- **تمهيد (اللغة والعلوم):**

ما زال العقل البشري في حيرة من أمره إزاء اللغة من حيث ماهيتها وظيفتها ونشاطها داخل المخ الإنساني، إذ تعددت الرؤى والمفاهيم ما بين كونها ظاهرة اجتماعية أو نفسية أو ملكة عقلية أو غريزة بيولوجية عصبية، ومع ذلك فقد عجزت هذه االرؤى منفردةً عن احتواء إشكالية اللغة، وتقديم تفسيرات لإدراك حقيقتها وأبعادها.

وفي محاولة اللسانيين الغربيين للإحاطة بتلك الإشكالية اللغوية فقد تم استضافة اللغة -لتعدد جوانبها- من قبل معظم **العلوم الصورية Formal science** كالإحصاء والرياضيات والمنطق، **والعلوم الإخبارية Informative science** كعلم النفس، وعلم الإجتماع، وعلم المعرفة، **والعلوم الطبيعية** كالكيمياء والفيزياء والبيولوجيا.

وقد أثمر هذا التضافر بين هذه العلوم المعرفية **Cognitive science** إلى تطوّرات هائلة في تجسيد اللغة لا سيما في محاكاتها عبر ميادين الذكاء الإصطناعي التي تقوم على التمثيل الرياضي. وإذا كان تطبيق القوانين الرياضية والفيزيائية لإدراك الأشياء الممكنة في المحيط الخارجي، فإن وصف اللغة رياضيا يجعلنا قادرين على تجسيدها لدى الحاسوب الذي يتّخذ الأداء الإنساني نموذجا له يسعى إلى مناظرته.[i]

- **علاقة اللغة بالرياضيات :**

تمثل اللغة للرياضيات تحدّيا قاسيا، فكيف لعلم صارم وقاطع أن يتعامل مع غموض اللغة والتباساتها ؟ لقد ظلت اللغة من دون الخضوع للمعالجة الرياضية، حيث عجزت رياضيات **إقليدس (ت:ق.م)** عن تناول إشكالية إبداعية اللغة، المتمثلة في قدرة الناطقين بها على توليد عدد لا نهائي من الجمل.

ترجع هذه الظاهرة إلى خاصية أساسية للتعبير اللغوي، التي يُعبّر عنها رياضيّا بمصطلح "**التداخل الحَلَقي**" **recursion** ويقصد به أن الجملة الفعلية مثلا يمكن أن تتضمن شبه جملة اسمية، وشبه الجملة الإسمية هذه يمكن أن تتضمن جمل فعلية، وهكذا. نحو قولنا:

- **جاء الرجل الذي قابلْنا أخا له يعمل في المؤسسة التي...**

يلاحظ أن الجملة الفعلية المبدوعة بالفعل "**جاء**" تتضمن فاعلا في هيئة شبه الجملة الإسمية المبدوءة بالاسم "**الرجل**" التي تتضمن بدورها الجملة الفعلية المبدوءة بالفعل "**قابلنا**" التي تتضمن مفعولا في هيئة شبه الجملة الإسمية المبدوءة بالإسم "**أخا**" التي تتضمن بدورها جملة الوصف الفعلية المبدوءة بالفعل "**يعمل**".[ii]

بمعنى أنّ من خصائص اللغة الإنسانية كثافة العلاقات وتشابكها ما بين الجمل والتراكيب التي تتداخل ما بين المستوى المبنوي والمستوى المعنوي.

لقد ظل هذا التكرار الحلقي حجر عثرة أمام إقدام أهل الرياضيات على تناول إشكالية اللغوي، فظلت دون حسم، إلى أن وضع **برتراند راسل (ت:1970م)** أسس النظرية الصورية للغة **formal theory of language** التي

كانت مدخلا أساسيا لتطوير اللغات الإصطناعية لبرمجة الكمبيوتر، ومهدت الطريق لكي يقيم **تشومسكي** نموذجه الرياضي للغات الإنسانية.

لقد أحال النحو التوليدي الذي أقامه **تشومسكي** النحوي اللغوي إلى سلسلة من المعادلات التي يمكن من خلالها توليد جميع التعابير اللغوية الممكنة على عكس ما عهدناه سابقا في النحو التقليدي، الذي يكتفي بإعطاء أمثلة من حالات الإطّراد والشذوذ، والتي مهما تعددت لا يمكن أن تغطي لا نهائية اللغة.[iii]

- **آلية التكرار الرياضيّة للغة:**

يرى **تشومسكي** أن فرادة اللغة البشرية تتمثل في" **آلية التكرار**"، وقد كان الفيلسوف الإيطالي **غاليليو غاليلي** (1564-1642) من أوائل العلماء القرن 17 الذين أدهشتهم قدرة آلية إبداعية اللغة عند الإنسان في "**الاستخدام اللامحدود لأداة محدودة**". وكذلك **رينيه ديكارت** و**فون همبولت** (1767-1835).

غير أنّ تشومسكي أعاد رصد ظاهرة التكرار في بنية اللغة البشرية ضمن سياق رياضي معاصر؛ من الرياضي البريطاني **آلان تورينج** ويستخدمه في استنطاق قابلية البُنية الهرمية للغة على تكرار مستوياتها العميقة إلى ما لا نهاية. [iv]

وهذا دليل أنّ الفكرة المشرقة مثل الشمس تغيب ولا تموت ؛ حتى وإن حجبتها كل غيوم التجاهل والإقصاء فإنها تبقى على قيد الحياة ساكنة؛ فمتى تحرّرت تحرّكت ومتى تحرّكت تحققت. وهذا ما جرى لمنطق **أرسطو** (ت:322 ق.م) الذي وضع فلسفة تعصم العقل من الخطأ، واعتبرها **أورقانون** (آلة) **Organon** العلم ووسيلته إلى الصواب.

