مجلة

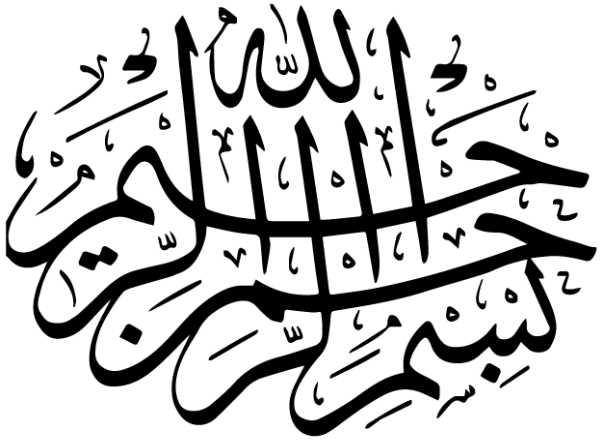
الجمعية العلمية السعودية للغة العربية

مجلة – علمية – محكمة

رقم الإيداع: (١٤٢٩/٣٣٠٢هـ بتاريخ ١٤٢٩/٦/٧هـ)
الرقم الدولي المعياري (ردمد): ١٦٥٨ – ٤١٥٥

كل بحث نشر في المجلة

يعبر عن رأي صاحبه


بسم الله الرحمن الرحيم

هيئة تحرير المجلة

المشرف العام على المجلة، رئيس مجلس إدارة الجمعية:

- د. بدر بن مُحَمَّد الراشد

رئيس التحرير:

- أ. د. عبد المجيد بن صالح الجار الله

مدير التحرير:

- د. سليمان بن صالح الزميع

أعضاء هيئة التحرير:

- أ. د. إبراهيم بن عبد العزيز أبو حيمد
- أ. د. أماني بنت عبد العزيز الداود
- أ. د. صالح بن عبد العزيز المحمود
- أ. د. عبد الرحمن بن رجا الله السلمي
- أ. د. عبد العزيز بن صالح العمري
- أ. د. فريد بن عبد العزيز الزامل

طبيعة المجلة وضوابط النشر

طبيعة المجلة:

١- مجلة الجمعية العلمية السعودية للغة العربية.

٢- مجلة علمية محكمة.

٣- تعنى بعلوم اللغة العربية وآدابها.

٤- تنشر البحوث والدراسات العلمية المحكّمة.

٥- دورية نصف سنوية، تصدر منتصف السنة الهجرية ونهايتها.

ضوابط النشر:

أولاً: الضوابط العامة لقبول البحث:

١- أن يكون البحث في علوم اللغة العربية وآدابها.

٢- أن يتّسم بالجِدّة والأصالة وسلامة الاتجاه.

٣- أن يلتزم البحث بالسلامة اللغوية، والدقة في التوثيق والتخريج.

٤- ألا يكون البحثُ منشوراً أو مقدّماً للنشر في مجلة أخرى.

٥- ألا يكون مستلاً من عمل علمي سابق للباحث.

ثانياً: ما يشترط في كتابة البحث وتوثيقه:

١- أن يُكتب البحث على ورق من مقاس (A4).

٢- أن يُكتبَ بخط (Traditional Arabic) بحجم (١٧) للمتن، وبحجم (١٤) للحاشية، وأن يكون تباعد المسافات بين الأسطر (مفرداً).

٣- أن تُكتبَ الهوامش أسفل كل صفحة على حدة.

٤- أن يُذيَّل البحث بثبت المصادر والمراجع.

٥- أن يكتب الباحث ملخّصاً لبحثه باللغتين العربية والإنجليزية لا تزيد كلماته على مائتي كلمة، ويتضمن الملخص موضوع البحث وأهدافه، ومنهجه، وأهم التوصيات، والكلمات المفتاحية.

٦- رومنة المصادر والمراجع.

ثالثاً: ما يشترط عند تقديم البحث:

١- يقدم الباحث طلباً بنشره، وإقراراً يتضمن امتلاكه لحقوق الملكية الفكرية للبحث كله، والتزاماً بعدم نشر بحثه المقدم إلا بعد موافقة هيئة التحرير.

٢- يقدم الباحثُ نسختين من بحثه على النحو التالي:

- نسخة من البحث خالية من اسم الباحث كاملة بصيغة (WORD).
- نسخة من البحث خالية من اسم الباحث كاملة بصيغة (PDF)

٣- يرفق الباحث ترجمة الملخص باللغة الإنجليزية.

٤- يرسل الباحث بحثه مع الملخصات إلى منصة مجلة الجمعية:
(https://imamjournals.org/index.php/josaa/index)

بناء مصفوفة قواعدية حاسوبية متدرجة لتحليل الأخطاء النحوية وتصحيحها في الكتابة العربية لدى متعلمي العربية بوصفها لغة ثانية

Developing a Gradual Computational Grammar Matrix for Analyzing and Correcting Grammatical Errors in Arabic Writing among Learners of Arabic as a Second Language

د.عبدالعزيز بن عبدالله صالح المهيوبي

الأستاذ المشارك في قسم الإعداد اللغوي– معهد تعليم اللغة العربية–

جامعة الإمام مُحَمَّد بن سعود الإسلامية

Abdulaziz Abdullah Saleh Almahyoubi - Associate Professor of Applied Linguistics - Imam Mohammad Ibn Saud Islamic University (IMSIU) - Institute of Teaching Arabic Language - Department of Language Preparation

المستخلص:

في ظل التقدم المتسارع في تقنيات الذكاء الاصطناعي وتزايد الحاجة إلى أدوات تعليمية ذكية، تبرز أهمية تطوير أنظمة قادرة على تحليل الأخطاء اللغوية وتصحيحها لدى متعلمي اللغة العربية بوصفها لغة ثانية، لا سيما في الجوانب النحوية التي تمثل تحديًا كبيرًا للمتعلمين والمعلمين على حدّ سواء. وانطلاقًا من هذه الحاجة، يهدف هذا البحث إلى بناء مصفوفة قواعدية حاسوبية متدرجة لتحليل الأخطاء النحوية وتصحيحها في كتابات متعلمي العربية من غير الناطقين بها، من خلال دمج التحليل اللغوي الوظيفي بالتقنيات الحاسوبية الحديثة.

اعتمدت الدراسة منهجًا وصفيًا تحليليًا تطبيقيًا، حيث تم تحليل عينة مكونة من ٣٠٠ نص واقعي مأخوذة من "المدونة اللغوية لمتعلمي اللغة العربية"، موزعة على ثلاثة مستويات تعليمية (مبتدئ، متوسط، متقدم). جرى تصنيف الأخطاء النحوية حسب نوعها وسياقها القاعدي، ثم بُنيت مصفوفة تعليمية تربط بين الخطأ والتصحيح المناسب وفق مستوى المتعلم. وتم تطوير خوارزمية برمجية بلغة Python بالاستعانة بمكتبات spaCy وCAMeL Tools لتنفيذ مهام التحليل النحوي، واكتشاف الأخطاء، واقتراح التصحيحات المناسبة من المصفوفة. وقد أظهرت نتائج التجريب أن الخوارزمية حققت أداءً مرتفعًا وفق مؤشرات الدقة (Precision) والاسترجاع (Recall) وF1-score، مما يدل على جدوى النموذج في تقديم تصحيحات تعليمية دقيقة تراعي المستوى اللغوي للمتعلمين وسياقاتهم التعبيرية. وتوصي الدراسة بضرورة تبنّي هذه المصفوفة في بناء أدوات تعليمية ذكية موجهة لمتعلمي العربية، والعمل على تطويرها ودمجها في بيئات التعليم الإلكتروني، بما يسهم في تحسين الكفاية النحوية لدى المتعلمين بصورة تدريجية ومفهومة.

الكلمات المفتاحية:

الأخطاء النحوية، تعليم العربية لغة ثانية، المصفوفة القواعدية، التصحيح الآلي، الذكاء الاصطناعي.

Abstract:

In light of the rapid advancements in artificial intelligence and the growing need for intelligent educational tools, the development of systems capable of analyzing and correcting linguistic errors—especially grammatical ones—has become a pressing priority in the field of Teaching Arabic as a Second Language (TASL). Grammatical errors pose significant challenges to both learners and instructors, necessitating context-sensitive and pedagogically informed solutions.

This study aims to develop a gradual computational grammatical matrix for analyzing and correcting grammatical errors in the writings of non-native learners of Arabic. The research adopts a descriptive, analytical, and applied methodology. A sample of 300 authentic texts was selected from the Arabic Learner Corpus, representing three proficiency levels (beginner, intermediate, advanced). The grammatical errors were categorized by type and

syntactic context, and a pedagogical matrix was constructed to map each error to its appropriate correction according to the learner's level.

A Python-based algorithm was implemented, utilizing libraries such as spaCy and CAMeL Tools to perform syntactic analysis, detect errors, and retrieve the appropriate corrections from the matrix. Experimental results demonstrated high performance across key metrics—precision, recall, and F1-score—confirming the algorithm's effectiveness in providing accurate, level-sensitive grammatical feedback.

The study recommends adopting this grammatical matrix as a foundation for developing intelligent educational tools in Arabic language instruction. Furthermore, it encourages the integration of such systems into digital learning environments to enhance learners' grammatical competence through a gradual, contextualized, and explainable correction approach.

Keywords:

Grammatical Errors, Teaching Arabic as a Second Language, Grammatical Matrix, Automated Correction, Artificial Intelligence.

الإطار العام للدراسة

المقدمة:

في ظل الثورة الرقمية والاتجاه العالمي نحو التعليم المدعوم بالتقنية، أصبحت المعالجة الحاسوبية للغة أحد أبرز الاتجاهات العلمية التي تربط بين علوم اللغة وتقنيات الذكاء الاصطناعي، لا سيما في مجال تعليم اللغات. وتعد اللغة العربية من اللغات ذات البنية التركيبية الغنية، والتعقيد النحوي البالغ، وذلك مما يجعل تطوير أدوات حاسوبية قادرة على تحليلها وتصحيحها تحديًا حقيقيًا، خاصة عندما يتعلق الأمر بتعليم العربية لغير الناطقين بها.

وفي هذا السياق شهد العقد الأخير اهتمامًا متزايدًا بمجال التصحيح النحوي الآلي (Grammatical Error Correction)، ضمن إطار أوسع لما يُعرف بمعالجة اللغات الطبيعية (NLP). ومع وجود أدوات مثل: (CAMeL Tools) أدوات كامل (Obeid et al., 2020) و(Farasa) فراسة (Abdelali et al., 2016) و(MADAMIRA) مادااميرا (Pasha et al., 2014)، وغيرها، إلا أنَّ معظم هذه الأدوات صُممت من منطلق لغوي عام، ولا تأخذ في الاعتبار خصوصية أخطاء متعلمي العربية بوصفها لغة ثانية، ولا مستوى المتعلم، أو السياق التعليمي.

في هذا السياق، تظهر الحاجة الملحة إلى بناء نظام تحليلي وتصحيحي حاسوبي يتماشى مع طبيعة الأخطاء النحوية الشائعة لدى المتعلمين، ويأخذ بعين الاعتبار السياق اللغوي والوظيفي للجملة، ومستوى المتعلم، ومدى تطوره في اكتساب القواعد. فمثلاً، لا يمكن التعامل مع خطأ في "رفع الفاعل" بالطريقة نفسها لدى متعلم في المستوى المبتدئ وآخر في المستوى المتقدم، بل ربما يلزم أن يكون التصحيح موجهًا تعليميًا، وبسيطًا لغويًا، ومحفزًا تعلميًّا.

وتذهب كثيرٌ من الدراسات في هذا الجانب إلى أن الأخطاء النحوية لا تنبع فقط من ضعف التأسيس النحوي القائم على حفظ القواعد الشكلية، بل كثيرًا ما ترتبط بالفهم السياقي والوظيفي للجملة (Ferris, 2011؛ Ellis, 2008). ومن هنا تأتي أهمية المزج بين التحليل النحوي التقليدي، والتحليل الوظيفي لبناء أدوات تعليمية ذكية، قادرة على التفسير اللغوي للأخطاء، وتقديم مقترحات تصحيحية مدعومة بالسياق.

ولأن التعليم الحاسوبي الفعال يقوم على قواعد بيانات مصممة بعناية، فإن الهدف الرئيس من هذه الدراسة هو بناء مصفوفة قواعدية حاسوبية متدرجة، تتضمن تصنيفًا للأخطاء النحوية حسب نوعها، وسياقها، ومستوى المتعلم، وتقديم آلية تصحيحية قابلة للتنفيذ داخل نموذج حاسوبي أولي. وتكمن أهمية

هذا النموذج في قدرته على التطور المستقبلي ليتكامل مع أدوات تحليلية كبرى مثل Farasa وCAMeL Tools، أو يُدمج في بيئات تعليمية إلكترونية تقدم تصحيحًا نحويًا ذكيًا وفوريًا.

ولأنَّ هذا البحث يربط بين النظرية النحوية والتطبيق البرمجي، فإن هذه الدراسة تستند إلى منهجية متعددة الأبعاد، تجمع بين التحليل اللغوي الوصفي، والبناء القاعدي الحاسوبي، والتحليل الإحصائي التجريبي، بما يجعلها إسهامًا جديدًا في تطوير المحتوى الرقمي العربي الموجه للناطقين بغيرها.

١-١ مشكلة الدراسة:

مع التطور الملحوظ في أدوات المعالجة الحاسوبية للغة العربية، خصوصًا في الجوانب المتعلقة بالتحليل الصرفي والتركيبي، فإن مجال تحليل الأخطاء النحوية وتصحيحها ما يزال يعاني من نقص حاد في النماذج المتخصصة التي تراعي خصوصية متعلمي العربية من غير الناطقين بها. فالعديد من الأدوات القائمة، مثل: Farasa وMADAMIRA وCAMeL Tools، صُممت للتعامل مع النصوص العربية العامة أو الإخبارية، دون مراعاة سياقات المتعلم أو الأخطاء الناتجة عن انتقال لغوي (Interlanguage) أو تأثير اللغة الأم (L1 transfer) (Ferris, 2003؛ Gass & Selinker, 2008).

ومما يُفاقم المشكلة، أن غالبية نماذج التصحيح النحوي المتوفرة تعتمد على أنماط إحصائية أو شبكات عصبونية (مثل Transformers) جرى تدريبها على نصوص سليمة لغويًا، وذلك مما يجعلها أقل كفاءة في التعامل مع الأخطاء النحوية الواقعية في كتابات المتعلمين، لا سيما إذا جاءت هذه الأخطاء في سياقات معقدة أو مرتبطة بمستويات تعلمية محددة (Rozovskaya & Roth, 2016).

ولأجل ذلك تزداد الحاجة إلى نماذج تحليل وتصحيح تكون قادرة على:

- اكتشاف الأخطاء النحوية بناءً على السياق وليس فقط الشكل.
- تقديم تفسيرات تعليمية مفهومة للمتعلمين.
- مراعاة المستوى اللغوي للدارس (مبتدئ، متوسط، متقدم).
- اقتراح تصحيحات تتناسب مع المرحلة التعليمية، بدلاً من التصحيحات "الجامدة" المعتمدة على النحو المعياري فقط.

