مؤسسة النهضة العلمية

الاقتصاد اللساني والذخيرة اللغوية في القاموس المحيط
دراسة تحليلية بأدوات اللسانيات المحوسبة

حسن مظفّر الرّزو

عضو أقدم

المعهد العالمي لحوسبة القرآن والعلوم الإسلامية

halrizzo@gmail.com

الخلاصة: تستكشف هذه الدراسة الذخيرة اللغوية والاقتصاد اللساني المستوطن في "القاموس المحيط" للفيروزآبادي باستخدام أدوات اللسانيات المحوسبة ونظرية المعلومات، ساعيةً إلى تجاوز المقاربات التقليدية التي ظلّت تحكم دراسة المعاجم العربية دون استثمار كافٍ للأدوات الكمّية المحوسبة. اعتمدت الدراسة منهجاً، كمّياً، وصفياً، وتحليلياً متعدّد التخصصات جمع بين اللسانيات النظرية والمحوسبة، ونظرية المعلومات، والإحصاء اللغوي، وشملت أدوات التحليل استخراج الجذور والأوزان بواسطة حزمة الخوارزميات الذكية، وحساب مؤشّرات الإنتروبيا والتكرار المعلوماتي، وتطبيق قانون زيبف، وبناء الشبكات الدلالية. كشفت النتائج عن أنماط اقتصادية لسانية، محكمة تحكم بنية القاموس، تتجلّى في إنتاجية صرفية تتبع توزيعاً لوغاريتمياً شديد التفاوت تستأثر فيه نسبة محدودة من الجذور بالحصّة الكبرى من المشتقات، وفي هيمنة البنية الثلاثية التي تعكس التوازن الأمثل بين الاختصار والتمييز، وفي أنماط صوتية تُجسّد مبدأ الجهد الأدنى عبر تفضيل الحركات الخفيفة والمقاطع البسيطة. وأظهرت مؤشّرات الإنتروبيا تنوّعاً استثنائياً على مستويَي الجذور والأوزان، فيما أفصحت الشبكات الدلالية عن كثافة تفاضلية تتباين حسب المجال المعرفي. وقد استُخلصت أربعة قوانين أساسية للاقتصاد اللساني العربي: التركيز الذكي، والثلاثية المُثلى، والطاقة الأدنى، والتوازن الإنتروبي. تُقدّم الدراسة أنموذجاً تكاملياً لسبر الذخيرة اللغوية والاقتصاد اللساني مع توفير بيانات كمّية دقيقة قابلة للاستثمار في تطبيقات الذكاء الاصطناعي ومعالجة اللغات الطبيعية، وتُرسي أسساً منهجية لحقل اللسانيات المعجمية الكمّية في الدراسات العربية.

الكلمات المفتاحية: الذخيرة اللغوية، الاقتصاد اللساني، القاموس المحيط، نظرية المعلومات، الإنتاجية الصرفية، اللسانيات المحوسبة، قانون زيبف، الإنتروبيا.

1. مقدمة في الذخيرة اللغوية والاقتصاد اللساني بمعاجم اللغة العربية

تُعد المعاجم اللغوية **مستودعات معرفية** تحفظ ذاكرة الأمة وتراثها الفكري، و**أدوات تواصلية** تضمن استمرارية اللغة وحيويتها عبر الأجيال.[1] وقد اكتسبت المعاجم أهمية مضاعفة بوصفها **أدوات لخدمة النص القرآني** وفهم دقائق معانيه، و**وسائل لحفظ اللغة** من اللحن والفساد، و**مراجع لسانية تدعم العلوم الإسلامية** من تفسير وحديث وفقه[2].

شهد التأليفُ المعجمي في دائرة علوم اللغة العربية مساراً تطوّرياً لافتاً، امتدّ على مدى أربعة عشر قرناً، تعاقبت فيه المناهج، وتنوّعت المقاصد، وتراكمت الخبرات العلمية واللسانية، غير أنّ دراسةَ المعاجم العربية ظلّت، إلى عهدٍ قريب، حبيسةَ مناهجَ تقليديةٍ تُعنى أساساً بدراسات نأت بنفسها عن المناهج اللسانية المحوسبة، دون أن تُفَعِّل على نحوٍ كافٍ ما أتاحته اللسانيات الحديثة وأدواتُ التحليل الكمي من إمكاناتٍ منهجية ومعرفية.[3]

ومع ظهور اللسانيات المحوسبة (Computational Linguistics) في النصف الثاني من القرن العشرين، وتطور أدوات معالجة اللغات الطبيعية(Natural Language Processing) ، ونظرية المعلومات(Information Theory) ، أصبح من الممكن دراسة المعاجم من منظور جديد يجمع بين **العمق اللساني** و**الدقة الكمية**، مما يتيح الكشف عن **الأنماط الخفية** و**القوانين الكامنة** التي تحكم بنية الذخيرة اللغوية وتنظيمها وتعلن عن عناصر الاقتصادي اللساني المودعة في نصوص المعاجم وشواهدها اللغوية[4].

1.1. مفهوم الذخيرة اللغوية(Lexical Inventory)

يطلق اصطلاح الذخيرةُ اللغوية على مجموع المفردات والوحدات المعجمية التي تحوزها لغةٌ ما في طورٍ زمنيٍّ محدّد، والتي تمثّل المادةَ الأولية التي يستثمرها المتكلّمون في بناء الخطاب وتحقيق فعل التواصل اللغوي[5]. وتتسم الذخيرةُ اللغوية بثلاث خصائصَ جوهرية هي: الاتساع (Extensiveness)، بما تعبّره قدرتها على استيعاب مختلف الحقول الدلالية والمعرفية؛ و**الدينامية** (Dynamism)، التي تعكس قابليتها المستمرة على النمو والتجدّد عبر آليات الاشتقاق والنحت والاقتراض؛ و**البنيوية** (Structuredness)، التي تعلن عن قدرتها على الانتظام ضمن نسقٍ محكَم تحكمه شبكةٌ من العلاقات الصرفية والدلالية والنحوية[6].

---

[1] . عمر، أحمد مختار .*صناعة المعجم الحديث* .القاهرة: عالم الكتب، 2009، ص 23-28.

[2] . نصّار، حسين .المعجم العربي: نشأته وتطوره .4th ed .القاهرة: مكتبة مصر، 2008، ص52،44،33.

[3] . زهرة، طاهر جبّار. "واقع الصناعة المعجمية العربية بين التراث والحداثة". جسور المعرفة 9. (3). (2023)، .271.

[4] . سعيدي، محمد. "اللسانيات المحوسبة وتطبيقاتها في دراسة التراث اللغوي العربي ".*المجلة العربية للعلوم الإنسانية* 38، رقم 150 (2020): 15.

[5] . Aitchison, Jean. *Words in the Mind: An Introduction to the Mental Lexicon*. 4th ed. Oxford: Wiley-Blackwell, 2012, 8-15.

[6] . Ray Jackendoff, "Lexical Semantics," in *The Oxford Handbook of the Mental Lexicon*, ed. Lila R. Gleitman, Anna Papafragou, and John C. Trueswell (Oxford: Oxford University Press, 2012), 1–3.

وتنفرد اللغة العربية بتميز ذخيرتها اللغوية، بالنظام الجذري (Root System)، حيث يمكن أن تتولد آلاف المشتقات والسبائك اللغوية من عدد محدود من الجذور، مما جعل لغتنا العربية، لغة عالية الاقتصاد، من حيث الموارد المعجمية الأساسية، وعالية الإنتاجية من حيث القدرة التوليدية[7].

1. 2. مفهوم الاقتصاد اللساني(Linguistic Economy)

يُقصد بالاقتصاد اللساني قدرةُ اللغة على بلوغ أعلى درجات الكفاءة التواصلية بأدنى قدرٍ ممكن من الجهد، وذلك من خلال تحقيق التوازن بين كثافة المعنى وبساطة البنية، سواء على مستوى **إنتاج الخطاب** عبر الاختزال والترميز، أو على مستوى **الفهم** من خلال تقليل الغموض وتسريع المعالجة الإدراكية، أو على مستوى **الاكتساب** عبر أنماط منتظمة وقواعد قابلة للتعميم، بما يضمن استدامة التواصل وفاعليته عبر الاستعمال الفردي والجماعي[8].

وقد نجح عالم اللسانيات الأمريكي جورج كينغسلي زيبف (George Kingsley Zipf) بالصياغة الرياضية لهذا المبدأ في كتابه الرائد" الطبيعة البشرية ومبدأ الجهد الأدنى" (1949)، بطرح افتراضٍ مفاده أنّ المتكلّم يميل إلى خفض الجهد المبذول في إنتاج الخطاب، في حين يحرص المتلقّي على تقليل الكلفة المعرفية اللازمة لفكّ الترميز واستيعاب الدلالة، وأنّ اللغة، في مسار تطوّرها، تعمل على إقامة توازنٍ دقيق بين هذين المطلبين اللذين يبدوان متعارضين في الظاهر[9].

وينصبغ الاقتصاد اللساني، في السياق العربي، بصور متعددة، تتناسب مع البنية الصرفية المميزة للغة، حيث يتحقق الاقتصاد من خلال :النظام الجذري الذي يختزل آلاف المفردات في عدد محدود من الجذور؛ والأوزان الصرفية القياسية التي تتيح توليد مشتقات جديدة دون الحاجة لحفظ كل مفردة على حدة؛ ونظامُ **التشكيل الإعرابي** الذي يضطلع بحمل المعلومات الصرفية والنحوية بأدنى قدرٍ من الكلفة الصوتية[10].

1. 3. نظرية المعلومات والتحليل اللساني

شهدت العقود الأخيرة تطوراً ملحوظاً في تطبيق نظرية المعلومات - (Information Theory)التي وضع أسسها كلود شانون (Claude Shannon) من خلال تطبيقاتها المتعددة على الدراسات اللسانية. حيث شهدنا استعارة الكثير من المفاهيم المستخدمة في فضاء المعلومات، مثل: الإنتروبيا[11] (Entropy)كمقياس لعدم اليقين أو التنوع، والتكرار[12] (Redundancy)

---

[7] . Boudelaa, Sami, and William D. Marslen-Wilson. "Morphological Structure in the Arabic Mental Lexicon: Parallels between Standard and Dialectal Arabic." *Language and Cognitive Processes* 26, no. 10 (2012): 1455.

[8]. Vicentini, Alessandra. "The Economy Principle in Language." *Anglisticum Journal (IJLLIS), Volume: 5 | Issue: 4, April 2016|32.*

[9] . Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley, 1949. Reprint, Mansfield Centre, CT: Martino Publishing, 2012,19-21.

[10] . Versteegh, Kees. *The Arabic Language*. 2nd ed. Edinburgh: Edinburgh University Press, 2014, 53.

[11] . الإنتروبيا المعجمية مقياس إحصائي مستمد من نظرية المعلومات، يُستخدم في اللسانيات الكمية لتحليل التوزيع الاحتمالي للوحدات اللسانية داخل المعاجم، بما يسمح بتقدير الثراء المعجمي ومستوى الاقتصاد البنيوي وعدم التمركز الدلالي في النظام المعجمي.

[12] . يُعرَّف التكرار (*Redundancy*)في اللسانيات، ولا سيما في اللسانيات الكمية ونظرية المعلومات اللغوية، بأنّه مقدار الفائض المعلوماتي في البنية اللغوية، أي تلك العناصر الصوتية، أو الصرفية، أو التركيبية، أو الدلالية التي لا تضيف معنىً جديداً بقدر ما تعيد تأكيده أو تسهّل استرجاعه وفهمه. ويُقاس التكرار بمدى

كمقياس للفائض المعلوماتي، وكفاءة القناة[13] (Channel Efficiency)، التي أصبحت أدوات تحليلية لفهم كيفية تنظيم المعلومات اللغوية وسبل استثمارها.[14]

ويمكن استخدام مؤشر الإنتروبيا، في مجال التحليل المعجمي، لقياس درجة التنوع في توزيع المشتقات على الجذور، أو في توزيع المفردات على المجالات الدلالية؛ كما ويمكن استخدام التكرار لفهم كيفية استثمار اللغة للفائض المعلوماتي في زيادة الوضوح ومقاومة الضوضاء اللسانية؛ وكذلك يمكن قياس الكفاءة المعلوماتية بحساب نسبة المعلومات المنقولة في الخطاب اللغوي إلى الجهد المبذول في الإنتاج والفهم[15] .وقد أظهرت الدراسات التطبيقية أن توزيع المفردات والأنماط اللغوية يتبع قوانين إحصائية منتظمة يمكن نمذجتها رياضياً، ويُعد قانون زيبف - (Zipf's Law)الذي ينص على أن تواتر الكلمة يتناسب عكسياً مع رتبتها - من أشهر هذه القوانين وأكثرها ثباتاً عند تطبيق صيغته الرياضية على اللغة العربية واللغات المختلفة[16].

2. الإشكاليات والغايات:

تواجه الذخيرة اللغوية المستودعة في معاجم لغتنا العربية، حزمة متنوعة من تحديات معرفية ومنهجية جوهرية، نتيجة لانغلاق جزء كبير من الدراسات التي أجريت حولها على المقاربات التقليدية، دون أن تستفيد بشكل كافٍ من التحليلات اللسانية المحوسبة التي باتت ترتكز الى نماذج رياضية مدعومة بمعالجات اللغة الطبيعية، التي تتعامل مع اللغة العربية بوصفها نظامًا لسانيًا معقّدًا قابل للتحليل المحوسب بواسطة نماذج لسانية تدار بواسطة خوارزميات الذكاء الاصطناعي[17]. وعلى صعيد آخر، لا زال المجال اللساني العربي يفتقر إلى دراسات شاملة تستكشف الاقتصاد اللساني في المعاجم العربية باستخدام مفاهيم نظرية المعلومات (Information Theory)والقوانين الإحصائية التي أثبتت فعاليتها في تحليل اللغات الطبيعية[18].

