

اللغة العربية في زمن الترجمة الآلية وتحديات الذكاء الاصطناعي

مراجعة مقال □ Subject Rview

م.د. علي محمد زغير

Ali.m.888@ircoedu.uobaghdad.edu.iq

جامعة بغداد/ كلية التربية ابن رشد للعلوم الإنسانية

الملخص

يهدف هذا المقال إلى تحليل موقع اللغة العربية في عصر الترجمة الآلية الحديثة، مع التركيز في أبرز التحديات التقنية واللغوية التي تواجه الباحثين والمطورين، واستعراض الجهود البحثية الأكثر أهمية التي سعت إلى تحسين أداء النماذج اللغوية في العربية، وصولاً إلى استنتاجات عملية وتوصيات مستقبلية تخدم تطوير الذكاء الاصطناعي اللغوي في سياق عربي. الكلمات المفتاحية: اللغة العربية، الترجمة الآلية، الذكاء الاصطناعي .

Arabic language in the age of machine translation and the challenges of artificial intelligence

Dr. Ali Mohammed Zghair

University of Baghdad / Ibn Rushd College of Education for Humanities

Abstract

This article aims to analyze the position of the Arabic language in the era of modern machine translation, focusing on the main technical and linguistic challenges faced by researchers and developers, and reviewing the most significant research efforts that have sought to improve the performance of Arabic language models, leading to practical conclusions and future recommendations to support the development of linguistic artificial intelligence in the Arab context.

Keywords: Arabic language, machine translation, artificial intelligence

في ظل الثورة الرقمية الكبرى، أصبحت الترجمة الآلية والذكاء الاصطناعي من التقنيات البارزة التي تعيد تشكيل مشهد التواصل الإنساني، إذ لم تعد الترجمة حكراً على المترجم البشري، بل أصبحت الخوارزميات قادرة على تحليل اللغات ونقلها بدقة متزايدة، غير أن هذا التقدم يثير تساؤلات جوهرية حول مكانة اللغة العربية: كيف تحافظ على هويتها وثنائها في زمن السرعة الآلية؟ وهل يمكن للذكاء الاصطناعي أن يكون أداة لخدمة العربية أم تهديداً لبقائها في المشهد اللغوي العالمي؟

في عصر رقمي سريع التطور، تُشكّل تقنيات الترجمة الآلية والذكاء الاصطناعي منعطفاً مهماً في علاقة الإنسان باللغة، إذ تُعدّ اللغة العربية إحدى أقدم اللغات السامية وأكثرها ثراءً من الناحية الصرفية والدلالية، وهي اللغة الرسمية لأكثر من عشرين دولة ويتحدث بها ما يزيد عن أربعمائة مليون شخص حول العالم، هذا الامتداد الجغرافي والتنوع اللهجي الكبير جعلها مادةً لغوية خصبة لكنها في الوقت ذاته تطرح تحديات كبيرة على أنظمة المعالجة الحاسوبية، وخصوصاً في عصر الترجمة الآلية القائمة على الذكاء الاصطناعي، فبينما حققت اللغات الأوروبية نتائج مبهره في أنظمة الترجمة العصبية (Neural Machine Translation - NMT)، ما زالت العربية تواجه عقبات تتعلق بالبنية الصرفية، غياب التشكيل، وتعدد اللهجات، وهي عوامل تؤثر في دقة الترجمة وجودتها .

