

## التحليل الآلي للنصوص التراثية العربية: برنامج AntConc نموذجاً

## Automated Analysis of Arabic Heritage Texts: A Case Study of AntConc

د. عبد الغاني سي موسى\*

جامعة وهران 1 أحمد بن بلة، [simoussa.abdelghani@edu.univ-oran1.dz](mailto:simoussa.abdelghani@edu.univ-oran1.dz)

أ. د. سعد الله زهرة

جامعة وهران 1 أحمد بن بلة، [saadallahchahra@gmail.com](mailto:saadallahchahra@gmail.com)

تاريخ النشر: 2026/10/14

تاريخ القبول: 2026/07/25

تاريخ الاستلام: 2026/06/20

**ملخص:** في ظل التحول الرقمي المتسارع، شهدت طرائق التعامل مع النصوص التراثية والمعاجم تطوراً ملحوظاً، حيث لم تعد مقتصرة على الأساليب التقليدية في الفهم والتحليل، بل أصبحت تعتمد على أدوات حاسوبية متطورة تتيح مقارنة هذه النصوص وفق منهجية دقيقة، وقد أسهم هذا التحول في فتح آفاق جديدة أمام الباحثين؛ من خلال تمكينهم من معالجة كميات كبيرة من النصوص، واستخراج معطياتها اللغوية بكفاءة، الأمر الذي عزز من إمكانيات البحث في ميدان اللغة والأدب العربي، ومن هذا المنطلق تسعى هذه الدراسة إلى معالجة وتحديد مفهوم التحليل الآلي للنصوص التراثية، مع إبراز أهميته في تطوير البحث اللغوي، واستكمالاً لهذا الطرح تم التعريف ببرنامج (AntConc)، باعتباره وسيلة تحليل حاسوبي تُعتمد في معالجة النصوص واستخراج معطياتها اللغوية وفق إجراءات علمية دقيقة ومنهجية، علاوة على ذلك، تقف هذه الدراسة عند أبرز التحديات والإشكالات التي قد تعترض استخدام هذا البرنامج، سواء على مستوى معالجة النصوص العربية أو على صعيد القيود التقنية والمنهجية التي قد تؤثر في دقة التحليل وتفسير نتائجه.

**كلمات مفتاحية:** اللسانيات الحاسوبية، التحليل الآلي، النصوص التراثية العربية، المدونات النصية، برنامج (AntConc)، المعالجة الحاسوبية للنصوص.

**Abstract:** Advances in computational technologies have significantly transformed approaches to the study of heritage texts and lexicons. Rather than relying solely on traditional methods, researchers now employ computational tools that enable the efficient processing of large textual corpora and the extraction of linguistic data with greater accuracy. This development has expanded research possibilities in the fields of Arabic language and literature. Accordingly, this study explores

the concept of computational analysis of heritage texts and highlights its role in supporting linguistic research. It also introduces AntConc as a text-analysis tool used to process texts and extract linguistic information through systematic procedures. Furthermore, the study discusses the main challenges associated with its use, particularly in relation to Arabic text processing and the technical and methodological limitations that may affect analytical accuracy and the interpretation of results.

**Keywords:** Computational Linguistics; Automated Text Analysis; Arabic Heritage Texts; Text Corpora; AntConc; Computational Text Processing.

## 1. مقدمة:

شهدت العقود الأخيرة تطورًا متسارعًا في مجال التقنيات الرقمية، الأمر الذي أدى إلى ظهور حقول معرفية جديدة تسعى إلى الإفادة من الحاسوب في معالجة الظواهر اللغوية وتحليلها، ويأتي في مقدمة هذه الحقول علم اللسانيات الحاسوبية؛ الذي يُعنى بدراسة اللغة الطبيعية وتطوير الآليات والتطبيقات الحاسوبية القادرة على معالجتها وتحليلها وفق أسس علمية ومنهجية دقيقة، وقد أحدث هذا المجال نقلة نوعية في طرائق البحث اللغوي، من خلال توفير أدوات قادرة على التعامل مع كميات كبيرة من البيانات اللغوية بسرعة وكفاءة تفوق ما تتيحه الأساليب التقليدية.<sup>1</sup>

تُعد النصوص التراثية من أغنى المصادر المعرفية واللغوية التي تزخر بمادة علمية وأدبية واسعة غير أن التعامل معها بالوسائل التقليدية يظل عملية شاقة تستلزم وقتًا وجهدًا كبيرين، خاصة عند الرغبة في تتبع الظواهر اللغوية أو دراسة التكرارات المعجمية أو تحليل السياقات النصية المختلفة ومن هنا برزت أهمية التحليل الحاسوبي الآلي بوصفه مقارنة حديثة تُمكن الباحث من استثمار الأدوات الرقمية في فحص النصوص التراثية واستخراج خصائصها اللغوية بصورة أكثر دقة وموضوعية،<sup>2</sup> ومن بين الأدوات الحاسوبية التي أثبتت فاعليتها في هذا المجال برنامج (AntConc) الذي يُعد من أشهر البرمجيات المخصصة لتحليل المدونات النصية، حيث يوفر مجموعة من الوظائف التي تسمح باستخراج التكرارات اللفظية ودراسة السياقات النصية، وتحليل المتلازمات اللفظية والكشف عن الأنماط اللغوية المختلفة، الأمر الذي جعله أداة مهمة في العديد من الدراسات اللسانية المعاصرة.