تجاوز الرسالة اللغوية الحدَّ الأدنى من المعلومات الضرورية لتحقيق الفهم، بحيث يسمح هذا الفائض بمقاومة التشويش، وتقليل الالتباس، وتعزيز الاستقرار التداولي في التواصل.

[13] . تُعرَّف كفاءة القناة (*Channel Efficiency*)، في المنظور اللساني المستند إلى نظرية المعلومات، بأنّها درجة استغلال القناة التواصلية اللغوية في نقل أكبر قدر ممكن من المعلومات ذات الصلة بأقل قدر من الفاقد أو التشويش، وذلك من خلال التوازن بين حمولة الرسالة اللغوية (Information Load) وقدرة القناة الصوتية أو الكتابية أو الإدراكية على استيعابها ومعالجتها. وتقاس كفاءة القناة لسانيًا، بمدى تقليل الهدر المعلوماتي الناتج عن الغموض، أو الإفراط في التكرار غير الوظيفي، أو التعقيد البنيوي الزائد، مقابل تعظيم وضوح الإشارة اللغوية وسرعة معالجتها لدى المتلقي. وتتحقق هذه الكفاءة عبر آليات متعددة، من قبيل الانتظام الصرفي، والاقتصاد التركيبي، والتوزيع الأمثل للتكرار، والتكيّف السياقي للخطاب، بما يضمن انتقال المعنى بدقة واستقرار عبر القناة التواصلية.

[14] . Shannon, Claude E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27, no. 3 (1948): 379-423.

[15]. Piantadosi, Steven T. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychonomic Bulletin & Review* 21, no. 5 (2014): 1116.

[16] . Manin, Dmitrii Yu. "Mandelbrot's Model for Zipf's Law: Can Mandelbrot's Model Explain Zipf's Law for Language?" *Journal of Quantitative Linguistics* 16, no. 3 (2009): 275.

[17] . موريده، عبد السلام، "المعجم الحاسوبي اللغوي ودوره في ضبط وتوحيد المصطلح التراثي العربي (الذخيرة اللغوية العربية للدكتور عبد الرحمن الحاج صالح نموذجًا)، "*مجلة التعليمية* 13، العدد 1 (20 مارس 2023): 366.

[18] . Lupyan, Gary, and Rick Dale, "Language Structure Is Partly Determined by Social Structure," *Proceedings of the National Academy of Sciences* 107, no. 41 (2010): **15094**.

وعلى الرغم من أن الاقتصاد اللساني قد حظي باهتمام واسع في الدراسات اللسانية الغربية(منذ أعمال جورج زيبف الرائدة)في منتصف القرن العشرين إلا أن تطبيقاته على المعاجم العربية لا تزال محدودة ومتفرقة، ولم ترقَ بعد إلى مستوى النماذج النظرية الشاملة المدعومة ببيانات كمية دقيقة[19].

وتكمن المشكلة البحثية الجوهرية :في كيفية قياس الذخيرة اللغوية، وتفسير الاقتصاد اللساني في معجم عربي كبير مثل "القاموس المحيط" باستخدام أدوات اللسانيات المحوسبة ونظرية المعلومات؟

وتتفرع من هذه المشكلة الأساسية عدة إشكاليات فرعية:

- الأولى: على المستوى الصرفي، كيف تتوزع الإنتاجية الاشتقاقية على آلاف الجذور العربية؟ وما هي طبيعة العوامل، مثل: (الطول، البنية الصوتية، المجال الدلالي) التي تحدد درجة إنتاجية الجذر؟ وهل يتبع هذا التوزيع قوانين إحصائية منتظمة يمكن نمذجتها رياضياً؟
- الثانية :على المستوى الصوتي-الفونولوجي، كيف تتوزع الأنماط الحركية (الشكل) والبنى المقطعية في القاموس؟ وهل تعكس هذه التوزيعات مبدأ الجهد الأدنى الذي افترضه زيبف؟ وما درجة الكثافة الصوتية والاقتصاد الفونولوجي في المعجم؟
- الثالثة :على المستوى الدلالي، كيف تتشكل الشبكات الدلالية في القاموس المحيط؟ وما درجة الكثافة والترابط في هذه الشبكات؟ وكيف يتوزع التشابه الدلالي والترادف عبر المجالات المعنوية المختلفة؟
- الرابعة :على المستوى المعلوماتي-الإحصائي، ماهي أهم مؤشرات نظرية المعلومات (الإنتروبيا، التكرار، الكفاءة) في القاموس؟ وكيف تختلف هذه المؤشرات عبر المستويات اللسانية (الجذري، الصرفي، الصوتي، الدلالي)؟ وهل يمكن بناء أنموذج لساني تكاملي يجمع هذه المستويات في إطار نظري موحد؟

من أجل هذا تسعى هذه الدراسة إلى تحقيق الأهداف التالية:

- أولاً :رسم خريطة كمية شاملة للذخيرة اللغوية في "القاموس المحيط"، تكشف عن حجمها وتوزيعاتها وأنماطها البنيوية.
- ثانياً :قياس مؤشرات الاقتصاد اللساني في القاموس، باستخدام مفاهيم نظرية المعلومات والتحليل الإحصائي المتقدم.
- ثالثاً :استكشاف العلاقات بين المستويات اللسانية المختلفة (الجذري، الصرفي، الصوتي، الدلالي، المعلوماتي) وكيفية تفاعلها في تشكيل البنية الاقتصادية للمعجم.

وسنسعى لتحقيق هذه الغايات عبر توظيف الخوارزميات الذكية في تحليل محتوى القاموس المحيط، وإجراء سلسلة من التحليلات اللسانية الدقيقة على المستويات الصرفية والنحوية والدلالية، تمهيداً لسلسلة من المعالجات المحوسبة المتقدمة،

[19] . زهرة، طاهر جبّار. "واقع الصناعة المعجمية العربية بين التراث والحداثة". جسور المعرفة 9. (3). (2023)، .271.

بهدف تقديم دراسة رائدة، ومتعددة الأبعاد، تكشف عن حجم الذخيرة اللغوية المستودَعَة في مفردات القاموس المحيط، مع تحديد وقياس أهم مؤشرات الاقتصاد اللساني التي تميّز هذا المعجم الكبير.

3. الدراسات اللسانية المعاصرة في تحليل الاقتصاد اللساني والذخيرة المعجمية في معاجم اللغة العربية

شهدت الدراسات اللسانية التطبيقية خلال العقد الأخير تحوّلًا نوعيًّا في مقاربة التراث المعجمي العربي، إذ تجاوزت المناهج التقليدية الوصفية إلى توظيف الأدوات الحاسوبية والمؤشرات الكمية في سبر البنى اللغوية العميقة. وقد حظيت ظاهرة الاقتصاد اللساني (Linguistic Economy) والذخيرة المعجمية (Lexical Richness) في المعاجم العربية الكلاسيكية باهتمام متزايد من الباحثين، لا سيما في ظل التطورات الهائلة في مجال اللسانيات الحاسوبية ومعالجة اللغات الطبيعية.

استهلّ الباحثون دراساتهم بتحليل البنية الاشتقاقية للمعاجم الكبرى، حيث كشفت دراسة صالح والعمري[20] عن الإنتاجية الصرفية في لسان العرب لابن منظور، مؤكدة أن المعدّل الاشتقاقي للجذور الثلاثية يبلغ 327 مشتقًّا للجذر الواحد، وهو ما يعكس كفاءة النظام الصرفي العربي في توليد المفردات من قاعدة محدودة من الجذور. وتوسّعت الأبحاث اللاحقة لتشمل التحليل المقارن بين المعاجم، فقدّم عبد الرحمن وآخرون[21] دراسة إحصائية موازنة بين الصحاح للجوهري وأساس البلاغة للزمخشري، مبيّنين أن معامل التنوع المعجمي (Type-Token Ratio) يتراوح بين 0.68 و0.74، ممّا يشير إلى ثراء لغوي معتبر في كلا المعجمين.

أمّا القاموس المحيط للفيروزآبادي، فقد ظلّ بعيدًا عن أدوات التحليل اللساني الكمي حتى وقت قريب، رغم مكانته المرموقة في التراث المعجمي العربي. فقد تفرّد الحمداني[22] بدراسة استكشافية أولية حول البنية الصرفية في القاموس المحيط، مستخدمًا عيّنة محدودة من المداخل، وخلص إلى أن المعجم يتميّز بتنوّع بنيوي استثنائي في الأوزان الصرفية، إلّا أنّه أشار إلى الحاجة الماسّة لدراسة شاملة تغطّي كامل المعجم. وعزّز النعيمي والقاسمي[23] هذا التوجّه بتحليل شبكة العلاقات الدلالية في عيّنة من المداخل، مؤكّدين أن القاموس المحيط يُظهر كثافة دلالية عالية في الحقول المفاهيمية المركزية. غير أنّ هذه الدراسات اقتصرت على عيّنات محدودة أو جوانب جزئية، ولم تُقدّم تحليلًا لسانيًّا شاملًا يغطي كامل البنية المعجمية.

---

[20] . محمد صالح، وأحمد العمري، «مؤشرات الإنتاجية الصرفية في *لسان العرب* لابن منظور: دراسة كمية»، *مجلة اللسانيات العربية* 12، رقم 2 (2016): ص. **145–178**.

[21] . عبد الله عبد الرحمن، ومحمد صالح، وأحمد العمري، «التحليل الإحصائي المقارن للثراء المعجمي في *الصحاح* و*أساس البلاغة*»، *مجلة الدراسات المعجمية* 5، رقم 1 (2018): 67–94.

[22] . سعيد الحمداني، «البنية الصرفية في *القاموس المحيط*: دراسة استكشافية»، *حوليات الجامعة الإسلامية بالمدينة المنورة*. *28 (2019)*، *الصفحات*: 201-229.

[23] . يوسف النعيمي، وفاطمة القاسمي، «شبكة العلاقات الدلالية في *القاموس المحيط*: مقاربة لسانية حاسوبية»، *مجلة العلوم الإنسانية*، جامعة الكويت، 38، رقم 3 (2020): 112–145.

بالمقابل تُعدّ الدراسة الحالية محاولة جادّة لسدّ الفجوة البحثية الواسعة في حقل اللسانيات العربية التطبيقية، إذ تُقدّم أوّل تحليل كمي شامل للقاموس المحيط بالاستناد إلى قاعدة بيانات ضخمة تستوعب جميع المفردات المعجمية في القاموس المحيط، موظِّفةً مؤشرات إحصائية معيارية في اللسانيات الحاسوبية. وتكتسب هذه الدراسة أهميتها من عدّة أوجه, كونها:

- أوّلًا، تملأ فراغًا علميًّا حادًّا في دراسة القاموس المحيط من منظور لساني كمي، بعد أن اقتصرت الدراسات السابقة على مقاربات تاريخية ووصفية تقليدية.
- ثانيًا، توفّر منهجية تحليلية قابلة للتطبيق على معاجم عربية أخرى، ممّا يُمكّن الباحثين من إجراء مقارنات موضوعية مبنية على معطيات لسانية، كمية دقيقة.
- ثالثًا، تكشف عن آليات الاقتصاد اللساني في التراث المعجمي، بإظهار كيف يُولّد النظام الصرفي العربي عددًا كبيرًا من المشتقات لكل جذر، يتجاوز بكثير ما سُجّل في المعاجم الأخرى.
- رابعًا، تُقدّم أدوات تقييمية موضوعية للثراء المعجمي، من خلال حساب مؤشرات التنوع البنيوي، والتوازن النحوي، والبنية الصوتية، ممّا يُسهم في فهم أعمق للبنية اللسانية للمعجم.
- خامسًا، تُرسي أسسًا علمية لحقل بحثي ناشئ يمكن تسميته بـ«اللسانيات المعجمية المحوسبة» في الدراسات العربية، وهو حقل شبه غائب في المكتبة اللسانية العربية المعاصرة، رغم ثراء التراث المعجمي العربي.

ويُضاف إلى ذلك أن غياب الدراسات المنهجية الشاملة حول الاقتصاد اللساني في المعاجم العربية الكبرى قد خلق فجوة معرفية أدّت إلى استمرار الأحكام الانطباعية غير المؤسَّسة على معطيات موضوعية. فلطالما تحدّث الدارسون عن «ثراء» العربية و«اقتصادها اللساني» دون تقديم براهين كمية محكمة، ممّا أضعف الخطاب اللساني العربي وجعله عُرضة للنقد المنهجي. وتسعى الدراسة الحالية إلى تجاوز هذا القصور من خلال توفير بيانات إحصائية دقيقة، وبناء نماذج رياضية قابلة للاختبار والتطوير، ومقارنة النتائج بالمعايير العالمية في اللسانيات الحاسوبية. وبهذا، تُسهم الدراسة في تعزيز مكانة اللسانيات العربية التطبيقية على الخريطة البحثية العالمية، وتفتح آفاقًا واعدة لبحوث مستقبلية في تحليل المدونات اللغوية الكبرى، ودراسة التطوّر التاريخي للبنى المعجمية، وتطوير أدوات ذكاء اصطناعي متخصّصة لمعالجة مدوّنات المعاجم العربية التراثية.

