مُعْجَمٌ تَكْراريٌّ لِلُّغَةِ العَرَبيَّةِ المُعاصِرَةِ "مُعالَجَةٌ لُغَويَّةٌ حاسوبيَّةٌ"*

مُحَمَّد مَجدي لَبيب حامِد*
Mohamed.labib@must.edu.eg

الكلمات المفتاحية: معالجة اللغات الطبيعية آليًّا – معجم تكراري – المدونات اللغوية – اللسانيات الحاسوبية – أدوات التحليل – العينات الإحصائية.

سعت الدِّراسةُ إلى بناء نمط معجمي جديد ينظرُ إلى واقع اللُّغة وشيوع مفرداتها في المجتمع اللُّغويّ؛ حيث وضعت منهجية صالحة لبناء "معجم تكراري للغة العربية المعاصرة"، عن طريق الدَّمج بين معايير الصناعة المعجمية العربيَّة والمناهج التربوية؛ بهدف الوصول إلى المعجم المنشود القائم على الاحتكام إلى تحليل معلومات التكرار الخاصة بمفردات المستوى اللغوي المدروس؛ لوضع يد المستخدم على أكثرها شيوعًا وأهمية، إلى جانب أن هذه الدراسة تُعَدُّ محاولة للإسهام في مضمار تعليم اللغة وتطوير آلياته، سواء للناطقين بالعربية أو للناطقين بغيرها.

* أُطْروحَةٌ ماجِستير في اللسانيات الحاسوبية – كلية دار العلوم – جامعة القاهرة.
* مدرس (م) اللغويات التطبيقية – جامعة مصر للعلوم والتكنولوجيا.

واستفاد البحث من مناهج بناء المعاجم التعليمية في اللغة الإنجليزية مثل معجم نوح ويبستر Noah Webster dictionary، ومعجم مريام ويبستر للمتعلمين المبتدئين Merriam-Webster's Advanced Learner's Dictionary، وسلسلة معاجم كولينز كوبيلد Collins COBUILD، خصوصًا Collins COBUILD Advanced Learner's English Dictionary.

وقد وُسِم البحثُ بعنوان "معجم تكراريٌّ للُّغة العربيَّة المعاصرة.. دراسةٌ لغويَّة حاسوبيَّة"؛ حيث قصد البحثُ العربيةَ الفصحى المعاصرة بالدِّراسة، محاولًا إقامة معجم لغوي تعليمي، متخذًا من المدونات اللغوية وسيلةً لجمع المادة اللغوية وتخزينها بما يسمح بتعامل أدوات التحليل الآلي معها؛ لاستخلاص المعلومات المؤمَّل أن يحتويها المعجم المنشود، ومجاراةً للتقدم التكنولوجي الحادث في مجال علم اللغة الحاسوبي.

وتم جمع مادة المدونة من ثلاثة مصادر مختلفة هي لغة الصحافة وموسوعة التدوين الحر (ويكيبيديا)، ولغة التأليف والكتب، في ضوء نظرية العينات الإحصائية، وكان النصف الثاني من القرن العشرين وبدايات القرن الحادي والعشرين (1950–2015م) حدًّا زمنيًّا تدور فيه المادة المجموعة، والوطن العربي حدًّا مكانيًّا للدراسة، وبلغ عدد مفردات المدونة (5.032.831)، بعدد كلمات فريدة بلغ (285.630).

وقد قام المعجم بتقسيم كلمات مدونته على خمس مجموعات تتدرج حسب درجة الشيوع والتكرار، حيث احتوت المجموعة الأولى المفردات الأكثر استخدامًا وشيوعًا في المستوى اللغوي المدروس، والثانية ضمت المفردات الأقل استخدامًا إلى المجموعة الخامسة التي احتوت الكلمات الأقل شيوعًا أو نادرة الاستخدام.

وتم ترتيب مواد المعجم التكراري وفقًا لمنهج مجمع اللغة العربية المتَّبع في معاجمه، القائم على الترتيب الألفبائي للجذور، والترتيب حسب الصيغ للكلمات المعجمية بداية من الأفعال الثلاثية المجردة ثم مزيد الثلاثي، والأفعال الرباعية المجردة ثم مزيد الرباعي، والأسماء فتم ترتيبُها ألفبائيًّا حسب أوائل الحروف.

وقد خلصت الدِّراسة إلى إمكانية بناء منهج صالح لبناء معجم تكراري للغة العربية يركن إلى معلومات التكرار في تقديم مواده لمتعلمي اللغة العربية من أبنائها أو من غير أبنائها، في ظل الاستفادة من المدونات اللغوية، إلى جانب توظيف أدوات التحليل الآلي في معالجة اللغة العربية والتفاعل مع نظامها الصرفي والتركيبي.

وتمثَّلت أهمية الدِّراسة في أنها تمكَّنت من تأسيس منهج علمي متكامل يصلح للتطبيق على اللغة العربية؛ لبناء معجم تكراري للغة العربية المعاصرة يكون معينًا لأرباب العربية من معلمين ومتعلمين، من أبنائها أو من غير أبنائها؛ لتيسير عملية التعليم والتعلم؛ حيث يعتمد المعجم في عرض مادته على درجة شيوع المفردات وواقع استخدامها الفعلي، ما يعني توفير وقت الدَّارسين وجهودهم

المبذولة في تعلُّم مفردات قليلة الاستخدام أو نادرة الشيوع بين أبناء اللغة الأصليين.

ويقدم المعجم لمستخدميه العديد من المعلومات الهامة عن المفردات المراد دراستها، متمثلة في الجذر المعجمي، والكتابة الصوتية للمفردة، ومعدل تكرار المفردة، ودرجة شيوعها في المجتمع اللغوي، ونسبة شيوعها، والمعنى المعجمي مأخوذ من ثلاث معاجم عربية هي المعجم الوسيط ومعجم الغني ومعجم اللغة العربية المعاصرة، والكلمات المعجمية وهي جميع الأشكال اللغوية المنبثقة عن الجذر اللغوي، والمعلومات الصرفية، والشواهد اللغوية هي تمثيل للاستخدام الفعلي للمفردة بين أبناء المجتمع اللغوي.

وقد تمت معالجة مادة الدراسة على مستويين، الأول وهو مستوى الجذر، والمستوى الثاني خاص بكل مفردة منبثقة عن الجذر.

***