وهي فلسفة صورية مجردة كتبها وظلت حبيسة في كتبها **24 قرنا** إلى أن تحولت هذه الفلسفة على يد العالم الرياضي البرطاني **آلان تورينج** (1912- 1954) إلى فلسفة مادية آلية، بداية من أول سؤال طرحه سنة **1950** هل يمكن للآلة أن تفكر ؟ وهو أول تساؤل في الذكاء الاصطناعي، ثم أثمر هذا التساؤل نتيجة محققة وأصبح الحاسوب يحاكي تفكير الإنسان ممكنا.

- **لماذا تأثّر تشومسكي بـ:"آلان تورينج" ؟**

إنّ الرياضي البريطاني "**آلان تورينج**"، لديه آلة تسمى **"آلة تورنج"**، برهن على خصائصها سنة **1936** وهي عبارةٌ عن نموذجٍ رياضيٍ يُعبّرُ عن آلةٍ لها القدرة على القيام بعملياتٍ حَوسبيةٍ متعدّدةٍ كجمعِ الأرقامِ وتمييزِ النّصوص. وهي آلة تعمل بشكل منطقي رياضي بدون تدخّل بشري طبقًا لمدخلات وأوامر محددة مسبقًا. فهي قادرة على حل أي معضلة رياضية قابلة للحل، وهي تعتبر الأساس الرياضي لأيّ حاسوب، ويعتبر أي حاسوب يعمل ما هو إلا نجاح جديد لـ: **"آلة تورنج"**، بحيث تقوم بنفس مهامه في معالجة البيانات.[v]

إنّ آلة تورينج" آلة الحوسبة الشاملة وهذه النظرية بالذات اخترق بها جهاز تشفير"**إنيجما**" (Enigma) للقوات النازية[vi] لأنه فقه كلّيات الرياضية وخورزميتها الموحدة في الآلات، ومن مبادئ هذه الآلة أنها آلة عامة تشبه فكرة تشومسكي المتمثلة في **"الكليات اللغوية"** في اللغات البشرية، أمّا لماذا تأثّر تشومسكي **بآلان تورينج** لأنّ من أفكار **آلة تورينج** آلة عامة ، تحاكي فكرة تشومسكي الاستخدام اللامحدود لأداة محدودة.

ولأنّ معالجة الحاسوب للبيانات تشبه معالجة العقل للمعلومات، **تأثّر تشومسكي بآلان تورينج** لأنه حدس أنّ هناك تشابه ديناميكي بين آلة تورينج العالمية التي تعتبر القاعدة الرياضية التي تعمل بموجبها جميع الحواسيب، وديناميكية اللغة مع العقل الموحدة بين جميع اللغات البشرية[vii]، وبهذا أعطى تشومسكي للغة مفهوما رقميا عندما حصر فرادة اللغة البشرية في " **آلية التكرار**"، و استثمر أفكار تورينج الرياضية في استنطاق البنية الهرمية للغة.

يقول **أبو حامد الغزالي** (ت:505ه):" **العلوم متعاونة، وبعضها مرتبط ببعض**"، ويقول **ابن الجوزي** (ت:597ه) "**لكل علم بعلم تعلُّق**"، فاللغة ظاهرة حسابية في جانبها الإبداعي وهذا الطابع الإبداعي الحسابي هو الذي جعل اللسانيات تلتقي بالرياضيات، ثم ولد ما يسمى بعلم اللغة الرياضي.

- **حتميّة التقاء اللغة بالرياضيات :**

إنّ لقاء اللغة مع الرياضيات كان أمرا محتوما، وذلك لخمسة عوامل؛ وهي:[viii]

1- **تعقّد اللغة**: وهو تعقّد نابع من داخلها، ومن شدّة العلاقات التي تُربط بخارجها. والتعقيد وهو يسعى إلى التبسيط، عادة ما يلوذ بالتجريد ليوقعه في شباك الرياضيات التي من مهامها تبسيط المعقّد.

2- **لا نهائية التعبيرات اللغوية**: فمن أبجدية محدودة العدد (30 حرفا في معظم اللغات)، يمكن تكوين مئات الآلاف من الكلمات، والجمل. ما يضاعف من هذه اللانهائية ما يعرف بخاصيّة "**التداخل الحلقي**" التي تتصف بها اللغة، وليس هناك ما هو أفضل من الرياضيات لاحتواء خاصية لا نهائية اللغة، وذلك بفضل قدرتها التوليدية **generative** الهائلة. بحيث يمكن من معادلة رياضية واحدة للخط المستقيم **(أ) س + (ب) ص + (ج) = صفر**. توليد العدد اللانهائي لجميع الحالات الممكنة للخط المستقيم، وذلك بمجرد تغيير قيم عوامل المعادلة؛ **أ، ب، ج**.[ix]

3- **كثافة العلاقات**[x]: بحيث تتمتع اللغة بشبكة كثيفة من العلاقات الصوتية والصرفية والتركيبية والدلالية والمنطقية والرياضيات علم قادر تمثيل العلاقات وإبراز عواملها الحاكمة. ويحتاج ذلك إلى درجة عالية من الصورية (**التجريدية**) على مستوى الحدود اللغوية، وعلى مستوى العلاقات التي تربط بين هذه الحدود.

4- **صورية الحدود اللغوية**: حيث تصنف الحدود اللغوية إلى أسماء وأفعال وحروف وصفات، وهو تصنيف يمكن إدراكه بالحدس، ولكي تنمذج اللغة رياضيا يجب أن تتخلّص من حدسيتها بما يسمح بصياغة حدودها بصورة صورية موحدة يمكن تطبيقها على جميع الحدود والمقولات اللغوية المركبة منها. وهو ما أتبعه بالفعل **تشومسكي** في نظريته عن المعمولية والرابط التي تستخدم حاليا في تمثيل نحو أي لغة تمثيلا رياضيا.

