المشرق الرقميّة
٢٨
العدد الثامن والعشرون
أيلول ٢٠٢٦

الترجمة الآليَّة بين الاحتمالات الإحصائيَّة والتمثيلات السياقيَّة
دراسة تحليليَّة مقارنة

الدكتورة ندى مرعشلي *

يتعيَّن على المُترجِم الاختيار بين الالتزام بالنصِّ الأصليّ أو إعادة الصياغة لضمان الفهم.

مستخلص

يُشير عنوان المقالة إلى تطوُّر طرائق الترجمة الآليَّة من أسلوبٍ يعتمد أساسًا على الاحتمالات والإحصاء إلى أسلوب أحدث يعتمد على فهم الكلمات ضمن سياقها، بالتركيز على دور الترجمة الآليَّة الإحصائيَّة Statistical Machine Translation، ولاسيَّما نموذجَي اللغة والترجمة في اختيار الترجمة الأنسب للُّغة الهدف. تناقش الدراسة تحدِّيات اللغة العربيَّة التي تعترض عمليَّة الترجمة من ناحية ثرائها الصرفيّ،

* **الدكتورة ندى مرعشلي:** أستاذة اللسانيَّات والأسلوبيَّات وتحليل الخطاب وعلم الصوتيَّات وتحقيق المخطوطات في الجامعة اللبنانيَّة، كلِّيَّة الآداب والعلوم الإنسانيَّة، الفرع الأوَّل. باحثة ومؤلِّفة، وعضو في لجنة تحكيم الأبحاث في منصَّة "Arid" الدوليَّة.
nadamerashly66@gmail.com

والنحويّ، واللواحق المضافة، وتناوب علامات التشكيل، وتنوُّع ترتيب الكلمات، والتعدُّد المعجميّ، موضِحةً دور الترجمة الآليَّة العصبيَّة Neural Machine Translation والتمثيلات السياقيَّة Contextual Representation في تذليل بعض التحدِّيات بالنظر إلى الكلمات ضمن سياقها اللغويّ الأشمل. تتبنَّى المقالة منظورًا تحليليًّا مقارِنًا لتسليط الضوء على الاختلافات بين الترجمة الآليَّة الإحصائيَّة والعصبيَّة، وتخلص إلى أنَّ التحوُّل نحو المناهج العصبيَّة والسياقيَّة يمثِّل تطوُّرًا مهمًّا في الترجمة الآليَّة، خصوصًا عند التعامل مع الغموض والمعنى السياقيّ، مع استمرار البحث عن معالجة فعَّالة للتحدِّيات المتعلِّقة بتعقيد اللغة العربيَّة اللغويّ.

إنَّ تعدُّد الترجمات الممكنة، والغموض اللغويّ، واختلاف التراكيب بين اللغات، والحاجة إلى المفاضلة بين البدائل واختيار الترجمة الأكثر احتمالًا اعتمادًا على البيانات والسياق، سرَّعت عمليَّة الانتقال من المقاربة الإحصائيَّة القائمة على الاحتماليَّة إلى المقاربة العصبيَّة القائمة على الشبكات العصبيَّة والتمثيلات السياقيَّة. فهناك ضرورة علميَّة تحتِّم فهم الآليَّات التي تعتمدها أنظمة الترجمة الآليَّة في اختيار معنى الكلمات متعدِّدة الدلالة المناسب، بخاصَّةٍ في ظلِّ تطوُّر الذكاء الاصطناعيّ المتسارع، وتقنيَّات معالجة اللغة الطبيعيَّة. إنَّ فهم حدود الترجمة الآليَّة وإمكاناتها، بخاصَّةٍ بالنسبة إلى المترجِم البشريّ، يرتبط بتبيان الفرق بين الترجمة الآليَّة الإحصائيَّة والترجمة الآليَّة العصبيَّة، وبالتعرُّف إلى كيفيَّة عمل أنظمة كلٍّ منها، فعبر آليَّة العمل وبمساعدة السياق يمكن تحديد معنى الكلمات متعدِّدة الدلالة، كما يمكن الجمع بينها من خلال الترجمة الهجينة Hybrid Translation؛ وهنا تبرز المشكلة وتدعو إلى السؤال الآتي:

إلى أيِّ مدًى أسهم الانتقال من الاحتمالات الإحصائيَّة إلى التمثيلات السياقيَّة في تحسين دقَّة الترجمة الآليَّة؟ ما المقصود بالترجمة الآليَّة الإحصائيَّة، وما آليَّة عملها؟ ما دور نموذج اللغة ونموذج الترجمة في المقاربة الإحصائيَّة؟ ما المقصود بالترجمة الآليَّة العصبيَّة والتمثيلات العصبيَّة؟ كيف تعالِج الترجمة العصبيَّة الكلمات متعدِّدة المعنى، مثل: Fan وStar؟ وهل أسهمت التمثيلات السياقيَّة في تجاوز بعض أوجه القصور في الترجمة الإحصائيَّة؟ إذ إنَّ المفترض بالترجمة العصبيَّة، وبفضل اعتمادها على السياق، أنَّها تتعامل مع الكلمات متعدِّدة المعنى بصورة أفضل من الترجمة الإحصائيَّة، وأنَّ الانتقال من الاحتمالات إلى التمثيلات سيُسهم في تحسين قدرة أنظمة الترجمة الآليَّة على اختيار الترجمة المناسبة؛ مثل: Fan التي قد تعني (مروحة) أو (مشجِّع)؛ وبذلك تتجاوز الترجمة العصبيَّة بعض أوجه القصور في الترجمة الإحصائيَّة، بخاصَّةٍ في معالجة السياق والعلائق بين الكلمات في الجملة؛ وسيكون الجمع الهجين بين المعلومات المستخلَصة من البيانات والتمثيلات السياقيَّة قابلًا لترجمةٍ أكثر ملاءمة للسياق. تمَّ استثمار إجراءات المنهج المقارن لأنَّ الترجمة لغويَّةٌ طبيعيَّة لديها قابليَّة التكرار والتجانس، عدا عن إمكانيَّة إخضاعها كلِّها للمقارنة والوصف في كلِّ ما يتعلَّق بالنموذج الترجميّ من بياناتٍ وإحصاءات؛ "المنهج المقارن هو منهج متعدِّد الأدوات، يُستخدم في مجالات الوصف والتفسير والتحليل والتنبُّؤ، ولكن وفق حاجات الدراسة المقارنة؛ كما أنَّه لا ينفصل عن مناهج البحث المعروفة: المنهج الوصفيّ، والمنهج التاريخيّ، والمنهج التحليليّ، والمنهج التجريبيّ." (بكر، ٢٠٠٣، ص. ٧). هذه المقارنة تحتاج إلى أدوات الوصف والتحليل والتفسير لاكتشاف ما في النظام اللغويّ من مشكلات تواجه تنفيذ أهدافه وسياسته. ويتحدَّد الهدف من دراسة

الترجمة الآليَّة بإنتاج تطبيقاتٍ قادرة على ترجمة نصّ (مكتوب) أو كلام (منطوق) من لغةٍ إلى أخرى؛ فعمليَّة الترجمة "تتمُّ بين لغتَين مكتوبتَين مختلفتَين، بتحويل نصّ مكتوب أصليّ (النصّ المصدر) في اللغة المنطوقة الأصليَّة (لغة المصدر) إلى نصّ مكتوب (النصّ الهدف) في لغة منطوقة مختلفة (لغة الهدف) (Munday, 2016, P.8)؛ إلَّا أنَّ لكلّ لغة تعقيدات تكمن في إنتاج ترجمة تولّد جملًا صحيحة في اللغة الهدف، وتحافظ على المعنى الأصليّ في اللغة المصدر، من دون ارتباطٍ ببِنية الجملة الأصليَّة. فهدف الترجمة الآليَّة "هو الحصول على جودةِ ترجمة تُضاهي جودة الترجمة البشريَّة." (Poibleau, 2017, P.19).

جدول يبيّن الفرق بين الترجمة الآليَّة الإحصائيَّة والترجمة الآليَّة العصبيَّة

الترجمة الإحصائيَّة	الترجمة العصبيَّة
تعتمد على الاحتمالات	تعتمد على الشبكات العصبيَّة
تعتمد على نماذج منفصلة نسبيًّا	لديها القدرة على تعلُّم تمثيلات متكاملة
تعتمد على البيانات الإحصائيَّة	تعتمد على التعلُّم العميق
سياقها محدود	تستثمر كامل السياق

بدايات الترجمة الآليَّة: شهدت الأنظمة الأولى، القائمة على القواعد في فترة ما بعد الحرب، ظهور أولى الحواسيب، وسرعان ما حظيت الترجمة الآليَّة باهتمامٍ شديد، وعُدّت تطبيقًا رئيسًا ومطلبًا ملحًّا لترجمة النصوص تلقائيًّا من مصادر أجنبيَّة في سياق الحرب الباردة Cold War التي اقتُصرت على التنافس السياسيّ، والعسكريّ، والاقتصاديّ، والإيديولوجيّ بين الولايات المتَّحدة والاتِّحاد السوفييتيّ واستمرَّت حتَّى العام ١٩٩١. وقد أسهم التقدُّم في مجال علم التشفير في الحرب في تذليل قيود الحواسيب الأولى للكشف عن مضمون اللغة الأجنبيَّة التي عُدَّت من أدوات التشفير؛ فطوَّر الخبراء منهجًا عمليًّا قائمًا على القواميس ثنائيَّة اللغة وقواعد النقل، وهذا الأمر أتاح تغيير ترتيب الكلمات وفقًا لخصوصيَّات اللغة الهدف؛ "هذا النهج المعروف باسم النهج القائم على القواعد، كان النهجَ السائد لعقود، ولا يزال شائعًا إلى اليوم." (Poibleau, 2017, p.40).