لأن اللغة العربية من أكثر اللغات تعقيداً من حيث البنية الصرفية والنحوية، وهو ما يجعلها من اللغات "عالية التعقيد الحاسوبي" في بيئة المعالجة الآلية للغة الطبيعية، ويكمن التحدي الأساسي في التنوع الصرفي الغزير الذي يسمح بتوليد عدد كبير من المشتقات من الجذر الواحد، الأمر الذي يخلق صعوبات في التحليل الصرفي وفي بناء قواعد بيانات لغوية دقيقة، فمثلاً، الكلمة العربية الواحدة قد تحمل أكثر من عشرة احتمالات صرفية بحسب السياق، وهو ما يضع عبئاً إضافياً على أنظمة الترجمة العصبية التي تتعامل غالباً مع تمثيلات خطية للنصوص، يضاف إلى ذلك أن غياب التشكيل في النصوص العربية المكتوبة الحديثة يجعل من الصعب على الخوارزميات التمييز بين المعاني المختلفة للكلمة الواحدة، فكلمة مثل "علم" قد تُقرأ بمعنى "أدرك" أو "درّس" أو "عرّف"، بحسب التشكيل الغائب، ولأن معظم النصوص العربية الرقمية غير مشكّلة، فإن خوارزميات الترجمة والذكاء الاصطناعي تواجه صعوبة في تحديد المعنى المقصود بدقة، إذ تظهر أيضاً إشكالية الازدواجية اللغوية (Diglossia)، أي الفجوة بين الفصحى والعاميات، وهي من أبرز خصائص اللغة العربية، فالذكاء الاصطناعي يواجه صعوبة في معالجة النصوص المكتوبة باللهجات المحلية، والتي تختلف جذرياً في المعجم والبنية عن العربية الفصحى، كما أن غياب الموارد المعيارية لكل لهجة (corpora) يؤدي إلى ضعف الأداء في معالجة اللغة الطبيعية للعاميات، رغم وجود مبادرات مثل "Arabic Dialect

Identification” لتحسين هذه النماذج ، ومن التحديات التقنية أيضاً قلة البيانات المشروحة (annotated data)، إذ تعتمد معظم نماذج التعلم العميق في اللغات الأخرى على ملايين الجمل المترجمة والموسومة يدوياً، بينما تفنقر اللغة العربية إلى قواعد بيانات ضخمة مماثلة، وقد أشارت دراسات عدة إلى أن الموارد اللغوية العربية المتاحة لا تغطي سوى نطاق محدود من المجالات الأسلوبية والمعجمية، مما يؤثر سلباً في جودة الترجمة الآلية وفي قدرة النماذج اللغوية على التعميم ، كما أن تعدد أنظمة الكتابة مثل استخدام الحروف اللاتينية في “العربيزي” أو الاختصارات الإلكترونية في وسائل التواصل الاجتماعي يخلق طبقة إضافية من التحدي أمام الذكاء الاصطناعي، فهذه الصيغ غير المعيارية تجعل من الصعب على الأنظمة تعلم الأنماط اللغوية بشكل منهجي، إذ تدمج بين أنظمة ترميز مختلفة لا تخضع لقواعد ثابتة ، إضافة إلى ذلك، تواجه اللغة العربية مشكلة نقص أدوات التقييم المعيارية لجودة الترجمة الآلية، فمعظم أدوات التقييم (مثل BLEU أو METEOR) صُممت أساساً للغات ذات بنية تركيبية أبسط كالإنجليزية أو الفرنسية، ولا تأخذ في الاعتبار طبيعة التركيب الحر في العربية، وقد أوصت دراسات حديثة بتطوير أدوات تقييم خاصة باللغة العربية تراعي الاختلافات الصرفية والإعرابية .

ومن التحديات الثقافية والمعرفية في الترجمة الآلية للغة العربية إلى جانب التحديات التقنية، تبرز قضايا ثقافية ومعرفية تتعلق بقدرة أنظمة الذكاء الاصطناعي على نقل المعاني الثقافية والدلالات الضمنية في النصوص العربية، فاللغة العربية غنية بالإحالات الدينية والرموز الثقافية التي يصعب تمثيلها بدقة في اللغات الأخرى، فعلى سبيل المثال، تحتوي نصوص الأدب العربي على طبقات من المجاز والتضمين البلاغي التي لا يمكن ترجمتها ميكانيكياً دون فقدان جزء من القيمة الدلالية الأصلية ، إذ تُظهر الدراسات أن الترجمة الآلية الحالية تميل إلى “تحييد” المعاني الثقافية أو استبدالها بمكافئات عامة، وهو ما يؤدي إلى فقدان العمق المعنوي للنص العربي، ويُلاحظ ذلك في الترجمات الآلية للقرآن الكريم أو الشعر العربي، حيث تُختزل الدلالات البلاغية في تراكيب لغوية مبسطة لا تعبر عن المقاصد الأصلية ، كما تطرح الترجمة الآلية إشكالية السيطرة اللغوية والمعرفية، إذ تميل الأنظمة المدعومة بالإنجليزية إلى فرض نماذج فكرية غربية على عملية الفهم اللغوي، وقد نبه باحثون عرب إلى خطر ما يعرف بـ “الهيمنة المعجمية” (Lexical Imperialism)، حيث تتراجع الألفاظ والتراكيب العربية الأصيلة لصالح مكافئات إنجليزية هجينة تستخدم في النصوص التقنية والإعلامية.