5- صورية العلاقات النحوية: إنّ العلاقات النحوية في مستوياتها السطحية حدسية بامتياز، كعلاقة الفعل بفاعله مثلا، بحيث يتطلب التنظير اللغوي صياغة أكثر صورية فمن علاقة العامل والمعمولية، كحرف يعمل في مجروره، (س يعمل في ص)، حتى تصبح التعبيرات اللغوية سلسلة من الرموز ترتبط فيما بينها بعلاقات مجردة.

إذن؛ فعلاقة اللغة بالحاسوب كان بسبب الرياضيات، والرياضيات بمعناها الواسع هي علم التعامل مع المفاهيم المجردة والقضايا الذهنية البحتة والعلاقات التي تربط بينها،

وفي نفس الوقت نجد أنّ النظام اللغوي يزخر بالمجردات، مثل علاقة الفاعلية والمفعولية مثلا، وصلة الرياضيات بالحاسوب وهو آلة التعامل مع الرموز ليست بحاجة إلى توضيح، ويمكن أن نختصرها بمقولة **آلان تورينج** :

"ما يمكن تمثيله رياضيا يمكن برمجته آليا وبشكل قاطع".[xi]

وأهم فروع الرياضيات التي تتعامل معها اللغة والحاسوب، وهي:

- **الجبر**: المجموعات، العلاقات، نظرية الدوال، جبر التمثيلات المنطقية.
- **الهندسة**: نظرية الأشكال (طوبوغرافيا)، الشجريات.

وإذا كان تطبيق القوانين الرياضية والفيزيائية لإدراك الأشياء الممكنة في المحيط الخارجي، فإن وصف اللغة رياضيا يجعلنا قادرين على تجسيدها لدى الحاسوب الذي يتخذ الأداء الإنساني نموذجا له يسعى إلى مناظرته. ويقتضي تمثيل اللغة رياضيّا من خلال الوصف والإستقراء الإستناد إلى مدونة لغوية تعكس الواقع اللغوي، بحيث يمكن استنباط القوانين والأحكام التي ترسم الخصائص العامة للظاهرة اللغوية.[xii]

- **عملية الرقمنة " digitization" المفهوم والآليات :**

لا يختلف الحاسوب الذي يوجه الصواريخ عن الحاسوب الذي يستخدمه الأطفال في ألعابهم، ولا فرق بين الحاسوب المتخصص في الكشوفات الحسابية وبين الحاسوب الذي يظهر الأشكال أو يحلل النصوص، فالحاسوب منظومة موحدة الخورزميات.

إنّ جبرية برامج الحاسوب وقطعيتها في أعلى مستويات التجريد البحت، فإنه يمكن رقمنة تضاريس الواقع بكل تعقيداته وغموضه وسحق جميع مشاكله رقميا، وكما نعرف أنه لا يوجد ما هو أكثر تجريدا من الأرقام وهي الحقيقة التي أسبغت على الكمبيوتر صفة الرقمية (**digital computer**) ولكونه رقميا يلزم تحويل كل ما يغذى له إلى أرقام.[xiii]

هناك أشياء بحكم طبيعتها هي أرقام مثل عدد صفحات الكتاب أو عدد السكان، وهناك خصائص يمكن أن نعبر عنها بقيم رقمية باستخدام طرق القياس المختلفة كالمسافة والطول والوزن مثلا. غير أنه يبدو الأمر أكثر صعوبة إذا أردنا أن نحيل النصوص والكلام المنطوق والموسيقى والأشكال إلى أرقام. وهو ما سنحاول الإجابة عنه هنا في إيجاز.

ترتكز عملية الرقمنة بصورة أساسية على عدة أساليب تستخدم مفردة أو متضافرة، وهي:[xiv]

1- **التشفير codification** : يستخدم لتمثيل النصوص المكتوبة، حيث يعطي لكل حرف من حروف الألفباء كودا رقميا، لتحل سلاسل الأرقام محل سلاسل الحروف في الكلمات، ومن ثم الجمل وما عداها من نصوص.
2- **التبسيط simplification** : يستخدم مثلا في تمثيل الصور الملونة رقميا بحيث يعطي كل لون أصلي رقما معينا، وهو الشيء نفسه بالنسبة لدرجة اللون حيث يتم تصنيفها في تسلسل رقمي من الشدة حتى الخفوت.
3- **التوصيف بدلالة السّمات features -based specification** : يستخدم مثلا في توصيف الرموز اللغوية، بحيث يتم تمثيل الأصوات اللغوية بدلالة عدد محدود من السمات الصوتية مثل الهمس والجهر والشدة والرخاوة وهكذا.
4- **الصياغة الصورية formalism** : بحيث يتم صياغة قواعد اللغة في صورة قواعد رياضية أو منطيقية، حيث يسهل بعد ذلك رقمنتها وتكويدها.

تعتبر الرقمنة أحد مصادر قوة تكنولوجيا المعلومات بحيث نجحت في تحويل أنماط المعلومات من نصوص وأصوات وصور ورموز إلى سلاسل رقمية قوامها "**الصفر والواحد**" وهو لغة الحاسوب التي يقوم عليها. بمعنى أنّ الحاسوب آلة رقمية بامتياز والرقمنة هي جوهر الوظيفة الأساسية التي تقوم بها وحدات الإدخال input devices التي تحول ما يغذى إلى الحاسوب مهما كان أصله إلى أرقام، في حين تقوم وحدات الإخراج output devices برد الأرقام إلى الصورة الطبيعية من نصوص وأشكال وأرقام وغير ذلك.[xv]

غير أنّ يرجع الفضل في اكتشاف ثنائية الصفر والوحد إلى عبقرية الهنود القدامى الذين اخترعوا نظام العد العشري، ثم توّجوه باكتشافهم لرقم الصفر، الذي يعتبر من أعظم اكتشافات البشر فهو الذي أطلق العد العشري ولولاه ما تقدمت علوم الرياضيات.

