حوسبة اللغة العربية

بين المقاربة اللغوية والمقاربة الإحصائية

د. سهيلة بربارة♥

جـــامعة الجزائر2

تاريخ الإرسال:07-07-2017 تاريخ القبول: 3-05-2018

Abstract Arabic language Processing
Between the linguistic approach and the statistical approach

In all times, translation has always been a powerful tool for communication between people and has played a key role in the rapprochement among civilizations. Its importance, in all dimensions was enhanced as demand for translation services increased sharply in the current century and, at the threshold of the twenty-first century, we will be inevitably witnessing a rush for translation work in record time. Moreover, the phenomenon of acculturation experienced by the present era makes translation a more complex and multifaceted process.

The persistent demand for the exchange of information and communication in different languages as a result of technological breakthrough and the enrollment of all active actors in the global network, resulted in a lot of pressure for the exchange of information in different languages and in record time in order to facilitate meeting the most urgent needs in the various aspects of human life (commercial, military, technical ,cultural, social, etc ...)

If human translation managed in the last time run to fulfill these requirements, it could no longer do it , due to the increasing

♥ **lynatrad@gmail.com**

performance and efficiency of computers connected to the Internet. Therefore, translators had to resort to digital tools to speed up the process of the completion of translation work, especially in the technical fields and the world of business services characterized by repeated terminology texts, free of creative and rhetorical figures which cannot be subject to automatic representation.

ملخص: شكّلت الترجمة على مرّ العصور، أداة قوية للتواصل بين الناس وللاحتكاك في ما بين الحضارات كما تضاعفت أهميتها بجميع أبعادها وتأجّج الطلب عليها في القرن الحالي وها نحن مقبلين على القرن الواحد والعشرين الذي سيشهد، لا محالة، تزايدا في الطلب على عمل الترجمة في وقت قياسي، ثمّ إنّ ظاهرة المثاقفة(Acculturation) التي يعيشها عصرنا الحالي، تجعل من الترجمة عملية أكثر تعقيدا وتعدّدا.

تمخّض عن المطالب المتواصلة لتبادل المعلومات والاتصال في لغات مختلفة نتيجة التسارع التكنولوجي، والتحاق جميع الجهات الفاعلة بالشبكة العنكبوتية العالمية ضغط كبير لتبادل المعلومات في لغات مختلفة وفي وقت قياسي قصد تسهيل تلبية الاحتياجات الأكثر إلحاحا في مختلف مناحي حياة الإنسان (التجارية والعسكرية والتقنية، والثقافية والاجتماعية ... الخ)

إذا كانت هذه التبادلات تتم في الوقت الماضي البعيد، من خلال الترجمة البشرية فقد أصبحت منذ تزايد سرعة الحواسيب وكفاءتها وارتباطها بشبكة الإنترنت تستعين بأدوات رقمية للتسريع من عملية إنجاز أعمال الترجمة ولا سيما في المجالات التقنية وعالم الخدمات التجارية المُتّسم بالنصوص ذات المصطلحات المتكرّرة والخالية من الصور الإبداعية والبلاغية التي يعجز الحاسوب عن تمثيلها.

الكلمات الدالة: الترجمة الآلية، المعالجة الحاسوبية، المقاربة اللغوية المقاربة الإحصائية، المقاربة الهجينة، وحدة نمطية للترجمة الآلية.

مقدمـــــة: يتناول بحثنا جانبا هاما من الرهانات التي تواجه المعالجة الحاسـوبية للغة العربية لأغراض الترجمة تحديدًا، والتي تعرف تأخرا ملموسا بالمقارنـة مـع اللغات التي تحتل الصدارة في المجال التكنولوجي.

نستهل دراستنا بلمحة تمهيدية عن نشأة الترجمة الآلية و روادها ونسترسـل فـي الحديث عن أبرز الفاعلين في مجال تصميم برمجيات للترجمـة الآليـة و عـرض العوامل التاريخية الداعية لحوسبة اللغة العربية، والعوامل العمليـة الدوليـة انتهـاءً بالعوامل الخاصة المرتبطة بخصائص هذه اللغة التي تؤهّلها للتمثيل الحاسـوبي فـي ضوء المبادرات التي شُرع فيها في هذا المجال، مما يسوقنا للتطرق إلى المسـاعي المبذولة في مجال حوسبة اللغة العربية لأغراض الترجمة و لا سيما منتجات شـركة صـخر وشـركة آي تـي آي (ATA Software Technology) التـي انفـردت بتصميمها لبرمجيات للترجمة الآلية من اللغة العربية وإليها تعتمـد علـى المقاربـة اللّغوية في ظل توفّر مقاربات متباينة لبناء برمجيات الترجمة الآلية.

غايتنا من هذه الدراسة هي التعريف بأبرز المقاربات المتاحة في هـذا المجـال واقتراح المقاربة الأنسب، إمّا لتصميم برمجية آلية للترجمة من اللّغـة العربيـة، ولا سيما إليها أو اقتراح تحسينات على الوحدة النمطية التي تُوفّرها شركة "آي تـي آي" في ضوء النتائج التي حصلنا عليها عند تقييم نتاج برمجية "آي تي آي" ذات المقاربة اللغوية وبرمجية "غوغل" (Google) ذات المقاربة الإحصائية.

يتعين التذكير بأنّ موضوع الدراسة ينحصر في البرمجيات المُصـمَّمة لترجمـة النصوص والشاملة للغة العربية على وجه الخصوص، وننطلق من فكـرة أنّ هـذه المنظومات أدوات مساعدة على الترجمة (CAT tools) بحاجة دائمة لتلقيح بشـري لاحق (Post-editing).

علاوة على ذلك، تجدر الإشارة إلى أن هذا الجانب التحليلي من هذه الدراسة هي عرض لعيّنة فقط من النماذج التي شكّلت موضوع دراسة تحليلية شـاملة وواسـعة

مضمونة في رسالة الدكتوراه التي تناولنا فيها هذا الموضوع من زوايا مختلفة.

يعقب وصفنا لمقاربات الترجمة الآلية ولا سيما اللغوية والإحصائية منها عرض حالة تطبيقية لترجمة نموذج نصي في الاتجاه الأحادي انجليزي ــ عربي بواسطة منظومة "الوافي الذهبي" اللّغوية ومنظومة "غوغل" الإحصائية مع تحديد مواطن ضعف وقوة كلتا المنظومتين وفقًا للطريقة الموصوفة في العرض التطبيقي، ومثّلنا النتائج التي في جداول ورسوم بيانية توضيحية قُمنا على أساسها باقتراح الوحدة النمطية الأنسب لمعالجة اللغة العربية في شكل توصيات أُدرِجت ضمن خاتمة هذا البحث.

منذ الإرهاصات الأولى من ولادة مشروع آلة تُترجم سواء على يد المهندس الروسي بيتر سميرنوف ترويانسكي (Petr Smirnov Troyanskij)، والمهندس الروسي المنشأ، جورج أرترستروني (Georges Artsruni)، أم من خلال النظرية التي طوّرها وارن ويفر (Warren Weaver[1]) عن الترجمة الآلية في عام 1949 والتي نقل عنها الخبير الأمريكي جون هتشينز (John Hutshins,2001 :2) في مدوّناته الشهيرة وكذا الاختبار الأول لبرنامج وضعه معهد ماساتشوستس (Massachusetts) للترجمة الآلية في عام 1954 في الولايات المتحدة الأمريكية سرعان ما أصبحت لغات عديدة تخضع لمعالجة حاسوبية على أساس تمثيل مستويات التحليل اللغوي التقليدي لتنتقل بعملية الترجمة من مجرد ترجمة مفردة بمفردة إلى مستوى ترجمة الجمل والنصوص بالاستناد إلى النظريات اللسانية ولا سيما نظرية تشومسكي في النحو التوليدي و التحويلي التي مكنت بيتر توما (Peter Toma[2]) من تصميم نظام سستران (Systran) عام 1968 الذي استغلته شركة غوغل (Google) العملاقة في برنامجها للترجمة الآلية إلى غاية سنة 2007 حيث أحدث فريق بحث "غوغل" في الترجمة الآلية وعلى رأسه فرانز جوزيف أوتش (Franz Josef Och) منعرجا في المقاربة المعتمدة لتصميم برامج

الترجمة الآلية من خلال ابتكاره وحدة نمطية للترجمة الآلية لا تقوم على أدوات التحليل اللغوي، ولكن على المنهج الإحصائي الذي أحدث ثورة في عالم الترجمة الآلية باستبدال أدوات التحليل اللغوية بما يُعرف بذاكرة الترجمة لتخزين كميات ضخمة من المتون المُترجَمة بشريا بين أزواج من اللغات للوصول إلى ترجمة سياقية أفضل و لقد كانت هذه الطريقة منعدمة في نظام سستران القديم.