4. الأسس الرياضية للتحليل اللساني المحوسب لمعجم القاموس المحيط

تعد اللسانيات المحوسبة Computational Linguistics نقلة منهجية نحو التحليل الكمّي المنظّم، للأنماط البنيوية، والذخيرة اللغوية الكامنة في المدوّنات اللغوية والمعاجم العربية، مستفيدةً من التقدّم في معالجات الحوسبة الذكية، ومعالجة اللغات الطبيعية[24]. ويُعدّ القاموس المحيط للفيروزآبادي مادّة استثنائية لتطبيق هذه المنهجيات، نظرًا لشموليته المعجمية ودقّة تنظيمه البنيوي[25].

[24] . Tony McEnery and Andrew Hardie, *Corpus Linguistics: Method, Theory and Practice* (Cambridge: Cambridge University Press, 2012), **1–29**.

[25] . مجد الدين محمد بن يعقوب الفيروزآبادي، *القاموس المحيط*، المقدمة (بيروت: دار الفكر، د.ت.)، **7–12**.

4. 1. النماذج التوزيعية والاقتصاد اللساني

يستند التحليل اللساني الكمّي للمعاجم إلى فرضية أساسية مفادها أنّ اللغات الطبيعية تخضع لقوانين توزيعية منتظمة تعكس مبدأ الاقتصاد اللساني[26]. ويُعدّ قانون زيبف (Zipf's Law) من أبرز هذه القوانين، والذي يفترض وجود علاقة عكسية بين تواتر استعمال الوحدة اللغوية وطولها أو تعقيدها البنيوي[27]. وينصّ هذا القانون على أنّ الكلمات الأكثر استعمالاً تميل إلى أن تكون أقصر، بينما تكون الكلمات النادرة أطول، وهو ما يعكس استراتيجية معرفية لتقليل الجهد الإنتاجي في الاستعمالات المتكرّرة[28]. ويمكن التعبير عن هذه العلاقة رياضيًّا باستخدام دالّة القوى(Power Function) ، حيث تتناقص القيم بشكل غير خطّي مع ازدياد الرتبة[29]. وقد أثبتت الدراسات التطبيقية انطباق هذا القانون على اللغة العربية بشكل عامّ، مع بعض الانحرافات المرتبطة بطبيعة البنية الصرفية الجذرية[30].

4. 2. نظرية المعلومات وقياس التنوّع اللساني

تُوفّر نظرية المعلومات (Information Theory) إطارًا رياضيًّا لقياس التنوّع والتوازن في الأنظمة اللغوية[31]. ويُعدّ مقياس إنتروبيا شانون (Shannon Entropy) الأداة الأساسية في هذا المجال، حيث يقيس درجة عدم اليقين أو التنوّع في توزيع الفئات النحوية داخل نظام معيّن[32]. وتتراوح قيمة الإنتروبيا بين الصفر (في حالة االهيمنة الكاملة لبنية نحوية واحدة دون غيرها)، ويقارب القيمة القصوى (في حالة التوزيع المتساوي بين جميع الفئات)[33]. ويتيح هذا المقياس تقييم مستوى التوازن النحوي في المعجم من خلال قياس كيفية توزّع المداخل المعجمية على الفئات النحوية المختلفة (أسماء، وأفعال، وصفات، وحروف)[34].

4. 3. النمذجة الاحتمالية للبنية الصوتية

تخضع البنية الصوتية للجذور اللغوية لقيود منتظمة تُعرف بقيود التعاقبية الصوتية(Phonotactic Constraints) ، والتي تحدّد أيّ الأصوات يمكن أن تتجاور، وأيّ المواقع تُفضّل فيها أصوات معيّنة دون غيرها. وتحوسب قيم هذه القيود باستخدام الاحتمالات الشرطية(Conditional Probabilities) ، حيث يُحتسب احتمال ظهور صوت معيّن في موقع محدّد بالنظر إلى

---

[26] . George Kingsley Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Cambridge, MA: Addison-Wesley Press, 1949), **19–55**.

[27] . George Kingsley Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Cambridge, MA: Addison-Wesley Press, 1949), **32**.

[28] . Ramon Ferrer i Cancho and Richard V. Solé, "Least Effort and the Origins of Scaling in Human Language," *Proceedings of the National Academy of Sciences* 100, no. 3 (2003): **788–791**.

[29] **.** Mark E. J. Newman, "Power Laws, Pareto Distributions and Zipf's Law," *Contemporary Physics* 46, no. 5 (2005): 323-325.

[30] . Steven T. Piantadosi, "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions," *Psychonomic Bulletin & Review* 21, no. 5 (2014): **1112–1130**.

[31] . Claude E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal* 27, no. 3 (1948): **390**.

[32] . Claude E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal* 27, no. 3 (1948): **403.**

[33] . Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, 2nd ed. (Hoboken, NJ: Wiley-Interscience, 2006), **13–55**.

[34] . Robert M. Gray, *Entropy and Information Theory*, 2nd ed. (New York: Springer, 2011), 32.

الأصوات التي تتجاور معه[35]. وتكشف التحليلات المحوسبة لتوزيع الأصوات في مواقع الجذر المختلفة عن أنماط احتمالية واضحة تعكس التفاعل بين العوامل الإدراكية والإنتاجية[36].

4.4. حوسبة التشابه الدلالي والعلاقات الدلالية:

يُعدّ قياس التشابه الدلالي (Semantic Similarity) بين الوحدات المعجمية من التطبيقات المهمّة على صعيد التحليلات اللسانية المحوسبة[37]. وتكشف مقاييس التشابه الدلالي عن البنية الهرمية للعلاقات الترادفية، حيث توجد درجات متفاوتة من التقارب المعنوي بين الوحدات المعجمية[38]. وتسهم تحليلات التوزيع الهرمي في التعبير عن مبدأ الاقتصاد اللساني: لأن ندرة الترادف التامّ تشير إلى أنّ اللغة تتجنّب الإسراف في تخزين كلمات متطابقة دلاليًّا، بينما تنبئ الكثافة العالية في مستويات التقارب الدلالي عن وجود شبكات دلالية كثيفة في المجالات الحيوية[39].

أظهرت دراسة الأسس الرياضية، للتحليل اللساني للقاموس المحيط، أنّ النماذج التوزيعية، ونظرية المعلومات، والاحتمالات التعاقب الصوتي، ومقاييس التشابه الدلالي، مجتمعةً، يمكن أن توفر لنا فهمًا متعدّد الأبعاد للبنية المعجمية يتجاوز ما يمكن تحقيقه بالملاحظة المباشرة[40]. ولا يسهم هذا الفهم بإثراء النظرية اللسانية فحسب، بل يفتح آفاقًا تطبيقية في مجالات مثل معالجة اللغات الطبيعية، والمعجمية المحوسبة، وتعليم اللغة. ويتوقع مع تطوّر الأدوات المحوسبة وتوافر المدوّنات الرقمية، أن تشهد المقاربات الرياضية للدراسات المعجمية، مزيدًا من التطوّر والعمق التحليلي اللساني، ممّا سيعزّز قدرتنا على فهم الآليات العميقة التي تحكم تنظيم اللغة واستعمال مفرداتها.

5. التحليل اللساني المعمّق للذخيرة اللغوية في القاموس المحيط

يُعدُّ القاموس المحيط للفيروزآبادي (ت. 817هـ) أحد أهمّ المعاجم العربية التي شكَّلت مرجعية لسانية أساسية في تاريخ المعجمية العربية. وتكمن أهمية هذا المُنجَز المعجمي في كونه يمثّل ذروة التطوُّر في صناعة المعاجم اللغوية العربية خلال العصر الوسيط، إذ جمع بين الدقة في الترتيب والشمولية في التغطية اللغوية، مع الاقتصاد في العرض الذي يميّز منهجه عن المعاجم الموسوعية التي سبقته[41].

---

35 . Sharon Goldwater and Mark Johnson, "Learning OT Constraint Rankings Using a Maximum Entropy Model," in *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl (Stockholm: Stockholm University, 2003), 111-120.

36 . Paul Boersma and Bruce Hayes, "Empirical Tests of the Gradual Learning Algorithm," *Linguistic Inquiry* 33, no. 1 (2002): **45–51**.

37 . **Graeme Hirst & David St-Onge, "Lexical Chains as Representations of Context," *Computational Linguistics*, Vol. 24, No. 2 (1998), pp. 315.**

38 . George A. Miller and Walter G. Charles, "Contextual Correlates of Semantic Similarity," *Language and Cognitive Processes* 6, no. 1 (1991): 7.

39 . Halliday, M. A. K., and Ruqaiya Hasan. *Cohesion in English*.London: Longman, 1976, **pp. 275**.

40 . Rens Bod, Jennifer Hay, and Stefanie Jannedy, eds., *Probabilistic Linguistics* (Cambridge, MA: MIT Press, 2003), 12.

41 . الفيروز آبادي، مجد الدين محمد بن يعقوب، القاموس المحيط، تحقيق مكتب تحقيق التراث في مؤسسة الرسالة (بيروت: مؤسسة الرسالة، 2005)، المقدمة، 7-1.

ويُشكِّل القاموس المحيط، من منظور اللسانيات الحاسوبية المعاصرة، مدوَّنة لغوية (Corpus) غنية تتيح فرصة استثنائية لفهم البنية العميقة للعربية الفصحى، والكشف عن القوانين الاقتصادية التي تحكم استعمال اللغة وتطوُّرها[42]. إذ إنَّ تحليل الذخيرة اللغوية (Lexical Inventory) المتضمَّنة في هذا المعجم لا يقتصر على الجانب الإحصائي السطحي، بل يمتدّ ليشمل فهم الأنماط الصرفية الكامنة، والتوزيعات الصوتية، والعلاقات الاشتقاقية التي تربط بين المفردات، وهو ما يكشف عن "الاقتصاد اللساني (Linguistic Economy) "الذي يحكم بنية اللغة العربية.

يقوم الاقتصاد اللساني على مبدأ أساسيّ مفاده أنَّ اللغات الطبيعية تسعى إلى تحقيق التوازن بين قطبين متعارضين: الوضوح التواصلي من جهة، والاختزال في الجهد الإنتاجي والإدراكي من جهة أخرى[43]. وفي اللغة العربية تحديدًا، يتجلّى هذا المبدأ بوضوح في نظام الاشتقاق من الجذور الثلاثية والرباعية، حيث تُنتَج آلاف المفردات من عدد محدود من الجذور عبر أوزان صرفية منتظمة. وتمثّل هذه الآلية الاشتقاقية استراتيجية اقتصادية مُثلى لتوليد ثروة معجمية هائلة من موارد لسانية محدودة[44].

تأتي هذه الدراسة لتقدِّم تحليلاً لسانيًّا معمّقًا، متعدِّد المستويات للقاموس المحيط، مستفيدةً من أدوات معالجة اللغات الطبيعية (Natural Language Processing) والذكاء المحوسب، والتحليل الإحصائي المتقدّم. وقد تمَّ تقسيم هذا التحليل إلى عدة مراحل متدرّجة في العمق على صعيد التحليل اللساني للذخيرة اللغوية والاقتصاد اللساني.

5 . 1 . المرحلة الأولى: البنية الأساسية والتوزيعات اللسانية

تركّز هذه المرحلة على البنية الأساسية والتوزيعات اللسانية العامة، وتشمل تحليل الأطوال الصوتية للمفردات، والإنتاجية الاشتقاقية للجذور، والوظائف النحوية للمداخل المعجمية. هذه المرحلة تُرسي الأساس الكمّي لفهم البنية السطحية للذخيرة المعجمية.

5. 1. 1. توزيع الأطوال الصوتية

يُعدُّ طول الكلمة الصوتي (Phonological Length) من المؤشّرات الأساسية في دراسة الاقتصاد اللساني. فالنظريات اللسانية المعاصرة، وخاصّة قانون زيبف (Zipf's Law) ومبدأ الاقتصاد اللغوي، تفترض وجود علاقة عكسية بين تواتر استعمال الكلمة وطولها الصوتي: فالكلمات الأكثر استعمالاً تميل إلى أن تكون أقصر، بينما الكلمات الأقلّ تواترًا تكون أطول[45]. ويعكس هذا النمط التوزيع استراتيجية معرفية لتقليل الجهد الإنتاجي في الاستعمالات الأكثر تكرارًا[46].

---

[42] . Tony McEnery and Andrew Hardie, Corpus Linguistics: Method, Theory and Practice (Cambridge: Cambridge University Press, 2012), 1-29.

[43] . **Holes, Clive.***Modern Arabic: Structures, Functions, and Varieties*, rev. ed.Washington, DC: Georgetown University Press, 2004, **100–125**.

[44] . Abbas, N."Productivity of Some Morphological Patterns in Modern Arabic Lexicographical Dictionaries." *Al-Lisaniyyat* 18, no. 2 (2012): 91–.95

[45] . George Kingsley Zipf, Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology (Cambridge, MA: Addison-Wesley Press, 1949; repr., New York: Hafner, 1965), 19-55.

[46] . Ramon Ferrer i Cancho and Richard V. Solé, "Least Effort and the Origins of Scaling in Human Language," Proceedings of the National Academy of Sciences 100, no. 3 (2003): 788-791.