ورغم ما تتيحه هذه الأدوات من إمكانيات واسعة في معالجة النصوص وتحليلها، فإن توظيفها في دراسة النصوص التراثية العربية ما يزال يثير جملة من التساؤلات المرتبطة بآليات استخدامها وحدود فاعليتها، ومدى قدرتها على تقديم معطيات لغوية دقيقة يمكن الاستفادة منها في البحث اللساني.<sup>3</sup>

وتأسيساً على ما سبق، تتبلور الإشكالية الرئيسية لهذه الدراسة في التساؤل الآتي:

- كيف يمكن توظيف برنامج (AntConc) في تحليل للنصوص التراثية العربية، وما مدى إسهامه في استخراج المعطيات اللغوية ودعم الدراسات اللسانية العربية المعاصرة؟
- ويتفرع عن هذا التساؤل الرئيسي عدد من الأسئلة الفرعية، من جملتها:
- ما المقصود بالتحليل الآلي للنصوص التراثية، وأين تكمن أهميته؟
- ما هي أبرز وظائف برنامج (AntConc) وإمكاناته في معالجة النصوص وتحليلها؟
- كيف يمكن الاستفادة من البرنامج في استخراج المعطيات اللغوية من النصوص التراثية العربية؟
- ما هي أبرز التحديات التقنية والمنهجية التي قد تواجه الباحث عند توظيف برنامج (AntConc) في معالجة النصوص العربية؟

## 2. ما المقصود بالتحليل الآلي للنصوص التراثية؟

التحليل الآلي للمعاجم والنصوص التراثية هو عملية منهجية تقوم على توظيف الأدوات الحاسوبية المتطورة لتحويل النصوص التراثية إلى صيغ رقمية قابلة للمعالجة، ثم تحليلها آلياً لاستخراج معطياتها اللغوية، ويشمل هذا التحليل دراسة البنية الصوتية والصرفية والتركيبية والدلالية... للنصوص، مما يُسهّل عملية البحث فيها، والكشف عن أنماطها اللغوية وتتبع تطور الألفاظ واستعمالاتها...، فضلاً عن الإسهام في بناء موارد لغوية تحفّظ التراث العربي وتيسّر استثماره.<sup>4</sup>

## 1.2 أهميته:

تبرز أهمية التحليل الآلي للنصوص التراثية فيما يلي:<sup>5</sup>

### 1.1.2 تسهيل البحث والوصول إلى المعلومات:

يُمكّن التحليل الرقمي للنصوص التراثية من البحث السريع داخلها، واستخراج الألفاظ والتراكيب في زمن وجيز، مما يوفر الجهد والوقت مقارنة بالبحث في المصادر الورقية.

### 2.1.2 دعم الدراسات الإحصائية والأسلوبية:

يتيح هذا النوع من التحليل إجراء دراسات كمية تعتمد على الإحصاء اللغوي، مثل حساب تواتر الكلمات، وتحليل توزيعها داخل النصوص، وهو ما يعزز موضوعية النتائج العلمية.

### 3.1.2 تتبع تطور الألفاظ والدلالات:

يساعد التحليل الرقمي للنصوص التراثية الباحث على دراسة تطور اللغة العربية عبر العصور من خلال تتبع استعمال الألفاظ وتغير معانيها في سياقات مختلفة.

### 4.1.2 بناء مدونات ومعاجم رقمية:

يسهم التحليل الرقمي للنصوص التراثية في إنشاء قواعد بيانات لغوية ومدونات نصية (Corpora)، تُستثمر في البحث العلمي، وفي تطوير تطبيقات المعالجة الآلية للغة العربية.

### 5.1.2 الحفاظ على التراث اللغوي:

يسهم التحليل الآلي للنصوص التراثية في رقمنتها وحفظها من الضياع مع إمكانية الوصول إليها بسهولة، الأمر الذي يعزز نشر المعرفة التراثية في البيئة الرقمية.

وفي ضوء ما تقدم، نخلص إلى أنَّ أهمية التحليل الآلي للنصوص التراثية لا تقتصر على الجوانب الإجرائية المرتبطة بسرعة البحث واستخراج المعطيات اللغوية فحسب، بل تتجاوز ذلك إلى الإسهام في تجديد مناهج دراسة التراث العربي ذاته؛ فمن خلال التحليل الآلي للنصوص التراثية ينتقل الباحث من القراءة الجزئية المحدودة إلى القراءة الشمولية القائمة على استثمار المدونات النصية الضخمة وتحليلها وفق معايير كمية دقيقة، الأمر الذي يفتح آفاقاً جديدة لفهم الظواهر اللغوية والأدبية، ويعزز التكامل بين المناهج التراثية والمقاربات اللسانية الحديثة.