أما الذي وهب ملكة العقل للكمبيوتر رياضيا فهو الرياضي البريطاني[xvi] **جورج بوول** (ت:1864م)؛ وذلك عندما قدم للبشرية الجبر المنطقي (فكرة ثنائية: 0و1) المستخدم حاليا في الكمبيوتر، في كتابه الموسوم بـ: "**التحقيق في قوانين الفكرّ**"، الذي نشره سنة **1854**.

ومهما يكن؛ فإنّ نجاح ثنائية الصفر والوحد في صناعة تكنولوجيا المعلومات لم تأت من فراغ ولم تصدر عن خواء، وليست نزهة أفكار، بل هي فلسفة علمية مثالية تؤمن بالقانون العلمي الصلب فـ:"**ليس هناك حظ بل هناك نتيجة، وليس هناك صدفة بل هناك أسباب**".

فثنائية الصفر والواحد فكرة رياضية ذهبية جبارة، بحيث استطاعت تجريد كل أنماط معلومات العالم الخارجي الواقعي في عالم رقمي الكتروني افتراضي، أكثر تفاعليةً ومرونة وأعمق تواصلا وسهولة؛ بحيث حوّلت المحسوس إلى مجرد؛ والمجرد إلى محسوس؛ وكأنها تحاكي لوغاريتم العقل البشري في تعامله مع اللغة.

- **المدوّنات اللغوية Linguistics corpora (المصطلح والمفهوم والنشأة) :**

إنّ كلمة **corpus** هي كلمة لاتينية تعني الجسد وجمعها **corpora** ومن مقابلاتها في العربية ؛ منها: المدونات اللغوية، الذخائر اللغوية، المتون اللغوية ...الخ.[xvii]

أم مفهوم **المدوّنة اللغوية Linguistic corpus :** فهي رصيد ضخم من نصوص اللغة في صورة إلكترونية، تُجمع اعتمادا على معايير خارجية (**التوازن والتمثل**)؛ لتمثل قدر المستطاع اللغة أو أحد صورها لتكون مصدرا للأبحاث اللغوية.

- فمعنى أنها في صورة إلكترونية، أي أنها بصيغة نصية بسيطة، وقابلة للمعالجة الآلية المباشرة من الحاسوب، وليست بصيغة صور أو بصيغة **Pdf**.
- أما معنى أنها تجمع اعتمادا على معايير خارجية أي **متوازنة** لا تطغى نصوص ذات طبيعة معينة على باقي نصوص المدونة كأن تطغي مؤلفات كاتب معيّن مثلا، وكذلك من شروط المدوّنة قدرتها على **تمثيل و اقع اللغة** في مجال الدراسة، أو الغرض الذي من أجله جمعت النصوص.[xviii]

إذن فمجمل القول أن المدوّنات اللغوية تحوي نصوصا شفوية أو مكتوبة موثقة من حيث المصدر والتاريخ والنوع؛ تعكس الاستعمال الحقيقي للغة في مجال معين في شكل مقروءٍ آليّ **machine readable** لغرض الدراسة والتحليل.

- **نشأة المدوّنات اللغوية Linguistics corpora :**

إن المدوّنات اللغوية بمفهومها الحديث لم تبدأ إلا في الستينيّات من القرن الماضي، حيث جُمعت أول مدوّنة لغوية محوسبة سنة **1961**وهي مدونة **براون** نسبة إلى جامعة **براون** الأميركية التي كانت تحوي مليون كلمة من الإنجليزية الأميركية المعاصرة في ذلك الوقت. ثم تبعتها مدوّنات عدة بعد ذلك بنفس الحجم والتصميم، مثل مدونة **LOB** التي جمعت سنة **1980**، والتي كانت نتيجة تضافر جهود جامعة **لانكستر** في بريطانيا وجامعة **أوسلو في النرويج**.

ومع تزايد قدرات الحاسوب وإمكانية رقمنة النصوص توالت المدوّنات اللغوية الغربية وبالأخص الإنجليزية منها؛ فظهرت مدوّنات عدة مثل :[xix]

1- **مدونة كوبيلد Cobuild** التي تحوي ما يقرب من ثلاثة مليارات كلمة استخدمت في بناء معاجم **كوبيلد** المتعددة، جامعة برمنغهام في بريطانيا سنة 1987.

2- **المدونة اللغوية الوطنية البريطانية** التي تحوي مئة مليون كلمة، في مطلع **ق 20** .

3- **مدونة أكسفورد اللغوية** التي تحوي 2 مليار كلمة وتستخدم خصيصا في بناء معاجم أكسفورد.

لقد استفادت صناعة المعاجم من استثمار الامكانيات الهائلة التي يوفرها الحاسوب من طاقة تخزين للمعلومات وإتاحة منظومات معلوماتية متطورة لبناء قواعد بيانات كبيرة الحجم ومعالجتها آليا وتقنيات لتنقيب النصوص التي لها أحجام هائلة لاستخراج معلومات لغوية مهمة.

إنّ فكرة تجميع النصوص في ديوان واحد موجودة في الثقافة العربية قديما، **فمصطلح "مدوّنة" لغة** : اسم مفعول مشتقٌّ من فعل دوّن **بمعنى كتب**، وهو بدوره مشتقّ من كلمة فارسية معربة هي "**ديوان**" التي استعملها العرب لتدل على الدفتر التي تكتب فيه أسماء العمال والجند، وكذلك على المكان التي تُحتفظ فيه هذه الدفاتر، ويعتبر **الخليفة عمر بن الخطاب** أول من دوّن الدواوين في الدولة الإسلامية، أي أنشأها ونظّمها.

ومادام أنّ المدونة هي مجموعة النصوص المكتوبة والموثقة من حيث **المصدر** و**التاريخ** و**النوع** كحد أدنى تكون في وعاء الكتروني؛ يمكن اعتبار المصنفات القديمة مدوّنات ورقية مثلا؛ كتاب"**العين**"، للخليل (ت:175ه) فهو مدونة معجمية ورقية. وللعلم أنّ أنّ هناك كتاب موسوم، بـ: **المدوّنة**: وهو مجموعة أحكام فقهية للإمام مالك (ت:179ه)، التي جمعها سحنون (ت:240ه).