يُقاس الطول الصوتي، في اللغة العربية، بعدد الأصوات الصامتة (Consonants) في الجذر، حيث إنَّ البنية الصرفية العربية تعتمد على هيكل جذريّ ثابت تُضاف إليه حركات وسوابق ولواحق. تُشكّل الجذور الثلاثية النواة الأساسية للعربية، تليها الجذور الرباعية ثمّ الخماسية، ويرتبط هذا التدرُّج، في الطول الصوتي، ارتباطًا وثيقًا بالإنتاجية الاشتقاقية والتواتر الاستعمالي[47].

يُظهر الجدول (1) أنماط توزيع الأطوال الصوتية في معجم القاموس المحيط، حيث تظهر فيه قيم مؤشّرات: الطول الصوتي، وعدد المداخل، مع النِّسَب المئوية لكلّ طول من الأطوال، والنِّسَب التراكمية المقابلة لكل منها.

الجدول (1): بيانات توزيع الأطوال الصوتية في القاموس المحيط.

الطول الصوتي	عدد المداخل	النسبة المئوية	النسبة التراكمية
حرف واحد	28	0.03%	0.03%
حرفان	556	0.66%	0.69%
ثلاثة أحرف	60,244	71.49%	72.18%
أربعة أحرف	19,889	23.61%	95.79%
خمسة أحرف	3,394	4.03%	99.82%
ستة أحرف	147	0.17%	99.99%
سبعة أحرف	5	0.01%	100.00%

أظهرت التحليلات اللسانية .المحوسبة أنَّ الجذور الثلاثية تُشكّل العمود الفقري للذخيرة المعجمية في القاموس المحيط، حيث تمثّل 71.49% من إجمالي المداخل (60,244 مدخلاً). وتعكس هذه الهيمنة خاصّيّة بنيوية أساسية في اللغة العربية.

فالجذر الثلاثي يمثّل البنية المُثلى من حيث الاقتصاد اللساني:

- أوّلاً :من الناحية الإنتاجية، يتيح الجذر الثلاثي أكبر قدر من الإمكانات الاشتقاقية دون إفراط في التعقيد الصوتي. فمن جذر واحد مثل "كَتَبَ" يمكن اشتقاق عشرات المفردات (كاتب، مكتوب، مكتبة، كتابة، استكتاب، إلخ).
- ثانيًا :من الناحية الإدراكية، يحقّق الجذر الثلاثي التوازن بين الوضوح الدلالي والسهولة في المعالجة الذهنية. فثلاثة أصوات صامتة تكفي لتمييز جذر عن آخر دون غموض، مع عدم إثقال الذاكرة بتتابعات صوتية طويلة.
- ثالثًا :تشير النظريات المقارِنة في فقه اللغات السامية إلى أنَّ البنية الثلاثية هي الأصل التاريخي، وأنَّ الجذور الرباعية والخماسية نشأت لاحقًا عبر عمليات الإلحاق والتركيب[48].

[47] . السامرائي، فاضل صالح، معاني الأبنية في اللغة العربية، الطبعة الثانية، دار عمار، 2007، 45-67.

[48] . قدور، أحمد محمد .*مبادئ اللسانيات* .طبعة مزيدة ومنقحة. دمشق: دار الفكر، 2008.187-201.

وتأتي الجذور الرباعية في المرتبة الثانية بنسبة 23.61% (19,889 مدخلاً)، وهي نسبة كبيرة تعكس حيوية هذا النمط في البنية الصرفية العربية. إذ إنَّ الجذور الرباعية غالبًا ما تنشأ من[49]:

- التركيب : من خلال دمج جذرين ثنائيين أو إلحاق صوت بجذر ثلاثي (مثل: دَحْرَجَ من "دَحَّ" و"رَجَّ").
- الاقتراض :عبر استيعاب مفردات مع تكييفها صرفيًّا (مثل: بَرْهَنَ، تَرْجَمَ).
- الأصوات التعبيرية :التي تحاكي أصوات طبيعية (مثل: قَرْقَرَ، خَشْخَشَ).

ورغم أنَّ الإنتاجية الاشتقاقية للجذور الرباعية أقلّ من الثلاثية عمومًا، إلّا أنّها تُشكّل مصدرًا مهمًّا للتوسُّع الدلالي، خاصّة في المجالات التعبيرية والحِسّية[50].

وتمثّل الجذور الخماسية نسبة متواضعة لا تتجاوز 4.03% (3,394 مدخلاً)، ويمكن تبرير ذلك بأنَّ معظم الجذور الخماسية[51]:

- تنتمي إلى مجالات دلالية محدَّدة (أصوات الطبيعة، الحركات المعقَّدة، الأوصاف المركَّبة).
- تتّسم بإنتاجية اشتقاقية محدودة جدًّا، فقد لا تتجاوز مشتقّاتها المدخل المعجمي نفسه.
- كثير منها يحمل طابعًا تعبيريًّا أو محاكيًا.(Onomatopoetic)

أمّا بالنسبة للأطوال القصيرة والطويلة، فنلاحظ أنَّ الجذور الأحادية والثنائية لا تشمل سوى 0.69% مجتمعة، وتعبّر عن حالات خاصّة، وهي في الغالب[52]:

- حروف معانٍ (حروف الجرّ والعطف والنفي).
- أسماء أصوات وأفعال مبنيّة (صَهْ، مَهْ، قَدْ، لَمْ).
- أسماء إشارة وضمائر منفصلة.

من جهة أخرى، فإنَّ الجذور السداسية والسباعية لا تشكّل سوى 0.18% مجتمعة (152 مدخلاً فقط)، وتُعدّ شاذّة من حيث البنية الصرفية العربية المعيارية، وغالبًا ما تكون:[53]

- مصطلحات معرّبة أو دخيلة لم تُهضم صرفيًّا.
- تراكيب وصفية متجمّدة معجميًّا.
- أسماء أعجمية للنباتات والحيوانات والأماكن.

إنَّ التوزيع الذي نراه يؤكّد المبدأ الاقتصادي في اللغة العربية، حيث تتركّز الذخيرة المعجمية في الأطوال المتوسّطة (3-4 أحرف) التي تحقّق التوازن الأمثل بين الوضوح والاختصار[54]. كما أنَّ النسبة التراكمية لهذين الطولين تبلغ 95.79%، ممّا يعني أنَّ

---

[49] . إبراهيم أنيس، من أسرار اللغة (القاهرة: مكتبة الأنجلو المصرية، 1966)، 156-162.

[50] . رمضان عبد التواب، التطور اللغوي: مظاهره وعلله وقوانينه، ط3 (القاهرة: مكتبة الخانجي، 1997)، 151-153.

[51] . حول الخصائص الصوتية والصرفية للجذور الخماسية، يُنظر: عبد الصبور شاهين، المنهج الصوتي للبنية العربية: رؤية جديدة في الصرف العربي (بيروت: مؤسسة الرسالة، 1980)، 178-195.

[52] . إبراهيم أنيس، من أسرار اللغة (القاهرة: مكتبة الأنجلو المصرية، 1966)، 140-148.

[53] . أحمد محمد قدور، المدخل إلى فقه اللغة العربية (دمشق: دار الفكر، 2008)، 95-102.

[54] . إبراهيم أنيس، من أسرار اللغة (القاهرة: مكتبة الأنجلو المصرية، 1966)، 78-88.

الغالبية الساحقة من المفردات تقع ضمن هذا النطاق الاقتصادي. أمّا بالنسبة لندرة الجذور الطويلة (5+ أحرف) فإنّها تعكس قيودًا إدراكية على معالجة التتابعات الصوتية الطويلة[55].

5.1.2. توزيع الإنتاجية الاشتقاقية

يعبّر مؤشّر الإنتاجية الاشتقاقية (Derivational Productivity) عن قدرة الجذر اللغوي على توليد مشتقّات متعدّدة عبر آليات الاشتقاق الصرفي. ويُعدّ الاشتقاق من أهمّ آليات النموّ المعجمي، حيث تتولّد آلاف المفردات من عدد محدود من الجذور الأساسية[56].

يُظهر الجدول (2) نتائج تحليل مؤشّر الإنتاجية الاشتقاقية في معجم القاموس المحيط، حيث تظهر فيه فئة الإنتاجية للمشتقّات قبالة عدد الجذور والنسبة المئوية لعدد المشتقّات.

الجدول (2): نتائج تحليل مؤشّر الإنتاجية الاشتقاقية في معجم القاموس المحيط.

فئة الإنتاجية	عدد الجذور	النسبة المئوية	النسبة التراكمية
مشتقّ واحد	29,506	58.44%	58.44%
2-5مشتقّات	16,338	32.36%	90.80%
6-10مشتقّات	3,327	6.59%	97.39%
11-20مشتقًّا	1,064	2.11%	99.50%
21-50مشتقًّا	223	0.44%	99.94%
51-100مشتقّ	26	0.05%	99.99%
أكثر من 100 مشتقّ	4	0.01%	100.00%

تُشكّل الجذور ذات المشتقّ الواحد 58.44% من إجمالي الجذور (29,506 جذرًا)، وهي نسبة تتجاوز النصف، الأمر الذي يثير كثيرًا من التساؤلات المهمّة. ولعلّ أهمّ التفسيرات المحتملة لذلك هي:

- أنَّ كثيرًا من هذه الجذور تشير إلى معانٍ محدَّدة جدًّا أو نادرة الاستعمال (أسماء نباتات نادرة، أوصاف خاصّة، أفعال محدودة الاستعمال).
- يمكن أن تكون هذه الجذور موروثة من مراحل مبكّرة للعربية بحيث لم تشهد توسُّعًا اشتقاقيًّا لاحقًا.
- القيود الصرفية التي تحكم بعض الجذور نتيجة عدم توافقها بنيويًّا إلّا مع عدد محدود من الأوزان الصرفية بسبب خصائصها الصوتية (وجود حروف علّة، تكرار أصوات، إلخ).
- الندرة الاستعمالية لهذه الجذور بسبب قلّة التواتر في الاستعمال، الأمر الذي يحول دون أن تحظى بفرص التوسُّع الاشتقاقي.

[55] . كمال بشر، علم الأصوات (القاهرة: دار غريب، 2000)، 548-556.

[56] . عبده الراجحي، التطبيق الصرفي (بيروت: دار النهضة العربية، 1984)، 78-95.

وتعكس هذه الظاهرة مبدأً اقتصاديًّا يؤكّد أنَّ اللغة العربية تحتفظ بجذور محدودة الإنتاجية لتغطية مجالات دلالية متخصّصة، دون أن تستثمر فيها طاقة اشتقاقية كبيرة.[57]

أمّا الجذور متوسّطة الإنتاجية (2-10 مشتقّات) فتمثّل هذه الفئة 38.95% من الجذور (19,665 جذرًا)، وهي تشكّل الكتلة الوسطى من حيث الإنتاجية. وتتميّز هذه الجذور بأنّها:

- تغطّي المعاني الأساسية اليومية :لأنَّ معظمها يتعلّق بأفعال وأوصاف شائعة في الحياة اليومية.
- تتمتّع بتوازن اقتصادي :لأنَّ إنتاجيّتها كافية للتعبير عن تنويعات المعنى الأساسي دون إفراط.
- تستجيب للأوزان الصرفية القياسية :لتوافقها مع الأوزان الشائعة (فاعِل، مفعول، فَعيل، إلخ).

أمّا الجذور عالية الإنتاجية (11+ مشتقًّا) فتمثّل 2.61% فقط من الجذور (1,317 جذرًا)، لكنّها تُشكّل نواة الثراء المعجمي في العربية.. وتتميّز هذه الجذور بخصائص دلالية، كونها:

- تدور حول معانٍ محورية وأساسية (الحركة، القول، الوجود، الصنع، الأخذ والعطاء).
- تتّسم بقابلية عالية للتوسُّع الدلالي والمجازي.
- ترتبط بمجالات حيوية في الثقافة العربية (الدين، الأدب، الحرب، التجارة).

وبخصائص بنيوية، كونها:

- تتوافق مع معظم الأوزان الصرفية العربية.
- تتحمّل التصريفات المعقَّدة (الصيغ المزيدة، المصادر المتعدّدة، إلخ).
- تُنتج مشتقّات في كافّة الفئات النحوية (أسماء، أفعال، صفات).

بالمقابل، هناك 30 جذرًا فقط (0.06%) فائقة الإنتاجية تُنتج أكثر من 50 مشتقًّا، و4 جذور فقط تتجاوز 100 مشتقّ. وتمثّل هذه الجذور الاستثنائية قمّة الهرم الاشتقاقي في العربية[58].

3.1.5. <u>توزيع الوظائف النحوية</u>

يعبّر مؤشّر الوظيفة النحوية (Grammatical Function) للمدخل المعجمي عن الفئة الصرفية أو النحوية التي ينتمي إليها: الاسم، أو الفعل، أو الصفة، أو الحرف، أو غيرها من أقسام الكلام في اللغة العربية. وتتداخل الحدود بين الفئات النحوية بسبب طبيعة الاشتقاق والتصريف، بينما يعتمد التصنيف المعجمي على الاستعمال الأساسي أو الأكثر شيوعًا للمدخل[59].

---

[57] . إبراهيم أنيس، من أسرار اللغة (القاهرة: مكتبة الأنجلو المصرية، 1966)، 78-88.

[58] . محمد حسن جبل، المعجم الاشتقاقي المؤصل لألفاظ القرآن الكريم (القاهرة: مكتبة الآداب، 2010)، 1: 79-95.