### 3. تجربة تحليل مقطع نصي باستخدام برنامج (AntConc)

#### 1.3 التعريف بالبرنامج:

يُعدّ (AntConc) من الأدوات الرقمية المجانية المخصصة لتحليل النصوص (Corpus Analysis)، وقد طوّره الباحث (Laurence Anthony)، حيث يضم مجموعة من الأدوات التي تُمكن الباحث من استخراج الكلمات الأكثر تكرارًا داخل المدونات النصية، وعرضها في سياقاتها المختلفة، فضلاً عن الكشف عن المتلازمات اللفظية والتراكيب المتكررة، وتحليل الأنماط اللغوية الشائعة، كما يوفر إمكانات إحصائية تساعد على وصف الظواهر اللغوية بصورة دقيقة، الأمر الذي يجعله أداة فعالة في الدراسات اللسانية وتحليل النصوص التراثية والأدبية.<sup>6</sup>

يمتاز هذا البرنامج بواجهة بسيطة وسهلة الاستخدام، كما يعمل على مختلف أنظمة التشغيل مثل: (Windows ، MacOS ، Linux)، ويُصنّف ضمن قائمة البرمجيات المجانية (Freeware) حيث يمكن استعماله دون مقابل، وهذا ما يجعله أداة مرنة ومناسبة لمختلف المستخدمين.<sup>7</sup>



الصورة 1: شكل أيقونة البرنامج عل سطح المكتب

رابط تحميل برنامج (AntConc):

<https://www.laurenceanthony.net/software/antconc//>

### 2.3 اختيار النص:

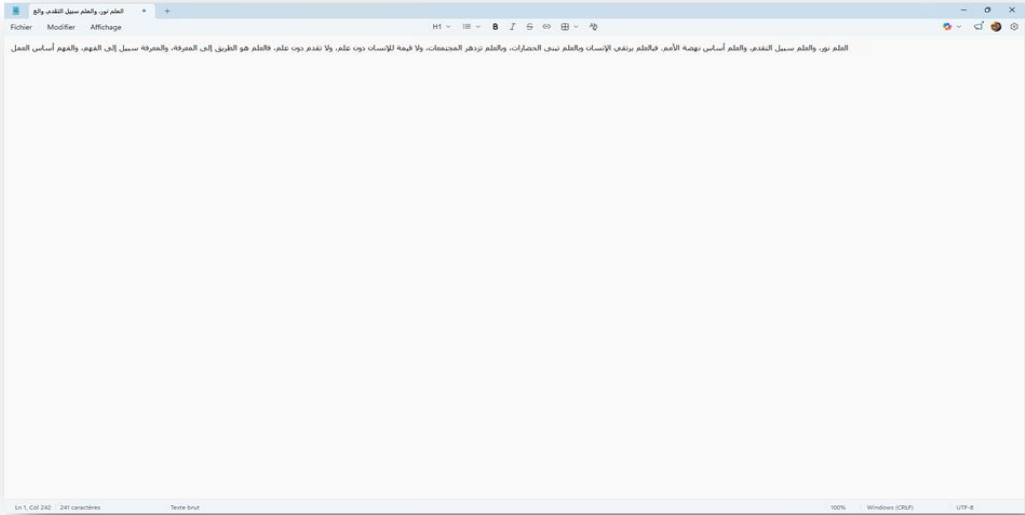
من أجل استعراض أبرز وظائف برنامج (AntConc) وإبراز قدرته على استخراج المعطيات اللغوية وتحليلها، تم اعتماد المقطع النصي التالي:

#### النص المعتمد:

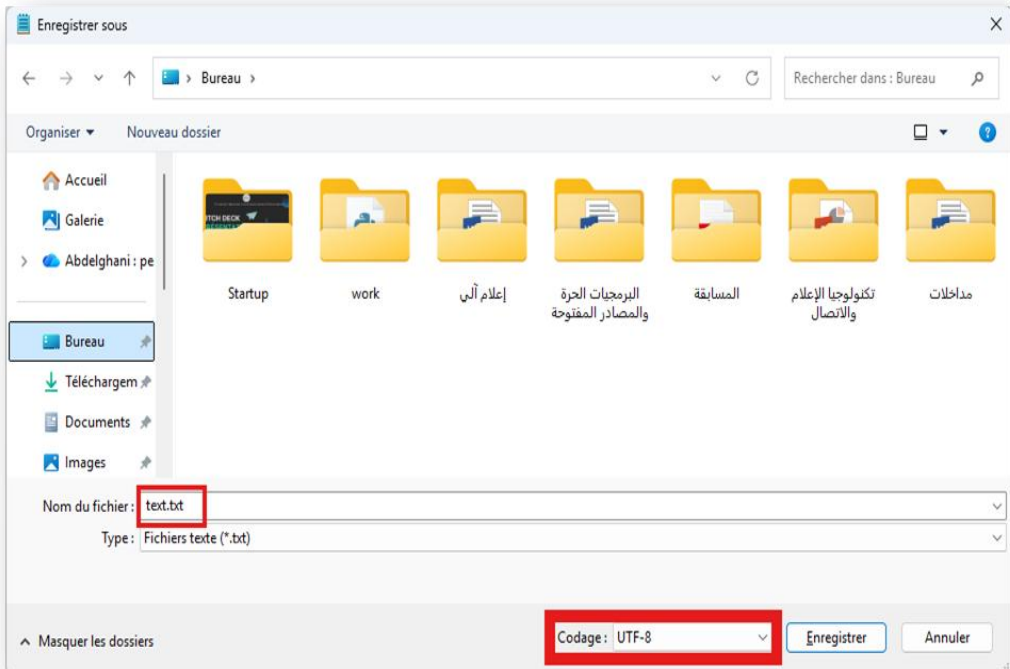
"العلم نور، والعلم سبيل التقدم، والعلم أساس نهضة الأمم. فبالعلم يرتقي الإنسان وبالعلم تبنى الحضارات، وبالعلم تزدهر المجتمعات، ولا قيمة للإنسان دون علم، ولا تقدم دون علم، فالعلم هو الطريق إلى المعرفة، والمعرفة سبيل إلى الفهم، والفهم أساس العمل".

### 3.3 تجهيز النص للتحليل والمعالجة:

بعد اختيار النموذج اللغوي، نقوم بإدخاله في ملف نصي حتى يكون قابلاً للمعالجة داخل البرنامج، ويتم ذلك بفتح محرر نصوص بسيط مثل (Bloc-notes)، ثم لصق النص بداخله (الصورة 2) مع الحرص على حفظ الملف بصيغة UTF-8 (الصورة 3)، حتى تظهر الحروف العربية بشكل صحيح داخل البرنامج، ويمكن تسمية الملف مثلاً: (text.txt).



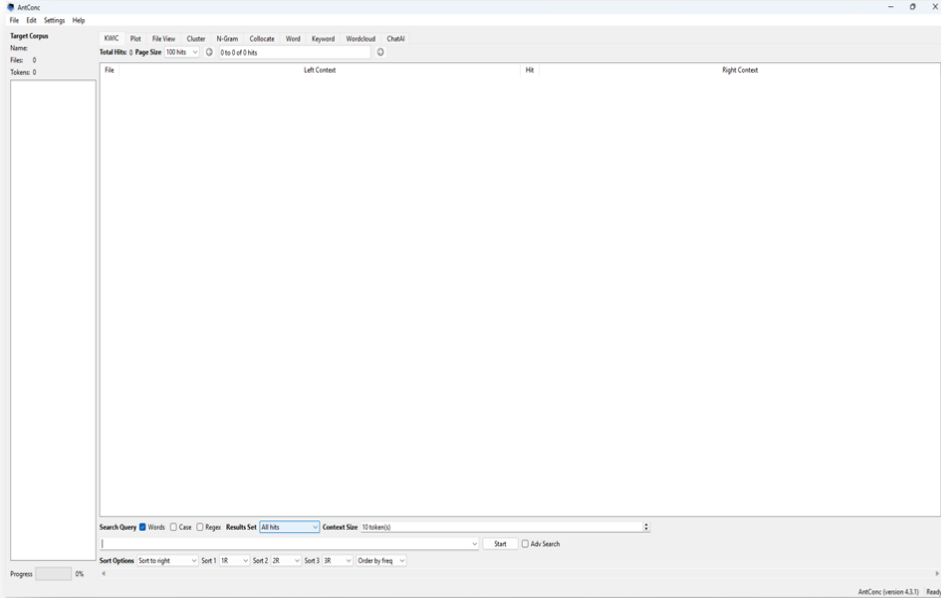
## الصورة 2: لصق النص داخل محرر النصوص



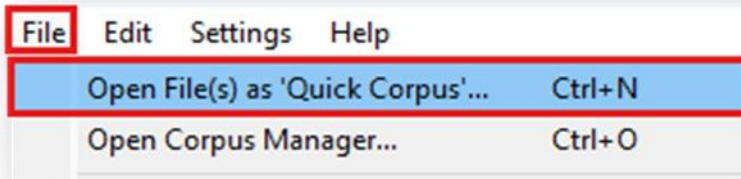
## الصورة 3: حفظ الملف بصيغة UTF-8

### 4.3 إدخال النص في البرنامج:

بعد تشغيل برنامج (AntConc) تظهر لنا الواجهة المبينة في الصورة رقم 4، فنقوم بتحميل الملف النصي من خلال الأمر: (File → Open File(s) as 'Quick Corpus')، ثم اختيار الملف الذي قمنا بحفظه سابقا (الصورة 5)، وبمجرد تحميل النص، يُصبح جاهزاً للتحليل، ويمكن الانتقال بين أدوات البرنامج المختلفة لدراسة محتواه.