إذن؛ فرغم الاختلاف بين المدونات اللغوية المحوسبة والمدونات اللغوية التراثية الورقية من حيث الكم والكيف وآلية المعالجة والتكشيف Indexing إلا أنّ كلاهما يعدّ مصدرا للبحث اللغوي؛ غير أن الثقافة الغربية استثمرت الحاسوب في صناعة المدونات في جيله الأول، وأول محاولة لإنشاء مدونة لغوية متخصصة اعتمدت على الحاسوب كانت على يد **روبرتو بوسا** حينما قاما بتجميع أعمال القديس توما الأكويني سنة **1949**[xx] وتكشيف نصوصها وفحص محتواها باستخدام البطاقات **المُثقَّبة**. بينما كانت أول مدونة لغوية محوسبة هي مدونة صخر في مصر سنة **1982**.

- **نماذج من المدوّنات اللغوية العربية المحوسبة:**

يمكن أن نصنف المدونات اللغوية العربية الحديثة المحوسبة إلى نوعين؛

أ- **المدوّنات اللغوية من إنشاء مؤسسات عربية** : سوف نختار ثلاث أمثلة عنها والتي لها صفحة رئيسية على الويب، وهي كالآتي :

1- **المدوّنة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية:**[xxi]

وهي من أكبر المدوّنات اللغوية العربية حيث يتجاوز حجم المدونة **المليار** كلمة، نشأت سنة **2012**، بحيث تغطي **481** موضوع في الدراسات اللغوية بمستوياتها المتنوعة، وتعتمد على خمسة ركائز أساسية في اختيار نصوص المدوّنة وهي: البعد الزمني، والبعد الجغرافي، والوعاء المعلوماتي، والمجال المعرفي، والتصنيف الموضوعي.

2- **مدوّنة المعجم التاريخي للغة العربية:**

وهي مدوّنة معجمية لغوية محوسبة انطلقت في قطر في ديسمبر **2018**، تضم كمّا وافيًا من النصوص التي تعكس واقع اللغة العربية في بيئاتها ومراكزها الثقافية والعلمية والحضارية التي شهدت نموّها وتطوُّر دلالات ألفاظها وتراكيبها.

فهي ديوان معجمي للعربية يُؤَرِّخ لألفاظها ومعانيها، ويُبَيِّن ما طرأ على تلك الألفاظ من تَحوُّل وتغيُّر، عبر تاريخها المديد، ما بين **480** قبل الهجرة إلى العام **1431** هجري، وتستمدّ مادّتها من التراث العربي المكتوب عبرَ المراحل الزَّمنيَّة المُتعاقبة للُّغة العربيَّة، وتضمّ ما يزيد على 116 مليون كلمة مجمّعة في ثمانمائة وتسع وستين (869) وثيقة.[xxii]

3- **منصة "فلك" للمدونات اللغوية** :[xxiii]

وهي مدوّنة معجمية لغوية من إنشاء مجمع الملك سلمان العالمي للغة العربية بالسعودية تأسست في 1 جانفي **2024** وتحتوي على أكثر من **مليار ونصف** كلمة، وثلاث مدونات لغوية هي :(**مدونة المجمع، ومدونة كتب العربية، والمدونة اللغوية لمتعلمي اللغة العربية**) وسيبلغ عددها في نهاية سنة 2024 عشر مدونات، هدفها المقارنة بين المدونات لاستقراء الظواهر اللغوية.

- **المدوّنات اللغوية العربية التي فشلت :**

- **مدوّنة صخر: وهي أول مدونة لغوية عربية** تأسست سنة **1982** في مصر تابعة لشركة صخر للبرمجيات التي أسسها محمد الشارخ، وهي تحتوي على 500 مليون كلمة وكان من أهدافها تطوير اللغة العربية ودعمها تكنولوجيا، وبعدما توقفت استثمرت **مدوّنة الرياض** السعودية مادتها ولها صفحة على الويب.
- **مشروع الذخيرة اللغوية العربية :** هو فكرة عبد الرحمن **الحاج صالح** بحيث عرض مشروع **الذخيرة** في عمان سنة **1986**، وهو عبارة عن بنك آلي من النصوص القديمة والحديثة ومن الجاهلية إلى وقتنا الحاضر، ويقوم تصوّر مشروع "**الذخيرة العربية = قوقل عربي**" على إدراج أجود الإنتاج العلمي العربي القديم والمعاصر، والإنتاج العلمي العالمي بعد ترجمته إلى العربية، في بنك آلي محوْسب يمكن لأي قارئ. ولكنه بقى مشروع نظري ورقي لم يتسنّ بعد فتح موقعا له على الشبكة العنكبوتية.

ب- **المدوّنات اللغوية من إنشاء مؤسسات أجنبية :**

1- **مدونة القرآن بالألمانية Corpus Coranicum :** هي مشروع بحثي بدأ سنة **2007** من قبل أكاديمية برلين، هدفها توثيق نص القرآن في صيغتيه المخطوط والشفوي، ومقارنته بالنصوص اليهودية والمسيحية؛ تقول مديرة المشروع أنجيليكا نويورث:"أن القرآن الذي بين يدي المسلمين لا دليل على اتصاله بزمن محمد لذلك تريد مع فريقها إخضاعه للتحقيق".

2- **المدونة العربية للقرآن الكريم :** مشروع تعاوني دولي قامت بها جامعة **ليدز**[xxiv] في بريطانيا عام **2009**. والغرض منها هو سد الفجوة بين قواعد اللغة العربية التقليدية وتقنيات اللسانيات الحاسوبية الحديثة. يقول عبد الرازق بن هرماس الغرض هذا من المشروع فهو حسب أصحابه **" تحقيق متن القرآن ؟؟؟"**.