ومن هذا الجانب تنبع مشكلة الدراسة في غياب أداة عربية تعليمية حاسوبية تتسم بـ:

- تحليل دقيق للأخطاء النحوية الشائعة في كتابات متعلمي العربية.
- تمثيل هذه الأخطاء في مصفوفة تعليمية قابلة للمعالجة الحاسوبية.

- تصميم خوارزمية مرنة تعتمد على هذه المصفوفة في اكتشاف وتصحيح الأخطاء بطريقة تعليمية تراعي السياق والمستوى.

١-٢ أسئلة الدراسة:

انطلاقًا من مشكلة الدراسة والسياق العلمي الذي تعمل ضمنه، تسعى هذه الدراسة إلى الإجابة عن السؤال التالي: كيف يمكن بناء مصفوفة قواعدية حاسوبية متدرجة لتحليل الأخطاء النحوية وتصحيحها في كتابات متعلمي العربية لغة ثانية، تأخذ بعين الاعتبار السياق والمستوى اللغوي للمتعلمين؟ ويتفرع عنه الأسئلة التالية:

- ما أكثر الأخطاء النحوية شيوعًا في كتابات متعلمي العربية بوصفها لغة ثانية، وكيف تتوزع حسب السياقات اللغوية والمستويات التعليمية؟
- ما الأسباب النحوية واللغوية الكامنة وراء هذه الأخطاء؟ وهل تعود إلى ضعف في استيعاب القاعدة، أم إلى اضطراب في السياق أو تركيب الجملة؟
- كيف يمكن بناء مصفوفة قواعدية حاسوبية متدرجة تمثل هذه الأخطاء بصورة منهجية، تأخذ بعين الاعتبار نوع الخطأ، وسياقه، والسبب النحوي، والتصحيح المناسب، ومستوى المتعلم؟
- إلى أي مدى تنجح الخوارزمية المقترحة (المبنية على هذه المصفوفة) في اكتشاف الأخطاء وتصحيحها بدقة، وتقديم تصحيحات مناسبة سياقيًا وتعليميًا، مقارنة بأدوات التصحيح التقليدية أو العامة؟

وتهدف هذه الأسئلة إلى تغطية كل من الجانب التحليلي للأخطاء والجانب التطبيقي البرمجي، وربط نتائج الدراسة بالواقع التعليمي الحاسوبي.

١-٣ أهداف الدراسة:

تهدف هذه الدراسة إلى تحقيق هدف رئيس يتمثل في بناء مصفوفة قواعدية حاسوبية متدرجة لتحليل الأخطاء النحوية وتصحيحها في الكتابة العربية لدى متعلمي العربية بوصفها لغة ثانية، تكون قابلة للتنفيذ ضمن نموذج خوارزمي ذكي وتعليمي.

وينبثق عن هذا الهدف العام عدد من الأهداف الفرعية المحددة، تشمل:

- رصد الأخطاء النحوية الشائعة وتصنيفها في كتابات متعلمي العربية، وذلك حسب نوع الخطأ، والسياق اللغوي، والسبب النحوي المحتمل.

- بناء مصفوفة تعليمية قواعدية حاسوبية (Matrix of Grammatical Errors) تشمل المعلومات اللازمة لكل خطأ، مثل: نوع الخطأ (مثل: نصب الفاعل بدل رفعه، حذف أداة النفي... إلخ)، والسياق اللغوي للجملة، والقاعدة النحوية المرتبطة بالخطأ، والتصحيح المقترح، ومستوى المتعلم (مبتدئ، متوسط، متقدم).
- تصميم خوارزمية تعليمية أولية بلغة Python (باستخدام أدوات مثل spaCy أو NLTK) تستند إلى هذه المصفوفة، وتعمل على استكشاف الأخطاء النحوية، واقتراح تصحيحات تعليمية مناسبة.
- اختبار الخوارزمية على عينة من النصوص المكتوبة من متعلمي العربية بوصفها لغة ثانية، وتقييم فعاليتها باستخدام مقاييس علمية مثل: الدقة (Precision)، الاسترجاع (Recall)، درجة التوازن (F1-score)
- اقتراح آلية لدمج هذه الأداة داخل بيئات تعليمية تفاعلية، يمكن للمتعلم الاستفادة منها في تحسين كفايته اللغوية النحوية من خلال التفاعل مع التصحيحات والتفسيرات المقترحة.

١-٤ أهمية الدراسة:

تنبع أهمية هذه الدراسة من تلاقيها مع عدد من التوجهات المعاصرة في تعليم اللغة العربية بوصفها لغة ثانية، والمعالجة الحاسوبية للغة الطبيعية (NLP)، وتتمثل في جوانب متعددة:

١-٤-١ الأهمية النظرية:

تسهم الدراسة في ربط قواعد النحو العربي بالمعالجة الآلية التعليمية، وتدفع باتجاه بناء نماذج نحوية قابلة للحوسبة تراعي السياق والمستوى اللغوي، وهو مجال ما يزال يعاني من ندرة واضحة (Habash, 2010; Obeid et al., 2020). كما تقدم هذه الدراسة إطارًا لتصنيف الأخطاء النحوية من منظور وظيفي وتعليمي، يتجاوز التصنيف التقليدي الجامد الذي لا يراعي طبيعة المتعلم وحاجاته.

١-٤-٢ الأهمية التطبيقية:

يمكن استخدام نتائج الدراسة في بناء أدوات تصحيح لغوي تعليمية تُدمج ضمن المنصات التعليمية، وتقدم تصحيحات موجهة تراعي مستوى المتعلم وتشرح له سبب الخطأ، بدل الاكتفاء بالإشارة إليه فقط. كما تمهّد الدراسة الطريق نحو تطوير نظم كتابة ذكية للمتعلمين، تُمكّنهم من التدرب الذاتي على الكتابة العربية واستقبال تغذية راجعة نحوية واضحة.

١-٤-٣ الأهمية البحثية:

تسهم الدراسة في سدِّ فجوة واضحة في البحوث المتعلقة بتحليل الأخطاء النحوية حاسوبيًا لدى غير الناطقين بالعربية، خاصة في ظل غياب مصفوفات معيارية معتمدة في هذا المجال. كما يمكن أن تفتح آفاقًا جديدة للباحثين في مجال نحو المتعلمين (Learner Grammar) واللغويات التطبيقية التعليمية، إذ توفّر نموذجًا قابلًا للتطبيق والتطوير.

١-٥ حدود الدراسة:

حرصت هذه الدراسة على تحقيق أهدافها ضمن نطاق علمي ومنهجي محدد، وفيما يلي أبرز الحدود التي تحكم هذه الدراسة:

١-٥-١ الحدود الموضوعية:

تركز الدراسة على الأخطاء النحوية فقط دون التطرق إلى الأخطاء الإملائية أو الأسلوبية أو الدلالية؛ وذلك بغرض تعميق المعالجة النحوية الدقيقة. وتُعنى الدراسة بكتابات متعلمين غير ناطقين بالعربية، أي أن النموذج غير موجه لمعالجة النصوص المكتوبة من الناطقين الأصليين باللغة العربية.

تنحصر هذه الدراسة في معالجة الأخطاء **النحوية** بالمعنى التركيبي(Syntactic Errors) ، أي الأخطاء المرتبطة ببنية الجملة وترتيب عناصرها النحوية ووظائفها، دون التطرق إلى الأخطاء الصرفية أو الدلالية أو الإملائية أو الأسلوبية. ويُقصد بـ"النحوية" هنا ما يوافق التراكيب القواعدية الصحيحة وفق النظام النحوي الوظيفي.

١-٥-٢ الحدود اللغوية:

سيكون تحليل الأخطاء ضمن النحو المعياري المعتمد في تعليم العربية لغير الناطقين بها، وفقًا للكتب المرجعية والقواعد المتفق عليها في البرامج التعليمية.

١-٥-٣ الحدود الجغرافية والزمنية:

تعتمد الدراسة على مدونة نصوص تعليمية (المدونة اللغوية لمتعلمي اللغة العربية)، دون امتداد إلى بيانات حية أو حديثة يجري جمعها أثناء الدراسة. ولا تشمل الدراسة مقارنة بين متعلمي العربية من خلفيات لغوية مختلفة (مثل: الإنجليزية، الفرنسية، الصينية)، ومع الإقرار بأهمية هذا العامل، إلا أن محدودية الدراسات العربية السابقة في هذا الجانب – كما أُشير آنفًا – إلى جانب القيود المتعلقة بالوقت

والإمكانات المتاحة، دفعت الباحث إلى التركيز على الهدف الرئيس للدراسة، وهو بناء المصفوفة القواعدية، دون التوسع في التفصيلات المصاحبة. ومع ذلك، تُعدّ هذه الدراسة تمهيدًا لأعمال مستقبلية يمكن أن تُبنى عليها، وتتوسع في معالجة الجوانب الأخرى التي من شأنها تعزيز دقة النتائج وتطوير النموذج المقترح.

١-٥-٤ الحدود التقنية:

تقتصر الخوارزمية المطورة في هذه الدراسة على نموذج أولي تعليمي، وقد لا ترقى إلى مستوى التصحيح الشامل المعتمد في أدوات متقدمة مثل Grammarly أو Microsoft Editor.

جرى تنفيذ الخوارزمية باستخدام Python ومكتبات spaCy/NLTK، وذلك مما يجعل تعميمها مشروطًا ببيئات دعم مماثلة.

١-٦ مصطلحات الدراسة الإجرائية:

توضح هذه الفقرة المعاني الإجرائية للمفاهيم الرئيسة المستخدمة في البحث، كما جرى اعتمادها لأغراض هذه الدراسة:

الخطأ النحوي (Grammatical Error):

هو انحراف عن البنية النحوية السليمة للجملة في اللغة المستهدفة، يظهر غالبًا في السياقات التي تتطلب توافقًا نحويًا بين عناصر الجملة من حيث الإعراب أو التركيب، مثل استخدام زمن غير مناسب، أو اختلال العلاقة النحوية بين الفاعل والفعل. ويُعد هذا النوع من الأخطاء مؤشرًا على نقص في الكفاءة اللغوية، وقد ينجم عن تأثير اللغة الأم، أو قصور في التعليم، أو اضطرابات معرفية في اكتساب القواعد (James, 1998).

ويُقصَدُ به في هذه الدراسة: انحراف منهجي في بنية الجملة عن القواعد النحوية المعيارية للغة العربية، ويشمل صورًا مثل الخلط بين الرفع والنصب، أو غياب الرابط النحوي الملائم، أو الاستخدام غير الصحيح لأدوات الشرط، ويُنظر إليه في هذه الدراسة بوصفه ناتجًا عن تداخل لغوي بين اللغة الأم والعربية، أو عن ضعف في التكوين التعليمي القواعدي، أو عن اضطراب معرفي لغوي في تنظيم البنية النحوية.

مصفوفة الأخطاء القواعدية (Matrix of Grammatical Errors):

تمثيل منظم ومشروح لأخطاء المتعلمين النحوية، يُستخدم لتصنيف الأخطاء وفق أبعاد متعددة تشمل: نوع الخطأ، والسياق التركيبي والدلالي الذي ظهر فيه، والقاعدة النحوية ذات الصلة، والصيغة

الصحيحة المقترحة، ومستوى المتعلم. ويهدف هذا التمثيل إلى تيسير بناء نماذج آلية لتصحيح الأخطاء، من خلال تدريب الخوارزميات على أنماط الأخطاء الشائعة واستراتيجيات المعالجة المقترحة (Rozovskaya & Roth, 2016).

الخوارزمية النحوية (Grammatical Algorithm):

نظام برمجي يُصمَّم لاكتشاف الأخطاء النحوية في النصوص المكتوبة وتصحيحها، بالاعتماد على قواعد لغوية محددة أو بيانات مُعلمة. وغالبًا ما يُكتب بلغة Python باستخدام مكتبات مثل NLTK أو spaCy، ويُدمج مع مصفوفة معرفية تحتوي على أنماط الأخطاء وتصحيحاتها المحتملة، وذلك مما يسمح بالمعالجة الآلية للنصوص بدقة وفعالية (Leacock et al., 2014).

متعلِم العربية لغةً ثانيةً (Arabic L2 Learner):

هو الشخص الذي يتعلم اللغة العربية في سياق لا تُستخدم فيه كلغة أولى، ويُعدّ جزءًا من فئة متعلمي اللغة الثانية الذين يكتسبون مهارات الاستماع والتحدث والقراءة والكتابة بالعربية لأغراض تعليمية، أو دينية، أو مهنية، أو تواصلية، وغالبًا ما يجري تعليمه من خلال برامج رسمية في بيئات غير ناطقة بالعربية (Leacock et al., 2014).

المعالجة الحاسوبية للغة العربية (Arabic NLP):

هي فرع من فروع الذكاء الاصطناعي يركز على تطوير أنظمة قادرة على محاولة تحليل نصوص اللغة العربية لفهم أنماطها وتوليدها، من خلال مهام تشمل التحليل الصرفي والنحوي، والتعرف على الكيانات، واكتشاف الأخطاء وتصحيحها، والتوليد الآلي للنصوص (Habash, 2010).

الإطار النظري والدراسات السابقة.

الإطار النظري:

يمثّل الإطار النظري ركيزة أساسية في البحوث العلمية عمومًا، واللغوية الحاسوبية منها على وجه الخصوص، إذ يُوفّر الخلفية المفاهيمية والمنهجية التي تُسهم في تفسير الظواهر اللغوية قيد الدراسة، وبيان أبعادها وتعقيداتها، وتحديد العوامل المؤثرة فيها. كما يُعدّ هذا الإطار حلقة وصل بين التجريد النظري وبين التطبيق العملي، وذلك مما يمنح البحث عمقًا علميًا، ويُعزّز من قيمته التفسيرية والتطبيقية على حدّ سواء (Creswell, 2014).

وفي سياق هذا البحث، يتقاطع الإطار النظري مع مجالين معرفيين متكاملين: أولهما علم النحو العربي، بما يتضمنه من مفاهيم إعرابية وبنائية تنظم الجملة وتُوجّه العلاقات بين مفرداتها، وثانيهما المعالجة الحاسوبية للغة الطبيعية (NLP)، التي تمثل فرعًا من علوم الحاسوب والذكاء الاصطناعي يُعنى بتمكين الحواسيب من فهم اللغة البشرية، وتحليلها، ومعالجتها بصورة آلية.

ويتطلب هذا التداخل المعرفي فهمًا عميقًا لآليات تمثيل البنية النحوية داخل السياقات التعليمية من جهة، وآليات النمذجة الحاسوبية للقواعد من جهة أخرى. فالنحو العربي لا يُعدّ مجرد نظام رمزي تقليدي، بل منظومة معرفية تُبنى على مقولات دقيقة كالموقع، والإعراب، والبناء، والتعدية، والمطابقة، وذلك مما يجعل تمثيله حاسوبيًا مهمة شاقة تتطلب تجاوز التعقيد الصوري إلى بناء قواعد سياقية مرنة تواكب خصائص اللغة وتنوع أخطاء المتعلمين.