في خضم تلك التحوّلات التي عرفها عالم الترجمة، لم يُشرع الاهتمام باللغة العربية إلاّ في الثمانينات من جانب بعض المؤسسات الأجنبية ولأغراض تجارية.

بفضل شركة البرمجيات والتكنولوجيا العالمية (ATA)، الكائن مقرها في لندن حظيت اللغة العربية بدراسة مطابقة لما أسماه الأستاذ ماثيو غيدير (Mathieu Guidère) في حديثه عن اللغة العربية، بالمقاربة الخاصة (Particularistic) مقارنةً بالمقاربة العامة (Universalist).

بالتالي، طوّرت شركة "آي تي آي" (ATA) برمجيات عديدة بما في ذلك مجموعة "الوافي" وهي أحدث التطورات الهامة في مجال الترجمة الآلية من اللغة العربية وإليها وتُعدّ مرجعا قيّما يتسنى الاقتداء به وتحسين مقاربته اللغوية لأنّه في الوقت الحاضر لا تزال وحدة "غوغل" لترجمة النصوص هي المنظومة التي تقدّم نتاجا مقبولا من اللغة العربية وإليها.

في الواقع، ومع مطلع القرن الواحد والعشرين، أصبحت الضرورة ملحة للارتقاء باللغة العربية في مجال الترجمة الآلية من المعالجة الحاسوبية السطحية (Shallow) إلى المعالجة العميقة من أجل تصميم وحدة ترجمة تشارك نظيراتها الأجنبية في المقاربات المتاحة من المنهج اللغوي والإحصائي والهجين في غياب منظومة مثلى لها وطرح الساعة هو: ما هي المقاربة الأمثل التي يمكن إسقاطها على الوحدة النمطية العربية للترجمة الآلية، وما هي آفاق حوسبة اللغة العربية؟

قبل الولوج في تحديد المقاربة الأنسب للترجمة الآلية من اللغة العربية وإليها

يتعيّن تخصيص جانب من الدراسة لعرض العوامل الداعية لحوسبة اللغة العربية والخصائص التي تؤهّلها لذلك، وهذا بغية دحض الرأي السائد لدى بعض الجهات العربية والقاضي بعجز لغة الضاد عن أن تكون موضوع تمثيل حاسوبي يضاهي من حيث النتاج اللغات الأخرى.

العوامل الداعية لحوسبة اللغة العربية: هناك عدة عوامل تستدعي معالجة اللغة العربية حاسوبيا وتمثيلها، نذكر منها ما يأتي:

العوامل التاريخية لحوسبة اللغة العربية: في صدد تناول العوامل التاريخية التي تشكّل الدوافع الداعية لبعث اللغة العربية، لقد حصر عبد الرحمن الحاج صالح[3] في قوله في هذا الشأن: " ...لولا أنّ اللغة العربية لغة الإسلام ولولا أنها تحمل من المفاهيم الحضارية والدينية السابقة الوجود والكثير من المفاهيم العلمية التي كانت أساسًا لانطلاق الحضارة الغربية، لاندثرت منذ زمان أو انزوت إلى لغة تخاطب كبقايا اللّهجات[4]..."

يكفي أن نشير ،أيضا، إلى ما ذكره الباحث الجزائري، **محمد بطّاز**، مدير شبكات وأنظمة المعلومات والاتّصال الجامعيّة في محاضرة ألقاها بمجمع اللغة العربية الأردني في موسمه السادس في 2008 حول مسألة بناء الأجهزة الحديثة وفق خصائص اللّغة العربيّة إذ شرح قابلية اللغة العربية للحوسبة موضّحا أنه "من المعروف عن الحاسوب أنّه لا يفهم إلا لغة الأرقام وبالذّات الأرقام العربيّة، وهذا يعني بطريقة أدقّ أنّ الأجيال المتتالية للحاسوب ولغات البرمجة وشبكات الحاسوب والتّكنولوجيّات المختلفة المُستخدَمة لبنائها تعتمد أساسًا على نموذج جبريّ محض وهذا يعني بكلّ بساطة أنّه لولا المدرسة العربيّة لما وجد الحاسوب الرّقميّ بالنّمط المعروف حاليًّا على الأقلّ" كما أكّد في باب آخر، في مداخلة له في ندوة علمية عام 2010 حول موضوع "اللغة العربية في مجال التشبيك الالكتروني" بأنه ""يجب أن ندرك أن الأمر ليس أمر لغة بقدر ما هو مسألة قدرة على الدخول في معركة الإنتاج

وقدرة على تشريح المنتجات المتوفرة لتكييفها وجعلها تتوافق مع لغتنا وثقافتنا ورؤانا في عالمنا هذا. نقول إن الأمر ليس أمر لغة ونكرر هذا لأنه من المعروف عن الحاسوب أنه لا يتكلم إلا لغة الأرقام. وإذا أردنا أن ننسب صفة لهذه الأرقام فإننا سنقول الأرقام العربية وإذا أردنا أن نصف عمليات الحاسوب فإننا نقول عمليات جبرية. ولكن إذا نظرنا إلى البلاد العربية فإننا نجدها في الواقع تستهلك الكثير ولا تنتج إلا القليل من البرامج الحاسوبية وهذا بالرغم من البحوث المتطورة والمجهودات الجبارة التي يبذلها الباحثون العرب في هذا المجال بالذات".

العوامل العملية والدولية لحوسبة اللغة العربية: في ضوء ما سبق، يمكن تلخيص العوامل التاريخية المحفّزة على إعادة بعث اللغة العربية في ما وصل إليه الباحثان الجزائريان المذكوران أعلاه، ولكن استذكار مجد اللغة العربية عبر التاريخ غير كفيل لوحده بإعادة بعث دورها العلمي في ظل التطور المعرفي والتكنولوجي الراهن بل يتعيّن تفعيل دورها انطلاقا من الحقائق التالية– اللغة العربية هي لغة مهمة في جميع الجوانب. يختلف تصنيفها من مصدر إلى آخر حيث هناك من يضعها في المرتبة الثامنة[5] في العالم من حيث عدد المتحدثين بها في حين صنّفها الباحث الأمريكي جورج فيبر (George Weber) في ورقة بحثية له[6] أعدّها عام 1997 في المرتبة الخامسة مستندا إلى معايير خاصة، ويرى هذا الأخير أن هذا التصنيف بقي ساريا إلى غاية 2008 غير أنّ الدراسات الغربية الحالية تُصنفها في المرتبة الثامنة عالميا.

إنّ اللغة العربية موضع اهتمام كبير من جانب الدول الغربية والولايات المتحدة الأمريكية لا سيما بعد أحداث 11 سبتمبر 2001 لدوافع أمنية وسياسية أساسًا واقتصادية، أيضا، لمَا يشكّله العالم العربي كسوق واسعة تستقطب الاهتمام، فضلا عن أن منظمة الأمم المتحدة قد اعتمدتها لتكون واحدة من اللغات الستّ الرسمية المُتعامل بها.