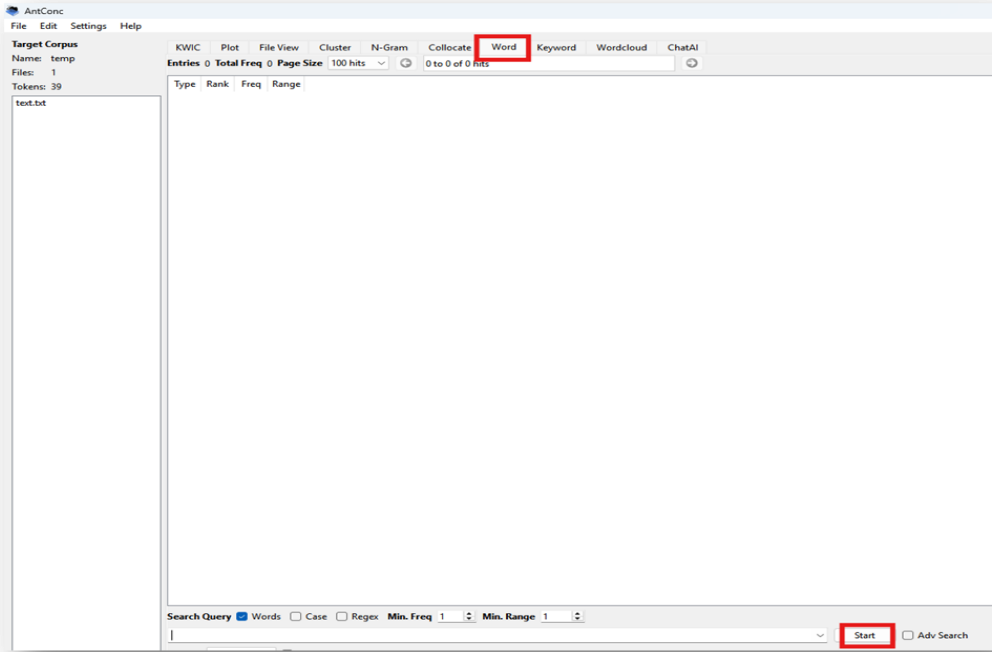
الصورة 4: الواجهة الرئيسية لبرنامج (AntConc)



الصورة 5: خيار تحميل الملف النصي داخل برنامج (AntConc)

### 5.3 استخراج قائمة الكلمات من النص (Word List):

في هذه المرحلة، نحدد خيار (Word) الموجود في أعلى الواجهة الرئيسية، وهو المخصص لعرض قائمة الكلمات الواردة في النص مرتبة حسب عدد مرات تكرارها، ثم نضغط على زر (Start) لبدء التحليل (الصورة 6).



الصورة 6: خطوات بدء عملية استخراج الكلمات من النص داخل البرنامج

### 1.5.3 قراءة نتائج التحليل:

بعد معالجة النص وتحليله، يعرض البرنامج جدولاً يتضمن مجموعة من المؤشرات الإحصائية (الصورة 7)، أهمها:

- **Rank**: يشير إلى ترتيب الكلمة مقارنة بغيرها من حيث التكرار داخل النص، فالكلمات ذات الترتيب (1) هي الأكثر وروداً، ثم تتدرج بقية الكلمات حسب عدد مرات ظهورها.
- **Freq (Frequency)**: يُمثل عدد المرات التي تكررت فيها الكلمة الواحدة داخل النص فمثلاً إذا كانت القيمة (2)؛ فهذا يعني أن الكلمة تكررت مرتين .

- **Range:** يدل على عدد الملفات (النصوص) التي ظهرت فيها الكلمة، وبما أننا نشتغل على نص واحد فقط فإن القيمة تكون (1)، أي أن الكلمة وردت في هذا النص وحده.

|    |           |   |   |   |
|----|-----------|---|---|---|
| 17 | المجتمعات | 9 | 1 | 1 |
| 18 | المعرفة   | 9 | 1 | 1 |
| 19 | تبنى      | 9 | 1 | 1 |
| 20 | تزدهر     | 9 | 1 | 1 |
| 21 | تقدم      | 9 | 1 | 1 |
| 22 | فالعالم   | 9 | 1 | 1 |
| 23 | فبالعلم   | 9 | 1 | 1 |
| 24 | قيمة      | 9 | 1 | 1 |
| 25 | للإنسان   | 9 | 1 | 1 |
| 26 | نهضة      | 9 | 1 | 1 |
| 27 | نور       | 9 | 1 | 1 |
| 28 | هو        | 9 | 1 | 1 |
| 29 | والفهم    | 9 | 1 | 1 |
| 30 | والمعرفة  | 9 | 1 | 1 |
| 31 | يرتقي     | 9 | 1 | 1 |