٢-١ تعليم النحو العربي بوصفه مكونًا وظيفيًا:

يُعدّ تعليم النحو العربي لغير الناطقين بالعربية أحد أبرز التحديات التي تواجه برامج تعليم اللغة العربية لغة ثانية؛ لما ينطوي عليه من إشكالات متعلقة بتجريد القاعدة النحوية، وتعدد مصطلحاتها، وكثرة استثناءاتها. وقد كشفت دراسات عدة (طعيمة، ٢٠٠٤؛ الصاعدي، ٢٠١٩؛ العنزي، ٢٠٢٠) عن صعوبة إدراك المتعلمين للبنى النحوية المجردة، لا سيما حين تُقدَّم لهم بأسلوب تقعيدي صرف، دون ربطها بالسياق التواصلي أو الوظيفة اللغوية.

ومن هذا الباب تبرز الحاجة إلى اعتماد مناهج أكثر تفاعلًا مع طبيعة المتعلم وبيئته اللغوية، إذ تؤكد الاتجاهات التربوية الحديثة، ولا سيما النحو الوظيفي التعليمي (Functional Grammar Approach)، على ضرورة تقديم القاعدة النحوية في إطار من السياق الواقعي والتواصل الفعّال. ويُنظر إلى الخطأ النحوي، ضمن هذا المنظور، لا بوصفه إخفاقًا لغويًا يجب إقصاؤه، بل كفرصة تعليمية تساعد في تطوير وعي المتعلم بالنظام النحوي للغة (Celce-Murcia & Larsen-Freeman, 1999).

وتقوم الفلسفة التعليمية للنحو الوظيفي على ربط التراكيب بالقصدية اللغوية، وتفسير القواعد في ضوء استعمالها داخل السياقات المختلفة، وذلك مما يوفّر بيئة أكثر دعمًا لاكتساب الكفاءة النحوية الفعلية. كما أن المعالجة التعليمية للنحو تستفيد من تحليل الأخطاء المتكررة، وتوظفها لتصميم أنشطة داعمة تنطلق من الخطأ ذاته، بدلًا من تجاهله أو معاقبة المتعلم عليه (Ellis, 2008).

إن استيعاب هذه المنطلقات يُعدّ تمهيدًا ضروريًا للانتقال إلى تمثيل هذه الظواهر – وخاصة الأخطاء النحوية – في بنى حاسوبية قابلة للمعالجة، وهذا أمر يتطلب الانتقال من النظر في القاعدة بوصفها بنية

وصفية إلى تصورها كإجراء تصحيحي قابل للتنفيذ ضمن خوارزمية أو مصفوفة تحليلية.

٢-٢ تحليل الأخطاء اللغوية: المفهوم والنماذج:

يُعدّ تحليل الأخطاء (Error Analysis) فرعًا حيويًا من فروع البحث في مجال تعليم اللغات، نشأ في ستينيات القرن العشرين كرد فعل على مناهج السلوكية التقليدية التي كانت تركز على تقليد المتعلم للنموذج اللغوي الصحيح فقط، دون النظر في طبيعة الأخطاء التي يرتكبها. وقد كان اللغوي البريطاني Stephen Pit Corder من أوائل من أسسوا لهذا الاتجاه حين بيّن في دراسته الشهيرة (Corder, 1967) أن الأخطاء ليست مجرد انحراف عن النموذج اللغوي، بل تمثل مؤشرًا مهمًا على مراحل تطور اللغة لدى المتعلم، وتوفر نافذة لفهم استراتيجيات التعلّم.

وقد ميّز كوردَر بين نوعين من الانحرافات:

الخطأ (Error): وهو الناتج عن نقص في المعرفة القاعدية للغة الهدف، ويكون ممنهجًا ومتكررًا، وهو ما يجعله قابلاً للتحليل والتفسير.

الزلّة (Mistake): وهي ناتجة عن اضطرابات مؤقتة في الأداء اللغوي، مثل النسيان أو التردد أو التسرع، ولا تعكس قصورًا معرفيًا.

تطورت هذه الرؤية لاحقًا على يد Larry Selinker (١٩٧٢) الذي صاغ مفهوم "اللغة الوسيطة" (Interlanguage)، وعرّفها بوصفها النظام اللغوي المؤقت الذي يُنتجه المتعلم خلال رحلة انتقاله من لغته الأم إلى اللغة الهدف. ويمثل هذا النظام خليطًا من قواعد اللغة الأم، والعناصر المكتسبة من اللغة الجديدة، والاستراتيجيات التعميمية التي يطوّرها المتعلم بنفسه.

وقد أظهرت الدراسات والتجارب التي أجريت في ميدان تحليل الأخطاء فاعلية هذه النماذج في تحليل أنواع الأخطاء وتفسيرها، ولا سيما في الميادين التطبيقية لتعليم اللغة، حيث بات يُنظر إلى الخطأ بوصفه معطًى إجرائيًا ينبغي تحليله وتوظيفه في تصميم المناهج والأنشطة.

٢-٣ تحليل الأخطاء النحوية في تعليم العربية:

في ميدان تعليم العربية للناطقين بغيرها، أظهرت دراسات حديثة (مثل الفيفي، ٢٠١٧؛ مُحَمَّد وسكباني، ٢٠٢٠) أن الأخطاء النحوية التي يرتكبها المتعلمون تتوزع على عدة مستويات بنيوية، أهمها:

- أخطاء في ترتيب المكونات الإسنادية، مثل: تقديم المفعول وتأخير الفاعل في غير موضعه النحوي.

- أخطاء في استخدام أدوات الربط، كحروف الجر، والعطف، والشرط.
- أخطاء في الإعراب، ناتجة عن الجهل بالعلامات أو الالتباس في التقدير.
- أخطاء في صيغة المبني للمجهول، نتيجة صعوبة فهم البنية وتطبيقها.

وتشير هذه الدراسات إلى أن غالبية هذه الأخطاء ليست عشوائية، بل تخضع لقواعد توليدية خاصة باللغة الوسيطة التي يكوّنها المتعلم، وذلك مما يجعلها قابلة للتصنيف والبناء عليها في نماذج تصحيحية.

٢-٤ المعالجة الآلية للغة العربية (Arabic NLP):

شهدت العقود الأخيرة طفرة ملحوظة في مجال معالجة اللغة الطبيعية (NLP)، وذلك مما ساهم في تطوير أدوات ذكية قادرة على التعامل مع النصوص اللغوية من حيث تحليلها الصرفي والنحوي والدلالي، والقيام بمهام متقدمة مثل الترجمة، والتلخيص، واكتشاف الأخطاء وتصحيحها.

إلا أن اللغة العربية – على خلاف بعض اللغات الأخرى – تمثل تحديًا فنيًا وبنيويًا خاصًا في هذا المجال؛ نظرًا لما تتصف به من نظام صرفي اشتقاقي غني، وعلامات إعرابية تؤثر في البنية، ووجود التشكيل الإلزامي للفهم، والتباس الضمائر المتصلة والمنفصلة.

وقد طُوِّرت عدة أدوات متخصصة لمواجهة هذه التحديات، أبرزها:

CAMeL Tools

طوّرها فريق من جامعة نيويورك أبوظبي بقيادة Nizar Habash، وتقدم حزمة متكاملة تشمل تحليل الصيغ الصرفية، والتعرف على الكيانات، والإعراب، والتجزئة (Habash et al., 2021).

Farasa

أداة خفيفة وسريعة تعتمد على تقنيات التعلم الآلي، وتمتاز بقدرتها على تجزئة النصوص وتحليلها دون الحاجة إلى تشكيل (Abdelali et al., 2016).

MADAMIRA

تجمع بين قدرات Buckwalter التحليلية وتقنيات Farasa، وتُستخدم في مشاريع لغوية كثيرة لمعالجة البيانات العربية آليًا.

غير أن معظم هذه الأدوات تعتمد في تدريبها على نصوص عربية فصيحة سليمة، وهو ما يُضعف

أداءها في التحليل وتصحيح النصوص التي تحتوي على انحرافات منهجية وغير نمطية لدى متعلمي العربية لغة ثانية. وقد أشار Habash (٢٠١٠) إلى الحاجة الملحة لتدريب النماذج الحاسوبية على بيانات تعليمية تحتوي على أخطاء واقعية؛ وذلك لتعزيز دقة أنظمة التصحيح، وتوجيهها تربويًا نحو الأداء اللغوي الفعلي للمتعلمين.

٢-٥ النماذج الحديثة لتصحيح الأخطاء النحوية الآلية (GEC):

تُعدُّ مهمة تصحيح الأخطاء النحوية الآلية (Grammatical Error Correction - GEC) من أبرز تطبيقات الذكاء الاصطناعي التي تخدم ميدان تعليم اللغات، وخاصة في البيئات التعليمية الرقمية. وتتمثل هذه المهمة في الكشف عن الأخطاء اللغوية (الصرفية، النحوية، الإملائية، الأسلوبية) وتصحيحها بطريقة آلية تراعي السياق والدقة اللغوية. وقد تطوّرت منهجيات GEC عبر ثلاث مراحل رئيسة:

٢-٥-١ النماذج القائمة على القواعد (Rule-Based Systems):

تمثل هذه النماذج الجيل الأول من أنظمة التصحيح الآلي، وتعتمد على قوائم من الأخطاء الشائعة وقواعد مكتوبة يدويًا لتحديد الخطأ وتقديم البديل الصحيح. وتُبنى هذه القواعد عادةً على أساس مصفوفات تُربط فيها الأخطاء بالتصحيحات المقترحة ضمن سياقات محددة، وهو ما يتيح درجة من الدقة لكنها تفتقر إلى التكيف مع السياقات الجديدة أو العبارات غير النمطية (Leacock et al., 2014). كما تُعد هذه النماذج محدودة في قدرتها على التعلُّم من البيانات الجديدة أو التنبؤ بالأخطاء غير المعروفة.

٢-٥-٢ نماذج التعلم الآلي التقليدية (Statistical Learning):

مع تطور تقنيات التعلم الآلي، بدأ استخدام نماذج مثل Conditional Random Fields (CRFs) و Support Vector Machines (SVMs) في بناء مصنفات آلية قادرة على التعرف على الأخطاء استنادًا إلى خصائص سياقية وإحصائية. وتُدرّب هذه النماذج على بيانات تحتوي على جمل صحيحة وأخرى بها أخطاء، لتعلُّم الفروق الإحصائية وتطبيق التصحيحات. ومع تفوقها على الأنظمة اليدوية في بعض الجوانب، فإن دقتها تبقى محدودة، خصوصًا في اللغات ذات البنية الصرفية الغنية كالعربية.

٢-٥-٣ النماذج العميقة (Deep Learning Models):

أحدثت تقنيات التعلم العميق نقلة نوعية في مجال GEC، من خلال اعتماد شبكات Transformer المعتمدة على الانتباه الذاتي (Self-Attention). وتأتي نماذج مثل BERT وGPT وT5 في طليعة هذه النماذج، حيث تُدرَّب على كميات ضخمة من البيانات اللغوية، وهو ما يمنحها قدرة فائقة على فهم السياق واكتشاف الأخطاء الدقيقة، إلا أن هذه النماذج غالبًا ما تحتاج إلى عمليات إعادة توليد للجملة بأكملها عند التصحيح، وهو ما قد يُنتج عبارات غير مقصودة أو تغييرات كبيرة في المعنى.

نموذج GECToR: طفرة في التصحيح السياقي:

يُعد نموذج GECToR (Grammatical Error Correction: Tag, Not Rewrite) من أبرز التطورات الحديثة، حيث يقترح مقاربة جديدة تعتمد على "الوسم التصحيحي" بدلًا من إعادة توليد الجمل كاملة (Omelianchuk et al., 2020). يقوم هذا النموذج بتحليل كل كلمة داخل السياق، ثم يحدد الوسم المناسب لها: إبقاؤها كما هي، أوحذفها، أو استبدالها، أو إدراج كلمة جديدة. ويتميز GECToR بكونه أسرع وأكثر دقة في التنفيذ، كما يعمل على التقليل من احتمالات التغيير غير المرغوب فيه في النصوص الأصلية.

وتزداد فعالية هذا النوع من النماذج حين يُدرَّب على بيانات حقيقية تحتوي على أخطاء فعلية من إنتاج متعلمي اللغة، وليس فقط على نصوص مصطنعة أو مثالية. وقد بيّنت التجارب أن تضمين مدونات تعليمية واقعية مثل Lang-8 Learner Corpus أو Arabic Learner Corpus يرفع من دقة التصحيح، ويُقرّب الأداء الآلي من التصحيح البشري التربوي.

٢-٦ التحدي في اللغة العربية:

مع نجاح هذه النماذج في العديد من اللغات، تبقى اللغة العربية بحاجة إلى مزيد من العمل المخصص، إذ لا تزال مواردها التعليمية المؤشرة نحويًا شحيحة نسبيًا، ولا توجد حتى الآن نسخة عربية كاملة من GECToR مُدربة على مدونات تعلمية عربية تحتوي على أخطاء حقيقية. وربما تكون في هذا إشارة كافية إلى أهمية المشاريع البحثية التي تُعنى ببناء مصفوفة تعليمية معيارية للأخطاء وتصحيحها حاسوبيًا، كما في هذه الدراسة، التي تمثل خطوة تأسيسية نحو بناء أدوات تصحيح نحوي ذكية مخصصة لمتعلمي العربية كلغة ثانية.

٢-٧ الدراسات السابقة:

تُعد الدراسات السابقة ركيزة أساسية في البحوث التطبيقية، إذ تتيح استعراض الجهود العلمية ذات الصلة، واستخلاص الفجوات المعرفية التي يسعى البحث الحالي إلى سدّها. وبالنظر إلى طبيعة هذا البحث، فإن الدراسات السابقة تنقسم إلى ثلاثة محاور رئيسة:

٢-٧-١ دراسات تناولت تحليل الأخطاء النحوية لدى متعلمي العربية:

ركزت العديد من الدراسات العربية على وصف الأخطاء الشائعة في الكتابة لدى متعلمي اللغة العربية بوصفها لغة ثانية. ومن أبرز هذه الدراسات:

دراسة أحمد الكرمي عبد رمضان عبد الكريم (٢٠٢١) بعنوان: (الأخطاء اللغوية لدى متعلمي اللغة العربية الناطقين بغيرها: طلاب جامعة القصيم نموذجًا). هدفت إلى رصد وتحليل الأخطاء اللغوية (الصرفية، والنحوية، والإملائية) لدى عينة من الطلاب الأجانب الدارسين للعربية لغةً ثانيةً في جامعة القصيم. اعتمد الباحث المنهج الوصفي التحليلي، واستند إلى مدخل التحليل اللغوي التطبيقي، جامعًا بين مختلف اتجاهات دراسة الأخطاء (التقليدي، والتقابلي، والتحويلي، والمعرفي). تكونت عينة الدراسة من ٢٥ طالبًا موزعين على مستويين دراسيين (الثاني والرابع)، وقد استخدم الباحث أدوات تحليل الأخطاء وتصنيفها وتفسيرها في ضوء السياقات النفسية والتعليمية. أظهرت النتائج أن أكثر الأخطاء شيوعًا كانت من النوع النحوي، تلتها الإملائية ثم الصرفية، وأن نسبة هذه الأخطاء تنخفض مع تقدم المستوى الدراسي، وهذا مما يدل على فاعلية البرنامج التعليمي. كما أوصت الدراسة بضرورة بناء برامج تدريبية قائمة على تحليل الأخطاء وتصميم مناهج تستجيب للحاجات الفعلية للمتعلمين وتستثمر في التراكم المرحلي للمهارات اللغوية.