[59] . محمد حماسة عبد اللطيف، بناء الجملة العربية (القاهرة: دار غريب، 2003)، 89-107.

من أجل هذا، فإنَّ توزيع الوظائف النحوية في المعجم يعكس طبيعة اللغة نفسها: هل هي لغة اسمية تميل إلى الأسماء، أم فعلية تميل إلى الأفعال؟ وما هي نسبة الفئات الأخرى؟ كما يكشف هذا التوزيع أيضًا عن التوجّهات المنهجية للفيروزآبادي في قاموسه: هل أدرج المصادر والمشتقّات في القاموس كمداخل مستقلّة أم أنّه عمد إلى دمجها تحت الجذر؟

يُظهر الجدول (3) عدد المداخل للوظائف النحوية في القاموس المحيط مع النسبة المئوية لحضور كلّ واحدة منها.

الجدول (3): عدد المداخل للوظائف النحوية في القاموس المحيط.

الوظيفة النحوية	عدد المداخل	النسبة المئوية
اسم	47,412	56.27%
فعل	32,368	38.42%
صفة	3,628	4.31%
حرف	722	0.86%
أخرى	133	0.16%

وتعلن بيانات الجدول عن الطبيعة الاسمية للغتنا العربية، حيث شكَّلت الأسماء 56.27% من مداخل القاموس. وتحمل هذه الخاصّيّة دلالات لسانية، لعلّ أهمّها[60]:

- أنَّ اللغة العربية لغة اسمية في بنيتها العميقة، حيث يمكن بناء جمل اسمية كاملة دون أفعال.
- أنَّ الأسماء في العربية تتوزّع على فئات متنوّعة: أسماء الذوات، المصادر، أسماء الفاعل والمفعول، الصفات المشبّهة، أسماء الزمان والمكان والآلة، وغيرها.
- أنَّ كثيرًا من المصادر في العربية تتمتّع باستقلالية معجمية عن أفعالها، ممّا يزيد من عدد المداخل الاسمية.

وتأتي الأفعال في المرتبة الثانية وبنسبة بلغت 38.42% من المدخلات، وهي نسبة كبيرة تعكس الطبيعة الديناميكية للعربية. بالمقابل، بلغت نسبة الأفعال إلى الأسماء (حوالي 2:3) وهي تعكس مستوى التوازن بين الجانب الثابت (الأسماء) والجانب الديناميكي (الأفعال) في الذخيرة المعجمية.

أمّا الصفات فقد شكَّلت 4.31% من المدخلات. ويمكن تبرير هذه النسبة المتواضعة[61] لأن:

- أنَّ كثيرًا من الصفات في العربية يمكن تصنيفها كأسماء (الصفة المشبّهة، اسم الفاعل المستعمَل صفة).
- أو أنَّ الفيروزآبادي قد أدرج كثيرًا من الصفات تحت الأسماء أو تحت الجذور الفعلية.
- أو بسبب كون معظم الصفات في العربية مشتقّة من أفعال أو أسماء، ممّا يقلّل من سمة الاستقلال المعجمي عنها.

[60] . تمام حسان، اللغة العربية: معناها ومبناها (القاهرة: عالم الكتب، 2006)، 167-189.

[61] . عبده الراجحي، التطبيق الصرفي (بيروت: دار النهضة العربية، 1984)، 148-167.

ولا تتجاوز نسبة حضور الحروف 0.86% من المدخلات، وهي نسبة منطقية بالنظر إلى كونها[62]:

- فئة مغلقة (Closed Class) محدودة العدد في اللغات الطبيعية.
- ودورها وظيفيّ نحويّ أكثر منه معجميّ دلاليّ.

أمّا الهامش المعجمي فيضمّ الفئات الأخرى التي بلغت نسبة حضورها 0.16% من المدخلات فقط، والتي تشمل:

- أسماء الأصوات والأفعال الجامدة.
- الظروف والضمائر وأسماء الإشارة.
- وأخيرًا، المداخل التي يصعب تصنيفها تحت مظلّة الفئات الأساسية.

ويمكن إجمال الدلالات اللسانية العامّة لمداخل الوظائف النحوية:

- أوّلاً :حضور مستوى متوازن بين البنيتين الاسمية والفعلية، حيث بلغت النسبة بين الأسماء والأفعال (1.46:1)، والتي تؤكّد أنَّ اللغة العربية تجمع بين الاسمية في بنيتها الجملية والفعلية في ديناميكيّتها التعبيرية، الأمر الذي يمنحها مرونة تعبيرية استثنائية.
- ثانيًا :التميُّز بثراء تصنيفي يبدو واضحًا نتيجة للتنوُّع في الفئات النحوية، فكلّ فئة تؤدّي وظائف تعبيرية ونحوية محدَّدة، ممّا يُثري إمكانات التعبير.
- ثالثًا :نهج الفيروز آبادي في قاموسه المحيط في الترتيب والاختيار، فهو يُدرج كثيرًا من المشتقّات والصيغ كمداخل مستقلّة، ممّا أسهم في زيادة عدد الأسماء والأفعال على حساب الجذور المجرّدة.

5. 2. المرحلة الثانية: التحليل اللساني المعمّق للذخيرة اللغوية

بعد أن أرست المرحلة الأولى الأساس الكمّي لفهم البنية السطحية للقاموس المحيط من خلال تحليل التوزيعات الأساسية (الأطوال الصوتية، الإنتاجية الاشتقاقية، الوظائف النحوية)، ننتقل من خلال المرحلة الثانية إلى مستوى أعمق من التحليل الذي سنركّز من خلال معالجاته المحوسبة على الأنماط البنيوية الدقيقة والمؤشّرات المتقدّمة التي تكشف عن الآليات الكامنة في تنظيم الذخيرة المعجمية للقاموس المحيط.

وتمثّل هذه المرحلة الجسر بين التحليل الإحصائي العامّ والتحليل الصرفي، والتي نأمل أن تكشف عن "القوانين الخفية" التي تنظّم الاقتصاد اللساني في القاموس المحيط، وتُظهر كيف أنَّ الثراء المعجمي الظاهري ينبثق من عدد محدود من الآليات البنيوية المنتظمة. وستوفّر البيانات التي أنتجتها المعالجات اللسانية المحوسبة فضاءً مفتوحًا لفهم:

- قوانين التوزيع الإحصائي للوحدات اللغوية[36^].(Zipf's Law, Power Law Distribution)
- طبيعة الأنماط التعاقبية الصوتية (Phonotactic Patterns) التي استوطنت في المعجم.[37^]
- طبيعة العلاقات الاحتمالية بين البنية الصوتية والإنتاجية الاشتقاقية.
- كما أنّها ستفصح عن طبيعة التوازنات النحوية التي تعكس الخصائص النمطية (Typological Features) لمفردات معجم القاموس المحيط.

[62] . تمام حسان، اللغة العربية: معناها ومبناها (القاهرة: عالم الكتب، 2006)، 97-115.

5. 2. 1. الجذورفائقة الإنتاجية

في لغتنا العربية، شأن بقية اللغات، هناك عدد محدود من الوحدات المعجمية الأساسية التي تتمتّع بإنتاجية استثنائية، وتُشكّل النواة المركزية للنشاط اللغوي. وتتجلّى هذه الظاهرة في "الجذور فائقة الإنتاجية (Hyper-Productive Roots) "التي تُنتج عشرات أو مئات المشتقّات، وتغطّي مجالات دلالية واسعة، وتتمتّع بحضور قويّ في الاستعمال اللغوي عبر المستويات والأساليب المختلفة[63].

بصورة عامّة، هناك أكثر من عامل من العوامل المحدِّدة للإنتاجية الفائقة[64]:

1. المحورية الدلالية :(Semantic Centrality)الجذور التي تدور حول معانٍ أساسية في التجربة الإنسانية (الحركة، الإدراك، القول، الوجود، الفعل) تميل إلى إنتاجية أعلى.
2. التوافق الصرفي :(Morphological Compatibility)بعض الجذور تتوافق بنيويًّا مع معظم الأوزان الصرفية، ممّا يزيد من إمكاناتها الاشتقاقية.
3. القابلية للتوسُّع الدلالي :(Semantic Extensibility)الجذور التي تقبل التوسُّع المجازي والاستعاري تُنتج مشتقّات في مجالات دلالية متعدّدة.
4. التواتر التاريخي :(Historical Frequency)الجذور التي استُعملت بكثافة عبر التاريخ اللغوي تراكمت لها مشتقّات متعدّدة.
5. الحيوية الثقافية :(Cultural Vitality)الجذور المرتبطة بمجالات حيوية في الثقافة (الدين، الأدب، التجارة، الحرب) تشهد توسُّعًا اشتقاقيًّا مستمرًّا.

وقد أثمرت التحليلات اللسانية المحوسبة لنصّ القاموس المحيط عن بيان الجذور فائقة الإنتاجية، والتي سعينا إلى إيداع أهمّ عشرة جذور منها في الجدول.(4) .

الجدول (4): أعلى عشرة جذور على صعيد الإنتاجية في القاموس المحيط.

الترتيب	الجذر	عدد المشتقّات	المجال الدلالي الأساسي
1	ع.ل.م	247	المعرفة والإدراك
2	ق.و.ل	189	الكلام والتعبير
3	ع.م.ل	176	الفعل والصنع
4	ك.ت.ب	164	الكتابة والتدوين
5	ح.م.ل	158	الحمل والنقل
6	ج.م.ع	152	الجمع والتجميع
7	ف.ع.ل	149	الفعل العامّ
8	خ.ر.ج	144	الخروج والإخراج
9	د.خ.ل	141	الدخول والإدخال
10	ق.ط.ع	138	القطع والفصل

[63] . 39. نبيل علي ونادية حجازي، الفجوة الرقمية: رؤية عربية لمجتمع المعرفة (الكويت: المجلس الوطني للثقافة والفنون والآداب، 2005)، 287-289.

[64] . أحمد مختار عمر، علم الدلالة (القاهرة: عالم الكتب، 1998)، 203-235.

ويبدو واضحًا أنَّ الجذر ع.ل.م يتصدّر الجذور فائقة الإنتاجية بـ247 مشتقًّا، نتيجة للبُعد الدلالي الذي يتمتّع به، والذي يؤكّد الاهتمام بالعلوم والمعارف، إضافة إلى توافق الجذر مع معظم الأوزان الصرفية:

- فيُنتج أسماءً :(عِلْم، عَالِم، عَلَّامة، مَعْلَم، مَعْلُوم).
- ويُنتج أفعالاً :(عَلِمَ، عَلَّمَ، أَعْلَمَ، تَعَلَّمَ، اسْتَعْلَمَ).
- كما يُنتج صفات ومشتقّات متنوّعة) :عَلِيم، مُعَلِّم، مُتَعَلِّم، مُسْتَعْلِم.(

ويضاف إلى ذلك البُعد الاستعمالي المميّز له نتيجة لحضوره المكثّف في جميع المستويات اللغوية (الفصحى، المتوسّطة، حتّى العامّيّة)، وحضور استعمالاته في حقول معرفية متعدّدة (كالدين، والفلسفة، والتعليم، والعلوم).

من جهة أخرى، يظهر الجدول أنَّ ثمّة ثالوثًا معرفيًّا قد استأثر بالمراتب الثلاث الأولى بين الجذور فائقة الإنتاجية، هو ثالوث الجذور الثلاثة الأولى (ع.ل.م247 :، ق.و.ل189 :، ع.م.ل176) والذي يعكس المحاور الأساسية للتجربة الإنسانية: المعرفة، التعبير عنها، وتطبيقها. وليس مصادفة أن تكون هذه الجذور هي الأكثر إنتاجية، فهي تغطّي أوسع مساحة من الحاجات التعبيرية بالخطاب.

وتأتي بعدها مجموعة من الجذور فائقة الإنتاجية تتعلّق بالحركة المكانية، هي:

- خ.ر.ج :(144)الخروج والإخراج.
- د.خ.ل :(141)الدخول والإدخال.
- ن.ز.ل :(135)النزول والإنزال.
- ر.ف.ع :(132)الرفع والعلوّ.

والتي ترتبط مشتقّاتها بالمفاهيم المجرّدة التي يمكن أن تُفهَم من خلال استعارات، أو ترتبط بالتوسُّع الاستعاري الذي يُستخدَم بكثرة في الخطاب العربي.

تأتي بعدها الجذور التي ترتبط بالإدراك الحسّيّ والعقلي، وهي:

- س.م.ع :(103)السمع والاستماع.
- ب.ص.ر :(101)البصر والرؤية.
- ع.ق.ل :(105)العقل والفهم.

والتي يمكن أن تُعدّ قنوات المعرفة الأساسية، التي تعدّ إنتاجيّتها العالية مؤشّرًا على أهمّيّتها المعرفية والثقافية في الحضارة العربية الإسلامية.

أمّا المجال الديني والأخلاقي فقد ارتبط بالجذور:

- ح.ك.م :(124)الحكم والقضاء (مرتبط بالشريعة).
- ش.ر.ع :(111)الشرع والتشريع (جوهر الدين).
- ح.ف.ظ :(109)الحفظ (حفظ القرآن والدين).

- ذ.ك.ر :(107)الذكر (ذكر الله).
- ص.د.ق :(95)الصدق (فضيلة أساسية).
- ك.ذ.ب :(93)الكذب (رذيلة أساسية).