3- **مدونة عربي arabiCorpus** : هي مدونة لغوية عربية متاحة وقد صممها الباحث **ديل باركينسون** بجامعة **بريجهام يانج** الأمريكية سنة **2004** تتيح هذه المدونة اللغوية إمكانية استرجاع الكلمات والعبارات وفقا لتكرار ترددها في عدد من الفئات. وتضم هذه الفئات خمسة أنواع أدبية رئيسة، هي: الصحف، الأدب الحديث، الأدب غير القصصي، العامية المصرية، الأدب قبل العصر الحديث .

- **أنواع المدونات اللغوية:** يذكر **سنكلير** خمسة أنواع من المدونات، وهي:[xxv]

1- **المدوّنات المرجعيّة Reference Corpora** : وهي التي تعطينا معلومات مفصلة عن استخدامات اللغة المتنوعة من حيث المفردات والقواعد بين الصحافات العربية مثلا؛ نحو بنوك وأبناك، مليارات وملايير، عمل وشغل...الخ.

2- **المدوّنات الراصدة Monitor Corpora**: وهي التي ترصد ظهور ألفاظ وتراكيب جديدة نحو "مصطلح الربيع العربي" أو اضمحلال بعضها.

3- **المدوّنات المقارنة Comparable Corpora**: وهي النصوص من لغات عدة مثل التعليقات السياسية حول قضية ما بين مختلف الصحافات.

4- **المدوّنات المتوازية Parallel Corpora**: تشتمل على مجموعة من النصوص المتماثلة بلغتين مختلفين (مثلا أحد النصين ترجمة للنص من لغة أخرى).

5- **المدوّنات المتخصصة Technical Corpora**: وهي مجموعة من النصوص ذات طابع محدد، مثل المقالات العلمية في مجال الفيزياء أو القانون.

ولقد أضافت **سوزان هنستون** نوعين آخرين لما سبق من أنواع المدونات، وهي:

1- **المدونات التاريخية**: وهي لدراسة تطور اللغة عبر التاريخ.

2- **مدونات المتعلمين**: وهي لدراسة نصوص متعلمي اللغة لأجل اكتسباها.

- **الأسئلة التي يمكن أن تجيب عنها المدوّنات اللغوية**؛ وهي:[xxvi]

1- ما أكثر الكلمات أو العبارات تردّدا ؟

2- ما أوجه الاختلاف بين النصوص المكتوبة والنصوص المنطوقة ؟

3- ما الأفعال، أو الأسماء، أو الحروف التي يستخدمها أهل اللغة أو أهل التخصص أكثر من غيرها ؟

4- ما حروف الجر أو الأفعال، أو الأسماء التي تسبق أو تلي كلمة بعينها ؟

5- كيف يستخدم أهل اللغة أو أهل التخصص كلمة أو مصطلحا معيّنا ؟

6- كم مرة تستخدم فيها التعبيرات الاصطلاحية بين أهل اللغة أو أهل تخصص ما ؟

- **الأسئلة التي لا يمكن أن تجيب عنها المدوّنات اللغوية**، وهي:[xxvii]

1- ما البراهين السلبية حول استعمال كلمة، أو مصطلح، أو عبارة معيّنة ؟

2- لماذا هذه الظاهرة اللغوية منتشرة في لغة الصحافة مثلا ؟

3- ما كافة الاستخدامات الممكنة لمصطلح أو كلمة أو عبارة في اللغة على إطلاقها ؟

فهذه أهم الأسئلة التي لا تستطيع المدوّنات اللغوية أن تجيب عليها، والإجابة عنها تتم عن طريق أهل اللغة أو أهل التخصص أنفسهم.

- **مزايا المدوّنات اللغوية:** إن المدونات اللغوية هي مقاربة منهجية **approach** الكترونية وبفضل البرمجيات التي صُممت لمعالجتها، ومن مزاياها مايلي:[xxviii]

1- إنها عملية وتجريبية **empirical**، مبنية على نصوص حقيقية للاستعمال اللغوي، وليس على الحدس الشخصي، وتدرس نماذج واقعية للغة.

2- التنوع المبنى على أسس علميّة لنصوص المدوّنة لتمثل استخدامات اللغة المختلفة، وذلك بمراعاة التمثيل الجغرافي والتاريخي والأسلوبي للغة واستعمالاتها المختلفة.

3- تساعد الباحث اللغوي في الوصف والتحليل والإحصاء في شتى حقول المعرفة اللسانية في ضوء النصوص المتاحة في المدوّنة المحوسبة .

4- إمكانية الاستفادة من المدونات في شتى فروع المعرفة اللسانية، مثل اللسانيات النفسية والاجتماعية والدراسات المعجمية والنحوية وتحليل الخطاب والتداولية.

ومهما يكن؛ فإنّ المدونة اللغوية هي كتلة من النصوص المرقمنة التي تستخدم لدراسة جوانب اللغة، يمكن قراءتها والتعامل معها آليا والتحكم في بياناتها ومدخلاتها، بالحذف أو التعديل من خلال قواعد بيانات (**Databases**) التي صُممت للتعامل مع هذه النصوص. وتعد قاعدة البيانات الحاوية لنصوص المدونة اللغوية مخزنا كبيرا للغة، يرجع إليه وقت الحاجة، ويتحمل أي قدر من النصوص التي تضاف إلى المادة الأساسية مستقبلا.[xxix]

- **المدونات اللغوية بين رفّض التوليدية وقبول اللسانيات التطبيقية :**
- **المدونات اللغوية من منظور النظرية التوليدية :**

لقد اعترضت النظرية التوليدية بقيادة **نعوم تشومسكي** على لسانيات المدوّنة التي تسعى إلى صناعة المدونات اللغوية اعتمادا على الحاسوب وتكون مصدر للتحليل اللغوي العلمي، مؤكداً أن اللغة ليست سلوكًا خارجيًا مكتسبًا بل بنية فطرية عميقة في ذهن الإنسان تُعرف بـ:**"القدرة اللغوية"** فقد رأى أن اللغة لا تُدرس من داخل المدونة اللغوية المحوسبة وأن الاعتماد عليها كمصدر بحث للتحليل اللغوي مخل بمنهجية الباحث اللساني العلمية، وذلك لعدة أسباب في رأيه يمكن أن نسردها في العناصر التالية، وهي :[xxx]

1- إنّ وظيفة اللساني هي وصف الكفاية اللغوية للناطق الأصلي وليس للأداء؛ أما المدونة اللغوية فهي أداة غير مفيدة للتحليل اللغوي لأنه أداء خارجي وليس بنظام داخلي للغة.