دراسة عبد العزيز موسى درويش علي ومُحَمَّد عبد الله صالح أبو الرب (٢٠١٨) بعنوان: (الأخطاء النحوية لدى متعلمي اللغة العربية الناطقين بغيرها: دراسة تحليلية تطبيقية). هدفت إلى تحليل وتصنيف الأخطاء النحوية التي يقع فيها دارسو اللغة العربية في مركز اللغات بالجامعة الأردنية، وذلك في المستويين الثالث والخامس. اعتمد الباحثان المنهج الوصفي التحليلي، واستخدما مدخل تحليل الأخطاء التكاملي، الذي يتضمن أربع مراحل: تعرف الخطأ، وصفه، ثم تفسيره، ثم تقويمه. وقد اشتملت عينة الدراسة على ٧٠ طالبًا من جنسيات مختلفة، وارتكز التحليل على أخطاء الكتابة النحوية في اختبارات تحريرية. كشفت النتائج عن تكرار أخطاء نحوية متعددة، كان أبرزها: الخلط بين المرفوع والمنصوب، وإهمال التوافق بين الفعل والفاعل، وضعف الاستخدام الصحيح لأدوات النفي والشرط. كما أظهرت النتائج

تحسن الأداء النحوي لدى طلبة المستوى الخامس مقارنة بالمستوى الثالث، وذلك مما يعكس أثر الممارسة والتدرج في التعليم. وأوصت الدراسة بضرورة تطوير مناهج تعليم النحو في ضوء تحليل الأخطاء الفعلية للمتعلمين، واعتماد الأنشطة التطبيقية لتعزيز الفهم.

دراسة يحيى عبدالمبدي مُحَمَّد وخلود سكباني (٢٠٢٠) بعنوان: (تحليل أخطاء الكتابة لدى وارثي العربية في المستوى المتقدم). سعت إلى رصد الأخطاء اللغوية الشائعة في كتابات الطلاب من وارثي العربية، أي أولئك الذين نشأوا في بيئات تستخدم العربية لغةً تراثيةً دون أن تكون لغتهم الأم. استخدم الباحثان المنهج الوصفي التحليلي، وحلّلا عينة من كتابات طلاب في المستوى المتقدم ببرنامج العربية للوارثين، شملت موضوعات إنشائية حرة. كشفت الدراسة عن طيف واسع من الأخطاء اللغوية، أبرزها الأخطاء النحوية والصرفية والإملائية، إضافة إلى ضعف في بنية الجملة العربية، والاستخدام غير الدقيق للمفردات، وهو مما يدل على فجوة بين المهارات الشفوية والكتابية. وأرجعت الدراسة هذه الأخطاء إلى عوامل مثل: قلة التمرس على الكتابة، ونقص الوعي البنيوي بالنظام اللغوي العربي، وتأثير أنظمة اللغة الثانية المكتسبة. وأوصت الدراسة بتصميم مناهج خاصة بوضعية وارثي العربية تراعي احتياجاتهم الفعلية، مع دمج الكتابة في أنشطة أصيلة تفاعلية، وتعزيز التوازن بين الكفاءة التواصلية والمعرفية.

دراسة ماهر بن دخيل الله الصاعدي (٢٠٢١) بعنوان: (تحليل الأخطاء النحوية لدى متعلمي اللغة العربية الناطقين بلغات أخرى). تهدف إلى الكشف عن أكثر الأخطاء النحوية شيوعًا لدى طلاب المستوى الرابع في معهد تعليم اللغة العربية لغير الناطقين بها بالجامعة الإسلامية، وتحليل أسبابها وتقديم توصيات لمعالجتها. استخدم الباحث المنهج الوصفي التحليلي، واعتمد على تحليل نصوص كتابية لعدد من الطلاب غير الناطقين بالعربية، حيث جرى استخراج الأخطاء وتصنيفها وفق بنية الجملة الفعلية. أظهرت النتائج أن الطلاب يواجهون صعوبات في ضبط الفاعل والمفعول، واستخدام أدوات النصب والجزم، والتطابق بين الفعل والفاعل، وذلك مما يعكس أثر اللغة الأم وأساليب التعليم التقليدية في وقوع تلك الأخطاء. وأوصت الدراسة بضرورة اعتماد مدخل وظيفي في تدريس النحو العربي للناطقين بغيره، مع التركيز على التدريب السياقي وتصحيح الأخطاء بطريقة تحليلية واعية.

دراسة أحمد علي مُحَمَّد حسين، ومصطفى رسلان رسلان، وفتحي علي إبراهيم يونس (٢٠١٤) بعنوان: (تحليل الأخطاء كمدخل لعلاج الصعوبات والأخطاء اللغوية الشائعة في تعليم اللغات الأجنبية). هدفت الدراسة إلى بيان أهمية تحليل الأخطاء كأداة تربوية تشخيصية تُسهم في كشف أنماط الصعوبات اللغوية التي تواجه متعلمي اللغات الأجنبية، واقتراح سبل علاجها. استخدم الباحثون المنهج الوصفي التحليلي معتمدين على مراجعة الدراسات السابقة والنظريات اللغوية والتربوية، وتحليل نماذج من الأخطاء

الواقعية. وبيّنت الدراسة أن تحليل الأخطاء لا يُعد فقط وسيلة لتقويم المتعلم، بل مدخلًا لفهم سيرورة التعلم، ومعالجة مواطن الضعف البنيوي والدلالي والتواصلي. كما أكدت على دور المعلم في التعامل البنّاء مع الخطأ، وتحويله إلى فرصة تعليمية. وأوصت الدراسة بإدراج تحليل الأخطاء في برامج إعداد المعلمين، وتطوير مناهج تعليم اللغات لتشمل أنشطة تصحيحية قائمة على الأنماط الأكثر شيوعًا.

٢-٧-٢ دراسات ومشروعات التصحيح النحوي الآلي (Grammatical Error Correction – GEC):

ظهرت عدة بحوث عالمية تناولت نماذج التصحيح النحوي الآلي، بعضها ركز على اللغة الإنجليزية، إلا أن بعضها وُجِّه لتطوير أدوات عربية:

دراسة Rozovskaya and Roth (٢٠١٦): قدّمت تحليلًا معمقًا للأخطاء النحوية الشائعة في كتابات متعلمي اللغة الإنجليزية، مركزةً على تطوير نموذج تصحيح آلي يستند إلى تقنيات التعلم الآلي، لا سيما التصنيف الإحصائي. قامت الدراسة بتصنيف أنواع الأخطاء وتحديد العوامل التي تؤثر في حدوثها، مثل السياق اللغوي ونوع المتعلم، واقترحت إطارًا يستخدم مزيجًا من المصنفات (classifiers) لمعالجة كل نوع من الخطأ على حدة، بدلًا من التعامل مع النصوص ككل. وقد أظهر هذا النهج فاعليته في رفع دقة التصحيح مقارنة بأساليب الترجمة الآلية. ولأهمية منهجها؛ استلهمت بعض البحوث العربية اللاحقة، خاصة في تصحيح الأخطاء النحوية لدى متعلمي العربية من غير الناطقين بها، هذه المقاربة. ويمثّل نجاح تطبيق هذه المقاربة على النصوص العربية مؤشرًا على قابليتها للتطبيق على لغات أخرى، مما يفتح آفاقًا لتوسيع نطاق استخدامها في بيئات لغوية متعددة.

دراسة Omelianchuk et al. (٢٠٢٠): قدّمت نموذجًا مبتكرًا لتصحيح الأخطاء النحوية يُعرف بـ GECToR، يعتمد على إستراتيجية "الوسم لا التوليد" (Tag, Not Rewrite)، وهو ما يمثّل تحولًا في منهجيات تصحيح الأخطاء اللغوية من إعادة كتابة الجمل إلى تعديلها باستخدام وسوم محددة. استند النموذج إلى بنية BERT، وعمل على وسم كل كلمة في الجملة بعلامة تدل على نوع التعديل المطلوب (إن وجد)، وذلك مما جعله أكثر كفاءة وسرعة مقارنة بالأنظمة المعتمدة على الترجمة الآلية أو النماذج التقليدية. وقد أظهر النموذج نتائج متقدمة في اختبارات CoNLL-2014 وBEA-2019، سواء من حيث الدقة أم السرعة، حيث فاق أداؤه نماذج أعمق بكثير بتكلفة حوسبة أقل. وتوصي الدراسة باستخدام هذا النوع من النماذج في أنظمة الكتابة التعليمية الآلية، خاصة لما يتمتع به من سهولة في التطبيق وكفاءة في النتائج.

دراسة **Alkanhal, Al-Salman & Al-Ghamdi (٢٠٢٢)**: تناولت تصحيح الأخطاء النحوية في اللغة العربية باستخدام مدونات لغوية مشروحة لمتعلمي العربية من غير الناطقين بها. هدفت الدراسة إلى تطوير نموذج آلي أكثر دقة وفعالية لتصحيح الأخطاء النحوية من خلال الاعتماد على بيانات حقيقية تحتوي على أخطاء وتصحيحاتها البشرية، وهو ما يُعزز جودة التدريب للنماذج اللغوية. استخدم الباحثون مدونة Annotated Learner Corpus التي اشتملت على عدد كبير من النصوص التي كتبها متعلمون عرب، حيث جرى تصنيف أنواع الأخطاء وتوثيقها. واعتمدت الدراسة المنهج التحليلي القائم على استقراء الأنماط الشائعة للأخطاء النحوية. أظهرت النتائج أن الاعتماد على بيانات مشروحة بعناية يسهم في تحسين أداء أنظمة التصحيح النحوي تحسنًا ملحوظًا. وأوصت الدراسة بضرورة بناء مدونات معيارية عربية مشروحة على نطاق واسع؛ لدعم أبحاث تصحيح الأخطاء وتحليلها، بما يسهم في تطوير أدوات معالجة اللغة العربية.

٢-٧-٣ دراسات على مدونات أخطاء تعليمية عربية:

برزت الحاجة إلى تطوير مدونات لغوية تحتوي على أخطاء واقعية لمتعلمي العربية، ومن أبرز المشاريع:

دراسة الفيفي (٢٠١٧) بعنوان "تحليل الأخطاء في المدونات اللغوية للمتعلمين: المدونة اللغوية لمتعلمي اللغة العربية نموذجًا". هدفت إلى تسليط الضوء على إمكانات استخدام المدونات اللغوية الحاسوبية الخاصة بالمتعلمين بوصفها مصادر تحليلية لتحديد وتصنيف الأخطاء اللغوية؛ وذلك من أجل تحسين تعليم اللغة العربية لغير الناطقين بها. انطلقت الدراسة من تأصيل نظري لمجال تحليل الأخطاء، فناقشت الفرق بين الخطأ والغلط، ونشأة مجال تحليل الأخطاء وتطوره، وعلاقته بنظرية اللغة المرحلية، ثم انتقلت إلى تطبيق عملي على عينة واقعية مأخوذة من "المدونة اللغوية لمتعلمي اللغة العربية"، تضمنت نصوصًا مكتوبة ومنطوقة حررها ٩٤٢ طالبًا من ٦٧ جنسية مختلفة، و٦٦ لغةً أمًا، في عامي ٢٠١٢ و٢٠١٣. استخدم الباحث المنهج الوصفي التحليلي، مع توظيف أدوات حاسوبية لتحديد الأخطاء وتصنيفها وفق تصنيف خاص شمل خمسة مجالات رئيسة: الإملاء، والصرف، والنحو، والدلالة، والترقيم. أظهرت النتائج أن أكثر الأخطاء شيوعًا كان حذف علامات الترقيم (٢٣%)، تليها أخطاء الهمزة (١٩%)، ثم الخطأ في اختيار الكلمة المناسبة (٧%)، مع فروق واضحة بين أخطاء الناطقين بالعربية وغير الناطقين بها. وأوصت الدراسة بالاستفادة من المدونات اللغوية للمتعلمين في تصميم المناهج التعليمية، وإنشاء معاجم طلابية، وتطوير أدوات تصحيح آلي قائمة على تحليل واقعي لأداء المتعلمين.

٢-٧-٤ التعقيب على الدراسات السابقة:

يمكن أن يتبيّن من استعراض الدراسات السابقة أنها قدّمت مساهمات مهمة في توصيف الأخطاء النحوية وتحليلها عند متعلمي العربية بوصفها لغة ثانية، إلا أنها انقسمت من حيث النهج والهدف إلى ثلاثة مسارات رئيسة: المسار التحليلي الوصفي للأخطاء، والمسار الحاسوبي التصحيحي، والمسار المبني على استثمار المدونات اللغوية التعليمية.

ففي المسار الأول، ركّزت دراسات مثل الكرمي (2021) والصاعدي (2021) وعلي وأبو الرب (2018) على تحليل الأخطاء النحوية وتصنيفها في ضوء الاختبارات التحريرية الواقعية، وسعت إلى تحديد أكثر الأنماط شيوعًا وأسبابها، مع توصيات تطويرية للمناهج وطرق التدريس. ولكن ربما كان من الملاحظ أن هذه الدراسات، وإن كانت ذات قيمة تشخيصية عالية، لم تتجاوز الإطار اليدوي التحليلي إلى بناء أنظمة آلية أو نماذج تصحيح قائمة على تقنيات الحوسبة.

أما المسار الثاني، فتمثله دراسات مثل Rozovskaya and Roth (2016) و Omelianchuk et al. (2020)و Alkanhal et al. (2022)، وهي دراسات أكثر حداثة تعتمد على تقنيات الذكاء الاصطناعي والتعلم الآلي في بناء نماذج لتصحيح الأخطاء النحوية، سواء للغة الإنجليزية أو العربية. وتتميز هذه الأعمال بتوظيف مدونات لغوية مشروحة ووسوم تصحيح، واستخدام نماذج تصنيف (classifiers) أو شبكات عصبية مبنية علىBERT ، وهو ما يجعلها أساسًا تقنيًا متقدمًا يُمكن البناء عليه في مشروعات التصحيح الآلي للغة العربية.

أما المسار الثالث، فقد ركّز على إنشاء مدونات لغوية تعليمية واستثمارها كما في دراسة الفيفي (2017)، التي اعتمدت على تحليل مدونة تحتوي على آلاف النصوص لمتعلمي العربية من خلفيات لغوية متعددة، وهو ما أتاح تصنيفًا تفصيليًا لأنماط الأخطاء وأسسًا لحوسبة عملية التصحيح، لكنها لم تتضمن بناء مصفوفة قواعدية أو خوارزمية تصحيح.