|    | Type     | Rank | Freq | Range |
|----|----------|------|------|-------|
| 1  | أساس     | 1    | 2    | 1     |
| 2  | إلى      | 1    | 2    | 1     |
| 3  | دون      | 1    | 2    | 1     |
| 4  | سبيل     | 1    | 2    | 1     |
| 5  | علم      | 1    | 2    | 1     |
| 6  | والعلم   | 1    | 2    | 1     |
| 7  | وبالعلم  | 1    | 2    | 1     |
| 8  | ولا      | 1    | 2    | 1     |
| 9  | الأمم    | 9    | 1    | 1     |
| 10 | الإنسان  | 9    | 1    | 1     |
| 11 | التقدم   | 9    | 1    | 1     |
| 12 | الحضارات | 9    | 1    | 1     |
| 13 | الطريق   | 9    | 1    | 1     |
| 14 | العلم    | 9    | 1    | 1     |
| 15 | العمل    | 9    | 1    | 1     |
| 16 | الفهم    | 9    | 1    | 1     |

### الصورة 7: قائمة الكلمات الواردة في النص

#### 2.5.3 ملاحظات:

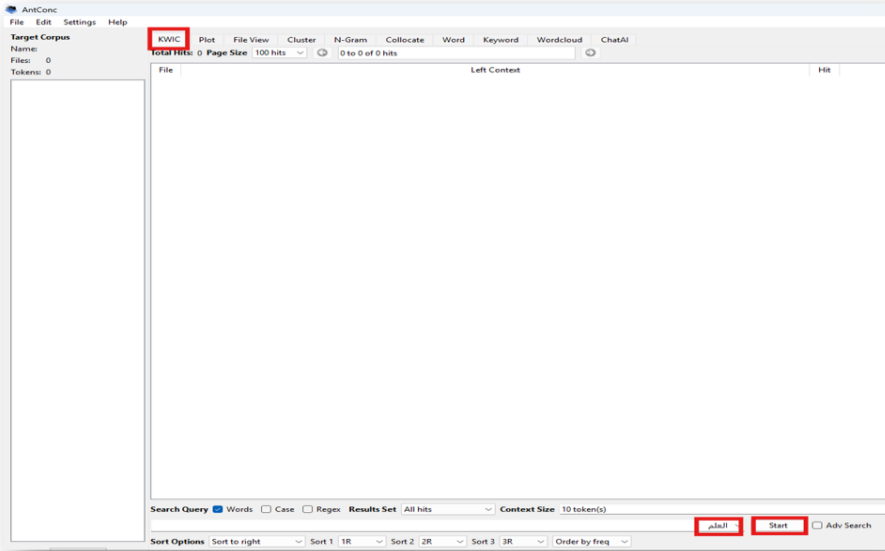
(أ) - يعتمد البرنامج على المطابقة الحرفية، إذ يميّز الكلمات بناءً على شكلها الظاهر كما كُتبت ولا يقوم بتصنيفها اعتماداً على بنيتها الصرفية (كالجذر أو الصيغة الأساسية) وعليه، فإن الكلمات مثل: علم، العلم، بالعلم، تُسجّل بوصفها وحدات مختلفة داخل قائمة النتائج، رغم اشتراكها في نفس الأصل المعجمي، ويرجع ذلك إلى أن البرنامج لا يتعرّف على العناصر المضافة مثل أداة التعريف (ال) أو حرف الجر (الباء) بوصفها مكونات منفصلة، بل يعالج الكلمة في صورتها الكاملة كما تظهر في النص.

(ب)- لا يقوم البرنامج بفصل الحروف المتصلة بالكلمات، إذ يتعامل معها بوصفها جزءاً من بنيتها الكتابية؛ فعلى سبيل المثال، تظهر صيغ مثل: [والعلم] كوحدة واحدة، ولا يتم تفكيكها (و + العلم)، ويرجع ذلك إلى أن البرنامج لا يعتمد آلية لتجزئة الكلمات في اللغة العربية، بل يعالجها في صورتها المتصلة كما ترد في النص.

تُبرز هذه الملاحظات أن التحليل الذي يقدمه برنامج (AntConc) قائم على الشكل الكتابي للكلمات لا على بنيتها اللغوية، مما يفرض على الباحث إجراء مراجعة لاحقة للنتائج.

### 6.3 تحليل الكلمات المفتاحية في سياقات مختلفة (KWIC)

لمعرفة السياقات التي وردت فيها كلمة معينة داخل النص، نحدد خيار (KWIC) من أعلى الواجهة الرئيسية، ثم نقوم بإدخال كلمة "العلم" مثلاً في مستطيل البحث الموجود في الأسفل ونضغط على زر (Start) (الصورة 8).