لمحة عامَّة عن تاريخ الاحتمالات في الترجمة الآليَّة الإحصائيَّة: يُعَدُّ Peter F. Brown وزملاؤه في IBM أشهر مرجع لفكرة استخدام الاحتمالات في الترجمة الآليَّة الإحصائيَّة؛ فقد نشروا بحثًا بعنوان "A Statistical Approach to Machine Translation" (١٩٩٠) وضعوا فيه الأساس لنماذج الترجمة الإحصائيَّة التي تعتمد على الاحتمالات لاختيار الترجمة الأكثر احتمالًا؛ والفكرة الأساسيَّة هي إذا كان هناك جملة في اللغة المصدر Source Language (S)، فيتمُّ البحث عن الجملة في اللغة الهدف Target Language (T) التي لها أعلى احتمال Probability $Pr(T-S)$، أي ما احتمال أن تكون (T) هي الترجمة المناسبة للجملة (S)؟ والجواب يرتبط بنموذجَين: (نموذج اللغة) يحسب مدى احتمال أن تكون الجملة أكثر سلاسة وطبيعيَّة في اللغة الهدف، و(نموذج الترجمة) يتعلَّم من البيانات الإحصائيَّة مدى احتمال أن تكون هذه الترجمة مقابلة للجملة الأصليَّة؛ "كلُّ جملة في لغةٍ ما هي ترجمة محتملة لأيّ جملة في اللغة الأخرى. نُسند لكلّ زوج من الجمل (S، T) احتمالًا، $Pr(T|S)$، يُفسَّر على أنَّه احتمال أن يُنتج المترجم الجملة (T)

في اللغةِ الهدف عند عرض الجملة (S) عليه في اللغة المصدر." (Brown, Cocke, and others, June 1990, N2, V16, p.79). ويرى موندايMunday (٢٠١٦) أنَّ دراسات الترجمة أدَّت في شقَّيها الكتابيّ والشفهيّ دورًا حاسمًا في التواصل بين البشر، ولاسيَّما في توفير الوصول إلى النصوص المهمَّة لأغراض القضايا الرئيسة البحثيَّة والدينيَّة؛ فمع نموِّ التجارة العالميَّة ازدادت أهمِّيَّة الترجمة؛ وبحلول العام ٢٠١٥ قُدِّر حجم السوق العالميّ للترجمة الخارجيَّة والترجمة الفوريَّة والتقنيَّات ذات الصلة بأكثر من ٣٨ مليار دولار أمريكيّ، بينما تترجِم منظَّمات دوليَّة مثل الاتِّحاد الأوروبيّ حوالى ٢٤ لغة، وتنفق حوالى ٤٥٦ مليون يورو سنويًّا على خدمات الترجمة والترجمة الفوريَّة؛ ومع ذلك، لم تبدأ دراسة الترجمة موضوعًا أكاديميًّا فعليًّا إلَّا في النصف الثاني من القرن العشرين (Poibleau, 2017, p.11). ويرى بوابلو أنَّ الترجمة تتطلَّب فهمًا عميقًا للنصّ المُراد ترجمته، وأنَّ الهدف من الترجمة الآليَّة ليس ترجمة الأدب أو الشعر، بل هو تقديم أدقِّ ترجمة ممكنة للنصوص اليوميَّة. ومع ذلك، تظلُّ المهمَّة بالغة الصعوبة؛ "تُتيح لنا الترجمة الآليَّة إعادة النظر في أسئلة قديمة وُظِّفت على نطاق واسع: ما معنى الترجمة؟ ما نوع المعرفة التي ينطوي عليها مسار الترجمة؟ كيف يُمكننا نقل نصٍّ من لغة إلى أخرى؟" (Poibleau, 2017, p.12). أمَّا كوهين ونايت Koehn and Knight (٢٠٠٢) فقد حدَّدا مكوِّنات نظام الترجمة الآليَّة الإحصائيَّة بنموذج اللغة Language Model المبنيّ على مدوَّنة أحاديَّة اللغة Monolingual Corpus، والمدرَّب على كميَّات كبيرة من البيانات الأحاديَّة في اللغةِ الهدف لضمان طلاقة النصّ الناتج، ونموذج الترجمة Translation Model المبنيّ على مدوَّنة ثنائيَّة اللغة Parallel/Bilingual Corpus. كما عملا على بناء معجمِ ترجمة على مستوى الكلمة، بالاعتماد كلِّيًّا على مدوَّنات أحاديَّة اللغة غير مترابطة، مثل: الكلمات المتشابهة Identical Words، والسياق المماثل Similar Context، والحفاظ على تشابه الكلمات وتكرارها Similarity. وتمَّ الإبلاغ عن نتائج تجريبيَّة لبناء معجم للأسماء باللغتَين الألمانيَّة والإنكليزيَّة، وتمَّ التحقُّق من دقَّة ترجمة الأسماء بنسبة ٣٩% مقارنةً بمدوَّنة اختبار موازية. (Koehn and Knight, July 2002, PP.9-16).