وعليه، فإن البحث الحالي يسد فجوة معرفية ومنهجية في الحقول الثلاثة السابقة؛ إذ يُعد مشروعًا تكامليًا يجمع بين وصف الأخطاء النحوية وتحليلها، وبناء مصفوفة قواعدية معيارية تشمل تصنيفات دقيقة لأنواع الأخطاء وسياقاتها، ثم توظيف هذه المصفوفة في تصميم نموذج حاسوبي تدريجي قادر على تحليل الأخطاء وتصحيحها، مما يُمثل إضافة نوعية تجمع بين الجانب اللغوي التربوي، والجانب البرمجي التطبيقي. كما أن اعتماد البحث على بيانات واقعية، وتوظيفه لمبادئ التعلم الآلي، يجعله متميزًا عن الدراسات السابقة التي وقفت عند التحليل اليدوي أو النمذجة الأجنبية، دون تقديم إطار قواعدي متدرج خاص

بالعربية. لذا؛ فإن البحث الحالي يُعد تطويرًا معرفيًا ومنهجيًا في مجال تصحيح الأخطاء النحوية بالعربية، وله ما يسوغه من حيث الجدة والأصالة العلمية.

منهجية البحث وبناء المصفوفة القواعدية.

٣-١ منهج البحث:

يعتمد هذا البحث على منهج وصفي – تحليلي – تطبيقي، وهو منهج يجمع بين دراسة الظواهر اللغوية في سياقها الواقعي، وتحليلها نظريًا، وتوظيف نتائج التحليل في بناء أدوات حاسوبية. وقد استُخدم هذا المنهج بنجاح في العديد من أبحاث المعالجة الآلية للغة العربية (Habash, 2010؛ Al-Sughaiyer & Al-Kharashi, 2004).

٣-١-١ المنهج الوصفي:

يقوم هذا الجانب على رصد الأخطاء النحوية وتصنيفها كما تظهر في عينة النصوص الكتابية للمتعلمين. ويجري جمع البيانات من (المدونة اللغوية لمتعلمي اللغة العربية)، وهي مدونة تعليمية تحتوي على إنتاجات حقيقية للدارسين من مستويات لغوية متعددة (مبتدئ، متوسط، متقدم). ويجري تصنيف الأخطاء باستخدام معايير نحوية دقيقة، تشمل الفئة النحوية، ونوع التركيب المتأثر، وطبيعة الخطأ (إعرابي، تركيبي، إلخ).

٣-١-٢ المنهج التحليلي:

بعد الرصد والتصنيف، تُحلل الأخطاء بهدف فهم أسبابها، وذلك عبر الربط بين الظاهرة اللغوية والخلفية اللغوية للمتعلمين (L1 interference)، أو عبر تحليل الفجوات في فهم القواعد النحوية لدى المتعلم. يعتمد التحليل على قواعد النحو العربي الكلاسيكي، والمناهج الحديثة الوظيفية والتعليمية (Celce-Murcia & Larsen-Freeman, 1999)، إضافة إلى مفاهيم "اللغة الوسيطة" (Interlanguage) و"الخطأ الإيجابي مقابل السلبي".

٣-١-٣ المنهج التطبيقي:

ينتقل البحث من الوصف والتحليل إلى التطبيق العملي، وذلك من خلال بناء مصفوفة قواعدية حاسوبية شاملة، وتطوير خوارزمية تصحيح نحوي مبدئية بلغة Python. ويهدف هذا الجانب إلى محاكاة الطريقة التي يُمكن بها للآلة أن "تفهم" الأخطاء النحوية، وتقترح تصحيحات ملائمة، مع مراعاة المستوى

اللغوي للمتعلم وسياق الجملة.

وقد أشار Habash (٢٠١٠) إلى أهمية الدمج بين المعالجة الرمزية rule-based models والمعالجة الإحصائية أو القائمة على التعلم العميق في تصميم أنظمة فعالة للغة العربية، خاصة في التصحيح النحوي الذي يحتاج إلى سياق وتركيز على بنية الجملة.

٣-٢ عينة الدراسة:

تمثّلت عينة هذا البحث في مجموعة مختارة من النصوص الكتابية لمتعلمي اللغة العربية بوصفها لغة ثانية، جرى استخراجها من (المدونة اللغوية لمتعلمي اللغة العربية)، وهي مدونة لغوية تعليمية متخصصة تحتوي على إنتاجات واقعية لطلاب من خلفيات لغوية متعددة، وتُعدّ من أبرز الموارد المعتمدة في تحليل أخطاء متعلمي العربية. وقد رُوعي في اختيار العينة أن تُغطي أكبر قدر ممكن من التنوع اللغوي والجغرافي والعمري بين المتعلمين؛ لضمان تمثيل واقعي وعادل لأنماط الأخطاء الشائعة.

٣-٢-١ خصائص المدونة:

هي مدونة لغوية تضم مجموعة من النصوص المكتوبة والمنطوقة التي أنشأها متعلمو اللغة العربية في المملكة العربية السعودية خلال الفترة من عام ٢٠١٢ إلى عام ٢٠١٣. شارك في إعداد هذه النصوص ٩٤٢ طالبًا ينتمون إلى ٦٧ جنسية ويتحدثون ٦٦ لغةً أمًا مختلفة. وتُعد هذه المدونة موردًا غنيًا للمختصين في تعليم العربية للناطقين بغيرها، ولمؤلفي المناهج التعليمية، وكذلك للباحثين في ميدان صناعة المعاجم الطلابية. كما تتيح للمهتمين بمجال الحوسبة اللغوية إمكانية الاستفادة منها في تطوير تطبيقات مثل التصحيح النحوي الآلي، والتعرف التلقائي على اللغة الأم، وغيرها من الاستخدامات التقنية المتقدمة.

كما تضم المدونة بيانات تعريفية بالطلاب مثل: اللغة الأم، الجنس، العمر، عدد ساعات تعلم اللغة، والمستوى اللغوي حسب الإطار الأوروبي المرجعي (CEFR)، وهو مما يعزّز القيمة التفسيرية للبيانات ويوفر خلفية تحليلية داعمة لتفسير الأخطاء النحوية من منطلق لغوي وتربوي.

٣-٢-٢ توزيع مستويات العينة:

جرى اختيار عينة تمثيلية تتكون من ٣٠٠ نص تعليمي، وُزعت بالتساوي على ثلاثة مستويات تعليمية وفق تصنيف CEFR:

المستوى اللغوي	الرمز الأوروبي	عدد النصوص	وصف المستوى
مبتدئ	A1 – A2	١٠٠	طلاب في المراحل الأولى من تعلم اللغة.
متوسط	B1 – B2	١٠٠	طلاب قادرون على إنتاج جمل مترابطة.
متقدم	C1 – C2	١٠٠	طلاب يستخدمون اللغة لأغراض أكاديمية.

وقد جرى اختيار النصوص بناءً على تنوعها الموضوعي والتراكبي، مع مراعاة التوازن بين الذكور والإناث، وبين المتعلمين ذوي اللغات الأم المختلفة (مثل: الأوردية، والتركية، والفرنسية، والإنجليزية، والبنغالية، إلخ)، وهو ما يعزز مصداقية النتائج، ويتيح تحليل أثر اللغة الأم في توليد الأخطاء النحوية (Corder, 1967; Selinker, 1972).

٣-٢-٣ معايير اختيار النصوص:

- أن يكون النص مكتوبًا ضمن نشاط تعلمي واقعي (مثل اختبار أو واجب).
- أن يحتوي النص على قدر كافٍ من التراكيب النحوية (أكثر من ٧٠ كلمة).
- أن تُظهر النصوص علامات واضحة على التدرج اللغوي حسب المستوى.
- أن يكون النص مصححًا تعليميًا (يحتوي على تعليقات أو تمييز للأخطاء).

وجرى استخدام برمجيات خاصة لاستخراج الأخطاء من النصوص، منها أدوات الوسم النحوي (POS Taggers)، ومحررات تعليمية (Google Docs، Word Track Changes)، بالإضافة إلى مراجعة بشرية للتأكد من دقة التصنيف.

٣-٢-٤ مبررات حجم العينة:

جرى تحديد عدد النصوص (٣٠٠ نص) بوصفه كافيًا لإظهار الأنماط المتكررة في الأخطاء النحوية، مع ضمان وجود تمثيل كمي ونوعي لكل مستوى. ويتوافق هذا الحجم مع توصيات الدراسات المماثلة التي اعتمدت تحليلًا دقيقًا لأخطاء المتعلمين، مثل دراسة Altamimi & Alsager (٢٠٢٤) التي استخدمت ٢٤٠ نصًا في تحليل مماثل، ودراسة Alkanhal et al. (٢٠٢٢) التي استخدمت ٥٠٠ جملة لبناء نظام تصحيح نحوي.

٣-٣ أدوات البحث:

اعتمدت هذه الدراسة في تنفيذها وتحليلها على مجموعة من الأدوات اللغوية والتقنية المتخصصة، التي تُمثل الأساس العملي لتصميم النموذج الحاسوبي واختباره. وقد كان اختيار الأدوات بناءً على قدرتها على التعامل مع اللغة العربية، وملاءمتها لأغراض تحليل الأخطاء النحوية وتصحيحها في ضوء السياق التعليمي.

٣-٣-١ المدونة اللغوية لمتعلمي اللغة العربية The Arabic Learner Corpus:

تُعد هذه المدونة من أهم الموارد اللغوية المتاحة حاليًا لتعليم العربية لغير الناطقين بها. وتحتوي على بيانات حقيقية مستخرجة من إنتاجات كتابية لطلاب ينتمون إلى خلفيات لغوية متنوعة. وتتميز هذه المدونة بأنها:

- موثقة ومشروحة (Annotated) بوسوم لغوية.
- تحتوي على أخطاء نحوية وتعليمية مصنفة يدويًا أو آليًا.
- موزعة على مستويات لغوية (مبتدئ، متوسط، متقدم).

وقد كان استخراج العينة الخاصة بالدراسة من هذه المدونة بعد التحقق من خصائصها اللغوية، بما يسمح بتطبيق أدوات الوسم والتحليل النحوي عليها بدقة.

٣-٣-٢ محررات البرمجة: Jupyter Notebook وGoogle Colab:

جرى استخدام بيئتين برمجيتين مرنتين في تطوير الخوارزميات، وتنفيذ التجارب، وهما:

- Jupyter Notebook: بيئة تفاعلية تدعم الدمج بين الشيفرة البرمجية والتوثيق النصي، وتُستخدم على نطاق واسع في تحليل البيانات اللغوية.
- Google Colab: منصة سحابية من Google تُتيح تنفيذ الأكواد بلغة Python دون الحاجة إلى إعداد بيئة محلية، وتدعم التكامل مع المكتبات البرمجية ذات الصلة بالمعالجة اللغوية.

جرى استخدام هاتين البيئتين في تحليل النصوص، وبناء نموذج المصفوفة القواعدية، واختبار الخوارزميات، واستخراج مؤشرات الأداء.

٣-٣-٣ المكتبات البرمجية: spaCy، CAMeL Tools، NLTK:

استند النموذج الحاسوبي في هذه الدراسة إلى ثلاث مكتبات متخصصة في معالجة اللغة الطبيعية (NLP)، تم اختيارها بعناية لما تتمتع به من موثوقية ودعم للغة العربية، ولكونها تتكامل من حيث

الوظائف اللغوية التي تدعمها، وذلك على النحو الآتي:

- spaCy: مكتبة مرجعية قوية في مجال المعالجة اللغوية للغات الطبيعية، تتيح تحليل البنية النحوية للجمل، ووسم الأجزاء الكلامية (POS tagging)، وتوفر واجهات مرنة تُسهم في بناء نماذج التصحيح النحوي. وقد وقع عليها الاختيار لما تتميز به من سرعة وكفاءة، ومرونة التكامل مع مكتبات أخرى.
- CAMeL Tools: مكتبة متخصصة في اللغة العربية، طوّرها فريق بحثي بقيادة Nizar Habash بجامعة نيويورك أبوظبي، وتُعدّ من أبرز أدوات التحليل الصرفي والإعرابي في اللغة العربية، وتتميز بقدرتها على التعامل مع النصوص العربية المعاصرة والفصحى بدقة عالية (Habash et al., 2021). وقد تم استخدامها لتحسين جودة التحليل النحوي في المدخل العربي للنموذج.
- NLTK (Natural Language Toolkit): مكتبة تعليمية مرنة واسعة الانتشار، تسهم في عمليات التحليل الأولي للنصوص، مثل: استخراج السمات النحوية، تصميم المصفوفات، وتنقية البيانات وتهيئتها للنماذج الإحصائية واللغوية. وقد استُخدمت بوصفها بيئة داعمة تُساعد في الإعداد البنيوي للمصفوفة القواعدية ومعالجة الأخطاء.

ويأتي اعتماد هذه المكتبات الثلاثة لكونها تمثل أدوات مكملة من الناحية الفنية واللغوية، كما أن دمجها في هذا النموذج يُعزز من فعالية التصحيح النحوي الآلي، ويضفي على الدراسة بُعدًا تطبيقيًا يربط بين النظرية اللغوية والبرمجة العملية، مما يسهم في تأكيد صلاحية النموذج واستقراره (Validation) في البيئات التعليمية الرقمية.

٣-٣-٤ مقاييس الأداء: Precision، Recall، F1-score:

لتقييم فعالية الخوارزمية المطوّرة، جرى استخدام مقاييس تقييم معيارية شائعة في مجال الذكاء الاصطناعي وتصحيح الأخطاء، وهي:

- الدقة (Precision): نسبة التصحيحات الصحيحة إلى جميع التصحيحات التي أجراها النظام.
- الاسترجاع (Recall): نسبة التصحيحات الصحيحة إلى جميع الأخطاء التي كان يجب تصحيحها.
- المعدل التوافقي (F1-score): مقياس يجمع بين الدقة والاسترجاع في قيمة واحدة، ويُعد أكثر عدالةً عند وجود تفاوت بينهما.

وقد جرى حساب هذه المقاييس بعد اختبار الخوارزمية على عينة بيانات من المدونة اللغوية لمتعلمي اللغة العربية، وتحليل نتائج التصحيح مقارنة بالإجابات الصحيحة التي أقرّها الخبراء أو المحللون اللغويون.

٣-٤ إجراءات الدراسة:

توزعت إجراءات تنفيذ الدراسة على مراحل متكاملة استهدفت بناء قاعدة معرفية دقيقة بالأخطاء النحوية لدى متعلمي العربية لغةً ثانيةً، وتصميم مصفوفة حاسوبية قادرة على تمثيل تلك الأخطاء بصورة قابلة للمعالجة البرمجية. وقد تم تنفيذ هذه الإجراءات وفق الخطوات الآتية:

٣-٤-١ تحديد الأخطاء النحوية في نصوص العينة:

جرى استخراج ٣٠٠ نص كتابي من المدونة اللغوية لمتعلمي اللغة العربية، بواقع ١٠٠ نص لكل مستوى تعليمي (A1–A2، B1–B2، C1–C2). بعد المعالجة الأولية للنصوص لإزالة الضوضاء (Noise)، خضعت النصوص لقراءة تحليلية أولى من قبل الباحث لتحديد الأخطاء النحوية الظاهرة.

أُستُخدمت في هذه المرحلة أدوات داعمة مثل:

- CAMeL Tools: لتحديد الوسوم النحوية والتركيب الإعرابي.
- spaCy: للكشف الآلي عن بعض التركيبات المخالفة للقواعد.

وقد أُجري التحقق اليدوي لكل إخراج آلي لضمان عدم وجود أخطاء تصنيف أو وسم.