العوامل الخاصة لحوسبة اللغة العربية: مما لا ريب فيه، هو أن العوامل الخاصة المتعلّقة بخصائص اللغة العربية، هي التي تؤهّلها وتحفّزها على تعزيز المبادرات لجعلها موضوع تمثيل حاسوبي من أجل مواكبة الركب الحضاري في احتواء المعارف بأنواعها والاستجابة للطلب المتزايد على المراجع العربية في الأوساط الأكاديمية وعلى مستوى المواقع الالكترونية العربية بالنظر إلى العدد الهائل من المتحدّثين بهذه اللغة في العالم. على الرغم من افتقار المدرسة العربية لدراسات حديثة مستوفية في مجال اللسانيات التطبيقية العربية، إلا أن المبادرات التي تجسّدت في مجال حوسبة اللغة العربية لأغراض الترجمة أو غيرها، قد ساهمت في تسليط الضوء على أهم الخصائص التي تتسم بها لغة الضاد والتي سمحت بمعالجتها الحاسوبية ونحن نحصرها في ما يلي:

– تتميّز اللغة العربية بأنها أقرب إلى قوانين المنطق من حيث أنساقها الصرفية ونظم أبنية الفئات النحوية، وقواعدها الإعرابية، وأنماط تركيب جُمَلها فضلا عن أنها لغة ثرية بضروب الاشتقاق مما يُكسبها خاصية المرونة في توليد الكلمات في اللغة ذاتها أو في تعريب المصطلحات الأعجمية في حالة ترجمتها دلاليا ولقد أجمع أهل اللغة على أن للغة العربية قياس ومعايير.

بالنظر إلى الخصائص المذكورة أعلاه، تشعّبت مجالات التمثيل الحاسوبي للغة العربية ولو أنها بقيت محدودة غير أن تلك المبادرات الضئيلة عرفت نجاحا كبيرا وشملت المجالات التالية:

– معالجة النص القرآني ، على خلفية تعريب نظم تشغيل الحواسيب الذي قضى على إشكالية نظام كتابة اللغة العربية من اليمين إلى اليسار، كما تمّ إدراج الحروف العربية ضمن لوحة المفاتيح التي صممت باللغة الانجليزية واللاتينية بحيث أُدخلت الحروف العربية الثماني والعشرين في جميع أشكالها وحسب موضع كتابتها في الكلمة والهمزة بأشكالها والتاء المربوطة، والألف المقصورة، وأدوات التشكيل

الثمانية، والأرقام العربية ،والألف الحنجرية وهمزة الوصل والهمزة العربية و تمّ إدخال اللام ألف جملة واحدة بل وبأشكال متعددة :لا لأ لإ لآ إلى جانب علامات التشكيل والشدّة.

– تصميم المعاجم الالكترونية، حيث برزت فكرة الاستعانة بالحاسوب في إحصاءات الحروف الأصلية لمواد اللغة العربية وفي تصميم معاجم الكترونية لإحصاء الأفعال العربية بأنواعها، إلى جانب أشهر مناهل اللغة مثل "الصحاح في اللغة"، و"تاج العروس"، و"لسان العرب"، و"القاموس المحيط"، و"معجم الفروق اللّغوية"، و"معجم اللغة العربية المعاصرة"، وما إلى ذلك.

المساعي في حوسبة اللغة العربية لأغراض الترجمة: أدى ظهور العولمة وانتشار وسائل الاتصال باللغة العربية إلى ضرورة تصميم عدد كبير من التطبيقات في مجال معالجة اللغة العربية. نتيجة لذلك، امتدّ النشاط البحثي لتناول جوانب أوسع من معالجة اللغة العربية بما في ذلك التحليل النحوي، والترجمة الآلية، وفهرسة الوثائق واسترجاع المعلومات وما إلى ذلك.

لم يبرز الاهتمام بمجال حوسبة اللغة العربية إلا في الثمانينات وتطوّر إلى حدّ ما في التسعينات على يد معاهد البحث في الولايات المتحدة الأمريكية وبعض الجامعات في أوروبا والشركات التجارية الغربية. لقد ركّزت تطبيقات الترجمة الآلية على الترجمة من اللغة الانجليزية إلى اللغة العربية وإليها دون سواها في بادئ الأمر.

بالنسبة للأدوات التكنولوجية المساعدة على الترجمة، فقد كانت شركة "صخر" (SAKHR) التي أنشئت عام 1982 من بين الشركات الرائدة في مجال تطوير العديد من التقنيات المتقدمة والتي تركت علامات بارزة في صناعة تقنية المعلومات. فقد قامت الشركة بتطوير جيلٍ جديدٍ من تقنيات المعالجة الطبيعية للغة العربية (NLP)، كما وضعت نظاما للترجمة أطلقت عليه تسمية "تَرجِم" (Tarjim) وهو ثنائي الاتجاه أي إنجليزي / عربي / انجليزي ومتوفّر على الموقع "عجيب"

(/http://tarjim.ajeeb.com/ajeeb) ، والبرنامج يترجم مواقع الإنترنت وصفحات الويب.

مع ذلك، تأتي شركة "آي تي آي" (ATA Software Technology) التي تأسست عام 1992 من قبل متخصصين ومبرمجين تعود خبرتهم إلى الثمانينات ويقع مقرها الرئيسي في لندن في صدارة الشركات المنتجة لبرمجياتٍ تستهدف اللغة العربية واللغة الإنجليزية في مجال الترجمة الآلية.

وفقا للمعلومات المستقاة من موقع هذه المؤسسة، فقد أصدرت الشركة في عام 1995 برنامجها للتعريب الذي أطلقت عليه تسمية "المترجم العربي" كأول برنامج للترجمة الآلية من اللغة الإنجليزية إلى العربية على الحواسيب الشخصية والماكنتوش وطوّرت الشركة حزمة برامج لتعريب النصوص الإنجليزية أطلقت عليها اسم «الوافي» التي لقيت إقبالا كبيرا من جانب المستخدمين في العالم العربي.

يذكر الصحفي ميرزا الخويلدي في مقاله الصادر على الموقع الالكتروني لجريدة الشرق الأوسط في 20 جوان 2002 والذي خصّه لتقديم منتجات شركة آي تي آي وأهم إسهامات هذه الأخيرة في مجال تطوير وتعريب البرمجيات، مشيرا إلى إصدار الشركة لثلاثة برامج متنوعة من برنامج الوافي: الأول: هو برنامج الوافي السريع وهو قاموس عربي/إنجليزي، وإنجليزي/عربي شامل يعمل داخل أغلب البرامج المتوافقة مع مايكروسوفت ونـدوز (Microsoft Windows). هو أكثر البرامج انتشارا، وسهل الاستخدام ومتوفر بسعر زهيد.

البرنامج الثاني هو "الوافي الجديد"، وهو برنامج متقدم لترجمة النصوص ترجمة سريعة وفورية، ويتميز البرنامج بتقديم ترجمة سريعة للنصوص الإنجليزية إلى العربية، كما يقدم ترجمة للنصوص الطويلة التي تحتوي على آلاف الكلمات بسرعة قياسية، مع إمكانية نطق الكلمات والعبارات الإنجليزية، ويعتمد هذا المترجم على قاموس حديث يشتمل على أكثر من 1.5 مليون كلمة إنجليزية وعربية، ويحتوي

على مدقق إملائي إنجليزي، وتقديم ترجمة صوتية لأسماء الأعلام.

أما البرنامج الثالث في حزمة الوافي، فهو برنامج الوافي الذهبي، للترجمة إلى اللغة العربية. من جهة أخرى، تمتلك الشركة عبر موقعها "المسبار" محركا الكترونيا لترجمة المواقع داخل شبكة الإنترنت بصفة مجانية وهو يعمل باتجاه واحد أي الترجمة من اللغة الانجليزية إلى اللغة العربية.

نظرا لأنه تظل هذه الشركة المؤسسة الرائدة في وضع وحدة نمطية لترجمة النصوص ترجمة آلية ثنائية الاتجاه انجليزي / عربي / انجليزي، تعتمد المقاربة اللغوية في تصميمها وتقدّم نتاجًا واعدا في حقل الترجمة الآلية، اخترنا أن نعرضها كنموذج في باب المقارنة بين المقاربة اللغوية والإحصائية التي تتزعّمها شركة "غوغل" التي تشمل اللغة العربية في قائمة اللغات التي تتعامل بها منظومتها.