وتعدّ الإنتاجية العالية لهذه الجذور انعكاسًا دقيقًا للاهتمام بالشريعة الإسلامية وعلومها وممارساتها الأخلاقية، والتي نشب عن تطوُّر العلوم الشرعية تطوُّرات ملموسة في دقّة اصطلاحاتها، مع تطوير مفردات متوازية للفضائل والرذائل من هذه الجذور.

في الوقت ذاته، نلاحظ أنَّ هناك جذورًا متقابلة دلاليًّا تتمتّع بإنتاجية متوازنة، مثل:

- ق.د.م (99)مقابل أ.خ.ر .(97)

وغيرها من الجذور التي تتنظّم من خلال مشتقّاتها المفاهيم في ثنائيات متقابلة.

5. 2. 2. مؤشّرات البنية الصوتية

بصورة عامّة، تخضع البنية الصوتية (Phonological Structure) للجذور اللغوية لقيود وأنماط منتظمة تُعرَف بـ"قيود التعاقبية الصوتية(Phonotactic Constraints) "، والتي يمكن من خلالها تحديد: أيّ الأصوات يمكن أن تتجاور، وأيّ المواقع تُفضَّل فيها أصوات معيّنة، وأيّ التتابعات الصوتية تعدّ محظورة أو نادرة[65]. ويمكن ملاحظة هذه القيود بجلاء في لغتنا العربية في:

1. قيود التتابع الصوتي :لأنَّ هناك بعض الأصوات لا تتجاور في الجذور (مثل: ندرة تتابع حرفي حلق متتاليين).
2. التوزيع المفضَّل للأصوات :لكون بعض الأصوات تُفضَّل في مواقع معيّنة من الجذر (الفاء، العين، اللام).
3. التماثل والتباين :لتحقيق التوازن بين تكرار نفس الصوت (المضاعَف) والتنويع الصوتي.
4. الوزن الصوتي :لأنَّ بعض الجذور تُعدّ "ثقيلة" صوتيًّا (تضمّ أصواتًا حلقية وتفخيمًا) وبعضها "خفيفة."

أثمرت عملية التحليل الكمّي للبنية الصوتية (في نصّ معجم القاموس المحيط) عن بيانات المؤشّرات الصوتية التي أودعناها في الجدول.(5)

الجدول (5): مؤشّرات توزيع أكثر الأصوات تواترًا في المواقع الثلاثة للجذر (ف.ع.ل).

الموقع	أكثر الأصوات تواترًا	النسبة المئوية
الفاء (الأوّل)	ع، ق، ح، ك، ب، س، ش، ف	متنوّع
العين (الوسط)	ر، ل، م، ن، و، ي	الأصوات السيّالة
اللام (الثالث)	م، ن، ل، ر، د، ت	الأصوات الذلقية والأسنانية

[65] . كمال بشر، علم الأصوات (القاهرة: دار غريب، 2000)، 478-512.

ففي موقع الفاء، تكثر الأصوات القوية والمفخَّمة (ع، ق، ح، خ، غ، ص، ض، ط، ظ)، ممّا يمنح الجذر قوّة صوتية في البداية، فيسهل تمييزه سمعيًّا[66]. أمّا في موقع العين، فتكثر الأصوات السيّالة أو الانزلاقية (ر، ل، م، ن، و، ي)، لكونها تشكّل جسرًا صوتيًّا بين الفاء واللام[67]. ويمكن أن تُعزى ندرة الأصوات الانفجارية القوية في هذا الموقع إلى تسهيل عملية النطق. بينما نلاحظ في موقع اللام كثرة الأصوات الذلقية (ل، ر، ن، م) والأصوات الأسنانية (د، ت، ذ، ث)، لأنَّ هذه الأصوات تُعدّ "نهايات" طبيعية للجذر، كما أنّها سهلة النطق والوقف[68].

على صعيد آخر، ظهر أنَّ الجذور المضاعَفة (مثل: ك.ر.ر، ق.ص.ص، م.د.د) تمثّل نسبة متواضعة لا تتجاوز (7.62%)، الأمر الذي يُظهر أنَّ اللغة العربية تُفضّل التنويع الصوتي في الجذور لتعظيم القدرة التمييزية. كما أنَّ كثيرًا من الجذور المضاعَفة تدلّ على التكرار أو الاستمرارية، مثل: (كرَّ = تكرار الهجوم، مدَّ = استمرار المدّ). إضافة إلى أنَّ كون تكرار نفس الصوت يتطلّب جهدًا نطقيًّا إضافيًّا، فإنَّ اللغة تحرص على تجنُّبه إلّا لضرورة دلالية[69].

أمّا بالنسبة للتوازن بين الأصوات المجهورة والمهموسة، فقد أظهرت التحليلات ميلاً واضحًا لصالح الأصوات المجهورة (التوازن 58:42 تقريبًا يحقّق تنوُّعًا صوتيًّا كافيًا دون إفراط في أيّ من الجانبين)، الأمر الذي يؤكد طبيعة اللغة العربية بوصفها لغة "صوتية(Sonorous) " ، ولأنَّ الأصوات المجهورة أسهل في الإدراك السمعي، ممّا يزيد من فعالية التواصل[70].

من جهة أخرى، أظهرت التحليلات وجود توازن شبه متكافئ (47:53) في نسبة الأصوات الحلقية (ء، ع، غ، ح، خ، ه) واللاحلقية في الجذور، وهو توازن يعكس طابع اللغة العربية الصوتي المميّز.

أمّا بالنسبة للجذور المعتلّة (التي تحوي حروف العلّة، مثل: ق.و.ل، ب.ي.ع، ج.ي.ء) فقد أظهرت التحليلات اللسانية أنّها تشكّل أكثر من الخُمس. وتعكس هذه النسبة الكبيرة أهمّيّة حروف العلّة في البنية الصرفية العربية[71] بسبب:

- شيوعها بكثرة بسبب أهمّيّتها الدلالية رغم أنّها تتّسم بتعقيد صرفيّ أكبر من غيرها بسبب الإعلال والقلب والحذف.
- كثير من الجذور عالية التواتر معتلّة (ق.و.ل، ك.و.ن، ج.ي.ء).

بصورة عامّة، يمكننا القول إنَّ الدلالات اللسانية العامّة للبنية الصوتية لمفردات معجم القاموس المحيط قد أظهرت أنَّ هناك قيودًا منتظمة تحكم بناء الجذور في اللغة العربية، فالتوزيعات تؤكّد:

- مبدأ التباين الأقصى :تفضيل التنويع الصوتي على التكرار.
- مبدأ السيولة المركزية :تفضيل الأصوات السيّالة في وسط الجذر.
- مبدأ القوّة الطرفية :تفضيل الأصوات القوية في بداية ونهاية الجذر.

---

[66] . عبد الصبور شاهين، المنهج الصوتي للبنية العربية: رؤية جديدة في الصرف العربي (بيروت: مؤسسة الرسالة، 1980)، 278-312.

[67] . إبراهيم أنيس، الأصوات اللغوية (القاهرة: مكتبة الأنجلو المصرية، 1975)، 95-118.

[68] . كمال بشر، علم الأصوات (القاهرة: دار غريب، 2000)، 567-595.

[69] . محمد حسن عبد العزيز، مدخل إلى علم اللغة (القاهرة: دار الفكر العربي، 2002)، 178-182.

[70] . كمال بشر، علم الأصوات (القاهرة: دار غريب للطباعة والنشر والتوزيع، 2000)، 412-425.

[71] . Holes, Clive. *Modern Arabic: Structures, Functions, and Varieties*. Rev. ed. Washington, DC: Georgetown University Press, 2004, 74–77.

كما أنَّ البنية الصوتية تحقّق توازنًا بين:

- الوضوح السمعي :استخدام أصوات متباينة ومجهورة لتسهيل الإدراك.
- الاقتصاد النطقي :تجنُّب التتابعات الصعبة والتكرارات المرهقة.

كما أنَّ النسبة العالية للأصوات الحلقية قد أكّدت الهوية الصوتية للعربية، حيث التوازن المدروس بين المجهور والمهموس اللذين يُعطيان العربية طابعها الصوتي المميّز الذي يميّزها عن اللغات السامية الأخرى[72].

5. 3. 2. مؤشّرات التوازن النحوي

يُستخدَم مؤشّر التوازن النحوي (Grammatical Balance) لقياس مستوى العلاقات الكمّية بين الفئات النحوية المختلفة في المعجم، وهو مؤشّر مهمّ يمكن أن نستنتج من خلاله[73]:

1. النمط اللغوي :(Language Typology) الذي يحدد هل أن اللغة العربية اسمية أم فعلية في توجُّهها الأساسي؟
2. طبيعة الاستراتيجيات التعبيرية :كيف تُفضّل اللغة العربية التعبير عن المعاني (بالأسماء، والأفعال، والصفات)؟
3. مستوى الإنتاجية الصرفية :أيّ الفئات أكثر قابلية للاشتقاق والتوليد؟
4. البنية المعجمية :كيف يُنظّم المعجم توزيع الموارد بين الفئات المختلفة؟

وقد أظهر التحليل المحوسب لمؤشّرات التوازن النحوي في القاموس المحيط طبيعة النِّسَب التي اتّسمت بها هذه الفئات النحوية - انظر الجدول.(6)

الجدول (6): نِسَب ظهور الفئات النحوية المختلفة في القاموس المحيط.

الفئة النحوية	العدد	النسبة	نسبة التحوُّل إلى فئات أخرى
أسماء	47,412	56.27%	18.3%منها صفات مشبّهة
أفعال	32,368	38.42%	12.7%منها يعمل اسميًّا
صفات	3,628	4.31%	65.4%منها اسمية الأصل
حروف	722	0.86%	-
أخرى	133	0.16%	-

لقد بلغت نسبة الأسماء إلى الأفعال 1.46:1، الأمر الذي يضع اللغة العربية في منطقة وسيطة بين اللغات الاسمية الخالصة (Noun-Heavy Languages)مثل اللغة الصينية (حيث تصل النسبة إلى 4:1)، واللغات الفعلية الخالصة (Verb-Heavy Languages)مثل بعض لغات السكّان الأصليين للأمريكتين (التي تنخفض النسبة فيها إلى 0.8:1)[74].

---

[72] . Sībawayhi. *Al-Kitāb*. Edited by ʿAbd al-Salām Muḥammad Hārūn. Vol. 4. Cairo: Dār al-Kutub al-Miṣriyya, 1988, 431–435.

[73] . Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman Grammar of Spoken and Written English*. Harlow: Longman, 1999, 42–48.

[74] . Dryer, Matthew S. "Word Order." In *The World Atlas of Language Structures*, edited by Martin Haspelmath et al., 330–333. Oxford: Oxford University Press, 2005.

وتُفصح الدلالة اللسانية لهذه الظاهرة عن حقيقة أنَّ اللغة العربية هي لغة "اسمية-فعلية متوازنة"، الأمر الذي يمنحها مرونة تعبيرية استثنائية، حيث توفّر للمتحدّث فرصة الاختيار بين الأسلوب الاسمي (للثبات والاستمرار) أو الأسلوب الفعلي (للحدث والتجدُّد)[75].

ويُظهر الجدول (7) نتائج مؤشّر الإنتاجية الاشتقاقية للفئات النحوية، حيث متوسّط المشتقّات لكلّ جذر، وقيمة الانحراف المعياري في هذه القيم.

الجدول (7): مؤشّر الإنتاجية الاشتقاقية لجذور الفئات النحوية في القاموس المحيط.

الفئة النحوية	متوسّط المشتقّات لكلّ جذر	الانحراف المعياري
أسماء	2.14	3.87
أفعال	1.89	2.43
صفات	0.87	0.64

ويبدو واضحًا أنَّ فئة الأسماء تتمتّع بأعلى إنتاجية اشتقاقية (2.14 مشتقّ لكلّ جذر)، والتي تعكس تنوُّع أنواع الأسماء في العربية (اسم الفاعل، المفعول، الزمان، المكان، الآلة، المصدر، إلخ)، وسهولة اشتقاق الأسماء من معظم الجذور. وأنَّ الأفعال تأتي في المرتبة الثانية (1.89)، مع انحراف معياري أقلّ (2.43)، ممّا يؤشّر إلى توزيع أكثر انتظامًا، وأنَّ هناك قلّة من الجذور التي تُنتج عددًا كبيرًا من الأفعال المزيدة (10-15 صيغة)[76].

أمّا الصفات فقد تميّزت بأنّها ذات إنتاجية محدودة (0.87)، مع انحراف معياري صغير جدًّا (0.64)، ممّا يعني أنَّ معظم الجذور تُنتج صفة واحدة أو قد لا تُنتج صفة مستقلّة، وأنَّ الصفات غالبًا ما تُشتقّ من الأسماء أو الأفعال.

أمّا بصدد نسبة المصادر إلى الأفعال، فقد بلغ عدد المصادر المستقلّة معجميًّا 18,643، ممّا يجعل من قيمة نسبة المصادر إلى الأفعال (0.58:1)، والتي تُفصح عن أنَّ هناك مصدرًا مُدرَجًا كمدخل مستقلّ لكلّ فعلين تقريبًا، الأمر الذي يعكس الاستقلالية المعجمية للمصادر، وأنَّ المصدر ليس مجرّد مشتقّ من الفعل، بل وحدة معجمية مستقلّة، وأنَّ كثيرًا من المصادر لها معانٍ خاصّة لا تُستنبَط مباشرة من الفعل.

من أجل هذا، نهج الفيروزآبادي في القاموس المحيط إلى إدراج المصادر كمداخل مستقلّة عندما تكون ذات أهمّيّة معجمية خاصّة. أمّا بالنسبة للمصادر القياسية البسيطة فقد تُذكَر تحت الفعل دون إفراد مدخل لها.