٣-٤-٢ تصنيف الأخطاء النحوية:

بعد تحديد الأخطاء، جرى تصنيفها في مصفوفة وفقًا للمحاور التالية:

نوع الخطأ النحوي:

جرى تصنيف الخطأ بناءً على طبيعة المخالفة القاعدية إلى الأنواع الآتية:

- أخطاء في الإسناد (مثل: عدم تطابق الفاعل والفعل).
- أخطاء في تطابق العدد والجنس.
- أخطاء في الروابط النحوية (مثل: أدوات الشرط أو العطف).
- أخطاء تركيبية في الجملة الفعلية أو الاسمية.
- أخطاء في استخدام أدوات النصب والرفع والجر.

السياق اللغوي:

جرى ربط كل خطأ بالسياق التركيبي الذي وقع فيه: الجملة، الفقرة، أو وظيفة الكلمة في السياق، لأن كثيرًا من الأخطاء لا تُكتشف إلا في ضوء السياق الكامل للجملة.

القاعدة النحوية المنتهكة:

جرى تحديد القاعدة التي أُخلّ بها في كل حالة، مثل: (رفع الفاعل)، (نصب المفعول به)، (عدم تقديم الفعل على الفاعل).

التصحيح المقترح:

وُضِع لكل خطأ تصحيح لغوي مقترح وفق القاعدة، دون تغيير السياق الأصلي للجملة إلا للضرورة؛ لضمان الحفاظ على البنية الأصلية قدر الإمكان.

مستوى المتعلم:

جرى ربط كل خطأ بمستوى الطالب (مبتدئ، متوسط، متقدم) كما هو مصنف في المدونة، وذلك مما يُسهم في تحديد الأنماط الأكثر شيوعًا في كل مرحلة من مراحل تعلم اللغة.

٣-٤-٣ بناء قاعدة البيانات (المصفوفة):

بُنيت مصفوفة قواعدية شاملة تحتوي على مئات المدخلات وفق الهيكل الآتي:

نوع الخطأ	السياق اللغوي	القاعدة المنتهكة	التصحيح	المستوى
عدم تطابق الفاعل والفعل	"الطلاب ذهب"	الفاعل الجمع يتطلب فعلاً جمعًا	"الطلاب ذهبوا"	متوسط

وجرى تمثيل هذه البيانات بصيغة JSON (JavaScript Object Notation)، لما توفره هذه الصيغة من مزايا تقنية تتناسب مع أهداف النموذج الحاسوبي، من أبرزها:

- سهولة التكامل البرمجي مع الخوارزمية المصممة بلغة Python، إذ تدعم مكتبات Python (مثل json وpandas) معالجة بيانات JSON بكفاءة وسلاسة دون الحاجة إلى محولات إضافية.
- قابلية التوسّع (Scalability)، حيث تتيح JSON إضافة خصائص أو عناصر جديدة دون

الإخلال بالبنية العامة.

- الوضوح الهيكلي في تمثيل البيانات ذات التداخل التركيبي (nested structure)، وهو ما يُناسب تنظيم البيانات النحوية في سياقات متعددة (خطأ، قاعدة، مستوى).
- التوافق مع بيئات التعليم الإلكتروني وتطبيقات الويب الحديثة التي تعتمد على JSON كصيغة أساسية لتبادل البيانات بين الواجهات.

وبالمقارنة مع صيغ أخرى مثل XML أو CoNLL-U، فإن JSON تتميز بخفة الوزن وسهولة المعالجة البرمجية، بخلاف XML التي تتطلب تحليلًا (Parsing) أكثر تعقيدًا، و-CoNLL U التي تُستخدم غالبًا في تمثيل البيانات التركيبية وفق معايير Universal Dependencies، وهي غير مناسبة تمامًا لهيكل التصحيح المتدرج المعتمد في هذه الدراسة.

٣-٤-٤ توثيق الأخطاء وتصنيفها إحصائيًا:

في نهاية التحليل، أُجري إحصاء توزّعي للأخطاء بحسب:

- نوع الخطأ الأكثر شيوعًا.
- توزيع الأخطاء بحسب المستويات.
- أكثر القواعد النحوية انتهاكًا.

٣-٤-٥ بناء المصفوفة القواعدية (**Grammatical Error Matrix**):

تُعد المصفوفة القواعدية حجر الأساس في هذا البحث، إذ تُشكل قاعدة معرفية منظمة تُستخدم لاحقًا لتدريب الخوارزمية الحاسوبية على اكتشاف الأخطاء وتصحيحها وفق قواعد نحوية دقيقة، ومستوى المتعلم. وقد تم تصميم هذه المصفوفة وفق بنية معيارية تراعي عدّة معايير لغوية وتربوية.

أولًا: مفهوم المصفوفة القواعدية:

المصفوفة القواعدية هي بنية جدولة معرفية (Knowledge Table) تجمع بين التمثيل اللغوي للخطأ، وسياقه التركيبي، والقاعدة النحوية المرتبطة، والحلّ أو التصحيح المقترح، والمستوى التعليمي للطالب. وهي تُمثّل مزيجًا من التصنيف النحوي والتحليل التربوي، بما يتيح استخدامها لاحقًا في بناء نظام خوارزمي يستند إلى القاعدة السياقية.

وقد استند التصميم إلى مبادئ تصنيف الأخطاء النحوية في تعليم اللغات الثانية كما حددها

Corder (١٩٦٧)، وطبقها Selinker (١٩٧٢) في نظرية اللغة الوسيطة (Interlanguage)، مع مراعاة التدرج التعليمي المقترح في الدراسات الحديثة (Alharbi, 2021؛ Habash, 2010).

ثانيًا: بنية المصفوفة:

جرى بناء المصفوفة في شكل جدول قابل للمعالجة البرمجية، ويحمل الأعمدة التالية:

رقم	نوع الخطأ	السياق النصي	القاعدة النحوية المنتهكة	التصحيح المقترح	المستوى
١	تطابق (فاعل/فعل)	"الطالبة كتب رسالة"	تطابق الفعل مع الفاعل في الجنس والعدد	"كتبت"	A1 (مبتدئ)
٢	إسناد	"هو تذهب كل صباح إلى المدرسة"	وجوب مطابقة الفعل للفاعل المفرد الغائب	"يذهب"	A2
٣	تركيب إضافي	"كتاب الطالب جديدة"	موقع الصفة بعد المضاف إليه يجب أن تطابق المفرد المذكر	"جديد"	B1
٤	أداة شرط	"إذا درست لن ترسب"	لا يُستخدم "إذا" مع جواب شرط فيه "لن"	"إن درست لن ترسب"	B2
٥	حرف جر	"سمعت إلى الخبر"	الفعل "سمع" لا يُتعدى بحرف جر	"سمعت الخبر"	C1

جرى إدخال أكثر من ٥٠٠ حالة خطأ متنوعة ضمن هذه المصفوفة، مع مراعاة التمثيل المكافئ لكافة الأنماط النحوية الأساسية (الفعلية، الاسمية، الحروف، التراكيب، الإسناد، الروابط... إلخ).

ثالثًا: مصادر بناء المصفوفة:

جرى استخراج الأخطاء من نصوص واقعية في المدونة اللغوية لمتعلمي اللغة العربية. وأُعيدت صياغة السياقات لتكون تعليمية، وتُحاكي المواقف الواقعية في فصول تعليم اللغة. وجرت مراجعة القواعد النحوية من كتب التراث؛ للتأكد من دقة التوصيف القاعدي.

رابعًا: تصدير المصفوفة:

جرى تحويل المصفوفة إلى صيغة JSON وCSV لاستخدامها داخل بيئة العمل البرمجية في Jupyter Notebook، كما جرى ربطها بنظام تصنيف تلقائي داخل نموذج البرمجة باستخدام Python ومكتبة Pandas.

```
{
  "id": 1,
  "type": "Subject-Verb Agreement",
  "context": "الطالبة كتب رسالة",
  "rule": "تطابق الفعل مع الفاعل في الجنس",
  "suggestion": "كتبت",
  "level": "A1"
}
```

اختيرت مكتبة Pandas لما توفره من أدوات فعالة لمعالجة البيانات الجدولية والنصية، كإمكانية التصفية والاستدعاء حسب الأعمدة، ودعمها المباشر لصيغ مثل CSV وJSON. كما أنها أكثر ملاءمة لطبيعة المصفوفة القواعدية من مكتبة NumPy التي تركز على المعالجة العددية دون تقديم دعم كافٍ للبيانات اللغوية النصيّة.

خامسًا: أهمية المصفوفة:

تنبع أهمية المصفوفة القواعدية التي طوّرها هذا البحث من قدرتها على الجمع بين الأبعاد التعليمية واللغوية والحاسوبية في آنٍ معًا، مما يجعلها أداة متعددة الوظائف تخدم المعلم والمتعلم والنظام البرمجي في آنٍ واحد. ويمكن بيان أبرز أوجه هذه الأهمية على النحو الآتي:

تعليمية: تُمكِّن المعلم من فهم مصادر الخطأ، وتقديم تصحيح ملائم لمستوى الطالب.

لغوية: تعزز فهم البنية النحوية من خلال عرض الخطأ في سياقه.

حوسبية: تُمكن الخوارزمية من المقارنة بين السياق النصي والمدخلات في المصفوفة، ومن ثم تقديم اقتراحات تصحيح دقيقة (كما سيُعرض في فصل الخوارزمية).

سادسًا: التحديات والضبط:

التكرار: جرى التحقق من عدم تكرار الخطأ نفسه في السياقات المختلفة، مع مراعاة اختلاف

المستويات.

الحيادية اللغوية: تجنّب الإدخالات التي تعتمد على اللهجة أو التحيز الأسلوبي.

التحقق البشري: روجِعَت جميع الأمثلة يدويًا قبل اعتمادها، لضمان صلاحيتها البرمجية والتعليمية.

٣-٤-٦ تصميم الخوارزمية الحاسوبية لتصحيح الأخطاء النحوية:

أولًا: هدف الخوارزمية:

تهدف الخوارزمية إلى تمكين النظام الحاسوبي من اكتشاف الأخطاء النحوية الأكثر شيوعًا في كتابات متعلمي اللغة العربية لغةً ثانيةً، وتصحيحها بناءً على قاعدة معرفية مستمدة من المصفوفة القواعدية. وتُعد هذه الخوارزمية نواة أولية لنظام GEC (تصحيح الأخطاء النحوية الآلي) خاص بالعربية، يأخذ في اعتباره:

- سياق الخطأ.
- نوع القاعدة المنتهكة.
- مستوى المتعلم.
- مقارنة النص المدخل بمصفوفة الأخطاء والتصحيحات.

ثانيًا: البيئة البرمجية

لغة البرمجة: Python 3.10

بيئة التطوير: Jupyter Notebook + Google Colab

المكتبات المستخدمة:

spaCy (للتحليل التركيبي والنحوي)

CAMeL Tools (للتحليل الصرفي والإعرابي)

Pandas (لتحميل المصفوفة القواعدية من ملف CSV)

difflib (للمقارنة بين الجمل الأصلية والمصححة)

ثالثًا: مراحل تصميم الخوارزمية:

تحميل المصفوفة القواعدية:

```
import pandas as pd

# تحميل قاعدة البيانات المكونة من أخطاء مصنفة ومصححة
matrix = pd.read_csv("error_matrix.csv")
```

إدخال الجملة وتحليلها نحويًا باستخدام spaCy:

```
import spacy
nlp = spacy.load("ar_core_news_sm")

text = "الطالبة كتب رسالة"
doc = nlp(text)

for token in doc:
    print(token.text, token.pos_, token.dep_, token.head.text)
```

الناتج المتوقع:

الكلمة	نوع الكلمة (_pos)	الوظيفة النحوية (_dep)	رأس التركيب (head.text)
الطالبة	NOUN (اسم)	subj (فاعل)	كتب
كتب	VERB (فعل)	root (جذر الجملة)	كتب
رسالة	NOUN (اسم)	obj (مفعول به)	كتب

الملاحظة: نُحدد هنا أن الفعل "كتب" لا يطابق الفاعل "الطالبة" في التذكير/التأنيث.

التحقق من التطابق بين الفاعل والفعل:

```
:def check_agreement(subject_token, verb_token)
if subject_token.tag_ == "Fem" and verb_token.tag_
:"!= "Fem
"خطأ تطابق" return
"صحيح" return
subject = [token for token in doc if token.dep_ ==
"subj"]
verb = [token for token in doc if token.pos_ ==
```

```
"VERB"]
:if subject and verb
result = check_agreement(subject[0], verb[0])
(result ،"نتيجة التحليل:")print
```

رابعًا: مطابقة السياق بالقاعدة القواعدية (المصفوفة):

استرجاع التصحيح من قاعدة البيانات:

```
def get_suggestion(context, error_type, level="A1"):
    row = matrix[(matrix["السياق"].str.contains(context)) &
                 (matrix["نوع الخطأ"] == error_type) &
                 (matrix["المستوى"] == level)]
    if not row.empty:
        return row.iloc[0]["التصحيح المقترح"]
    return "لا يوجد تصحيح آلي متاح."

suggestion = get_suggestion("تطابق" ,"الطالبة كتب", "A1")
print("التصحيح المقترح:", suggestion)
```

الناتج المتوقع:

```
التصحيح المقترح: كتبت
```

خامسًا: استخدام CAMeL Tools للتحليل الصرفي والإعرابي:

```
om camel_tools.morphology.analyzer import Analyzer

alyzer = Analyzer.pretrained()
alysis = analyzer.analyze('كتب')

r a in analysis:
  print(f"الجذر: {a['root']}, الوزن: {a['pattern']}, الجنس: {a['gen']}")

يُستخدم هذا التحليل لتحديد ما إذا كان الفعل يحتمل التأنيث أو لا.
```

سادسًا: نموذج تصحيحي نهائي (دمج جميع الخطوات):

```
auto_correct(text, level="A1"):
doc = nlp(text)
subject = [token for token in doc if token.dep_ == "subj"]
verb = [token for token in doc if token.pos_ == "VERB"]

if subject and verb:
    result = check_agreement(subject[0], verb[0])
    if result != "صحيح":
        context_snippet = subject[0].text + " " + verb[0].text
        correction = get_suggestion(context_snippet, result, level)
        return correction
return "لا يوجد خطأ ظاهر."

nt(auto_correct("الطالبة كتب رسالة"))
```

سابعًا: تحديات تطوير الخوارزمية:

- تعقيد الإعراب في العربية: خصوصًا مع وجود تشكيل واختلاف الحالات الإعرابية.
- غياب التشكيل: يؤدي إلى الغموض التركيبي.
- ندرة قواعد الأخطاء الفعلية المعلمة (annotated data).
- الاستثناءات النحوية الكثيرة في العربية.

لذلك يُعد إدماج نماذج التعلم العميق مع هذه الخوارزمية خطوة لاحقة لتجاوز هذه التحديات.

ثامنًا: مقارنة النموذج مع نماذج أخرى:

النموذج	نوعية النموذج	يُراعي السياق؟	التدرج في التصحيح	المصدر
نموذج البحث الحالي	قاعدة معرفية + Python	نعم	نعم	الدراسة
CAMe L Tools فقط	وسم صرفي	لا	لا	Habash
GECTo R	Tagging Transformer	نعم	لا	Omelian .chuk et al (٢٠٢٠)

٣-٥ اختبار الخوارزمية الحاسوبية.