قبل عرض دراستنا المقارنة للمقاربتين المُستخدمتين في الوقت الراهن في مجال الترجمة الآلية من اللغة العربية وإليها، يتعين عرض تقديم وجيز للمقاربات المُتاحة في تصميم منظومات الترجمة الآلية.

مقاربة الترجمة الآلية المعتمدة على قواعد اللغة (Rule-based Machine Translation): الترجمة الآلية القائمة على قواعد اللغة (RBMT) هي مقاربة تقليدية إذ تمثل النظم الأولى للجيل الثاني من الحواسيب وهي تعتمد على القواعد اللغوية المرتبطة باللغات المصدر واللغات الهدف والمُستخلصة أساسا من القواميس الأحادية اللغة أو الثنائية اللغة أو المتعددة اللغات وكذلك من المعاجم المتضمنة القواعد الأساسية للغة، لا سيما مظاهر الاطراد اللغوية في مجالات الدلالة والصرف والنحو.

ترتكز تطبيقات نظام الترجمة الآلية القائمة على قواعد اللغة على التحليل الصرفي والنحوي والدلالي لكل من اللغات المصدر واللغات المستهدفة باستخدام لغة تمثيل وسيطة حيث يكون النص المُخْرَج (output text) حصيلة مرحلتي تحليل اللغة المصدر نحويا وصرفيا ودلاليا (بصفة سطحية انطلاقا من التراكيب أو عميقة

انطلاقا من الدلالة) وتحويل (بواسطة اللغة الوسيطة) أو نقل التمثيل النحوي والصرفي والدلالي مع توليد المكافئ النحوي، والصرفي، والدلالي في اللغة الهدف التي يكون قد تَمَّ تمثيل قواعدها اللغوية في مرحلة التحليل من أجل الحصول على نص مُترجَم آليا. المقاربة اللغوية تلجأ لتطبيق القواعد اللغوية على المراحل الثلاث من عملية الترجمة، أي مرحلة التحليل ومرحلة التحويل.

لعلّ منظومة "سسترن" (Systran) هي أول برمجية صُمّمت باعتماد هذه المقاربة قبل تطوير الشركة للمقاربة الهجينة التي سوف نعرضها لاحقا.

مزايا مقاربة الترجمة الآلية القائمة على قواعد اللغة:

– لها الفضل في الجهود الحثيثة والإنجازات التي تحقّقت في مجال تصميم القواميس الالكترونية الأحادية اللغة والمتعدّدة اللّغات باعتمادها على إحصاء مترادفات في اللغة الهدف للّفظة الواحدة في اللغة المصدر.

– اعتمادها على قواعد اللغة المصدر واللغة الهدف سمح بتمثيلها حاسوبيا وعزّز من استغلال اللسانيات الحاسوبية في خدمة الترجمة.

– تترجم آلاف الكلمات في ثوان فقط.

مآخذ مقاربة الترجمة الآلية القائمة على قواعد اللغة:

– التكلفة الباهظة من الوقت والموارد المالية الناجمة عن تحليل اللغات المعنية بالترجمة الآلية.

– عدم تعميم استعمالها كنموذج للترجمة الآلية على جميع اللغات لانفراد كل لغة بخصائصها النحوية، والصرفية، والتركيبية، والدلالية ممّا يستدعي التطوير اليدوي للقواعد اللّسانية لكل لغة.

– اعتماد هذه المقاربة على ترجمة الكلمات (Word-based translation) التي لم تعد شائعة الاستعمال في الوقت الراهن عدا في تصميم القواميس الالكترونية.

النمط الأكثر شيوعا الآن هو نمط الترجمة الآلية القائمة على الجمل (Phrase-

based translation) التي لا تزال تدرج إلى حدّ ما، نظام الترجمة القائمة على المفردة الواحدة ولكن لأغراض ترتيب النّصوص بشكل خطّي قبل معالجة الجمل باستخدام نظام[7] (GIZA++)على سبيل المثال.

مقاربة الترجمة الآلية التجريبية :(Empirical MT approach) لقد أحدثت المقاربة التجريبية (Linarez,2008) ثورة في مجال الترجمة الآلية وهي تتميّز عن الترجمة الآلية القائمة على قواعد اللغة في كونها تعتمد على ذخيرة وافرة من نصوص تمّت ترجمتها فيما بين أزواج من اللغات أو أكثر وتخزينها في ذاكرة الحاسوب في شكل مدوّنات متوازية و لقد عملت اللسانيات الإحضار (Corpus linguistics) على تطوير مقاربتين ارتأينا تصنيفهما ضمن هذا المنهاج كتطبيقين يقومان على ما يطلق عليه بتسمية ذاكرة الترجمة[8] (Translation's memory).

الترجمة الآلية القائمة على الأمثلة (Example-based MT): مقاربة الترجمة الآلية القائمة على الأمثلة (EBMT) هي أسلوب الترجمة الآلية يعتمد على متون تشمل نصوصًا موازية يتم إحضارها في عملية الترجمة. هي أساسا ترجمة عن طريق القياس أو التناظر، ويمكن أن ينظر إليها على أنها تطبيق لمقاربة الاستدلال المبني على دراسة الحالات لعملية تعلّم الآلة.

مقاربة الترجمة الآلية الإحصائية SMT (Statistical Machine Translation): هي مقاربة من مقاربات الترجمة الآلية حيث يتم توليد الترجمات على أساس النماذج الإحصائية التي تعتمد على تحليل مجاميع من النصوص الموازية في زوج أو أكثر من اللغات. أدخل وارن ويفر (Weaver,1955) الأفكار الأولى عن الترجمة الآلية الإحصائية ولو أنّ تطبيقها اقتصر على إحصاء المفردات ومقابلها في اللغة الهدف.

كان فرانز جوزيف أوتش (Franz Joseph Och)، الذي ترأّس فريق البحث في مجال الترجمة الآلية في مؤسّسة "غوغل" (Google)، من الأوائل الذين انتقدوا الخوارزميات المصمّمة لنظم الترجمة الآلية القائمة على قواعد اللّغة وأوّل من وضع

المقاربة الإحصائية بالاستناد إلى الخوارزميات الرياضية كنظرية بايز للاحتمالات.

إن عملية انتقاء المفردة في اللغة الهدف والأكثر احتمالا بأن تؤدي الوظيفة التّرجمية المناسبة بحاجة إلى أدوات في غاية الدقة في البحث وأدوات تفكيك (Callison-Burch,2007)عناصر النص المُقرّر تقريره مع استهداف المفردة المعنية بالترجمة وانتقاء مكافئها ضمن عدة مترادفات في تواتر استعمالاتها وفي سياقات متعدّدة.

كانت بدايات الترجمة الإحصائية مقتصرة على المفردات، ثم تطوّرت لتشمل جملا وارتقت إلى إدراج البنى التركيبية النحوية.

مآخذ المقاربة الإحصائية للترجمة الآلية:

- التكلفة الباهظة لإعداد المدوّنات.
- عدم التنبؤ مسبقا بالأخطاء التي يمكن أن ترد وبالتّالي صعوبة معالجتها وتصحيحها.
- قد تبدو سلاسة الترجمة التركيبية سطحية وتنطوي على أخطاء في الترجمة.
- عدم نجاعتها في الترجمة بين أزواج من لغات مختلفة الأصل التي تتباين في نظامها الكتابي وترتيب الكلمات في الجمل. فعلى سبيل المثال، لا يمكن قياس النتائج الجيّدة لهذه المقاربة فيما بين اللغات الأوروبية على الأزواج الأخرى من اللّغات التي لم تحظ بتجميع ضخم للمدوّنات المُعالجة آليا والتي تتميّز بفروق شاسعة من الناحية اللغوية النحوية.
- ترتيب الجمل والكلمات، كما أسلفنا الذكر، يشكل معضلة تواجه منظومة الترجمة الإحصائية، فضلا عن الشذوذ الذي قد يشوب العملية الإحصائية بعدم انتقاء الاحتمال الأكثر مواءمة، وتبرز هذه المشاكل عند ترجمة أسماء الأعلام والعبارات الاصطلاحية إذ أنّ نجاعة الترجمة تكون دومًا رهينة ذاكرة الترجمة والبرامج المُعتمدة في تصميمها.