تتعدَّد أنواع الترجمة؛ فهناك الترجمة بالقواعد اللسانيَّة، أو الترجمة الآليَّة القائمة على القواعد (*Rule-Based Machine Translation – RBMT*)؛ وهي نظام محدَّد من قواعد الترجمة الآليَّة ومن أقدم مقارباتها، وتختلف عن المقاربة اللسانيَّة، إذ تقوم على استخدام قواعد لغويَّة وقواميس أعدَّها اللغويُّون، بدل الاعتماد أساسًا على الاحتمالات الإحصائيَّة أو الشبكات العصبيَّة؛ ويبدو أنَّ الترجمة الآليَّة الإحصائيَّة تسيطر على معظم المنشورات البحثيَّة العلميَّة، وتكاد تكون المعتمدة في أكثر شركات تطوير البرامج المعلوماتيَّة، مثل Google، إذ تعتمد على كمٍّ هائل من البيانات المنظَّمة وفق تقنيَّات الحاسوب من النصّ الموازي، أي من اللغة الأصليَّة إلى اللغة الهدف. ويرى الباحث في الذكاء الاصطناعيّ والإعلام الآليّ صديق بِسّو أنَّ البرنامج يتعلَّم من البيانات المقدَّمة إليه ما لا يمكن توقُّعه، حتَّى يصبح قادرًا على إنجاز ترجماتٍ لجمل ونصوص لم يسبق له رؤيتها من قبل، وليست مخزَّنة لديه؛ "هذا التطوُّر المذهل في خوارزميَّات التعلُّم الآليّ جعل الترجمة الإحصائيَّة تتحصَّل على نتائج مبهرة لتتفوَّق على باقي [بقيَّة] المقاربات." (بِسّو، ٢٠١٧، ص. ٦١). إنَّ معاينة القواميس ثنائيَّة اللغة لا يمكن أن تكون هدف الترجمة الآليَّة الأمثل، إذ إنَّ هذا النوع من الترجمة يؤدِّي إلى ضياع شكل الجملة وهيكلها، وبالتالي ضياع المعنى

لاسيَّما وأنَّ تعقيدات اللغة تكمن في إنتاج ترجمة تولِّد جملًا صحيحة في اللغة الهدف، وتحافظ على المعنى الأصليّ في اللغة المصدر من دون ارتباط ببِنية الجملة الأصليَّة. يطرح بِسّو مقاربات لتجاوز هذه العقبات، وهي:

١ – **المقاربة بالقواعد اللسانيَّة:** وتتمُّ فيها الترجمة عبر ثلاث مراحل يجري فيها تحليل النصّ المراد ترجمته إلى **تمثيلٍ مجرَّد** في اللغة المصدر، ومن ثمَّ **تحويل** هذا التمثيل إلى تمثيلٍ مجرَّد آخر في اللغة الهدف، إلى أن يتمَّ **توليد** نصٍّ في اللغةِ الهدف انطلاقًا من هذا التمثيل. (بِسّو، ٢٠١٧، ص. ٦٣). يمكن تقريب هذه العمليَّة من خلال تطبيق المراحل الثلاث (التحليل/التحويل/التوليد) بالأمثلة الآتية من الإنكليزيَّة إلى العربيَّة:

المثال Example:

- The teacher shut the door
- أغلقَ المعلِّمُ الباب.

أ – **التحليل Analysis:** يحلِّل النظام الجملة الإنكليزيَّة محدِّدًا عناصرها والعلاقة التي تجمع بين أجزائها.

- The teacher = فاعل
- Shut = فعل
- The door = مفعول به

التحليل حوَّل الجملة إلى تمثيلٍ مجرّد، مثل:

[the teacher: فاعل] + [shut: فعل] + [the door: مفعول به]

في مرحلة التمثيل المجرَّد لا يهتمُّ النظام بالكلمات الإنكليزيَّة فحسب، بل يستخرج البِنية، والعلائق النحويَّة، والمعنى الأساسيَّ من الجملة.

ب – التحويل Transfer: يعمل النظام على نقل هذا التمثيل إلى تمثيلٍ يناسب العربيَّة؛ على سبيل المثال:

[المعلِّم: فاعل] + [أغلق: فعل] + [الباب: مفعول به]

بحسب قواعد العربيَّة، ربَّما يقرِّر النظام أنَّ الصياغة الطبيعيَّة هي تقديم الفعل على الفاعل:

[أغلق: فعل] + [المعلِّم: فاعل] + [الباب: مفعول به]

ج – التوليد Generation: أغلقَ المعلِّمُ الباب.

في **التحويل** يتمُّ اختيار البِنية المناسبة للعربيَّة، وهذه المراحل الثلاث تختصُّ بالمقاربة اللسانيَّة بالقواعد للترجمة الآليَّة، إذ إنَّ النظام لا ينتقل مباشرةً من كلمات اللغة الإنكليزيَّة إلى كلمات عربيَّة، بل يمرُّ عبر

تمثيلٍ مجرَّد للمعنى والبِنية، ثمَّ يحوِّلُه إلى تمثيلٍ مناسب للُّغة الهدف، ثمَّ **يولِّد** النصَّ النهائيّ. المقاربة اللسانيَّة بالقواعد تفصل بين فهم النصِّ الأصليّ وبين صياغته في اللغة الهدف، بدل الانتقال المباشر من كلمات اللغة الأصليَّة إلى كلمات اللغة المترجمة.