أما على مستوى التواصل الإنساني، فإن الاعتماد المفرط على الترجمة الآلية قد يؤدي إلى فقدان الحس اللغوي والذائقة الأدبية لدى المستخدمين، خاصة في المجالات الأدبية أو الإعلامية، فالذكاء الاصطناعي يعتمد على أنماط إحصائية ولا يمتلك إدراكاً دلاليًا حقيقياً، مما يجعل نتائجه ميكانيكية في التعبير عن المشاعر أو الرموز الثقافية .

رغم التحديات السابقة، يمكن النظر إلى الذكاء الاصطناعي أيضاً كفرصة تاريخية لتجديد حضور اللغة العربية، إذ أدى انتشار أدوات المعالجة اللغوية العربية إلى تعزيز إعادة تموضع اللغة في الفضاء الرقمي، خاصة مع تطوير تطبيقات مثل "BERT Arabic" و "GPT Arabic Models" التي أظهرت كفاءة في فهم النصوص العربية وتوليدها، وقد ساهمت هذه التقنيات في إتاحة أدوات تحرير لغوية ذكية تحسن الكتابة العربية وتكشف الأخطاء النحوية والصرفية آلياً، مثل أدوات التدقيق النحوي والتوليد الآلي للنصوص، ومن ثم يمكن القول إن الذكاء الاصطناعي لم يقتصر على الترجمة، بل أصبح عنصراً مساعداً في تعليم اللغة العربية وفي تمكين الباحثين من تحليل النصوص التراثية الرقمية على نطاق واسع، ومع تطور المعالجة اللغوية العربية، بدأت تظهر مشروعات رقمية كبرى تهدف إلى توثيق التراث العربي رقمياً، مثل "مشروع الذخيرة العربية" و "Arabic WordNet"، والتي ساهمت في توسيع قاعدة البيانات المعجمية وتسهيل التكامل بين اللغة العربية والمنصات الذكية العالمية، وتعد هذه الجهود نواة لنهضة لغوية رقمية يمكن أن تجعل العربية لغة فاعلة في الفضاء المعرفي العالمي.

ومن المبادرات العربية في تطوير الذكاء الاصطناعي اللغوي إذ أطلقت عدة مؤسسات عربية برامج استراتيجية لدعم اللغة العربية رقمياً في إطار "رؤية التحول الرقمي"، ومن أبرزها مبادرة "مركز محمد بن زايد للذكاء الاصطناعي" في الإمارات، التي تسعى إلى تطوير نماذج لغوية عربية مفتوحة المصدر تنافس النماذج العالمية مثل GPT و BERT. كما أطلقت منظمة الألكسو مشروعات لتعريب تقنيات الترجمة والتعليم الآلي، وتهيئة بيئات حوسبة سحابية مخصصة للنصوص العربية، وفي المملكة العربية السعودية، قامت هيئة الذكاء الاصطناعي (SDAIA) بإطلاق برامج لتطوير المعاجم الرقمية والأنظمة الذكية للغة العربية، ضمن استراتيجية وطنية تهدف إلى جعل العربية لغة التكنولوجيا والبحث، كما أنشئت مبادرة "مشروع اللغة العربية الموحدة" لتوحيد المصطلحات التقنية وتطوير معايير الترجمة الآلية بين العربية واللغات الأخرى، تدل هذه الجهود على وجود وعي مؤسسي عربي متزايد بأهمية الذكاء الاصطناعي في حماية الهوية اللغوية وتعزيز مكانة العربية عالمياً ومع ذلك، تبقى الحاجة ملحة إلى مزيد من التعاون البحثي بين الجامعات العربية والمراكز التقنية الدولية لضمان استدامة هذه المبادرات.