مزايا المقاربة الإحصائية للترجمة الآلية:

-الاستعمال الناجع للموارد البشرية والبيانية.

– ضخامة المدونات الموازية في شكل مقروء آليا فضلا عن البيانات الأحادية اللغة.

– تعميم تطبيق هذا النمط على جميع اللغات وعدم تصميمه لزوج معيّن من اللغات.

-الحصول على ترجمة أكثر طلاقة وسلاسة بفضل استعمال نموذج اللّغة القائم على نظرية توزيع الاحتمالات بحيث أنّه يوفّر السياق للتّمييز بين الكلمات والعبارات التي تبدو مشابهة. تستخدم نماذج اللّغة في استخراج المعلومات باستخدام نموذج الاحتمالات المطلوبة بصفة أكثر.

باستثناء برمجية "غوغل" الموصوفة أدناه لأغراض هذه الدراسة، هناك برمجية أخرى تعتمد الترجمة القائمة على المقاربة الإحصائية وتتمثّل في نظام بينغ (Bing) للترجمة وهو محرّك بحث حديث النشأة والمعروف سابقا بتسميات مثل لايف سورتش (Live Search) وويندووز لايف سورتش (Windows Live Search) وأم أس أن سورتش (MSN Search) هو نظام طوّرته شركة مايكروسفت (Microsoft) وعرضته على الاستعمال العام في الفاتح جوان من عام 2009. إنه نظام يعتمد على المقاربة الإحصائية ويسعى لمنافسة نظام "غوغل".

بين هذه المقاربة وتلك، تظل الترجمة الآلية من اللغة العربية وإليها تعاني من ضعف النتاج بالمقارنة مع اللغات الأخرى ومن الأرجح أن ذلك راجع أساسا إلى محاولات تطبيق المقاربات المستهدِفة للغات الغربية على اللغة العربية. بالتالي ما هي المقاربة الأنسب التي يمكن تطويرها لتصميم وحدة نمطية للترجمة الآلية العربية؟

لكي تتّضح الرؤية حول هذه القضية، ارتأينا تقديم النتائج التي خلصنا إليها في إطار إعداد رسالة الدكتوراه[9] التي خصّصنا شقّا من جانبها التطبيقي لتقييم ومقارنة

أداء ترجمة منظومة "الوافي الذهبي" القائمة على قواعد اللغة وأداء ترجمة "غوغل" القائمة على المقاربة الإحصائية وأقدمنا على تقييم أداء المنظومتين على أساس الأخطاء الواردة في النتاج المُخرج باستخدام طريقتنا الخاصة التي استوحيناها من أبرز مستويات تحليل الخطاب في تطبيقات اللسانيات، أي المستوى التركيبي والنحوي والصرفي، والمعجمي.

فقد حدّدنا فئات الأخطاء وفقا للفئة اللغوية ولمستويات تحليل اللغات الطبيعية التي تنتمي إليها المفردة المعنية وهي على النحو التالي:

الأخطاء النحوية: تشمل الأخطاء الإعرابية، وموضع الكلمات وحركاتها الإعرابية حسب قواعد اللغة والوظيفة النّحوية لكل مفردة وعلاقة الأسماء فيما بينها من حيث الإفراد والتأنيث، والتّذكير، والجمع، والتّثنية، والتعريف، والتنكير والمتلازمات اللفظية وتوافق حروف الجرّ والضّمائر وما شابه ذلك مع الأسماء أو الأفعال التي تسبقها، الخ.

الأخطاء الصّرفية: تشمل علامات الصرف المرتبطة بالأفعال تحديدا وتوافقها مع الأزمنة ومع علاقتها بالأسماء التي تُنسَب إليها أو الضمائر من حيث الجمع والتأنيث والتذكير، والتثنية ...إلخ.

الأخطاء المُعجمية والسياقية: تشمل الأخطاء في إيجاد المكافئ من حروف وأسماء وأفعال، وأشباه جمل. تندرج ضمن هذه الفئة من الأخطاء، الأخطاء المقامية أو السياقية.

الأخطاء التركيبية: هي جميع الأخطاء المرتبطة بتراكيب الجمل والمتمثّلة في مواضع الكلمات وفي إضافة كلمات أو حروف لا أساس لها في النّص المستهدف أو حذف لحروف أو كلمات أو جمل معيّنة لم يقم المترجم الآلي بترجمتها.

الأخطاء الإملائية: هي كتابة الكلمة بشكل لا يتّفق مع قواعد الإملاء وقد ينتج ذلك عن جهل الكاتب أو عن خطأ مطبعي أو لمشاكل تقنية برمجية. تُعدّ الأخطاء

الإملائية والنّحوية ظاهرة مُنتشرة بكثرة أثناء الكتابة باللغة العربية دون غيرها من اللغات الأخرى.

لقد عمدنا إلى تحديد أنواع الأخطاء الواردة في نتاج ترجمة برمجيتي "الوافي الذهبي" و"غوغل" من خلال تحليل عميق لاستخلاص نسب النجاح المئوية للبرمجيتين والتّرجيح لإحداهما مع تمثيل النّسب المحصل عليها في جداول ورسوم بيانية.

النموذج المُقترح: النص الانجليزي الأصلي[10]

Artificial Intelligence : Artificial Intelligence (AI) is the area of computer science focusing on creating machines that can engage on behaviors that humans consider intelligent.

The ability to create intelligent machines has intrigued humans since ancient times and today with the advent of the computer and 50 years of research into AI programming techniques, the dream of smart machines is becoming a reality.

ترجمة "الوافي الذهبي" لعام 2015

الاستخبارات الاصطناعية: الاستخبارات الاصطناعية (AI) منطقة علم الحاسبات تركّز على خلق المكائن التي يمكن أن تشغل على سلوك ذلك البشر يعتبر ذكي.

القدرة لخلق مكائن ذكية فتنت بشرا منذ أوقات قديمة واليوم بوصول الحاسوب و50 سنة من البحث إلى تقنيات برمجة الاستخبارات الاصطناعية، حلم المكائن الذكية **يصبح** a حقيقة.

تحديد أخطاء الترجمة الآلية بمنظومة الوافي الذهبي لعام 2015[11]

الأخطاء التركيبية واللّغوية والإملائية:

ــ **الخطأ الأول:** العنوان: ... الاستخبارات الاصطناعية ...: كتابة خاطئة للمبتدأ "الاستخبارات" باستخدام همزة القطع. يتعيّن كتابتها بهمزة الوصل.

ــ **الخطأ الثاني:** الفقرة 01 الجملة 01: .... الاستخبارات الاصطناعية...: ورود نفس الخطأ الإملائي السابق ــ

الخطأ الثالث: الفقرة 01 الجملة 01: ... يمكن أن تشغل على السلوك ذلك البشر يعتبر ذكي ــ العبارة بحاجة إلى إعادة صياغة.

ــ **الخطأ الرابع:** الفقرة 02 الجملة 01: ... **القدرة لخلق** ــ حرف الجر "الــلام" غير مناسب والصواب تعويضه بالحرف "على".

ــ **الخطأ الخامس:** الفقرة 02 الجملة 01: ... **البحث إلى تقنيات** ــ حرف الجــر "إلى" غير مناسب والصواب تعويضه بحرف الجر " في".

ــ **الخطأ السادس:** الفقرة 02 الجملة 02: ... **الإستخبارات** .... تكــرار نفــس الخطأ الإملائي.

ــ **الخطأ السابع:** الفقرة 02 الجملة 02: ... يصبح **a** حقيقة ...: ورود الحــرف «a» في النص العربي.

الأخطاء النحوية:

ــ الخطأ الوحيد: الفقرة 01 الجملة 01: ...خلق **المكائن** ...: الصواب في التنكيــر بدل التعريف.