٢ – **الترجمة الإحصائيَّة:** تعتمد على التعلُّم الآليّ بتطبيق دوالٍّ رياضيَّة على مجموعات ضخمة من النصوص ترجمها بشر؛ فمعنى أنَّ الترجمة الإحصائيَّة قائمة على الاحتمالات، أن يتمَّ حساب احتمال أن تكون جملةٌ ما في اللغة الهدف هي ترجمة جملةٍ ما في اللغة المنطلَق. (بسّو، ٢٠١٧، ص. ٦٤) وللترجمة الإحصائيَّة نموذجان:

أ – **نموذج للُّغة:** يهتمُّ ببِنية الجملة في اللغةِ الهدف بحساب الاحتمالات الممكنة. أي يتعلَّم كيف تُبنى الجمل الصحيحة والطبيعيَّة في اللغة الهدف؛ مثلًا: إذا كانت اللغة العربيَّة هي الهدف يتعلَّم البرنامج من ملايين الجمل العربيَّة، فيلاحظ أنَّ جملة "حضرَ التلاميذ إلى الصفّ" هي جملة عاديَّة، بينما جملة "الصفُّ التلاميذ إلى حضرَ" هي أقلُّ احتمالًا في العربيَّة المعاصرة؛ فيسأل نموذج اللغة: "أيُّ صياغة تبدو أكثر طبيعيَّة في اللغة الهدف من بين الترجمات الممكنة؟"، وبهذا يتعلَّم من مجموعة النصوص المترجمة ليستنتج نموذج الترجمة، ويتعلَّم من مجموعة النصوص في اللغةِ الهدف ليستنتج نموذج اللغة؛ فالنصوص المترجمة Translation Model تحدِّد مدى احتماليَّة ترجمة عبارةٍ من اللغةِ المصدر إلى اللغة الهدف، والنصوص الأحاديَّة في اللغة الهدف Language Model تحدِّد مدى احتماليَّة أن تكون الجملة طبيعيَّة. إنَّ إضافة النصوص المترجمة إلى النصوص الأحاديَّة في اللغةِ الهدف يحدِّد الترجمة الأفضل. وقد ذكر كوهين Koehn ونايت (2002) Knight أنَّ المدوَّنة الأحاديَّة تُستخدم لنموذج اللغة، والمدوَّنة الثنائيَّة/الموازية تُستخدم لنموذج الترجمة.

ب – **نموذج للترجمة:** وهو يقترح لكلِّ جزءٍ من أجزاء الجملة في اللغةِ المصدر ترجمات محتملة في اللغة الهدف، ويتعلَّم العلاقة بينهما من نصوص مترجمة من اللغتَين ليستنتج نموذج اللغة. على سبيل المثال: يتدرَّب على آلاف الترجمات التي تحتوي على جملة "هو يحملُ الطابة He is carrying the ball"؛ فيتعلَّم أنّ:

- He = هو Is carrying = يحمل The ball = الطابة

لكنَّ الأمر ليس مجرَّد ترجمة كلمة بكلمة، إذ يتعلَّم أيضًا احتمالات العبارات والتركيبات. ويتعلَّم من النصوص بوساطة قاعدة بيانات ضخمة تحتوي على نصوص موازية بين العربيَّة والإنكليزيَّة:

English	Arabic
She is looking for a Job	هي تبحث عن وظيفة

من هذه الأمثلة يتعلَّم البرنامج نموذج الترجمة، ويتدرّب على الترجمات العربيَّة المحتملة للكلمات والعبارات الإنكليزيَّة، ويقدِّر مدى احتمال كلِّ ترجمة؛ فيصبح قادرًا على تعلُّم نماذج اللغة من التراكيب والجمل الأكثر احتمالًا وانسجامًا مع اللغة الهدف، حتَّى وإن أُعطيَ ملايين الجمل في تلك اللغة. والنتيجة

أنَّه عند ترجمة البرنامج جملة، فإنَّه يجمع بين النموذجَين (نموذج الترجمة الذي يقترح ترجمات ممكنة) و(نموذج اللغة الذي يختار الصياغة الأكثر احتمالًا وطبيعيَّة)، وهذا يُعدُّ من إيجابيَّات الترجمة الإحصائيَّة، لأنَّ نوعيَّة الترجمة تتحسَّن مع مرور الوقت؛ فكلَّما زادت كميَّة المعطيات في قاعدة البيانات ازدادت درجة التعلُّم وتحسَّنت نوعيَّة الترجمة؛ ولهذا، سُمِّيت ترجمة آليَّة إحصائيَّة لأنَّها تعتمد على احتمالات البيانات لاختيار الترجمة الأفضل.