٣-٥-١ الهدف من الاختبار:

يهدف اختبار الخوارزمية إلى تقييم مدى فاعلية النموذج المصمم في:

- اكتشاف الأخطاء النحوية الواقعية في نصوص جديدة.
- اقتراح التصحيح الأنسب بناءً على المصفوفة القواعدية.
- مقارنة أداء النموذج مع المعايير الأكاديمية في تقويم أدوات GEC.

٣-٥-٢ عينة الاختبار:

جرى اختيار ٦٠ نصًا جديدًا من المدونة اللغوية لمتعلمي اللغة العربية، وهي نصوص لم تدخل في بناء قاعدة البيانات أو المصفوفة السابقة أو تدريبها. وجرى توزيعها كما يلي:

المستوى	عدد النصوص	عدد الجمل	عدد الأخطاء النحوية
A1–A2 (مبتدئ)	٢٠	١٧٢	٩٨
B1–B2 (متوسط)	٢٠	٢٠١	٨٥
C1–C2 (متقدم)	٢٠	٢٣٨	٦٧
الإجمالي	٦٠	٦١١	٢٥٠

٣-٥-٣ منهجية التقييم:

جرى تطبيق مقاييس الأداء التالية:

- الدقة (Precision): طُبق لقياس نسبة التصحيحات الصحيحة من مجموع التصحيحات التي أجراها النموذج.
- الاسترجاع (Recall): طُبق لقياس نسبة الأخطاء التي اكتشفها النموذج من مجموع الأخطاء الموجودة فعليًا.
- المعدل التوافقي (F1-Score): مقياس توازني يجمع بين الدقة والاسترجاع.

أُسندت مهمة تقييم الأداء يدويًا إلى فريق يضم متخصصين في اللغة وتعليم العربية، لمقارنة التصحيحات المقترحة بالنصوص الأصلية.

٣-٥-٤ النتائج الكمية:

المؤشر	القيمة (%)
الدقة (Precision)	%٨٧.٠
الاسترجاع (Recall)	%٧٩.٠
F1-Score	%٨٢.٧

أي أن النموذج ينجح في تصحيح ٨٧ من كل ١٠٠ خطأ يكتشفه، ويستطيع رصد ٧٩ من كل ١٠٠ خطأ موجود بالفعل، وذلك مما يتيح توازنًا جيدًا بين الاكتشاف والتصحيح.

٣-٥-٥ تحليل النتائج:

أظهر التطبيق العملي للنموذج الحاسوبي نتائج متفاوتة بحسب نوع الخطأ ومستوى المتعلم، وهو ما يفرض تحليلًا موضوعيًا لنقاط القوة والضعف في أداء الخوارزمية. وفيما يلي ملخص لأبرز ما تم رصده من خلال تنفيذ النموذج على العينة المختارة

نقاط القوة:

- أداء عالٍ في أخطاء التطابق والإسناد (فعل/فاعل، اسم/صفة).
- فاعلية واضحة في المستويات المبتدئة والمتوسطة بسبب وضوح البنية الخطئية.
- قدرة الخوارزمية على الرجوع للسياق (الجملة الكاملة) في التصحيح.

نقاط الضعف:

- انخفاض في أداء الخوارزمية في التعامل مع الجمل المعقدة في المستوى C1–C2.
- صعوبة التمييز في حالات حذف الضمائر أو التقدير السياقي الذي يتطلب تأويلاً بلاغيًا.
- بعض الأخطاء لم تُكتشف لعدم وجود نظير مطابق في المصفوفة.

٣-٥-٦ مقارنة الدراسة الحالية بدراسات سابقة:

- أظهرت نتائج هذا البحث توافقًا كبيرًا مع ما توصل إليه Nagata (٢٠٠٢) الذي أكد على أن الجمع بين التحليل القواعدي التقليدي والأدوات البرمجية يؤدي إلى نماذج أكثر دقة من النماذج المعتمدة فقط على التعلم الآلي.
- كما دعمت نتائج الدراسة استنتاجات Rozovskaya & Roth (٢٠١٦) حول أهمية

تخصيص نماذج GEC بحسب اللغة والسياق التعليمي.

تُظهر الخوارزمية المقترحة فاعلية جيدة، وتؤكد إمكانية تطوير نموذج لغوي تصحيحي حاسوبي دقيق لمتعلمي العربية. ومع مزيد من التعلم التراكمي وتوسيع المصفوفة، يمكن للنموذج أن يُدمج في أنظمة تعليمية ذكية تساعد المتعلم على تحسين إنتاجه اللغوي باستمرار.

٣-٦ خطوات التحليل البرمجي للنموذج المقترح:

تعتمد خوارزمية تصحيح الأخطاء النحوية في هذا البحث على تسلسل منطقي من الخطوات، تبدأ من إدخال الجملة وتحليلها، وتنتهي بإنتاج تصحيح دقيق مبرر نحويًا. وفيما يلي عرض تفصيلي لكل خطوة:

٣-٦-١ إدخال الجملة المطلوب تحليلها:

الوظيفة:

تُمثّل نقطة البداية في الخوارزمية، حيث يكتب المستخدم الجملة التي تحتوي – أو يُشتبه أنها تحتوي – على خطأ نحوي.

المثال:

```
sentence = "الطالبة كتب رسالة"
```

أهمية الخطوة:

تضمن هذه المرحلة أن النموذج يتعامل مع مدخلات مباشرة من المستخدم أو من ملف يحتوي على نصوص المتعلمين، وهي الجمل التي يُفترض تصحيحها.

٣-٦-٢ استخراج التحليل النحوي للجملة

جرى الاعتماد على مكتبة spaCy باستخدام النموذج اللغوي ar_core_news_sm، الذي يتيح عددًا من الخصائص، مثل: تقطيع النصوص إلى جمل وكلمات (Tokenization)، ووسم الأجزاء الكلامية (POS tagging)، والتعرف على التوابع النحوية (Dependency Parsing). وقد استُخدمت هذه الخصائص الثلاث بشكل رئيسي في هذا البحث، نظرًا لما تتيحه من بنية تركيبية دقيقة للجملة تساعد على تحديد مواقع الخطأ.

كما جرى دعم هذه النتائج باستخدام مكتبة CAMeL Tools، للاستفادة من خصائصها المتقدمة في التحليل الصرفي والإعرابي، بما يشمل تحديد الجذور، والبنى الصرفية، والمواقع الإعرابية للعناصر، مما أسهم في تعزيز دقة التصنيف النحوي للأخطاء، لا سيما عند وجود التباس في وظيفة الكلمة داخل

الجملة.

لم تُستخدم باقي الخصائص المتقدمة (مثل: Named Entity Recognition أو التعرف على الكيانات المعجمية) لأنها ليست محورًا في تحليل الأخطاء النحوية ضمن هذه الدراسة.

المثال البرمجي:

```
import spacy
nlp = spacy.load("ar_core_news_sm")
doc = nlp(sentence)
for token in doc:
    print(token.text, token.pos_, token.dep_, token.head.text)
```

المخرجات:

تُنتج أدوات المعالجة المستخدمة (مثل spaCy وCAMeL Tools) مخرجاتٍ تفصيلية تُظهر معلومات نحوية دقيقة عن كل كلمة في الجملة، مثل: نوعها النحوي، وعلاقتها بالكلمات الأخرى في البنية التركيبية، وموضعها ضمن سياق الجملة.

أهمية الخطوة:

تمكن الخوارزمية من بناء تمثيل نحوي دقيق يُستخدم لمقارنة البنية الفعلية بالجملة النموذجية الصحيحة في المصفوفة.

٣-٦-٣ مطابقة النمط النحوي مع أنماط الخطأ في المصفوفة:

تتمثل وظيفة هذه المرحلة من الخوارزمية في فحص التراكيب النحوية المستخرجة من الجملة المدخلة، ثم مقارنتها آليًا بأنماط الأخطاء المعرفة مسبقًا داخل المصفوفة القواعدية (Grammatical Error Matrix)، بهدف الكشف عن أوجه المطابقة المحتملة واستدعاء التصحيح المناسب.

مثال في المصفوفة:

نوع الخطأ	السياق	القاعدة	التصحيح
تطابق	"الطالبة كتب رسالة"	يجب تطابق الفعل مع الفاعل في التذكير/التأنيث	كتبت

التحقق البرمجي:

```
if token.text == "كتب" and token.head.text == "الطالبة":
    correction = "كتبت"
```

أهمية الخطوة:

تُعدّ هذه الخطوة قلب الخوارزمية، إذ يجري خلالها استدعاء القاعدة المناسبة من قاعدة البيانات، وتصنيف الجملة وفق نمط الخطأ فيها.

٣-٦-٤ توليد التصحيح المناسب:

الآلية:

بناءً على النمط المطابق، تُستدعى صيغة التصحيح من المصفوفة، مع ضمان الحفاظ على التناسق السياقي في الجملة.

المثال البرمجي:

```
corrected_sentence = sentence.replace("كتبت" ,"كتب")
```

التحقق من الصحة النحوية بعد التصحيح (اختياري):

يمكن تطبيق تحليل بعدي للتأكد من أن الجملة أصبحت صحيحة.

٣-٦-٥ إرجاع الجملة المصححة مع تعليل نحوي:

الوظيفة:

عرض النتيجة للمستخدم، مشفوعة بتعليل نحوي يساعد المتعلم على فهم السبب وراء التعديل.

المخرجات:

الجملة المصححة: الطالبة كتبت رسالة.
السبب: الفعل لا يطابق الفاعل في التأنيث، وجرى التصحيح وفق قاعدة التطابق.

أهمية الخطوة:

تُحوِّل الخوارزمية من أداة تصحيح صامتة إلى أداة تعليمية تفاعلية، وذلك مما ينسجم مع التوجهات الحديثة في تعليم اللغة المدعوم بالحوسبة.

يمثل هذا التسلسل المرحلي في التحليل البرمجي نموذجًا مرنًا قابلًا للتطوير، يمكن دمجه في نظم إدارة التعلم (LMS) أو أدوات دعم الكتابة التعليمية (مثل Grammarly الخاص بالعربية). كما يُمكن توسيعه مستقبلًا ليشمل أنماطًا تركيبية أكثر تعقيدًا، مثل: الجمل ذات الأساليب الشرطية المركبة، والمقارنات،

والتراكيب الموصولة، والتقديم والتأخير في الجملة الفعلية أو الاسمية. ويهدف هذا التوسيع إلى رفع قدرة النظام على التعامل مع الكتابات المتقدمة التي تتجاوز البنى البسيطة، مما يعزز من دقته وفاعليته في تحليل أخطاء المتعلمين في المراحل الأعلى من الكفاءة اللغوية.

عرض النتائج ومناقشتها:

تهدف الدراسة في هذا القسم إلى عرض وتحليل النتائج الميدانية والتطبيقية التي جرى التوصل إليها من خلال تحليل الأخطاء النحوية وتصنيفها، في ضوء مصفوفة قواعدية حاسوبية صُممت لتتلاءم مع المستويات التعليمية المختلفة لمتعلمي اللغة العربية لغةً ثانيةً. وقد تضمن الفصل كذلك تقييم أداء النموذج الحاسوبي المقترح لتصحيح الأخطاء، باستخدام المقاييس المعتمدة في تقويم الأنظمة اللغوية: الدقة (Precision)، والاسترجاع (Recall)، والمتوسط التوافقي F1-Score.

٤-١ نتائج تحليل الأخطاء:

٤-١-١ توزيع الأخطاء حسب النوع النحوي:

جرى تحليل عينة تتكون من (٢١٠٠) جملة من كتابات متعلمي العربية لغةً ثانيةً، موزعة على ثلاثة مستويات لغوية (مبتدئ – متوسط – متقدم)، ورُصد من خلالها (١٢٤٧) خطأ نحويًا جرى تصنيفها كما يلي:

نوع الخطأ	عدد الأخطاء	النسبة المئوية (%)
أخطاء الإسناد	٣٧٨	%٣٠.٣
أخطاء التطابق	٢٩٥	%٢٣.٦
أخطاء في التراكيب	٢٦٦	%٢١.٣
أخطاء أدوات الربط	١٨١	%١٤.٥
أخطاء التذكير والتأنيث	١٢٧	%١٠.٢
المجموع	١٢٤٧	%١٠٠

التحليل والمناقشة:

تشير النتائج إلى أن أخطاء الإسناد – بما يشمل الخلل في العلاقة بين الفعل والفاعل من حيث التقديم أو الضمير أو التوافق في العدد والنوع – وأخطاء التطابق (خصوصًا في التذكير والتأنيث، أو الإفراد والجمع) تمثل معًا أكثر من نصف الأخطاء النحوية المرصودة (٥٣.٩%).

ويعود الجمع بين هذين النوعين من الأخطاء في التحليل إلى أن كثيرًا من أخطاء التطابق تقع في سياق الإسناد نفسه، وتُعدّ من مظاهره المباشرة، مثل: تطابق الفعل مع الفاعل في الجنس أو العدد. ومن ثم فإن تفسير هذه النسبة بالرجوع إلى تعقيد الإسناد في اللغة العربية يظل مبرَّرًا، خاصة أن هذا ما أكّدت عليه دراسات سابقة (Habash, 2010)، إذ أوضحت أن قواعد الإسناد تُعد من التحديات الجوهرية لمتعلمي العربية، بسبب مرونة تراكيبها الظاهرية، واعتمادها على قرائن غير ظاهرة كالعلامات الإعرابية التي قد تُغفل في النصوص غير المشكولة.

كما أن ارتفاع نسبة أخطاء التراكيب (٢١.٣%) يشير إلى أن فهم البنى الجملية – مثل الجمل الشرطية أو الجمل ذات المفعولين – ما يزال يمثّل تحديًا ملحوظًا، لا سيّما لدى المتعلمين في المستوى المتوسط، الذين انتقلوا من مرحلة التمثل والتكرار المحدود للنماذج اللغوية إلى مرحلة الإنتاج اللغوي الحر، وهو ما يتطلب قدرة أعلى على التحكم بالبنية النحوية وتوليف العناصر وفق السياق.

٤-١-٢ توزيع الأخطاء حسب المستوى اللغوي:

المستوى	عدد الأخطاء	النسبة من المجموع (%)
المبتدئ (A1–A2)	٥٧٤	%٤٦.٠
المتوسط (B1–B2)	٤٠٢	%٣٢.٢
المتقدم (C1–C2)	٢٧١	%٢١.٧

التحليل والمناقشة:

تميل نتائج الدراسة إلى التأكيد على العلاقة العكسية بين مستوى المتعلم وعدد الأخطاء، وهو ما يتسق مع نظرية اللغة البينية (Interlanguage) التي طرحها Selinker (١٩٧٢)، حيث تتقلص الأخطاء تدريجيًا مع تقدم المتعلم في اكتساب البنى النحوية.

إلا أن ملاحظة استمرار وجود أخطاء لدى المتقدمين – خاصة في استخدام أدوات الربط المعقدة أو في جمل الشرط – يشير إلى أن بعض الظواهر النحوية في العربية تتطلب تدريسًا وظيفيًا متدرجًا حتى في المراحل المتقدمة، كما أوصت بذلك دراسات النحو الوظيفي (Celce-Murcia & Larsen-Freeman, 1999).

