المعالجة الحاسوبية لمعاجم اللغةالعربية من حيث الجذرية والأصوات.

الأستاذ: الحبيب دحماني.
جامعة ابن خلدون – تيارت.

مقدمة:

تؤكد الدراسات المعجمية أن العرب قد تفوقوا على غيرهم في صناعة المعاجم من حيث المدارس والتنوع والاتقان والتأليف الفعلي، وقد أدهش هذا التفوق كثيرا من علماء الغرب فشهدوا لهم بالسبق والتميز، ومنهم العالم haywood إذ يقول في مقولته الشهيرة في كتابه arabic lexicography :" الحقيقة أن العرب في مجال المعاجم يحتلون مكان المركز سواء في الزمان أو المكان بالنسبة للعالم القديم أو الحديث، وبالنسبة للشرق أو الغرب" (1). والواقع أن هذا التفوق يظهر جليا في صناعة معاجم الألفاظ كما هو معروف في المدارس بنوعيها القديمة والحديثة.

نشأة الإحصاء اللغوي عند العرب.

إن المتتبع للإحصاء اللغوي عند العرب يجده بدأ مبكرا عندهم على يد العالم الكبير الخليل بن أحمد الفر اهدي (ت 175 هـ) الذي يعد رائدا فيه بلا منازع، ويتجلى في العمل الفذ الذي قام في إحصاء مفردات اللغة العربية بواسطة منهجه الرياضي بعد أن استقرأ كلام العرب، إذ وجده إمّا ثنائيا أو ثلاثيّا أو رباعيّا أو خماسيّا، وقد أجرى الاحتمالات الآتية: (2).

أولا: احتمال تقاليب الجذور:

الثنائي له صورتان: مثل: لبّ، بل.

الثلاثي له ست صور.

الرباعي له 24 صورة.

الخماسي له 120 صورة.

ثانيا: احتمالات إحصاء مفردات اللغة العربية، فكانت كالآتي:

1- احتمال الثنائي فهو حاصل ضرب 28 × 27 = 756 كلمة.

2- احتمال الثلاثي فهو حاصل ضرب 28 × 27 × 26 = 19656 كلمة.

3- احتمال الرباعي فهو حاصل ضرب 28 × 27 × 26 × 25 = 419400 كلمة.

4- احتمال الخماسي فهو حاصل ضرب 28×27×27×26×25×24 = 11793600 كلمة.

فكان مجموع الاحتمالات 12305412 كلمة بين المستعمل والمهمل، كما هو معروف في كتابه **العين** إذ كان يشير إلى الكلام المستعمل والكلام المهمل.

ويلاحظ مما سبق أن الإحصاء اللغوي قد عرف طريقه عند العرب حين بدأ التأليف الفعلي لصناعة المعاجم على يد الخليل، أما في العصر الحديث فقد تطور الإحصاء اللغوي تطورا هائلا نتيجة الاستعانة بالحاسوب، مما أصبح يصطلح عليه بـ (علم اللغة الحاسوبي) الذي اقتحم المجالات التطبيقية لمستويات اللغة ومنها، علم المعاجم الذي هو موضوع بحثنا، والذي يعد حقلا خصبا لعلم اللغة الحاسوبي.

ومن الأعمال الرائدة في الدراسات الإحصائية الحاسوبية في علم المعاجم التي تعد فتحا مبينا، وتفتخر بها جامعاتنا، هو العمل الذي قام به الدكتور **علي حلمي موسى** أستاذ في علم الفيزياء في جامعة الكويت، ويعد أول من اقتحم المعالجة الإحصائية لجذور اللغة العربية في المعاجم بدقة متناهية، وتشمل ثلاثة معاجم: الصحاح للجوهري (ت 400 هـ) واللسان لابن منظور (ت 711هـ) والتاج للزبيدي (ت 1205هـ) باعتبارها موسوعات.

وقد تناولت هذه الدراسة الإحصائية الجذور والأصوات:

أولا: الجذور.

لقد وازن العالم بين المعاجم الثلاثة: الصحاح واللسان والتاج لإحصاء حجم المادة اللغوية في كل معجم فكانت النتيجة كما يوضحها الجدول الآتي: (3)

المعجم	الثلاثي	الرباعي	الخماسي	المجموع
الصحاح	4814	766	38	5618
اللسان	6538	2548	187	9273
تاج العروس	7597	4081	300	11978

ويلاحظ من هذه الموازنة أن التاج أكثرهم مادة، إذ يظهر ما جاء في اللسان يبلغ 75 من جذور التاج، وما جاء في الصحاح يساوي 50 من التاج.