### الصورة 8: طريقة تحليل كلمة "العلم" في سياقات مختلفة داخل البرنامج

يُفترض أن تظهر النتائج في جدول منظم يضع كلمة "العلم" في المركز (الوسط)، بحيث يستطيع الباحث قراءة "ما قبلها" و "ما بعدها" لاستنتاج المعنى (الجدول 1)، لكننا لاحظنا أن البرنامج لم يُظهر النتائج بالشكل المثالي (الصورة 9).

| السياق السابق | الكلمة المفتاحية | السياق اللاحق   |
|---------------|------------------|-----------------|
| (بداية النص)  | العلم            | نور             |
| و             | العلم            | سبيل التقدم     |
| و             | العلم            | أساس نهضة الأمم |
| فـ            | العلم            | يرتقي الإنسان   |
| وبـ           | العلم            | تُبني الحضارات  |

الجدول 1: النموذج المعياري لتحليل كلمة "العلم" في سياقات مختلفة



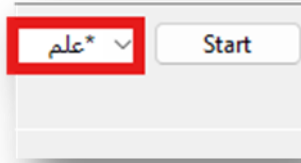
الصورة 9: نتائج تحليل كلمة "العلم" في سياقات مختلفة داخل برنامج (AntConc)

### 1.6.3 شرح وتفسير:

يبحث البرنامج عن الكلمة كما يكتبها المستخدم في مستطيل الإدخال، وقد تجاوز في هذه المرحلة الكلمات التالية: (علم، بالعلم، فالعلم) لأنها بالنسبة له مختلفة تماماً (بسبب الحروف المضافة).

### 2.6.3 الحل المقترح:

بعد أن لاحظنا أنَّ البرنامج لم يُوفَّق في العثور على جميع السياقات التي وردت فيها كلمة "العلم" قمنا باستخدام أمر البحث التالي: [\*علم] (الصورة 10).



**الصورة 10:** استخدام أمر [\*علم] للبحث عن جميع السياقات التي وردت فيها كلمة "العلم"

### 3.6.3 ما معنى الرمز (\*) في اللسانيات الحاسوبية؟

يُسمى هذا الرمز (WildCard)، ووظيفته إخبار البرنامج بما يلي: "ابحث عن الكلمات التي تنتهي بالحروف (ع، ل، م) بغض النظر عما يسبقها من سوابق لغوية"؛ فالحاسوب لا يفهم بالفطرة أن "بالعلم" هي نفسها "علم" فقط سُبقت بالباء و (ال)؛ بل هو يراها سلسلة حروف مختلفة تماماً (ب-ا-ل-ع-ل-م)، والنجمة قبل الكلمة [\*علم] تجعله يقتنص كل الكلمات التي تنتهي بـ "علم" مهما كان ما يسبقها.<sup>8</sup>

### 4.6.3 نتيجة:

بفضل هذا الرمز (\*)، استطاع البرنامج اصطياد جميع السياقات التي وردت فيها كلمة "العلم" ووضعها في جدول واحد (الصورة 11).

| File      | Left Context                                                        | Hit     | Right Context                                                    |
|-----------|---------------------------------------------------------------------|---------|------------------------------------------------------------------|
| 1 text.bt | العلم نور، والعلوم سبيل التقدم،                                     | والعلم  | أساس نهضة الأمم، فبالعلم يرتقي الإنسان والعلم تبنى الحضارات،     |
| 2 text.bt | والعلم سبيل التقدم، والعلم أساس نهضة الأمم، فبالعلم يرتقي الإنسان   | والعلم  | تبنى الحضارات، وبالعلم تزدهر المجتمعات، ولا قيمة للإنسان دون     |
| 3 text.bt | والعلم أساس نهضة الأمم، فبالعلم يرتقي الإنسان والعلم تبنى الحضارات، | والعلم  | تزدهر المجتمعات، ولا قيمة للإنسان دون علمي ولا تقدم              |
| 4 text.bt | العلم نور،                                                          | والعلم  | سبيل التقدم، والعلم أساس نهضة الأمم، فبالعلم يرتقي الإنسان       |
| 5 text.bt | تزدهر المجتمعات، ولا قيمة للإنسان دون علمي ولا تقدم دون             | علمي    | فبالعلم هو الطريق إلى المعرفة، والمعرفة سبيل إلى الفهم           |
| 6 text.bt | المجتمعات، ولا قيمة للإنسان دون علمي ولا تقدم دون علمي،             | فبالعلم | نور، والعلم سبيل التقدم، والعلم أساس نهضة الأمم، فبالعلم         |
| 7 text.bt | والعلم تبنى الحضارات، وبالعلم تزدهر المجتمعات، ولا قيمة للإنسان دون | علمي    | هو الطريق إلى المعرفة، والمعرفة سبيل إلى الفهم والفهم            |
| 8 text.bt | العلم نور، والعلوم سبيل التقدم، والعلم أساس نهضة الأمم،             | فبالعلم | يرتقي الإنسان والعلم تبنى الحضارات، وبالعلم تزدهر المجتمعات، ولا |

**الصورة 11:** نتائج البحث عن السياقات التي وردت فيها كلمة "العلم" باستخدام أمر [\*علم]

### 5.6.3 أهمية تحليل الكلمات المفتاحية في سياقات مختلفة

يساعد تحليل الكلمات المفتاحية في سياقات متنوعة على تحديد المعنى بدقة، إذ لا يتضح مدلول الكلمة إلا من خلال موقعها داخل الجملة وعلاقتها بما يحيط بها من ألفاظ، كما يتيح هذا التحليل أيضاً الكشف عن التصاحب اللفظي؛ أي الكلمات التي ترد معها باستمرار، مما يساعد

على فهم الروابط الدلالية وبناء الحقول المعجمية داخل النص، وبذلك ينتقل الباحث من مجرد إحصاء التكرارات إلى فهم أعمق لطبيعة الاستعمال اللغوي.