[75] . Ryding, Karin C. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press, 2005, 59–62.

[76] . Ryding, Karin C. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press, 2005, 528–534.

5. 4. 2. مؤشّر التوازن النحوي الشامل(Grammatical Balance Index)

يستخدم مؤشّر التوازن النحوي الشامل لتقدير درجة الانتظام والتناسق النسبي في توزيع الفئات النحوية داخل مدوّنة لغوية أو معجمية، من حيث تنوُّعها، وكثافتها، واستقرار علاقاتها الوظيفية، وذلك عبر تحليل إحصائي بنيوي يقيس التناسب بين الفئات النحوية الكبرى (اسم، فعل، حرف)، وأنماط الإسناد، والفئات التركيبية، ومستويات التعقيد النحوي، مقارنةً بمعيار مرجعيّ يُمثّل الاستعمال اللغوي الطبيعي أو النمط المثالي للغة[77].

ويُستخدَم مقياس إنتروبيا شانون (Shannon Entropy) لحساب مؤشّر التوازن النحوي في اللسانيات المحوسبة، بواسطة الصيغة الرياضية[78]:

$$H = -\Sigma\, p_i \ln(p_i)$$

حيث يعبّر المتغيّر:

- H = قيمة إنتروبيا شانون.
- p_i = النسبة المئوية لكلّ بنية من الفئات النحوية.

وتُظهر القائمة أدناه قيم مؤشّر إنتروبيا شانون لكلّ فئة من الفئات النحوية في معجم القاموس المحيط:

الفئة النحوية	قيمة الإنتروبيا
الأسماء	0.527
الأفعال	0.527
الصفات	0.198
الحروف	0.057
بُنى أخرى	0.016
المجموع الكلّي	1.325

وفق الأسس النظرية، فإنَّ القيمة القصوى للبُنى اللغوية تكون 2.322 في حالة تساوي نسبة جميع الفئات النحوية في أيّ نصّ من النصوص (قيمة$\log_2 5 = 2.322$). وعلى هذا الأساس، فإنَّ نسبة التوازن النحوي في معجم القاموس المحيط هي 57.1% والتي حصلنا عليها من تقسيم المجموع الكلّي لمؤشّر إنتروبيا شانون على القيمة المثالية القصوى للبُنى اللغوية.

وتُعدّ هذه القيمة مؤشّرًا على:

- أنَّ هناك تركُّزًا كبيرًا في فئتين هما (الأسماء والأفعال) واللتين تُشكّلان 94.69% من المداخل.
- الفئات الأخرى (الصفات، الحروف، أخرى) تُعدّ هامشية نسبيًّا.

أمّا الدلالات اللسانية العامّة للتوازن النحوي الذي ظهر جليًّا في القاموس المحيط، فيمكن تفسيرها من خلال القواعد الآتية:

[77] . Ryding, Karin C. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press, 2005, 47–52.

[78] . Cover, Thomas M., and Joy A. Thomas. *Elements of Information Theory*. 2nd ed. Hoboken, NJ: Wiley, 2006, 13–18.

القاعدة الأولى :أنَّ اللغة العربية لغة "ثنائية النواة(Bi-Nuclear Language) " ، بلغت فيها نسبة النواة الاسمية (56.27%) التي تغطّي التسميات والتصنيفات والمفاهيم المجرّدة. أمّا النواة الفعلية فقد بلغت نسبتها (38.42%) والتي تغطّي الأحداث والأفعال والتحوُّلات. بينما تُعدّ بقية الفئات النحوية هامشية ووظيفية. وتُسهم هذه الظاهرة في منح اللغة العربية قدرة تعبيرية مزدوجة تتأرجح بين: الثبات والحركة، الكينونة والصيرورة[79].

القاعدة الثانية :تتميّز اللغة العربية بمرونة تصنيفية نتيجة لمرونة الحدود المقامة بين الفئات النحوية. والشواهد على ذلك أنَّ:

- اسم الفاعل يعمل اسمًا، وصفة، وفعلاً (في بعض السياقات).
- المصدر يعمل اسمًا، ومفعولاً مطلقًا.
- الصفة المشبّهة تتداخل مع الاسم.

وتُثري هذه المرونة الإمكانات التعبيرية لمفردات اللغة العربية، لكنّها تُورثها، بالوقت ذاته، تعقيدًا على صعيد التصنيف المعجمي[80].

القاعدة الثالثة :تتّسم اللغة العربية بسمة الاقتصاد في الفئات الوظيفية. ذلك لأنَّ كون الحروف تشكّل فئة صغيرة جدًّا (0.86%) يمكن أن يعكس مبدأً اقتصاديًّا، ذلك لأنَّ[81]:

- الحروف فئة مغلقة تؤدّي وظائف نحوية محدَّدة.
- اللغة لا تحتاج إلى توسيع هذه الفئة النحوية، بل تستثمر في الفئات المفتوحة (الأسماء والأفعال).
- الاقتصاد في الفئات الوظيفية (الحروف) مقابل الثراء في الفئات المفتوحة.

5. 3. تحليل التشابه الدلالي في القاموس المحيط

يُستخدَم مقياس التشابه الدلالي (Semantic Similarity) لقياس درجة التقارب في المعنى بين وحدتين لغويتين (كلمتين، جذرين، مفهومين) في اللسانيات الحاسوبية المعاصرة[82]. ويبرز في هذا الإطار التحليلي، مفهوم المعجم الذهني (Mental Lexicon)بوصفه البنية المعرفية الداخلية التي تُخزَّن فيها الوحدات المعجمية في ذهن المتكلّم، لا بوصفها قائمة كلمات جامدة، بل باعتبارها شبكة ديناميكية من العلاقات اللسانية المتداخلة. ولا يُخزّن المعجم الذهني الكلمات بصيغها الخطّيّة فقط، بل يخزّن المدخل المعجمي[83] (Lexical Entry) الذي يضمّ:

- الجذر أو الأساس الصرفي.
- البنية المورفولوجية الممكنة.
- السمات النحوية (النوع، التعدّي، الفئة الصرفية).
- والخصائص الدلالية الأساسية.

---

[79] . Ryding, Karin C. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press, 2005, 59–66.

[80] . Ryding, Karin C. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press, 2005, 112–118.

[81] . Ryding, Karin C. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press, 2005, 48–52.

[82] . Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd ed. Draft, 2023, 945.

[83] . Aitchison, Jean. *Words in the Mind: An Introduction to the Mental Lexicon*. 4th ed. Oxford: Blackwell, 2012, 56–61.

ويتميّز المعجم الذهني بكونه منظَّمًا وفق شبكات علائقية لا وفق ترتيب أبجدي، حيث ترتبط الوحدات المعجمية بعلاقات مثل: الاشتقاق، والترادف والتضادّ، والحقول الدلالية، والتقارب الاستعمالي(Collocational Proximity) . ولا ينفصل المعجم الذهني عن النظام النحوي، بل يحتوي كلّ مدخل معجميّ على معلومات انتقائية (Selectional Properties) تحدّد إمكاناته التركيبية، ممّا يجعل المعجم الذهني محرّكًا أساسيًّا للبناء النحوي لا مخزنًا ثانويًّا. كما أنّه يعمل وفق مبدأ الاقتصاد اللساني؛ إذ يُخزَّن الحدّ الأدنى من المعلومات، بينما تُولَّد الصيغ والتراكيب عند الاستعمال عبر قواعد اشتقاقية وتركيبية، لا عبر التخزين الكامل لكلّ صورة لفظية[84].

وتتوفّر مقاييس لسانية عدّة لقياس مستويات التشابه الدلالي، والتي تنقسم إلى:

1. مقاييس ترتكز على البنية المعجمية :تُمارَس فيها عمليات قياس عدد المشتقّات المشتركة، والتشابه في الجذور، والتشارك في الأوزان الصرفية.
2. مقاييس تعتمد على تحليل السياق :حيث تُمارَس فيها عمليات احتساب التوزيع المشترك(Distributional Similarity) ، والتضمين اللغوي(Word Embeddings) ، والنماذج السياقية.(Contextual Models)
3. مقاييس هجينة :توظَّف لدمج المعلومات الصرفية والسياقية، أو استخدام الشبكات الدلالية القائمة (WordNet, BabelNet).

ويرتبط التشابه الدلالي ارتباطًا وثيقًا بالاقتصاد اللساني، لأنَّ مبدأ الترادف يمكن أن يُعدّ اقتصادًا لغويًّا لأنّه يستثمر التنوّع الأسلوبي والظلّ المعنوي لكلّ مرادف في توسيع نطاق الدقّة الدلالية، واختيار الكلمة الأنسب للسياق[^85]. كما أنّه كلّما كانت الشبكة الدلالية أكثر كثافة (أي: كلمات أكثر ترابطًا)، كانت اللغة أكثر ترابطًا، الأمر الذي يوسّع سمة التوسُّع الاستعاري، والتعلّم اللغوي، وهي أمور تصبّ في مفهوم الاقتصاد اللساني[85].

5. 3. 1. <u>التشابه الدلالي في البنية المعجمية للقاموس المحيط</u>

تكشف البيانات التي أُودعت في الجدول (8) عن ظاهرة التوزيع التفاضلي للتشابه الدلالي (Differential Distribution of Semantic Similarity)عبر الفئات المعجمية المختلفة في القاموس المحيط .وتعكس هذه الظاهرة طبيعة البنية العميقة للنظام الدلالي في اللغة العربية، كما أنّها تُسهم في الكشف عن مبادئ التنظيم المعجمي التي تحكم العلاقات بين الوحدات المعجمية.

---

[84] . Aitchison, Jean. *Words in the Mind: An Introduction to the Mental Lexicon*. 4th ed. Oxford: Blackwell, 2012, 61–66.

[85] . Daniel Jurafsky and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*, 3rd ed. (online draft), released January 6, 2026, https://web.stanford.edu/~jurafsky/slp3.

الجدول (8): تحليل البنية الصرفية والكثافة الدلالية في القاموس المحيط.

البنية المعجمية	متوسّط التشابه	الانحراف المعياري	أعلى تشابه	أدنى تشابه	عدد العناصر
جذور ثلاثية	0.342	0.187	0.956	0.012	44,127
جذور رباعية	0.289	0.156	0.923	0.008	16,234
جذور خماسية	0.234	0.134	0.867	0.005	2,891
أسماء	0.387	0.203	0.981	0.015	47,412
أفعال	0.318	0.174	0.947	0.011	32,368
صفات	0.429	0.218	0.989	0.019	3,628
الأوصاف الجسمانية	0.512	0.241	0.993	0.034	8,427
الحركة والانتقال	0.467	0.226	0.978	0.028	6,891
المشاعر والعواطف	0.548	0.256	0.996	0.041	3,214
العلوم الدينية	0.491	0.234	0.987	0.037	2,986
النباتات والزراعة	0.376	0.198	0.952	0.022	5,734
الحيوانات والطيور	0.394	0.205	0.961	0.025	4,982

وتكشف البيانات عن علاقة عكسية بين التعقيد الصرفي (Morphological Complexity) والتشابه الدلالي، فالجذور الثلاثية تُظهر أعلى معدّل تشابه بين الجذور، الأمر الذي يعكس مبدأ الاقتصاد المعجمي[86] .(Lexical Economy) ولمّا كانت الجذور الثلاثية النواة الأساسية للاشتقاق، من أجل هذا فإنّها تحمل الحمولة الدلالية الأكبر، كما أنَّ قيمة الانحراف المعياري (0.187) تشير إلى تنوُّع دلالي واسع، ومرونة اشتقاقية مميّزة للجذر الثلاثي.

وتأتي بيانات الجذور الرباعية والخماسية لتعلن أنَّ التناقص المطّرد الذي يميّزها عن الجذور الثلاثية يكشف عن مبدأ التخصُّص الدلالي .(Semantic Specialization) فكلّما ازداد التعقيد الصرفي، ازدادت التحديدية الدلالية (Semantic Specificity)، كما أنّها غالبًا ما تكون معرّبة أو مولّدة لمعانٍ متخصّصة[87].

أمّا الترتيب التفاضلي بين الفئات النحوية: الصفات (0.429)، والأسماء (0.387)، والأفعال (0.318)، فإنَّ هذا التدرُّج يعكس طبيعة البنية المفهومية (Conceptual Structure) التي تتميّز بها لغتنا العربية[88]. فالصفات تُظهر أعلى تشابه دلالي، وتعمل على متّصلات دلالية[89] (Semantic Scales) ، مثل: كبير/صغير، جميل/قبيح، مع امتلاكها لحقول دلالية كثيفة.

[86] . Haspelmath, Martin. "Frequency vs. Iconicity in Explaining Grammatical Asymmetries." *Cognitive Linguistics* 19, no. 1 (2008): 30.،

[87] . Haspelmath, Martin, and Andrea D. Sims. *Understanding Morphology*. 2nd ed. London: Routledge, 2010, 64–71.

[88] . فاضل السامرائي .*معاني الأبنية في اللغة العربية* .عمّان: دار عمّار، 2007، 41–48، 112–118.

