد كانت بالآلاف، ولكن قواعد البيانات أصبحت تتيح مادة أكبر بكثير.

2. هناك مراحل في صناعة المعجم في عصر الحاسوب لا تتم إلا بالعمل البشري:

آ. التخطيط للمعجم: الهدف، المستفيدون، المادة اللغوية، معايير اختيار المداخل، الكلمات المأخوذة من المداخل، نظام الشرح، المستوى اللغوي.

ب. إضافة المعلومات اللغوية عن الكلمة واستعمالها.

ج. إضافة الشروح الدلالية للكلمات في سياقاتها.

د. المراجعة النهائية لمتن المعجم.

هـ. إضافة المقدمات والملاحق.

3. هناك مراحل في صناعة المعجم يمكن أن يقوم بها الحاسوبيون:

آ. عمل البرامج الحاسوبية المناسبة تطبيقاً لخطة المعجم

ب. التحليل الصرفي للكلمات

ج. وضع المعلومات عن الاستعمال اللغوي ومناطق استعمال الكلمة

د. بناء قاعدة بيانات لغوية

هـ. وضع برامج للاختيار من قاعدة البيانات

و. إتاحة المادة اللغوية للمعجميين.

ز. عمل نظام الإحالات الدلالية بين الكلمات

وهناك زيادة مطردة في إمكانات الإفادة عن الحاسوب لتستوعب مجالات أخرى.

عاشراً: آفاق التعاون العربي والدولي

1. تبادل المعلومات عن المشروعات المنجزة والقائمة ليس بهدف الإعلان أو الفخر بل للإفادة منها والبناء عليها وتعرف جوانب النقص لعمل مشروعات جديدة. وهناك مشروعات منجزة في بعض الدول لعمل قواعد بيانات لغوية نفذتها جهات حكومية او شركات الحاسوب. وثمة

مشروعات طموحة في بعض البلدان في هذا المجال وكذلك في مجال المصطلحات.

2. تكامل المادة المتاحة مطلب مهم، هناك اهتمام بالفصحى المعاصرة في بعض المشروعات، وثمة مجال لإتاحة المادة اللغوية لمطبوعات عربية كثيرة لتدخل نسق قاعدة البيانات. أما الجهود التي تركز على اللغة في مراحلها التاريخية القديمة فقد تكررت في مشروعات المعجم التاريخي، وبعضها توقف وبعضها بدأ من جديد.
3. توزيع المشروعات الكبرى على مجامع اللغة العربية والمؤسسات البحثية المعنية باللغة العربية مطلب متجدد ومهم. وهناك بدايات واعدة في اتحاد المجامع اللغوية والعلمية العربية ومركز الملك عبد الله بن عبد العزيز لخدمة اللغة العربية. وفي هذا السياق لابد من تكوين وعي جديد بأن المشروعات الكبرى لا تكون جديرة بالقبول العربي العام إلا بمشاركة عدة دول عربية فيها تخطيطاً وتنفيذاً وإتاحة.
4. توسيع قاعدة الإتاحة ولا تكون المادة مقصورة على مجمع أو مؤسسة واحدة أو العاملين فيها. لقد أصبح التواصل عبر الأقطار بين المؤسسات المعنية ممكناً على نحو يوفر المال والجهد ويتيح عدم التكرار. ومن المناسب أن تتسع قاعدة الإتاحة للعمل المنجز للجامعات والوزارات والمعنيين.
5. عمل شروط المحافظة على حقوق الملكية الفكرية يقلل من شك بعض المؤسسات في أن عملها سيصبح مشاعاً ولا يستمر ملكاً لها. وقد لوحظ أن المعاجم المنجزة تنشرها المؤسسات التي أعدتها ولا تتاح على المستوى العربي العام للمستفيدين منها. لابد من قواعد تزيل هذه الشكوك والتحفظات من أجل الإتاحة الواسعة للمعاجم المعدة والتي تعد، ولا شك أن للتقنيات الحديثة دوراً في ذلك كله.
6. هذا كله يمهد لعمل علاقات منظمة مع المؤسسات الأوربية المعنية بالمعاجم والمصطلحات بعدة لغات. بنوك المصطلحات المتقدمة وتضم المصطلحات بلغتين أو أكثر، وتكون مصنفة طبقاً للتخصصات ثم في داخل كل تخصص طبقاً للمجالات مع تعريف كل مصطلح. أما الطرف الآخر فيأخذ هذا كله ويضيف المقابل العربي، ويعيدها إلى بنك المصطلحات الأول

فيكسب لغة جديدة تضاف إلى ما عنده، ويوفر بنك المصطلحات أعباء التصنيف إلى تخصصت ومجالات ويعتمد أيضاً على التعريفات التي تعد في بنك المصطلحات الأول. ومن الممكن تطبيق هذا النظام مع الدول العربية في حالة وجود عدة مراصد للمصطلحات تتعاون وتتكامل.