ويظهر من هذه الدراسة أن اللسان أقل مادة من التاج، غير أن الدراسات المعجمية تؤكد أن اللسان أكثر المعاجم العربية مادة إطلاقا، والزبيدي نفسه يقر بهذا، إذ يقول: " إن اللسان يشتمل على ثمانين ألف مادة، وتحت كل مادة كثير من المشتقات، وهذه المشتقات من الصعب تعدادها في اللغة العربية لكثرتها". (4) وهذا العدد تقريبا هو حجم المادة اللغوية المستعملة في اللغة العربية كما تؤكد الدراسات المعجمية.

و بعد هذا قام الدكتور **علي حلمي موسى** بإحصاء تردد الأصوات في الجذور الثلاثية والرباعية والخماسية في المواقع الثلاثة: الأول والوسط والأخير. وقد اقتصرنا في مداخلتنا على عينة وذلك لشساعة الموضوع وتشعبه وتعقيده.

فمثلا تردد الأصوات في الجذر الثلاثي، فقد تردد صوت الراء في الموقع الأول 342 مرّة، وفي الموقع الثاني 423 مرّة، وفي الموقع الثالث 440 مرّة، فالمجموع 1205مرّة بنسبة 6.144 ٪

ثم صوت النون الذي تردد في الموقع الأول 397، ثم يليه صوت الواو الذي تردد 356، ثم استنتج العالم نتائج من هذا التردد مبينا فيه أن أكثر الأصوات ترددا في الجذور كما يظهر في الجدول الثالث والعشرين:(5).

الجذر الثلاثي	الجذر الرباعي	الجذر الخماسي
ر 1205	ر 1102	ن 103
ن 1135	ل 793	ر 101
م 128	ب 754	ل 83

فاستنتج العالم أن أكثر الأصوات ترددا هي: الأصوات الذلقية (ر ل ن) والأصوات الشفهية (ب م ف) كما يظهر من الجدول الآتي:

الصوت	تردده	النسبة

ر	2408	٪ 7.883
ل	1907	6.203
ن	1887	6.137
ب	1825	5.937
م	1772	5.764
ف	1575	5.123

ومن هذه الإحصاءات استنتج الدكتور بعض الأحكام الخاطئة التي وقع فيها القدماء، ومنها ما قاله ابن منظور عن نسبة تردد بعض الأصوات كما يتضح في الجدول الآتي:(6)

	الفئة الأولى	الفئة الثانية	الفئة الثالثة
تقسيم ابن منظور	أ – ل – م – هـ و – ي – ن	ر– ع– ف– ت ب–ك–د– س ق–ح– ج	ظ–غ–ط–ز ث–خ–ض–ش ص–ذ
نتائج الحاسوب	ر– ل– ن – ب م – ع – ق	د– ف– س– ج ح– هـ – ش– ك ط – و– خ	ز–أ–ت–ص ث–غ– أ– ي ض– ذ – ظ

ويلاحظ من هذا الجدول أن **ابن مظور** لم يحالفه الصواب في معظم نتائجه، إذ نرى صوت الراء الذي هو أكثر الأصوات ترددا وضعه في الفئة الثانية وأيضا صوت الباء، وقليل من الأصوات التي تتقاطع بين تقسيم ابن منظور والحاسوب. فمثلا في الفئة الأولى الأصوات المتقاطعة هي: ل – ن، وفي الفئة الثانية هي: ف – ك– ح، وفي الفئة الثالثة هي: ظ – ز – ض – ذ.

كما لوحظ من هذه الإحصائيات أن الأصوات القليلة التردد هي أصوات الإطباق (ص ض ط ظ) و صوتا (غ خ) وهي أصوات التفخيم، ويستنتج من هذا أن هذه الأصوات لا تميل إليها كثيرا اللغة العربية.