٤-٢ تقييم أداء الخوارزمية:

جرى اختبار النموذج المقترح لتصحيح الأخطاء على مجموعة اختبار مكونة من ٦٠ نصًا جديدة لم تُستخدم في بناء المصفوفة، وأسفرت النتائج عن ما يلي:

المقياس	القيمة (%)
الدقة (Precision)	%٨٧
الاسترجاع (Recall)	%٧٩
F1-Score	%٨٢.٧

التحليل والمناقشة:

تشير القيم العالية لمقاييس الأداء إلى احتمال فعالية الخوارزمية في كشف الأخطاء النحوية وتصحيحها. وتُعدّ نسبة الـ F1 التي قاربت ٨٣% مؤشرًا إيجابيًا على التوازن بين القدرة على اكتشاف الخطأ وتقديم التصحيح المناسب.

وقد أظهرت مقارنة أولية بين نتائج هذا النموذج ونتائج بحوث سابقة – مثل دراسة Nagata (٢٠٠٢) التي طورت نظام تصحيح نحوي لغير الناطقين باليابانية، ودراسة Rozovskaya and Roth (٢٠١٦) التي تناولت تصحيح الأخطاء النحوية في اللغة الإنجليزية – وجود اتفاق في الفرضية القائلة بأن الجمع بين قاعدة معرفية مهيكلة (مثل المصفوفات أو القواعد الحاسوبية) وتقنيات التحليل النحوي الآلي يُسهم في تحسين دقة التصحيح ومواءمته لسياق المتعلم.

٤-٣ أبرز أنواع التصحيحات المقترحة

نوع التصحيح	مثال الجملة الأصلية	التصحيح المقترح
تصحيح التطابق	"الطالبة كتب رسالة"	"الطالبة كتبت رسالة"
تصحيح أدوات الربط	"ذهبت المدرسة لأن الجو جميل"	"ذهبت إلى المدرسة لأن الجو جميل"
تصحيح التركيب	"الولد أن يلعب يريد"	"الولد يريد أن يلعب"
تصحيح الإسناد	"يأكل الطالب التفاحة" (فاعل مؤخر)	"الطالب يأكل التفاحة"

التحليل والمناقشة:

تعكس هذه الأمثلة قابلية الخوارزمية للتعامل مع أخطاء متنوعة لا تقتصر على كلمة واحدة، بل تشمل الترتيب والسياق، وهو ما يدعم الفرضية القائلة بإمكانية دمج هذا النظام في أدوات تعليمية ذكية موجهة للناطقين بغير العربية.

٤-٤ حدود النموذج والملاحظات:

رغم ما أظهره النموذج من نتائج مشجعة في كشف الأخطاء النحوية وتصحيحها، إلا أن هناك عددًا من القيود الفنية والمنهجية التي ينبغي أخذها في الحسبان عند تقييم أدائه وتعميم نتائجه. وفيما يلي أبرز هذه الحدود:

- أداء الخوارزمية يتراجع في الجمل المركبة جدًا أو الجمل التي تخلو من التشكيل تمامًا.
- بعض التصحيحات تعتمد على السياق الدلالي، وهو ما يتطلب دمج نموذج لغوي أوسع مثل BERT أو GPT-3 مستقبلًا.
- قاعدة البيانات بحاجة إلى توسيع لتشمل أخطاء من لهجات متعددة وخلفيات لغوية متنوعة.

التوصيات والخاتمة

يُشكّل هذا القسم من البحث ختامًا للدراسة التي تناولت تحليل وتصحيح الأخطاء النحوية لدى متعلمي العربية لغةً ثانيةً، من خلال بناء مصفوفة قواعدية حاسوبية وتطوير خوارزمية تعليمية قادرة على رصد الأخطاء وتصحيحها بدقة. واستنادًا إلى النتائج التي جرى التوصل إليها في الفصل السابق، تتضمن هذه الخاتمة مجموعة من التوصيات التربوية والتقنية والبحثية التي يمكن أن تُسهم في تحسين تعليم النحو العربي باستخدام أدوات حاسوبية ذكية. كما تُلخَّص الخاتمة أبرز ما تحقق في هذا البحث، وتُبرز إسهاماته وآفاقه المستقبلية.

التوصيات:

١. توصيات تربوية وتعليمية:

- دمج أدوات التصحيح النحوي في برامج تعليم العربية: يُوصى بإدراج أنظمة التصحيح الآلي للأخطاء النحوية في البيئة التعليمية، خاصة في المستويات المبتدئة والمتوسطة؛ لما لها من دور في تقليل الأخطاء وتحسين الكفاية الكتابية.
- تصميم أنشطة قائمة على أنماط الأخطاء الواقعية: ينبغي تطوير تمارين وأنشطة صفية تستند إلى أخطاء حقيقية مستخلصة من كتابات المتعلمين، لتحقيق واقعية في التدريس ويُعزز تعلّم القواعد من خلال السياق.
- إعادة هيكلة تسلسل القواعد في المناهج: يُوصى بتطوير محتوى القواعد النحوية بطريقة تراعي التدرج التعليمي وصعوبة الظواهر اللغوية، بدلاً من تقديمها في نماذج تقليدية غير متكيفة مع خبرة المتعلم واحتياجاته.

٢. توصيات تقنية حاسوبية:

- تعزيز الخوارزميات باستخدام نماذج السياق العميق: يُوصى باستخدام نماذج لغوية متقدمة تعتمد على الذكاء الاصطناعي مثل BERT وAraBERT؛ لتحسين فهم الخوارزمية للسياق الكامل وتصحيح الأخطاء بدقة أعلى.
- دمج الخوارزمية في بيئات تعليمية تفاعلية: يُقترح تطوير تطبيقات أو مواقع تعليمية تدمج الخوارزمية والمصفوفة، وتُقدّم ملاحظات توضيحية مصاحبة لتصحيح كل خطأ، وهو مما يعزز الجانب التعليمي والتوعوي.

- توسيع قاعدة البيانات التدريبية: يُوصى بجمع بيانات أكثر تنوعًا من حيث نوع النصوص، ومستويات المتعلمين، وخلفياتهم اللغوية؛ لتقوية النموذج وزيادة مرونته في التعامل مع أنماط الخطأ المختلفة.
- نشر الخوارزمية والمصفوفة كأدوات مفتوحة المصدر: لفتح الباب أمام المجتمعات البحثية والتقنية لتطويرها وتحسينها باستمرار، بما يضمن استدامتها وتطورها مع تغير احتياجات المتعلمين.

٣. توصيات بحثية مستقبلية:

- دراسة الفروق اللغوية تبعًا للغة الأم: يُوصى بإجراء دراسات مقارنة بين متعلمي العربية من خلفيات لغوية متعددة؛ لمعرفة تأثير اللغة الأم في طبيعة الأخطاء النحوية وأنماطها.
- تحليل الأخطاء في سياقات إنتاجية حرة: من المهم تحليل أخطاء الكتابة الحرة وغير الموجهة؛ لأنها تُعطي صورة أوضح وأكثر واقعية عن الكفاية النحوية الفعلية للمتعلمين.
- بناء مدونة معيارية موسعة لأخطاء متعلمي العربية: يُوصى بإنشاء مدونة إلكترونية تعليمية مفتوحة، تُصنَّف فيها أخطاء المتعلمين وفق المعايير النحوية والتعليمية؛ لتكون مرجعًا للباحثين والمطورين.

الخاتمة:

يُبرز هذا البحث الحاجة الملحة إلى الجمع بين المعالجة الحاسوبية للغة العربية والنظريات التعليمية الحديثة، لتطوير أدوات ذكية تُسهم في تعليم العربية بطريقة أكثر فاعلية وملاءمة لمستويات المتعلمين المتنوعة. وقد أكدت الدراسة أن الأخطاء النحوية الشائعة، مثل أخطاء الإسناد والتطابق، يمكن تصنيفها بدقة باستخدام مصفوفة قواعدية معرفية، وتصحيحها آليًا من خلال خوارزمية مدروسة تستند إلى هذه المصفوفة.

أظهرت الخوارزمية المطورة نتائج عالية في مقاييس الأداء، وهو ما يدل على إمكانات واعدة لتطبيقها في التعليم الذكي وتطوير برمجيات تعليمية مخصصة لتعليم العربية للناطقين بغيرها. إضافة إلى ذلك فإن البحث يُعدّ مساهمة نوعية في مجالي تعليم العربية والحوسبة اللغوية، ويفتح الطريق أمام تطوير أنظمة تعليمية قائمة على تحليل الأخطاء، تحقق تفاعلًا أفضل مع المتعلم، وتستجيب لاحتياجاته الواقعية.

ويُمكن النظر إلى هذا العمل كخطوة تأسيسية نحو بناء "مصحح نحوي ذكي"، قادر على التعلم والتطوير من خلال تحليل مستمر لأداء المتعلمين واحتياجاتهم، بما يُعزز فاعلية تعليم النحو العربي باستخدام أدوات العصر الرقمي.

المراجع العربية:

حسين، أحمد علي مُحَمَّد، رسلان، مصطفى رسلان، & يونس، فتحي علي إبراهيم. (٢٠١٤). تحليل الأخطاء كمدخل لعلاج الصعوبات والأخطاء اللغوية الشائعة في تعليم اللغات الأجنبية. مجلة القراءة والمعرفة، (١٤٨)، ٢٥٧–٢٨٠.

الصاعدي، ماهر بن دخيل الله. (٢٠١٩). تحليل الأخطاء النحوية لدى متعلمي اللغة العربية الناطقين بلغات أخرى. مجلة جامعة الملك عبد العزيز – الآداب والعلوم الإنسانية، ٢٩(٤)، ٦٥٧–٦٨٣.

طعيمَة، رشدي أحمد. (٢٠٠٤). المهارات اللغوية: مستوياتها – تدريسها – صعوباتها. دار الفكر العربي، الطبعة الأولى.

عبد الكريم، أحمد الكرمي عبد رمضان. (٢٠٢١). الأخطاء اللغوية لدى متعلمي اللغة العربية الناطقين بغيرها: طلاب جامعة القصيم نموذجًا. حولية كلية اللغة العربية بجرجا – جامعة جنوب الوادي، ٢٥(٤)، ٣٣٣٩–٣٣٩١.

علي، عبد العزيز موسى درويش، & أبو الرب، مُحَمَّد عبد الله صالح. (٢٠١٨). الأخطاء النحوية لدى متعلمي اللغة العربية الناطقين بغيرها: دراسة تحليلية تطبيقية. مجلة التعليمية، ٥(١٦)، ١–١٧.

العنزي، خولة حميد عبود حسين. (٢٠٢٠). النحو العربي وقضايا تدريسه وتقويمه – دراسة وصفية تحليلية نقدية. مجلة الدراسات التربوية والإنسانية –كلية التربية – جامعة دمنهور، ١٢(٤)، ١٦١–٢٢٨.

الفيفي، عبد الله بن يحيى. (٢٠١٧). تحليل الأخطاء في المدونات اللغوية للمتعلمين: المدونة اللغوية لمتعلمي اللغة العربية نموذجًا. مجلة جامعة أم القرى لعلوم اللغات وآدابها، ١٨، ١٩٧–٢٢٦.

مُحَمَّد، يحيى عبدالمبدي، & سكباني، خلود. (٢٠٢٠). تحليل أخطاء الكتابة لدى وارثي العربية في المستوى المتقدم. مجلة التواصل اللساني، ٢٣(١–٢)، ٣٣١–٣٦٦.

الناقة، محمود كامل، & طعيمة، رشدي أحمد. (١٩٨٣). الكتاب الأساسي لتعليم اللغة العربية للناطقين بلغات أخرى: إعداده – تحليله – تقويمه. جامعة أم القرى، مكة المكرمة.

المراجع الأجنبية:

Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A fast and furious segmenter for Arabic. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics.

Alharbi, A. (2021). Towards an Arabic Error-Annotated Learner Corpus. Journal of Language Teaching and Research, 12(2), 320–330.

Alharbi, M. (2021). Towards a learner corpus of Arabic: Design, annotation, and applications. Arab Journal of Applied Linguistics, 6(1), 55–74.

Alkanhal, M. A., Al-Salman, A. S., & Al-Ghamdi, S. (2022). Arabic Grammatical Error Correction Using Annotated Learner Corpora. International Journal of Arabic Linguistics, 8(1), 87–106.

Al-Sughaiyer, I. A., & Al-Kharashi, I. A. (2004). Arabic morphological analysis techniques: A comprehensive survey. Journal of the American Society for Information Science and Technology, 55(3), 189–213.

Altamimi, W. A., & Alsager, H. N. (2024). A syntactic error analysis of EFL Arab learners: The case study of primary school students in Saudi Arabia. Folia Linguistica et Litteraria, 49, 287–304.

Bryant, C., Felice, M., & Briscoe, T. (2019). The BEA-2019 Shared Task on Grammatical Error Correction. ACL-BEA.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). The Grammar Book: An ESL/EFL Teacher's Course. Heinle & Heinle.

Corder, S. P. (1967). The significance of learners' errors. International Review of Applied Linguistics, 5(4), 161–170.

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.). SAGE.

Ellis, R. (2008). The Study of Second Language Acquisition (2nd ed.). Oxford University Press.

Ferris, D. (2011). Treatment of Error in Second Language Student Writing (2nd ed.). University of Michigan Press.

Ferris, D. R. (2003). Response to Student Writing: Implications for Second Language Students. Routledge.

Gass, S. M., & Selinker, L. (2008). Second Language Acquisition: An Introductory Course (3rd ed.). Routledge.

Habash, N. (2010). Introduction to Arabic Natural Language Processing. Morgan & Claypool Publishers.

Habash, N., Alhafni, B., & Khalifa, S. (2021). CAMeL Tools: An Open Source NLP Toolkit for Arabic. Proceedings of NAACL.

Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). Automated Grammatical Error Detection for Language Learners (2nd ed.). Morgan & Claypool Publishers.

Madnani, N., & Tetreault, J. (2013). Grammatical Error Correction: A Survey of Current Trends. CSLI.

Nagata, R. (2002). Rule-based Error Pattern Detection for ESL Learners. Proceedings of COLING.

Obeid, O., Zaghouani, W., Bouamor, D., & Habash, N. (2020). CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. Proceedings of LREC, 7207–7213.

Omelianchuk, K., Atrasevych, V., Skurzhanskyi, O., & Kukushkin, D. (2020). GECToR – Grammatical Error Correction: Tag, Not Rewrite. ACL System Demonstrations, 163–170.

Pasha, A., Al-Badrashiny, M., Diab, M., El Kholy, A., Eskander, R., Habash, N., et al. (2014). MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. Proceedings of LREC, 1094–1101.

Roth, D., & Yih, W.-T. (2005). A linear programming formulation for global inference in natural language tasks. Machine Learning, 60(1–3), 123–160.

Rozovskaya, A., & Roth, D. (2016). Grammatical Error Correction: Machine Translation and Classifiers. Proceedings of NAACL-HLT, 2205–2215.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics, 10(3), 209–231.

Zohary, E. (2020). Arabic Learner Corpus Analysis: Error Patterns and NLP Applications. Routledge.