[89] . **المتّصلات الدلالية** هي بُنى مفهومية منظَّمة تمثّل مجالًا دلاليًا مستمرًا تُرتَّب فيه القيم أو الصفات على امتداد طيفٍ واحدٍ بين قطبين متقابلين أو أكثر، بحيث يُفهم المعنى بوصفه **درجةً قابلةً للقياس النسبي** داخل هذا الطيف، لا قيمةً ثنائيةً منقطعة. وبعبارة تقنية: هي تمثيل رتبي (ordinal structure) لمجموعة من السمات أو المحمولات المتدرجة (gradable predicates)تُسنَد إلى كيانٍ ما بدرجات متفاوتة، وتخضع لعمليات المقارنة، والتكثيف، والكمّ، والتضمين التداولي.

أمّا الأسماء، فإنَّ احتلالها للموقع الوسيط يعكس مبدأ التنوُّع في المفاهيم، لكونها تغطّي مجالات دلالية متباينة (ملموسة/مجرّدة، ومحسوسة/معنوية). وتأتي الأفعال أخيرًا بأدنى تشابه دلالي لأنّها تعكس التنوُّع البنيوي للحدث، كما أنّها تمتلك بنية حجاجية (Argument Structure) معقّدة تميّزها دلاليًّا[90].

الجدول (9): بنية التدرُّج الترادفي اللغوي - الثنائيات الأكثر تشابهًا في القاموس المحيط.

الرتبة	الكلمة الأولى	الكلمة الثانية	درجة التشابه	نوع العلاقة	المجال الدلالي
1	فَرِحَ	سُرَّ	0.996	ترادف تامّ	المشاعر
2	حَزِنَ	كَئِبَ	0.994	ترادف تامّ	المشاعر
3	شُجاع	بَطَل	0.993	ترادف قويّ	الأوصاف
4	جَبان	خائِف	0.991	ترادف قويّ	الأوصاف
5	عَلِمَ	عَرَفَ	0.989	ترادف قويّ	المعرفة
6	قالَ	تَكَلَّمَ	0.987	ترادف متوسّط	القول
7	ذَهَبَ	مَضى	0.985	ترادف تامّ	الحركة
8	جاءَ	أتى	0.984	ترادف تامّ	الحركة
9	كَبير	عَظيم	0.982	ترادف قويّ	الأوصاف
10	صَغير	ضَئيل	0.981	ترادف قويّ	الأوصاف
11	حَسَن	جَميل	0.979	ترادف قويّ	الأوصاف
12	قَبيح	دَميم	0.977	ترادف قويّ	الأوصاف
13	سَريع	عاجِل	0.975	ترادف متوسّط	الأوصاف
14	بَطيء	مُتَمَهِّل	0.973	ترادف متوسّط	الأوصاف
15	أبيَض	ناصِع	0.971	ترادف قويّ	الألوان
16	أسوَد	قاتِم	0.969	ترادف متوسّط	الألوان
17	كَريم	جَواد	0.967	ترادف تامّ	الأوصاف الخُلُقية
18	بَخيل	شَحيح	0.965	ترادف تامّ	الأوصاف الخُلُقية
19	صادِق	أمين	0.963	ترادف قويّ	الأوصاف الخُلُقية
20	كاذِب	مُفْتَرٍ	0.961	ترادف قويّ	الأوصاف الخُلُقية

تكشف بيانات الجدول عن طيف متّصل من التدرُّج الترادفي(Synonymy Gradience) ، وأنَّ الكثافة الدلالية المجالية (Domain-Specific Semantic Density)قد تركّزت في: المجالات النفسية-الانفعالية، والأوصاف الأخلاقية، والأوصاف الجسمانية، الأمر الذي يعكس الأولويات المفاهيمية (Conceptual Priorities) في الثقافة اللسانية العربية[91].

[90] . أحمد مختار عمر .*علم الدلالة* .القاهرة: عالم الكتب، 2005، 233–241.

[91] . فاضل السامرائي .*معاني الأبنية في اللغة العربية* .عمّان: دار عمّار، 2007، 83–91.

ويمكن لهذه البيانات أن تشكّل أساسًا أوّليًّا لنظرية دلالية كمّية تسعى إلى الربط بين البنية الصرفية، والتنظيم المعجمي، وأنماط التوزيع الإحصائي للتشابه الدلالي، بواسطة نماذج محوسبة للبنية الدلالية تتجاوز الأنساق التقليدية في المعالجة والتحليل. كما أن القيم المقاسة تكشف عن أنماط عميقة في تنظيم المعجم العربي تتجاوز الوصف التقليدي إلى النمذجة الرياضية للبنية الدلالية.

ولتوسيع نطاق الكشف عن التشابه الدلالي، أعددنا الجدول (10) الذي أدرجنا فيه مستويات حضور التشابه الدلالي في القاموس المحيط.

الجدول (10): مستويات التشابه الدلالي.

المستوى	نطاق التشابه	الوصف	أمثلة
شبه تطابق	0.95 - 1.0	ترادف تامّ أو شبه تامّ	فرح/سرور، ذهب/مضى
ترادف قويّ	0.85 - 0.94	معانٍ متقاربة جدًّا	كريم/جواد، شجاع/بطل
ترادف متوسّط	0.70 - 0.84	معانٍ متقاربة	قال/تكلّم، كبير/عظيم
تقارب دلالي	0.50 - 0.69	حقل دلالي مشترك	فرح/ضحك، حزن/بكاء
تداعٍ بعيد	0.30 - 0.49	علاقة غير مباشرة	نار/دخان، ماء/عطش
تباعد	0.00 - 0.29	لا علاقة واضحة	حجر/سعادة

ويمكن استثمار نطاق ومستويات التشابه في تحديد سمة التوزيع الهرمي في القاموس المحيط -انظر الجدول.(11)

الجدول (11): مستويات التوزيع الهرمي في القاموس المحيط.

المستوى	النسبة من إجمالي الأزواج
شبه تطابق	2.3%
ترادف قويّ	8.7%
ترادف متوسّط	14.2%
تقارب دلالي	23.5%
تداعٍ بعيد	31.8%
تباعد	19.5%

ويعكس هذا التوزيع الهرمي سمة الاقتصاد اللساني في اللغة العربية، فقلّة الترادف التامّ (2.3%) تُعلن أنَّ اللغة لا تُهدر الموارد اللسانية في تكرار تامّ. كما أنَّ كثرة التقارب (46.4% في المستويات الثلاثة الأولى) تُعلن عن حضور شبكات دلالية كثيفة في المجالات الحيوية، بينما تشير نسبة التوازن (51.3% تداعٍ بعيد + تباعد) إلى حقيقة أنَّ معظم الكلمات متباعدة، ممّا يمنح الخطاب اللساني تنوُّعًا وخصوبة دلالية فائقة[92].

[92] . أحمد مختار عمر .*علم الدلالة* .القاهرة: عالم الكتب، 2005، 214–222.

6. خلاصة ختامية واستنتاجات:

تمثّل الذخيرة اللغوية المستودَعة في المعاجم العربية الكبرى سجلاً معرفياً تراكمياً يختزل التجربة الحضارية للأمة في بنى صرفية ودلالية وصوتية محكمة الانتظام، ويُفصح الاقتصاد اللساني الكامن في هذه البنى عن قدرة اللغة العربية على بلوغ أقصى كفاءة تواصلية بأدنى كلفة إنتاجية وإدراكية، وهي قدرة لا تتكشّف أبعادها الحقيقية إلّا حين تُخضَع المادة المعجمية لأدوات التحليل الكمّي المحوسب ونماذج نظرية المعلومات التي تتجاوز الأحكام الانطباعية إلى البراهين الرياضية القابلة للاختبار والتكرار. ويحتلّ القاموس المحيط للفيروزآبادي موقعاً استثنائياً في هذا السياق، لا بوصفه مرجعاً معجمياً تاريخياً فحسب، بل بوصفه مدوّنة لسانية تختزن في بنيتها العميقة منظومةً من القوانين التوزيعية التي تجسّد مبادئ الاقتصاد اللساني على مستوياته كافّة.

وقد أثبتت التحليلات المحوسبة التي أُجريت على 84,263 مدخلاً معجمياً و50,488 جذراً لغوياً أنَّ البنية الثلاثية تهيمن بنسبة 87.4% من الذخيرة الجذرية، مؤكّدةً أنَّ النظام الصرفي العربي قد بلغ نقطة التوازن المُثلى بين الاختصار الصوتي والقدرة التمييزية، إذ يتيح الجذر الثلاثي أوسع فضاء اشتقاقي بأدنى كلفة فونولوجية وفق مبدأ الجهد الأدنى الذي صاغه زيبف. وأفضت المعالجات الإحصائية إلى استخلاص نمط توزيعي لوغاريتمي شديد التفاوت، حيث تُنتج 20% من الجذور 72% من المشتقات بمعامل جيني بلغ 0.847، مما يكشف عن استراتيجية التركيز الذكي التي تستثمر الطاقة الاشتقاقية في نواة محدودة من الجذور المحورية دلالياً — يتصدّرها ثالوث المعرفة والتعبير والفعل (ع.ل.م: 247، ق.و.ل: 189، ع.م.ل: 176) — بينما تحتفظ بجذور متخصّصة لتغطية الحقول الهامشية.

وعلى المستوى الفونولوجي، أظهرت تحليلات التعاقبية الصوتية، حضور ثلاثة مبادئ بنيوية متكاملة: التباين الأقصى الذي يُفضّل التنويع الصوتي على التكرار (نسبة الجذور المضاعفة 7.62%)، والسيولة المركزية التي تُوطّن الأصوات الانزلاقية في موقع العين، والقوة الطرفية التي تُرسي الأصوات المفخّمة في مطلع الجذر ومنتهاه. وأكّدت هيمنة الفتحة (56.3%) والمقاطع البسيطة (77.1%) اشتغال النظام الصوتي وفق مبدأ الطاقة الأدنى. أمّا على صعيد نظرية المعلومات، فقد بلغت الإنتروبيا 95.1% على مستوى الجذور و84.4% على مستوى الأوزان، وهما قيمتان تقتربان من الحدّ الأقصى النظري وتدلّان على توازن إنتروبي منضبط يستثمر الفضاء المعلوماتي بكفاءة استثنائية. كما أظهر التوازن النحوي أنَّ العربية لغة ثنائية النواة تتأرجح بين القطب الاسمي (56.27%) والفعلي (38.42%)، ممّا يمنحها مرونة تعبيرية مزدوجة بين الثبات والصيرورة.

وأفصحت الشبكات الدلالية عن كثافة تفاضلية تعكس الأولويات المفاهيمية للثقافة العربية، إذ بلغت ذروتها في حقل المشاعر (0.548) والأوصاف الجسمانية (0.512)، فيما لم يتجاوز الترادف التامّ 2.3% من أزواج المفردات، مؤكّداً أنَّ اللغة تتجنّب الإسراف في تخزين وحدات متطابقة وتستثمر في شبكات التقارب الدلالي الكثيفة. وتأسيساً على هذه النتائج، نوصي بتوجيه الدراسات اللاحقة نحو: إجراء تحليل مقارن منهجي بين القاموس المحيط ولسان العرب والصحاح بالمؤشرات الكمّية ذاتها، وتوظيف نماذج التعلّم العميق لبناء شبكات دلالية ديناميكية تُحاكي المعجم الذهني العربي، وكذلك دراسة التطوّر التعاقبي للإنتاجية الاشتقاقية عبر المعاجم المتعاقبة للكشف عن قوانين النموّ والانقراض المعجميين، وتطوير نماذج احتمالية تنبُّؤية لقيود االتعاقب الصوتي، يمكن أن تُسهم في تحسين أنظمة التعرّف الآلي على الكلام ومعالجة النصوص التراثية.

Linguistic Economy and Lexical Inventory in Al-Qamus Al-Muhit

An Analytical Study Using Computational Linguistics Tools

Hasan M. Al-Rizzo

Senior Member

ABSTRACT

This study investigates the lexical repertoire and the intrinsic linguistic economy embedded in *al-Qāmūs al-Muḥīṭ* by al-Fīrūzābādī, employing computational linguistics and Information Theory as its principal analytical frameworks. It seeks to move beyond traditional approaches that have long governed the study of Arabic lexicography without adequate integration of quantitative computational methodologies. Adopting a quantitative, descriptive, and analytically integrative design, the research brings together theoretical and computational linguistics, Information Theory, and quantitative linguistic statistics. The analytical toolkit includes automated extraction of roots and morphological patterns through an intelligent algorithmic suite, computation of entropy and informational frequency indices, application of Zipf's law, and the construction of semantic networks. The findings reveal systematically organized patterns of linguistic economy governing the internal architecture of the lexicon. These patterns manifest in morphologically productive structures that follow a highly skewed logarithmic distribution, whereby a limited subset of roots accounts for a disproportionate share of derived forms. The predominance of triliteral root structures reflects an optimal balance between economy and distinctiveness. Phonological configurations further instantiate the principle of least effort through the preferential distribution of light vowels and simple syllabic structures.

Entropy measures indicate exceptional diversity at both the root and morphological pattern levels, while semantic network analysis uncovers differential density patterns across knowledge domains. From these results, four foundational laws of Arabic linguistic economy are formulated: intelligent concentration, optimal tri laterality, minimal energy, and entropic equilibrium.

The study advances an integrative model for probing lexical resources and linguistic economy, providing precise quantitative evidence that can be leveraged in artificial intelligence and natural language processing applications. It also establishes methodological foundations for the emerging field of quantitative Arabic lexicology.

Keywords: Lexical Repertoire; Linguistic Economy; *Al-Qāmūs Al-Muḥīṭ*; Information Theory; Morphological Productivity; Computational Linguistics; Zipf's Law; Entropy.