وقد يقال: ما الجدوى من هذه الدراسة التي قام بها الدكتور **علي حلمي موسى** ففيها فوائد كثيرة في الدراسة اللغوية منها:

أ – تساعد الدارس على ما قام به البلاغيون القدامى في تحديد شروط الفصاحة والبلاغة كما تحدث عنها ابن سنان الخفاجي (ت 466 هـ) الذي يعد من الأوائل الذين اهتموا بالفصاحة والبلاغة، وقد اشترط لها شروطا ثمانية منها، أن يكون تأليف اللفظة من حروف متباعدة المخرج (7). أي كلما كانت الأصوات متباعدة المخرج كانت فصيحة. وإن كان عبد القاهر الجرجاني لا يوافق ما ذهب إليه ابن سنان الخفاجي، فإن مرد الفصاحة والبلاغة عنده لا تعود إلى تباعد الأصوات، وإنما مردها إلى النظم أي إدخالها في السياق، فالكلمة

عنده من حيث الصوت لا وزن ولا قيمة لها في فصاحة أو بلاغة(8) وقد أثبتت هذه الإحصائيات أن أكثر الأصوات تنافرا ولا يجتمع بعضها ببعض هي:

1- الأصوات الحلقية وهي: (ع ح هـ خ غ).
2- الأصوات اللهوية: (ق ك).
3- الأصوات الأسلية: (ص س ز).

1 ـالأصوات الشفهية: (ف ب م)

أما بعض الأصوات الأخرى فيندر تتابعها في الأول كما يظهر من خلال الجدول الآتي:(9).

التجمع	عدد مرات وروده في لسان العرب	عدد مرات وجوده في تاج العروس
ب + ب	6 مرات	13 مرة
ت + ت	مرتان	3 مرات
ج + ج	-	3 مرات
د + د	5 مرات	6 مرات
ذ + ذ	مرة واحدة	3 مرات
ر + ر	مرة واحدة	مرتان
ز + ز	مرة واحدة	4 مرات
س + س	مرة واحدة	7 مرات
ش + ش	مرة واحدة	3 مرات
ص + ص	.	مرتان
ط + ط	.	مرة واحدة
ف + ف	.	مرة واحدة
ق + ق	7 مرات	6 مرات
ك + ك	.	مرة واحدة
ل + ل	.	مرة واحدة
م + م	مرة واحدة	3 مرات
ن + ن	مرة واحدة	4 مرات
هـ + هـ	.	مرة واحدة
و + و	مرة واحدة	مرة واحدة
ي + ي	مرة واحدة	مرة واحدة

وقد أظهرت أيضا هذه الدراسات حرية وقيود الأصوات في التتابع فمنها:

1- أن أصوات الذلاقة (ر ل ن) لا قيود لها في مجاورتها مع الأصوات الأخرى إلا نادرا، فالراء تقبل جميع الأصوات، أما اللاّم فلا تسبق بصوت النون ولا تتبع بصوت الشين أما صوت النون فلا يتبع بنون مثله، ويجوز أن يسبقه صوت من الأصوات العربية.

أما صوت الباء فمقيد بأنه لا يتبعه ولا يسبقه راء ولا نون ولا فاء.

وأما صوت الفاء فمقيد بأنه لا يتبعه واو أو فاء أو باء ولا يسبقه باء أو واو أو ميم.

وأما صوت الميم فمقيد بأنه لا يتبعه التوأمان الياء والفاء. ويمكن أن يسبق بأي صوت.

فإذا قارنا بأحرف أخرى لا حظنا قيود الاستعمال الكثيرة:

أ – فمثلا صوت التاء مقيد بأنه:

لا يتبعه: ث، ذ، ص، ض، ط، ظ.

ولا يسبقه: ج، د، ذ، ض، ط، ظ.

ب – وصوت الحاء مقيد بأنه:

لا يتبعه: أ، خ، ع، غ، هـ.

ولا يسبقه: ث، خ، ظ، ع، غ، هـ.

ج – و حرف الذال مقيد بأنه:

لا يتبعه: ت، ث، د، ز، س، ش، ص، ض، ظ، غ.

ولا يسبقه: ت، ث، د، ز، ص، ض، ط، ظ، ي.

ب – تعين الدارس على ما قاله اللغويون في مسألة الدخيل على اللغة العربية فهذه الدراسة تؤصل لهذا الرأي – وإن كان فيه خلاف بينهم – ومن خصائص الكلمة العربية الرباعية والخماسية أن تشتمل على أصوات الذلاقة والأصوات الشفهية (ر ل ن ، ف ب م) ما عدا كلمة عسجد بمعنى ذهب.

ولا يقع صوت النون وبعده صوت الراء في كلمة عربية مثل: نرجس فهي أعجمية ولا يجتمع صوتا القاف والجيم في كلمة عربية مثل: منجنيق فهي أعجمية ولا يجتمع صوت الجيم والصاد في كلمة عربية مثل: جصّ وصولجان فهما أعجميتان.

ج – وتساعد الدارس في دراسة علم العروض في معرفة قوافي الشعر وتردد أصوات القافية بكثرة أو ندرتها في أشعار أخرى.