2- إن الأداء اللغوي مرآة مظللة للكفاية، لأنه يتأثر بعوامل خارجية كعيوب النطق مثلا، كما لو استمعنا إلى شخصين مصابين بالحبسة مثلا فإننا سنصور الكفاية لديهما بشكل خاطئ اعتمادا على معطيات المدونة المشوهة الذي أنتجاه وهو في الحقيقة أداء.

3- إنّ متن المدونة اللغوية بطبيعته جزئي وناقص يشوه اللغة فهو يحاكي المصدر الخطأ أي الأداء اللغوي بدلا من الكفاية التي هي مناط التحليل اللغوي العلمي.

4- إن المدونة اللغوية بطبيعتها هي جزئية وناقصة تمثل جزءا من اللغة لا اللغة كلها، فالكفاية اللغوية تعتمد قواعد محدودة تولد جملا غير محدودة.[xxxi] فلا يمكن الاستدلال باللانهائي اللامحدود على النهائي المحدود.

5- إنّ دراسة أداء التفاعل الاجتماعي دراسة لتَمظْهُر القدرة وليس دراسة للقدرة ذاتها.

6- إنّ الهدف الصحيح للساني هو دراسة الكفاية اللغوية[xxxii] **(القواعد والمعجم)** كما تتجلى في عقل المتحدث الأصلي وتفسيرها. وليست إحصاء الظواهر الأدائية للغة وتوصيفها مستقلة عن خصائص كفاية المتكلم اللغوية.

7- إن الأصل في الجمل لدى **تشومسكي** هو الإبداع وليس التكرار، والجمل التي في المدونة تعكس نهائية الجمل لدى المتحدث وهذا تصور خاطئ.

8- إنّ اللغة الخارجية[xxxiii] لا تفيد في دراسة الطبيعة الداخلية للمقدرة اللغوية لدى البشر خلافا لما لو كانت الدراسة موجهة للملكة الإدراكية الداخلية فهو يركز على العضو الإدراكي لدى البشر والملكة نفسها وليس تمظهرها عبر الأداء اللغوي.

9- إنّ نظرية القواعد الكلية تهتم بالبناء الداخلي في عقل المتكلم أي كيف يربط النظام الرياضي (الحاسوبي) الصوت بالمعنى، وليس دراسة عمليات اكتساب كل لغة لوحدها، فعملية الاكتساب هي نفسها واحدة لا تتغير إلا في تمظهرها وتشكلها في اللغات التي تعد بالآلاف وبين المتكلمين الذين يختلفون بالمليارات.

10- إنّ اكتساب اللغة ليس مستقلا عن عضو الاكتساب المغروس في عقل المتكلم يقول تشومسكي:"**نحن لا نتعلم اللغة؛ و إنما تنمو القواعد في عقولنا**".

إن الهدف الأساس من اللغة ليس هو التواصل بحيث أن هناك مظاهر كثيرة من الفشل في التواصل، فمعظم استخدام اللغة هي"**للحوار الداخلي بين المرء ونفسه وللتفكير**"، فلو قارنا **الحوار الداخلي** لدى أي إنسان بمنطوقه ومكتوبه لوجدنا أن **حواره الداخلي** أضعاف ما ينتجه، وبالتالي فإن المتون اللغوية أقل من اللغة الفعلية التأملية كما تتجلى عبر **الحوار الداخلي** ولا تمثل اللغة وهذه ملحوظة ذكية من تشومسكي.

ومجمل القول إنّ لسانيات المدونة من منظور تشومسكي لا تعني شيئا، فالمنهج المتعارف عليه في العلم ليس جمع كمية هائلة من البيانات لاستنتاج تعميمات وأحكام من هذه البيانات؛ إنما العلم بناء مفاهيم وفرضيات وإجراء بعض التجربة داخل أطر نظرية.[xxxiv]

- **المدونات اللغوية من منظور اللسانيات التطبيقية :**

بحكم أنّ اللسانيات التطبيقية هو علم تطبيقي أكثر من نظري في حل المشكلات اللغوية فقد دعا إلى استثمار قدرات الحاسوب في صناعة المدونات اللغوية لتكون مصدرا للبحث اللغوي، ومن بين اللسانيين في هذا التوجه هما **ماكنري وويلسون**، اللذان كان تساؤلهم على النحو التالي؛ كيف نبني نظرية لغوية على التأمُّل أم على المعطيات الملاحظة ؟

إنّ هذا التساؤل ينطبق على تقعيد أحكام اللغة العربية بين البصرة والكوفة، بحيث مدرسة **البصرة** كانت تستند إلى **السماع**، بينما مدرسة **الكوفة** كانت تعتدّ **بالقياس**[xxxv]، ولقد كانت مقاربة تشومسكي التأملية الإدراكية قريبة من مذهب البصريين لبنائه على أمثلة مسموعة (أي الأخذ المباشر من المتكلم الأصلي) وعدم اعتداده بالقياس خلافا للكوفيين.