٣ – **الترجمة العصبيَّة: أو السياقيَّة** Artificial Neural Networks: هي طريقة أحدث من الترجمة الآليَّة الإحصائيَّة، تعتمد على **الشبكات العصبيَّة الاصطناعيَّة** لكي تتعلَّم الترجمة من كميَّات ضخمة من النصوص، ولا تنظر إلى الكلمات بمعزلٍ عن الكلمات المحيطة بها، بل تسعى إلى فهم الكلمة في سياق الجملة الكامل للجملة؛ مثل: كلمة Fan = مروحة؛ وكلمة Fan = معجب أو مشجِّع. مثال:

- شغِّل المروحة، الجوُّ حارٌّ جدًّا اليوم = Turn on the fan; it's very hot today
- هو معجب جدًّا بوائل كفوري = He is a huge fan of Wael Kfoury

فالترجمة العصبيَّة تحاول فهم الكلمة في السياق الكامل للجملة. أمَّا الشبكة العصبيَّة فتعتمد على جملة من الآليَّات التكنولوجيَّة لتحسين **خوارزميَّاتها**، منها: الانتباه Attention الذي له الدور الأكبر في تعيين تسلسل عناصر الجملة الصحيح في اللغة الهدف، "وهذا ما دفع الباحثين إلى تطوير آليَّة الانتباه لتمكين البرنامج من ترجمة الجمل الطويلة بسهولة، وتركيب عناصرها بحسب قواعد اللغة الهدف وخصوصيَّاتها" (جيلاني. هاجر، ٢٠٢٠، ٩١–١٠٨)؛ فمن خلال التدريب على عدد هائل من الأمثلة، تتعلَّم أنَّ معنى الكلمة يتغيَّر بحسب السياق؛ فعند ورود: (Fan + Air + Handheld)، ترجِّح أنَّ معنى Fan هو مروحة؛ أمَّا إذا رأت (Fan + Sport + Singer)، فسترجِّح أنَّ معنى Fan هو مشجِّع. في التمثيلات السياقيَّة النظام لا يمثِّل كلمة Fan على أنَّها كلمة لها ترجمة واحدة ثابتة، بل يأخذ في الحسبان السياق الذي وردت فيه، وهذا الأمر يفسِّر استمراريَّة المقاربات الإحصائيَّة وكيفيَّة تطويرها للترجمة العصبيَّة. ويمكن تلخيص التطوُّر الحاصل كالآتي:

الترجمة الآليَّة الإحصائيَّة	الترجمة الآليَّة العصبيَّة
تعتمد على الاحتمالات المستخلصة من النصوص	تستخدم الشبكات العصبيَّة لتتعلَّم من النصوص وتفهم العلائق بين الكلمات في سياقها، وتعتمد على جملة من الآليَّات التكنولوجيَّة لتحسين خوارزميَّاتها، نحو "التعلُّم العميق Deep Learning، والتوقُّع Prediction، والانتباه Attention لإنتاج ترجمة طبيعيَّة مترابطة".

٤ – **الترجمة الهجينة Hybrid Machine Translation:** بعد نجاح أنظمة الترجمة الإحصائيَّة، سعتْ أكثريَّة الأنظمة التقليديَّة (القائمة على معاجم ضخمة وقواعد نقل) تدريجيًّا إلى دمج المعلومات الإحصائيَّة في منهجها؛ فبعض النظم استغلَّت نقاط القوَّة في كلِّ مقاربة لتشكِّل نظامًا هجينًا يدمج بين

طريقتين للحصول على نظام يتجاوز نقاط الضعف ويحقّق نتائج أدقّ. ومن هذه الأنظمة نظام Systran الذي كان رائدًا في النهج القائم على القواعد في أوائل العقد الأوّل من القرن الحادي والعشرين؛ فبدايةً استخدم الإحصاء لتصحيح المخرجات وإنتاج ترجمات أسلَس، ولكن كان يعاني من تكييف الترجمة ديناميكيًّا في المجالات المختلفة، مثل: الطبيَّة والقانونيَّة وتقنيَّة المعلومات ديناميكيًّا. (Poibleau, 2017, P.114-115). والمقصود بالترجمة الهجينة أنَّ النظام يفيد من مزايا كلِّ طريقة ويقلِّل من عيوبها. ففي الطريقة الإحصائيَّة يمكن الإفادة من كميَّات كبيرة من الترجمات السابقة وحساب الاحتمالات، (لكنَّها تواجه صعوبة في فهم السياق الطويل). وفي الطريقة العصبيَّة يمكن فهم السياق وإنتاج جملٍ أسلَس، (ولكنَّها تحتاج إلى كميَّات كبيرة من البيانات والتدريب)؛ فالنظام يجمع بين تقنيَّات مختلفة للوصول إلى الترجمة الأنسب؛ مثال: كلمة **Fan**:

The fans were sitting on the ground to listen to their favorite singer.

كان المشجِّعون يجلسون على الأرض كي يستمعوا إلى مطربهم المفضَّل.

يحدِّد النظام أنَّ كلمة Fan مشجِّع وليست مروحة، اِعتمادًا على السياق: (*Sitting on the ground + Singer + Favorite*)، تدلُّ على أنَّ المقصود هو المشجِّعون أو المعجبون.

بينما في هذا المثال:

We turned on the fan because we were feeling very hot.

شغَّلنا المروحة لأنَّنا كنَّا نشعر بالحرِّ الشديد.