د – تبين هذه الإحصائيات الكلام المستعمل والمهمل في اللغة العربية، وقد توصل الدكتور علي حلمي موسى إلى النتائج الآتية:(11) .

الجذور	الصحاح	النسبة	اللسان	النسبة	التاج	النسبة

الثلاثي	4814	21.3	658	39.78	7597	65.55
الرباعي	766	1.30	2548	4	–	–
الخماسي	37	0.673	–	–	–	–

وقد لاحظنا في هذا الجدول أن المؤلف لم يكن مصيبا في هذه النتائج حسب الإحصائيات التي استنتجها الخليل في جمع اللغة العربية كما هو مبين في الجدول الآتي:

الجذور	إحصاء الخليل	الصحاح	النسبة	اللسان	النسبة	التاج	النسبة
الثنائي	756	–	–	–	–	–	–
الثلاثي	19656	4814	٪21.39	6538	٪39.78	7597	٪65.55
الرباعي	491400	766	٪1.3	2548	٪4	4081	–
الخماسي	11793600	38	٪0.674	187	–	300	–
المجموع	12305412 بين المستعمل والمهمل	5618		9273		11978	

فنلاحظ من الجدول أن مجموع المادة اللغوية في المعاجم الثلاثة يساوي 26869 من 12305412 أي بنسبة 0.21 ٪ وهو عدد ضئيل بالنسبة للعدد المستعمل في اللغة العربية.

وما زالت هذه الدراسة الإحصائية تحتاج على دراسة دقيقة في إحصاء المستعمل والمهمل في اللغة العربية، وهذا يحتاج إلى جهود جماعية تشترك في هذا العمل، لأن العمل الفردي في صناعة المعاجم قد ولّى، كما هو الشأن في هذه الدراسة، والدراسة التي أجراها الدكتور عبد القادر عبد الجليل على معجم الجيم لأبي عمرو الشيباني (ت 206هـ) باستخدام الحاسوب من مدونة حجمها ثلاث مجلدات وقد تعلقت هذه الدراسة بإحصاء جذور المعجم وتردد أصواتها في الأول والثاني والثالث.(12) ومن الأفضل أن يهتم به المشتغلون في صناعة المعاجم في بنوك المصطلحات العربية، وإن كان حتى الآن لم يوجد للعرب بنك مركزي موحد للمصطلحات، على غرار ما نجده عند الغربيين، كما هو الشأن في بنك المصطلحات (INFOTERM) أي: مركز الاستعلامات الدولي للمصطلحات، الموجود **بفينا** بالنمسا والذي يعالج جل اللغات العالمية بما فيها اللغة العربية.

وهذه الإحصائيات من شأنها أن تؤصل للغة العربية، وأن ترقى بها إلى العالمية، وأن تجعلها تنافس اللغات الأخرى، ولاسيما اللغة الإنجليزية التي أصبحت اللغة العالمية بلا منازع.

المصادر والمراجع:

1- الدكتور أحمد عمر مختار – محاضرات في علم اللغة الحديث – عالم الكتب – القاهرة – ط1 – 1995م – ص.56

2- ينظر المرجع السابق ص.104

3- المرجع نفسه ص.118

4- تاج العروس من جواهر القاموس – مكتبة دار يحيى – بيروت – 1958م – 9/1. والدكتور عبد اللطيف الصوفي – مصادر اللغة في المكتبة العربية – دار الهدى – الجزائر – ص.182

5- ينظر الدكتور أحمد مختار عمر- محاضرات في علم الحديث- مرجع سابق- ص109-.110

6- المرجع السابق – ص.114

7- ينظر سر الفصاحة – تحقيق محمد علي قودة- مكتبة الخانجي- القاهرة- ط1- 1932م –ص55. أو الدكتور حلمي خليل- الكلمة- دار المعرفة الجامعية للنشر والتورزيع- القاهرة- 1998م – ص28 وما بعدها.

8- ينظر دلائل الأعجاز – شرح وتعليق أحمد مصطفى المراغي- المكتبة العربية – القاهرة- ط1- 1952م- ص37-42-.68

9- الدكتور أحمد مختار عمر – محاضرات في علم اللغة الحديث- مرجع سابق- ص113-.114

10- المرجع السابق – ص.10

11- المرجع نفسه- ص.116

12- ينظر المدارس المعجمية- دار صفاء للنشر والتوزيع- عمان-ط1-1999م- ص94.