#### 4. خاتمة:

من خلال التحليل الذي أجريناه باستعمال برنامج (AntConc)، تمكّنا من التعرف على المعطيات اللغوية الخاصة بالنص والكشف عن موضوعه العام الذي يدور حول "العلم" ويتضح ذلك من خلال تكرار كلمة "العلم" في أكثر من سياق، كما كشفت هذه الدراسة عن بعض الصعوبات المرتبطة بقدرات البرنامج في معالجة اللغة العربية، من أبرزها اعتماده على المطابقة الحرفية؛ حيث تعامل مع الصيغ التالية: علم، العلم، بالعلم على أنها كلمات مختلفة إضافة إلى عدم فصل الحروف المضافة إلى الكلمات، وهو ما قد يؤثر على دقة النتائج الإحصائية.

ورغم هذه الملاحظات المسجلة، يظل البرنامج أداة تحليلية فعالة تتيح جملة من الوظائف المهمة، مثل استخراج قوائم الكلمات وتحديد تكراراتها، وتحليل السياقات اللغوية للكلمات المتكررة عبر خاصية (KWIC)، إلى جانب دراسة المصاحبات اللفظية والكشف عن التراكيب المتكررة (N-grams)، غير أن تحسين دقة نتائجه في معالجة النصوص العربية يقتضي تهيئة النصوص قبل إخضاعها للتحليل، من خلال توحيد صيغ الكلمات ومعالجة السوابق والعناصر المضافة إليها فضلاً عن الاستعانة بأدوات المعالجة الصرفية القادرة على رد الكلمات إلى أصولها المعجمية، كما تبرز الحاجة إلى تطوير حلول برمجية تراعي الخصائص اللغوية للغة العربية، بما يسمح بتجاوز حدود المطابقة الحرفية وتحقيق أكبر قدر من الدقة في التحليل الإحصائي واللساني، ومن شأن هذه الإجراءات أن تعزز فاعلية البرنامج وتوسّع إمكانات الإفادة منه في دراسة النصوص التراثية العربية وتحليلها وفق أسس علمية أكثر دقة وموضوعية.

## 5. قائمة الإحالات:

- <sup>1</sup> - يُنظر: خالد حوير الشمس، اللسانيات الحاسوبية دراسة في المُعطى البيئي، مقال ضمن كتاب جماعي بعنوان: اللسانيات الحاسوبية تنظيراً وتطبيقاً، مركز الكتاب الأكاديمي، عمان، الأردن، ط1، 2021م، ص 30، 31، 32، 33.
- <sup>2</sup> - ينظر: أحمد بن محمد الزهراني وآخرون، المخطوط والتراث اللغوي: خدمة التراث اللغوي المخطوط في مجال اللغة العربية بين الواقع والمأمول، مجمع الملك سلمان العالمي للغة العربية، الرياض، السعودية، ط 1، 1446هـ، ص 39.
- <sup>3</sup> - ينظر: مصطفى كداد، صادق شهيت أوغلو، دور الذكاء الاصطناعي في تحليل النصوص العربية وتفسيرها: دراسة في إشكالية الحمل على اللفظ والمعنى، مجلة إعجاز عربي، م 8، ع 2، جوان 2025م، ص 581.
- <sup>4</sup> - ينظر: محمد يزيد سالم، بناء المعجم الرقمي في ظل اللسانيات الحاسوبية، مجلة إشكالات في اللغة والأدب، م 10، ع 1، 2021م، ص 1149، 1150.
- <sup>5</sup> - ينظر: عيجولي حسين، المعالجة الآلية للغة، مجلة دراسات وأبحاث، م 13، ع 2، أفريل 2021م، ص 505، 506، 507.
- <sup>6</sup> - Laurence Anthony, Issues in the design and development of software tools for corpus studies: the case for collaboration, A chapter in a collective book titled: Contemporary Corpus Linguistics, Continuum International Publishing, London, UK, 1st edition, 2009, p 95.
- <sup>7</sup> - ينظر: وفاء الزهيلي، المدونة اللغوية: تطبيقات حاسوبية من خلال برنامج "أنتكونك" (AntConc)، مجلة مقامات، م 5، ع 2، 2021م، ص 181.
- <sup>8</sup> - Sandrine Zufferey, Introduction to Corpus Linguistics, John Wiley and Sons, Inc, Hoboken, United States of America, 1st edition, 2020, p 127.