الخاتمة :

يمكن القول إن اللغة العربية، رغم ما تواجهه من تحديات في زمن الترجمة الآلية والذكاء الاصطناعي، تمتلك إمكانات فريدة تؤهلها لأن تكون فاعلاً لغوياً عالمياً إذا ما توفرت البيئة البحثية والموارد التقنية اللازمة، فالتعدد الصرفي والدلالي الذي يشكّل عقبة اليوم، يمكن أن يصبح غداً مصدر ثراء للأنظمة الذكية إن أحسن استثماره علمياً،

إن تطوير أدوات المعالجة اللغوية العربية لا ينبغي أن يكون مشروعاً تقنياً فقط، بل مشروعاً حضارياً شاملاً يهدف إلى صون الهوية اللغوية وتمكين الأجيال الجديدة من التفاعل مع الذكاء الاصطناعي من موقع المشاركة لا التبعية.

وفي ضوء ما سبق، تبرز الحاجة إلى إنشاء مراكز بحث عربية مشتركة تجمع بين المتخصصين في اللسانيات والحوسبة والترجمة، لتطوير نماذج لغوية تراعي الخصوصيات الثقافية للعربية وتتنافس على المستوى العالمي، إن مستقبل العربية في زمن الذكاء الاصطناعي لن يصنع بالآلة وحدها، بل بالعقل العربي القادر على فهمها وتوجيهها نحو غايات معرفية وإنسانية راقية .

المصادر :

- الألكسو. (٢٠٢٢). التقرير العربي حول الذكاء الاصطناعي وتطوير اللغة العربية. تونس: المنظمة العربية للتربية والثقافة والعلوم.
- الخشاب، طارق. (٢٠٢٢). اللغة العربية في عصر الذكاء الاصطناعي. القاهرة: دار الفكر العربي.
- الخالدي، سامر. (٢٠٢٢). الهيمنة المعجمية في اللغة العربية الرقمية. بيروت: مركز دراسات اللغة والمجتمع.
- الهيئة السعودية للبيانات والذكاء الاصطناعي (SDAIA). (٢٠٢٣). مبادرات اللغة العربية والذكاء الاصطناعي. الرياض.
- Al-Jarf, R. (2021). Machine Translation and Arabic Texts: Challenges and Strategies. *Journal of Linguistic Studies*, 12(3), 11–22.
- Al-Saif, A., & Markert, K. (2010). The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. *Language Resources and Evaluation*, 44(1), 97–112.
- Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. arXiv preprint arXiv:2003.00104.
- Darwish, K. (2014). Arabizi Detection and Conversion to Arabic: A Survey. *ACM Transactions on Asian Language Information Processing*, 13(4), 1–11.
- Habash, N. (2010). *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool.

- Habash, N. (2021). Arabic NLP in the Age of Deep Learning: A Survey. *Journal of Computational Linguistics*, 47(2), 1–38.
- Hatim, B., & Mason, I. (1997). *The Translator as Communicator*. London: Routledge.
- Kilgarriff, A., et al. (2014). The Sketch Engine: Ten Years On. *Lexicography*, 1(1), 7–36.
- Post, M. (2018). A Call for Clarity in Reporting BLEU Scores. *Association for Computational Linguistics*, 209–218.
- Salameh, M., & Bouamor, H. (2018). Understanding Machine Translation Evaluation for Arabic. *Proceedings of the 2nd Arabic NLP Workshop*, 30–38.
- Zerrouki, T., & Balla, A. (2017). Tashkeela: An Arabic Diacritized Corpus. *Arabic Language Resources Journal*, 2(1), 42–50.
- Zaidan, O., & Callison-Burch, C. (2014). Arabic Dialect Identification. *Computational Linguistics*, 40(1), 171–202.