الأخطاء الصرفية

ــ الخطأ الوحيد: الفقرة 01 الجملة 01: ... ــ **يصبح** ...: ــ الصوب تصــريف الفعل في زمن الماضي.

الأخطاء المعجمية والدلالية:

ــ **الخطأ الأول:** العنوان: ... **الإســتخبارات الاصــطناعية** ...: وردت الترجمــة خاطئة سياقيا ودلاليا والصواب هو اختيار العبارة "الذكاء الاصطناعي".

ــ **الخطأ الثاني:** الفقرة 01 الجملــة 01: ... **الإســتخبارات الاصــطناعية** ...: تكرار نفس الخطأ السابق.

ــ **الخطأ الثالث:** الفقرة 01 الجملة 01: ... **منطقة** علم الحاســبات ...: الكلمــة "منطقة" غير مناسبة سياقيا ودلاليا ويتعيّن استبدالها بكلمة "مجال".

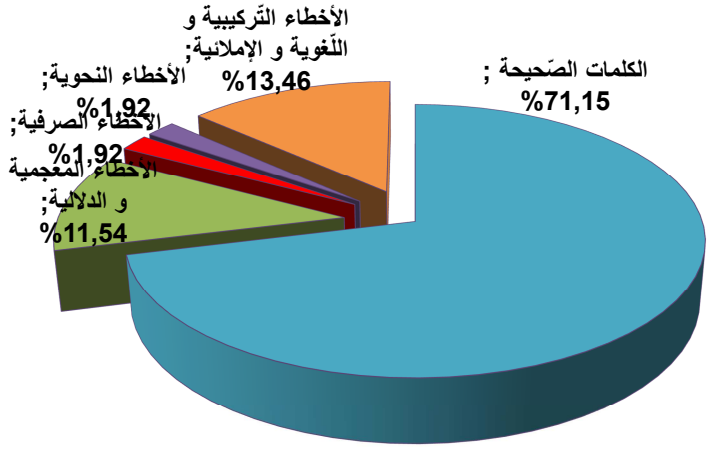
ـ **الخطأ الرابع:** الفقرة 01 الجملة 01: ... **خلق** المكائن ...: الاسم "خلق" غير مناسب سياقيا ودلاليا ويتعيّن استبداله بكلمة "ابتكار".

ـ **الخطأ الخامس:** الفقرة 02 الجملة 01: ... **الإستخبارات الاصطناعية** ...: تكرار الخطأ السابق نفسه.

ـ **الخطأ السادس:** الفقرة 02 الجملة 01: ... **بوصول** ...: الكلمة "بوصول" غير مناسبة سياقيا ودلاليا ويتعيّن استبدالها بكلمة "بظهور".

	كلمات النص المترجم	الأخطاء					الكلمات الصّحيحة
		الأخطاء التّركيبية واللّغوية والإملائية	الأخطاء النحوية	الأخطاء الصرفية	الأخطاء المعجمية والدلالية	مجموع الأخطاء	
العدد	52	07	01	01	06	15	37
النسبة المئوية %	100	13,46	1,92	1,92	11,54	28,85	71,15

تقييم أداء منظومة الوافي الذهبي 2015[12]



تمثيل نسب الأخطاء الواردة في ترجمة الوافي الذهبي لعام 2015[13]

ترجمة "غوغل" لعام 2015[14]:

الذكاء الاصطناعي:

الذكاء الاصطناعي (AI) هو مجال علوم الكمبيوتر **مع التركيز** على الأجهـزة **التي يمكن أن خلق الانخراط** في السلوكيات **بأن تنظر البشر ذكاء.**

القدرة على خلق آلات ذكية **لها البشر مفتون** منذ العصور القديمة، واليوم مـع ظهور الكمبيوتر و50 سنوات من البحث في تقنيات البرمجة **منظمة العفو الدولية** و**الحلم من الأجهزة الذكية أصبحت** حقيقة واقعة.

الأخطاء التركيبية واللّغوية والإملائية[15]:

ــ الخطأ الأول: الفقرة 01 الجملة 01: .... **مع التركيزــ** الصواب اسـتعمال تقنية استبدال الاسم بالفعل على النحو التالي: التي تركّز...

ــ الخطأ الثاني: الفقرة 01 الجملة 01: ... **التي يمكن أن خلق الانخـراط** ...: العبارة بحاجة إلى إعادة صياغة.ــ

ــ الخطأ الثاني: الفقرة 01 الجملة 02: ... **بأن تنظر البشر ذكاء** ...: العبـــارة بحاجة إلى إعادة صياغة.

ــ الخطأ الثالث: الفقرة 02 الجملة 01: ... **لها البشـــر مفتـــون** ...: العبـــارة بحاجة إلى إعادة صياغة.

الأخطاء النحوية:

الخطأ الوحيد: الفقرة 02 الجملة 02: الحلم **من** الأجهزة الذكية ــ يجدر إعـــادة صياغة الجملة نحويا على الشكل التالي: الحلم بأجهزة ذكية...

الأخطاء الصرفية:

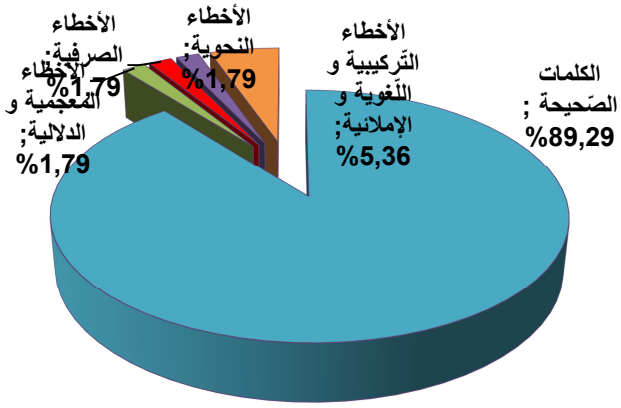
الخطأ الوحيد: الفقرة 02 الجملة 03: ــ **أصـــبحت** ...: يجـــدر التـــذكير بـــدل التأنيث.

الأخطاء المعجمية والدلالية:

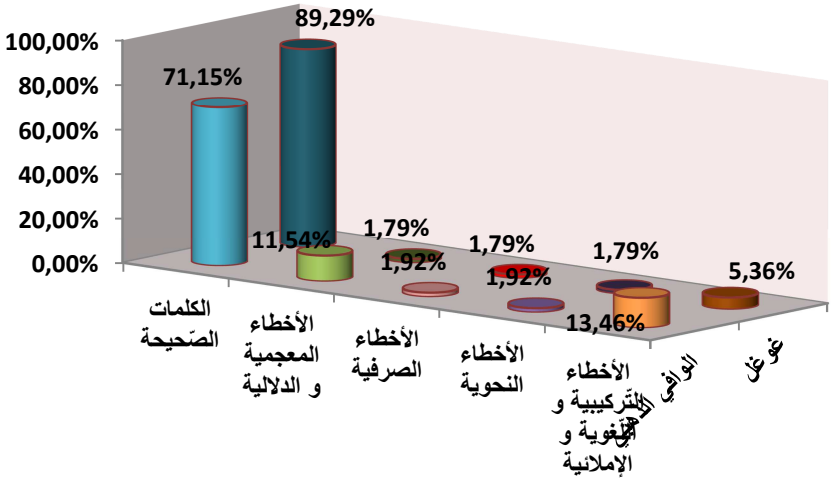
الخطأ الوحيد: الفقرة 02 الجملة 02: ...**منظمة العفو الدولية** ...: هي ترجمة خاطئة للعبارة المختصرة AIالتي تشير إلى الذكاء الاصطناعي.