يحدِّد النظام أنَّ (*turned on + feeling + hot*) تدلُّ على أنَّ الكلام يدور عن آلة؛ لذلك Fan هي مروحة. **في نظام الترجمة الهجين**، يمكن أن يجمع النظام بين المعرفة الإحصائيَّة من البيانات والترجمات السابقة التي تعلَّم منها، وقدرة النماذج العصبيَّة على استثمار السياق، ثمَّ يتمُّ اختيار "مشجِّعون" في الجملة الأولى، و"مروحة" في الثانية؛ وبهذا يحصل النظام الهجين على أفضل نقاط القوَّة في كلِّ مقاربة ترجميَّة، وتكون النتيجة كالآتي:

جاء المعجبون لأنَّهم كانوا متشوِّقين لسماع أغنية مطربهم الجديدة

فالنظام أخذ أفضل ما في كلِّ طريقة، وجمع أكثر من تقنيَّة في نظام واحد، كي يتجاوز نقاط الضعف الموجودة في كلِّ مقاربة، لتحسين جودة الترجمة.

اللغة العربيَّة وتحدِّيات الترجمة الآليَّة الإحصائيَّة

تُعدُّ اللغة العربيَّة ذات خصائص صرفيَّة ونحويَّة تجعل الترجمة الآليَّة الإحصائيَّة أكثر تحدِّيًا. ويمكن استعراض بعض خصائص اللغة العربيَّة التي تؤثِّر في أداء الترجمة الإحصائيَّة، ومنها:

الغنى الصرفيّ: ويظهر ذلك في جذر، نحو: (ض–ح–ك) الذي يصدر عنه: (ضحك – يضحك – اِضحك – ضاحك – ضحوك – ضحَّاك – مضحكة – ضحكوا – ضحكتُ – ضحكنا...)؛ فالنظام الإحصائيُّ يحتاج إلى رؤية عدد كبير من هذه الأشكال في بيانات التدريب كي يتعلَّم العلائق بينها. وهنا تظهر المشكلة إذا ما ظهر شكل نادر في البيانات، فقد لا يملك النظام المعلومات الإحصائيَّة الكافية عنه.

اللواصق: كأن تتَّصل عناصر كثيرة بالكلمة، نحو: (ولأطفالهم)، فيمكن تحليلها إلى: (و+ لـ + أطفال + هم)؛ فكلمة واحدة في الكتابة يمكن أن تحمل معلومات تؤدِّيها عدَّة كلمات في لغات أخرى، وهذا ما يجعل من عمليَّة محاذاة الكلمات *Word Alignment* في الترجمة الإحصائيَّة أكثر تعقيدًا. يرى بوابلو أنَّ محاذاة الكلمات هي مهمَّة أكثر تعقيدًا من محاذاة الجمل؛ فعلى مستوى الجملة توجد مطابقة تامَّة بين النصِّ الأصليِّ وترجمته (تقابل جملة واحدة من النصِّ الأصليِّ جملة واحدة في الترجمة)؛ إلَّا أنَّ هذا لا ينطبق على مستوى الكلمة لأنَّ اللغات تختلف وبعض الكلمات لا يمكن ترجمتها مباشرة، لذلك، فإنَّ معظم هذه المطابقات غير متناظرة *Asymmetrical*، أي إنَّ كلمة واحدة من اللغة المصدر تقابل صفرًا أو كلمة واحدة أو عددًا من الكلمات في اللغة الأخرى. (Poibleau, 2017, P.86).

تناوب الحركات: نحو: (عرف)، فيمكن أن تُقرأ (عَرَفَ – عُرِفَ – عَرَّفَ – عُرْفٌ)؛ فهذه الأشكال مختلفة في معناها، كلُّ الكتابة غير المشكولة تجعل التمييز بينها شاقًّا. (رباحي، ٢٠٢١، ص. ٦٠–٦١).

ترتيب الكلمات: نحو: (وقع الإبريق على الأرض)، أو: (الإبريق على الأرض وقع)؛ فقد تختلف البِنية والترتيب وفق السياق والأسلوب، وهذا ما يسبِّب صعوبة للنظام الإحصائيِّ عندما يحاول تعلُّم العلاقة بين ترتيب الكلمات في اللغة المصدر وترتيبها في العربيَّة. أو ترجمة هذه الجملة الإنكليزيَّة إلى الفرنسيَّة:

(This does not necessarily apply at word level) = (Cela ne s'applique pas nécessairement au niveau des mots).

التذكير والتأنيث والتثنية والجمع: تميِّز اللغة العربيَّة بين: (هو/هي)، و(المغنِّي/المغنِّية)، و(معلِّم/معلِّمان/معلِّمون/ معلِّمات)؛ لذلك، قد تحتاج الترجمة من لغة مثل الإنكليزيَّة إلى العربيَّة إلى تحديد معلومات لا تكون واضحة دائمًا في النصِّ الإنكليزيّ، مثل: (The singer shouted)؛ فمن دون سياق إضافيّ، لا يمكن معرفة هل المقصود (صرخ المغنِّي) أم (صرخت المغنِّيَة)، وكذا الأمر بين (صرخ المغنِّيان) و(صرخ المغنُّون)؛ فهنا يصبح السياق مهمًّا جدًّا.