إنّ التوجه التجريبي في اللسانيات ينزع إلى التأسيس على تحليل البيانات والمعطيات الخارجية مثل النصوص والشواهد والمتون مثل النحو الكوفي مثلا. في حين ينزع التوجه العقلي في اللسانيات إلى الاعتماد على التأمل الداخلي لنظام اللغة من ابن اللغة أكثر من اعتماده على المعطيات الخارجية مثل النحو البصري.[xxxvi]

ويمكن تقسيم الإنتقادات النظرية التوليدية إلى لسانيات المدونة إلى ثلاثة أقسام، وهي :[xxxvii]

1- انتقادات لا يمكن حلها :

جادل "**ماكنري وويلسون" تشومسكي** في رفضه معطيات الأداء اللغوي أن يكون مصدرا للدليل فهو نقد غير مناسب لدراسات اكتساب اللغة، بحيث أن الأداء لا الاستغناء عنه مهما اعتراه من قصور حيث لا يمكن سؤال المتحدث الأصلي لأنه غير موجود. فالمصادر المتاحة لدينا عن العربية الفصحى في العصر الجاهلي وما بعده هي مدونات فحسب، فليس لنا في اللغة العربية الفصحى إلا أن نقبل بما لدينا من متون لغوية لأنها المصدر الوحيد المتاح لدينا للغة، فلم يعد لدينا عرب قدامى ولا حدس ولا سليقة يمكننا مساءلتها.

وكذلك نقد **تشومسكي** للمتن اللغوي بأنه مرآة مضللة للكفاية لأن الأداء **اللغوي** يتأثر بعوامل أخرى غير الكفاية مما يعكس صورة مغلوطة عنها. وأن ملفوظات اللغة الطبيعة ليست نحوية. فهذا نقد غير دقيق بحيث أن معظم الملفوظات في كل السياقات في جمل تامة في المتون المنطوقة، بحيث أنّ المتن يحتوي على ما هو غيري نحوي ولكن ليس بحدة افتراض تشومسكي.

أما اعتراض تشومسكي على أهمية المعطيات الكمية بحيث لا فائدة فيها ووصفها حرفيا بالتافهة وهذا نقد غير مقبول بحيث أن المتن اللغوي مصدر مهم للمعطيات الكمية التي بدورها أثمرت نتائج جيدة للغويين، فالتعامل الحاسوبي مع إعراب الكلمة ووسمها أثمر نتائج لا يمكن لها أن تتحقق دون المدونات اللغوية.

2- انتقادات تغفل نقاط القوة في لسانيات المدوّنة :

لقد تجاهل **تشومسكي** موطن القوة التي تتحلى بها المتون ولا تتوفر في غيرها من الميادين، مثال ذلك ؛

1- كون المتن مقروءا آليا مما يمكن من الدقة في التحليل والبحث بل الموضوعية والبعد عن الذاتية التي تؤثر في محلل اللغة.

2- كما أن أسئلة البحث الموجهة معجميا تجيب عنها المدونات بجدارة فلا يمكن أن نتصور معجما لأي لغة دون أن يتأسس كليا على الأمثلة والشواهد.

أما الانتقادات **تشومسكي** التي يمكن الإفادة منها في تهذيب منهجية المدونات اللغوية وضبطها وفق معايير وأسس وهي التي في بعض المفاهيم التي أخذت بها المدونات اللغوية مثل **التوازن ودرجة التمثيل** فكانت بذلك إجابة مباشرة على انتقادات **تشومسكي**.

3- انتقادات وجهة نحو لسانيات المدوّنة :[xxxviii]

لقد تقبل **ماكنري وويلسون** بعض انتقادات تشومسكي للسانيات المدونة، فمن الاعتراضات التي لا يمكن رفضها وهي مسألة لا نهائية اللغة، والإبداعية الكامنة في الملكة العقلية، فلا يمكن للسانيات المدونة إلى أن تقبل بهذا النقد واقعا لا يمكن تجاهله، ومثل ذلك قضية العوامل الخارجية التي تؤثر في الأداء اللغوي وتشوه الكفاية اللغوية.

غير أنه نبغي أن نتفهم أن المدونات اللغوية لا تدعي أنها تعالج اللغة كاملة وإنما تتعامل مع معطيات تعترف بأنها تمثل جزءا من اللغة المدروسة وبهذا تكون قد أقرت بحدودها وقدرتها وعلى الباحث أن يتعامل معها على هذا الأساس.

بالإضافة إلى ذلك فقد اعترض **ماكنري وويلسون** على ما أثاره تشومسكي من أن المتن يعد مرآة مضللة للكفاية لأن الأداء (الإنتاج اللغوي) يتأثر بعوامل أخرى غير الكفاية مما يشوه الكفاية أو يعكس صورة مغلوطة عنها. فتشومسكي انتقد المعطيات الطبيعية نقدا لاذعا ووصفها بأنها ذات جودة منحطة. بحيث يرى أن ملفوظات اللغة الطبيعة ليست نحوية. وهو نقد غير دقيق بحيث أن معظم الملفوظات في كل السياقات في جمل تامة في المتون المنطوقة، فحتى وإن قلنا أنها غير نحوية فالمتن لابد أن يحتوي على ماهو نحوي في الحد الأدنى.[xxxix]

وختاما ؛ إنّ المناهج المبنية على الحدس ليست معادية للمناهج المبنية على المتون (المدونات اللغوية) بل ينبغي النظر إلى النوعين أنهما متكاملان، كما أن الإفادة من المتون تعتمد بشكل كبير على ما يطرحه الباحثون من أسئلة بحثية وما يستخدمونه من أدوات ومناهج للإجابة عن تلك الأسئلة، كما يجب أن نعلم أن المتون ليست بالضرورة قادرة على الإجابة عن كل ما نراه من المتون، وهذه المحدودية في القدرات لا ينبغي أن تكون سببا لطرحها جانبا، وأخيرا تبقى المتون منهجية ذات قيمة كبرى في التحليل اللغوي وكذلك في تعليم اللغة.