	كلمـــات النص المترجم	الأخطاء					الكلمات الصّحيحة
		الأخطـــاء التّركيبية واللّغويـــة والإملائية	الأخطاء النحوية	الأخطاء الصرفية	الأخطـــاء المعجمية والدلالية	مجموع الأخطاء	
العدد	56	03	01	01	01	06	50
النســـبة المئوية %	100	5,36	1,79	1,79	1,79	10,71	89,29

تقييم أداء منظومة غوغل 2015[16]



تمثيل نسب الأخطاء الواردة في ترجمة غوغل لعام 2015[17]



تمثيل مقارن للنسب المئوية لأداء منظومة الوافي لعام 2015 ومنظومة غوغل 2015[18]

الترجمة النموذجية للنموذج الخامس:

الذكاء الاصطناعي: يعدّ الذكاء الاصطناعي مجالاً من علم الحاسوب يركّز على إنجاز آلات تحقّق وظائف يصفها البشر بالذكيّة.

لقد فتنتّ إمكانية انجاز آلات ذكية العديد من البشر منذ الأزمنة الغابرة واليوم بفضل ظهور الحاسوب وبعد خمسين سنةٍ من البحثِ في مجال تقنيات البرمجةَ للذكاء الاصطناعي، بدأ حلم انجاز آلات ذكيّة يتحوّل إلى حقيقة.

استنتاج: بالنسبة للنموذج المقترح، وبالاستناد إلى النتائج المحصل عليها باستخدام المنظومتين، يتبيّن أن منظومة "غوغل" تتفوّق في توليد نتاج بعدد أخطاء تقل نسبها عن منظومة الوافي الذهبي، لا سيما فيما يتعلق بنسبة الأخطاء التركيبية واللغوية والإملائية التي أحرزت فيها برمجية غوغل **5.36 %** مقارنة بالوافي الذهبي الذي سجّل عدد أخطاء أكبر تقدر بنسبة **13.46%**، في حين نلاحظ تراجعا طفيفا لمنظومة الوافي الذهبي في نسبة الأخطاء النحوية التي بلغت **1.92%** مقارنة بنسبة **1.79 %** من جانب منظومة "غوغل".

بالنسبة للأخطاء المعجمية والدلالية، نلاحظ أن منظومة "غوغل" تفوّقت بفارق معتبر بنسبة أخطاء بلغت **1.79 %** مقابل 11.54. **%** لدى منظومة الوافي الذهبي.

لقد لاحظنا من خلال نتاج ترجمة العيّنة كيف أنّ برمجية "غوغل" تمكّنت من خلال تطبيق خوارزميات مختلفة تقوم على الاحتمالات والإحصاء، من توليد ترجمات مقبولة شكلا ومفهومة من حيث المضمون الدلالي. كما وفّقت منظومة "غوغل" في ربط ترجمتها الآلية بذاكرة الترجمة، التي ساعدت على تهذيب النص المُنتج بواسطة عبارات سبق ترجمتها من قبل الإنسان ومراجعتها. بالتّالي، نلاحظ أن الجمع بين الترجمة الآلية وذاكرة الترجمة يزيد من نسبة الاتّساق والجودة في الترجمة الآلية. بالطبع، نحن لا ننكر أن المنظومة لا تزال تكتنفها شوائب ولا سيما

في الترجمة إلى لغات لا تنتمي إلى عائلة اللّغات نفسها، كما هو الحال في العلاقة بين اللغة الانجليزية والعربية.

من جهة أخرى، لاحظنا من خلال العرض، بأنّ منظومة الوافي الذهبي تستخدم في ترجمتها الآلية المعلومات اللّغوية لتفكيك نصّي اللغة المنقولة واللغة المنقول إليها بصفة منطقية وهي بذلك تتفوق من ناحية النتاج على المنظومة الإحصائية في معالجة الجانب النحوي حيث تقلّ نسبة الأخطاء الناتجة عن ترجمتها الآلية بالمقارنة مع نتاج "غوغل" عدا في بعض الحالات التي حقّقت فيها منظومة "غوغل" تفوّقا في هذا الجانب. لتوضيح الفروق القائمة بين المقاربتين، قمنا بحصرها في الجدول أدناه:

المقاربة اللغوية (الوافي الذهبي)	المقاربة الإحصائية (غوغل)
هي مقاربة قائمة على قواعد اللغة	هي مقاربة قائمة على المتون الموازية أي ترجمات بشرية موازية وتتجاهل القواعد اللغوية
تطبق قواعد اللغة على مراحل التحليل والتحويل والتوليد	تعتمد قوانين الاحتمالات
هي بالتالي تولّد نصوصا تقل فيها الأخطاء اللغوية على العموم ولكنها تتضمن مشاكل دلالية وسياقية.	توفر أداء ذا جودة من حيث قلة الأخطاء الدلالية.
يستلزم بناء منظومة قائمة على قواعد اللغة تمثيل قواعد لغوية واسعة النطاق وجهد بشري، خاصة عندما لا تنتمي اللغتان، المصدر والهدف إلى نفس عائلة اللغات.	لا يستلزم بناء منظومة إحصائية على قواعد اللغة ولا على الجهد البشري لتمثيل تلك القواعد وتحديدها.

الاعتماد على قواميس ثنائية اللغة وعلى الأدوات اللغوية (المحلل الصرفي، المدقق الإملائي، المحلل الإعرابي...الخ)	لا تُستخدم القواميس ولا الأدوات اللغوية في هذه المقاربة.
لا تعتمد هذه المقاربة على ترجمات سابقة ولا على متون مُخزّنة.	تعتمد هده المقاربة على ترجمات سابقة من خلال متون كبيرة الحجم.
عدم الحاجة إلى ذاكرة الترجمة	قائمة أساسا على ذاكرة الترجمة
التراكيب بحاجة إلى إعادة صياغة على نطاق كبير.	غالبا ما لا ترد مشاكل على مستوى التراكيب.
تواجه المقاربة معضلات اللبس على اختلاف أنواعه	لا تواجه المقاربة معضلات اللبس في أغلب الأحيان
يكون النتاج ذا جودة كلما كانت الجمل قصيرة في النص.	يكون النتاج ذا جودة كلما كانت الجمل قصيرة في النص.
يبقى الأداء ضعيفا مقارنة بالترجمة البشرية.	يبقى الأداء ضعيفا في غياب المتون ذات الصلة بالنص المترجم.

مقارنة بين المقاربة اللغوية والمقاربة الإحصائية[19]

على الرغم من مواطن القصور التي تمت معاينتها في برمجيات ترجمة النصوص على اختلاف مقارباتها، إلا أنّها أدوات لا يمكن الاستغناء عنها في المجالات التبليغية وفي الترجمة التفسيرية، كما أنها تشكل أداة ثمينة لربح الوقت وتوفير العناء عن المترجمين المحترفين في البحث في القواميس الورقية وفي رقن النص المُترجم في اللغة المستهدفة.

لاحظنا أن المقاربة الإحصائية التي أثبتت نجاعتها وتفوّقها في مجال الترجمة الآلية لا سيما بين اللغات التي تنتمي إلى العائلة اللغوية ذاتها، بحاجة إلى أدوات لغوية إضافية للتقليص من مواطن ضعفها، خاصة لمّا يتعلّق الأمر بالتعامل مع اللغة العربية كلغة مُترجم منها أو إليها.

في ضوء الدراسة الاستقرائية التي سبقت، نحن نرى في خاتمة دراستنا أدناه ضرورة اقتراح برمجية نموذجية في تصميم وحدة نمطية عربية للترجمة الآلية في شكل توصيات تأخذ بعين الاعتبار المعايير الأربعة الأساسية لتقييم الأداء العام لأي منظومة كانت.

خاتمة : لمّا كانت جودة أي منظومة آلية ترتكز على تحليل الأخطاء الناتجة عن تلك المعالجة والتي توحي بالأساليب المعتمدة لتصميم برمجيات الترجمة أظهرت دراستنا التطبيقية بأنّ تباين النتاج من برمجية إلى أخرى يبقى مرتبطا بالأساليب المتّبعة في تصميم تلك البرمجية فكلّما كان الأسلوب المنتهج صائبا، كلّما كان النتاج أكثر جودة. لذلك فالبرمجيات المُعتمدة على تمثيل قواعد اللغة، غالبا ما لا يحتوي نتاجها على أخطاء نحوية وصرفية عديدة ويكون ذلك على حساب التراكيب وفي ذكرنا لهذا الجانب الهام من مستويات تحليل اللغة الطبيعية، يبدو لنا من وحي التجربة، أنّ التراكيب تشغل حيّزا كبيرا في مجال تأليف النصوص وتحليلها بالنّظر إلى الدور المحوري الذي تلعبه في الحصول على التجانس بين أجزاء الجمل والعبارات مع احترام القواعد البنيوية الخاصة بكل لغة. بالتّالي تقتضي هذه الأهمّية التي تكتسيها التّراكيب في بناء النّصوص، عنايةً كبيرة في التمثيل الحاسوبي للحصول على نتاج يتضمّن أقلّ عددٍ ممكنٍ من الأخطاء في التّراكيب.