تعالج الترجمة العصبيَّة هذه العوائق من خلال التمثيلات السياقيَّة، فتحوِّل الكلمات إلى تمثيلات عدديَّة (*Vectors*) وتُحدَّث هذه التمثيلات وفقًا للسياق الذي تظهر فيه الكلمة؛ مثل كلمة (نجم Star)، يمكن أن تعني: (النجم الموجود في السماء، أو: الفنَّان المشهور)؛ مثال: *The star appeared in the sky* = ظهر النجم في السماء؛ فالسياق يُشير إلى استقرار النجم في السماء. أو: *The player became a global star* = أصبح اللاعب نجمًا عالميًّا؛ السياق يُشير إلى بروز اللاعب واشتهاره بين الناس. فالنظام العصبيُّ يحاول بناء تمثيل للكلمة وفق الكلمات المحيطة بها. يُلحظ أنَّ كلمة نجم احتفظت بالترجمة نفسها Star في الإنكليزيَّة، لكنَّ معناها اختلف باختلاف السياق؛ ففي الجملة الأولى أشارت إلى جرم سماويّ، بينما أشارت

في الثانية إلى شخص مشهور، وهذا المثال يبرز أهمِّيَّة السياق في تحديد معنى الكلمة المناسب في عمليَّة الترجمة.

ختامًا

تختلف التوقُّعات من الترجمة اختلافًا جذريًّا تبعًا للسياق، والأشخاص، والعصر، وطبيعة النصِّ واستخداماته؛ فيتعيَّن على المترجم الاختيار بين الالتزام بالنصِّ الأصليِّ أو استخدام إعادة الصياغة لضمان الفهم. وقد أظهرت المقاربة الإحصائيَّة أهمِّيَّة الاحتمالات في اختيار الترجمة المناسبة، من خلال الاعتماد على نموذج الترجمة ونموذج اللغة المستمدَّين من المدوَّنات النصِّيَّة. إلَّا أنَّ الاعتماد على الاحتمالات لا يكفي، ولاسيَّما عند التعامل مع الكلمات متعدِّدة المعنى والخصائص اللغويَّة المعقَّدة، وهو ما ظهر من خلال الأمثلة في اللغة العربيَّة. لا يمكن الإقرار بأنَّ الترجمة العصبيَّة تحلُّ جميع مشكلات العربيَّة، فالأرجح أنَّها تساعد النماذج العصبيَّة والتمثيلات السياقيَّة على معالجة بعض الصعوبات بصورة أفضل من النماذج الإحصائيَّة التقليديَّة، ولاسيَّما الصعوبات وتعدُّد المعنى والعلائق بين الكلمات، لكنَّها لا تلغي جميع التحدِّيات التي تطرحها خصائص اللغة العربيَّة الصرفيَّة والنحويَّة.

المراجع العربيَّة والأجنبيَّة

- بِسّو، صديق (ديسمبر ٢٠١٧). "اللغة العربيَّة والترجمة الآليَّة". الجزائر: "العربيَّة". جامعة فرحات عبَّاس في صطيف. **مؤتمر اللغة العربيَّة والترجمة الجزائر**. تمَّت المراجعة في ٢٠٢٦/٢٠/٨.
https://www.researchgate.net/publication/325951688_allght_alrbyt_waltrjmt_alalyt_alahsayyt.
- بكر، عبد الجوَّاد (٢٠٠٣). **منهج البحث المقارن – بحوث ودراسات**. ط.١، الإسكندريَّة: دار الوفاء لدنيا الطباعة والنشر.
- جيلاني، ناصر؛ هاجر، اِبن ونَّان، مختار (٧ (١) ٢٠٢٠). "نحو ترجمة آليَّة بسِمات بشريَّة للنصوص المتخصِّصة من اللغة الإنكليزيَّة إلى العربيَّة – دراسة مقارنة." Aleph اللغة والإعلام والمجتمع.
https://aleph.edinum.org/2265#doc-authors-infos
- رباحي، محمَّد (٢٠٢١). تحدِّيات المعالجة الآليَّة للغة العربيَّة، "مجلَّة **اللغة العربيَّة**"، العدد: ٢. المجلَّد ٢٣. ص. ٢٤٩–٢٦٨.

- Brown, F, Peter. Cocke, John and others (June 1990). “A Statistical Approach To Machine Translation”. *Computational Linguistics*, Yorktown Heights: Thomas J. Watson Research Center. Vol.16. N2.
- Koehn, Philipp (2014). *Statistical Machine Translation*. U.K: Cambridge University Press. 5th print.
- Koehn, Phillipp. Knight, Kevin (June 2002). *Learning a Translation Lexicon from Monolingual Corpora*. California: University of Southern California. P.9-16.
- Munday, Jeremy (2016). *Introducing Translation Studies, Theories and Applications*. Fourth Edition. New York: Routeledge.
- Poibleau, Thierry (2017). Machine Translation. London: The MIT Press, Cambridge.