من جهة أخرى، وفي ضوء ما عرضناه في هذا البحث من مقاربات لبناء برمجيات ضرورية للاستخدام كوسائل مساعدة على الترجمة وكفيلة بتوليد نتاج لا يحتاج إلى جهد كبير لإعادة صياغته، خلصنا إلى أنّ خصائص اللغة العربية وضرورة تعاملها مع اللغات الغربية تستدعي برمجية نموذجية تعتمد المناهج التالية:

– التمثيل اللغوي للمستويات الأساسية من قواعدها اللغوية، لا سيما الصرفية والمعجمية منها والتراكيبية: بالتالي، فهي بحاجة إلى وحدات من المقاربة اللغوية.

– بالنسبة للمستويات السياقية والدلالية، البرمجية العربية بحاجة إلى حواسيب ذات سعة تخزين كبيرة لاحتواء حجم كبير من المتون الموازية الموضوعية، فكلما تقلّص عدد المجالات المُعالجة، كلما حقّقت منظومة الترجمة نسبة عالية من الدقة في الترجمة. في هذه الحالة، تحتاج البرمجية العربية إلى تزويدها بالخوارزميات المُطبّقة في المقاربة الإحصائية

– البرمجية العربية بحاجة إلى المزج بين الوحدات التي حقّقت مواطن القوة في المقاربة اللغوية والمقاربة الإحصائية.

– المنظومة العربية بحاجة إلى الاستيحاء من الجهود الغربية الرامية إلى تصميم برمجية تشمل المقاربتين وهي برمجيات قائمة على المقاربة الهجينة ((Hybrid على طريقة منظومة "سستران" التي لا تزال تعاني من قصور رغم تحقيقها لترجمات ذات مستوى مقبول فيما بين اللغات المتشابهة الخصائص.

– البرمجية العربية بحاجة إلى مقاربة هجينة خاصة بها.

– البرمجية العربية بحاجة إلى تطبيق خوارزميات في قانون الاحتمالات على أساس وحدات ن–غرام التي سيكون لها تأثيرا إيجابيا هاما إذا تمّ الاعتماد على المتلازمات اللفظية وعلى العبارات الاصطلاحية أو المصطلحات المتخصصة التابعة لمجال محدّد.

– البرمجية العربية بحاجة إلى تضافر جهود المختصين في علوم الحاسوب واللغة والمجالات المتخصصة حسب أهميتها العملية والاستراتيجية وذلك، في مرحلة التصميم.

– البرمجية العربية بحاجة إلى عناصر بشرية لتقييم أدائها بعد مرحلة التصميم. في هذه الحالة، دور المترجم المحترف حاسم في تقييم الأداء من منطلق

جودة النتاج وسرعة الأداء وكلفته، وتوجيه المصمّمين فيما يتعلق بالغرض المنشود من البرمجية وفي تجميع المتون اللازمة لتعبئتها بها، في حين يتولى عالم اللغة واللسانيات تحديد القواعد اللغوية والمستويات الأساسية التي يجب تمثيلها بعد القيام بدراسة عملية مسبقة.

– البرمجية النموذجية العربية تستلزم حسن تحديد الأغراض وانتقاء المادة اللغوية والإحصائية وتجنيد المتدخلين المؤهلين في تصميمها من أجل تحقيق الأهداف المرتقبة.

إنّ اقتراح برمجية عربية للترجمة مشروع عملي يمكن تجسيده على أرض الواقع لتوفر البلاد العربية على الكفاءات والمهارات في مجال المعلوماتية وغيرها من الشُّعَب التي هي بحاجة إلى أن تُقيم بشأنها المؤسسات الجامعية جسرا للتواصل فيما بينها كما يجب حث القطاع الخاص على بذل الجهد والمال في مثل هذه المشاريع ولعلّ أبسط المشاريع التي لم تكترث لها المؤسسات العربية والتي تنطوي على أهمية قصوى ولا تستدعي تجنيد موارد مالية ضخمة، هي تشجيع تصميم قواميس ومعاجم الكترونية مفتوحة المصدر ومتخصصة المجالات، بحيث تسعى كل أسرة جامعية إلى بناء مسرد رقمي خاص بمختلف تخصّصاتها لتعريب العلوم والمعارف وستقوم مقام بنوك معطيات.

تكمن فائدة القواميس الالكترونية المفتوحة المصدر في توفير المادة الأساسية للمجال المُختص وفتح المجال للمختصين وغيرهم في ميدان معيّن لتعبئة القاموس بصفة مستمرة وتحديثه والرجوع إليه عند الحاجة وخلال إعداد البحوث باللغة العربية كما يمكن استغلال محتوى القاموس كمادة لغوية لتعبئة برمجيات الترجمة.

تجدر الإشارة إلى أنّ الاتجاه إلى اعتماد مقاربات هجينة، أصبح محلّ دراسة لدى الشركات العملاقة المُصمّمة لبرامج الحاسوب وبرمجيات الترجمة. في هذا المجال نذكر جهود شركة "غوغل" في رعاية بحوث أكاديمية تُقدّم اقتراحات حول

تحسين النظم الإحصائية وإرفاقها بوحدات نمطية لغوية. في هذا الشأن، نحن ننوّه بجهود الطلبة والباحثين الشباب في سعيهم إلى ترقية اللغة العربية من خلال إدراجها بصفة فعالة في نظم الترجمة الآلية ونذكر على سبيل المثال العمل القيّم الذي أعدته الباحثة سهى سلطان (Soha Soltan) بمعهد علوم الحاسب في سويسرا، من خلال مناقشتها لمذكرة ماستر باللغة الانجليزية في عام 2011 بالتعاون مع شركة "غوغل" تحت عنوان: " *Applying Morphology to English-Arabic Statistical Machine Translation*" (تطبيق قواعد الصرف على منظومة الترجمة الآلية الإحصائية من اللغة الانجليزية إلى اللغة العربية") أشرفت عليها الدكتورة كيث هول (Keith Hall) وهي تعمل بفريق البحث العلمي في شركة "غوغل"، والأستاذ دونالد كوسمان (Donald Kausmann) ، باحث في علوم الحاسب أيضا في سويسرا. مثل هذه الأعمال الجليلة بحاجة إلى الإحصاء وإلى الترجمة إلى اللغة العربية لإثراء قائمة المراجع المعرفية في هذا المجال ولتكون نقطة انطلاق لإعداد بحوث في هذا المجال، ولا سيما أنّ العمل يُبلور بوضوح فكرة تقريب الباحث من ميدان البحث وتنسيق الجهود الأكاديمية مع عمل المؤسسات والشركات، بحيث يكون مضمون البحث العلمي الأكاديمي حتما موضوعيا وعمليا يستند إلى تجارب ملموسة ويطرح مشاكل واقعية، ويتقدّم باقتراحات عملية.

قائمة المراجع:

باللغة العربية:

1) ــ القحطاني، سعد بن هادي: *" تحليل اللغة العربية بوساطة الحاسوب"*. الرياض: مركز اللغة الإنجليزية – معهد الإدارة.

2) ــ بربارة، سهيلة (2006): الترجمة بمساعدة الحاسوب من الانجليزية إلى العربية: دراسة تحليلية ونقدية لنصوص مترجمة آليا. الجزائر: مذكرة لنيل شهادة الماجستير، قسم الترجمة، جامعة الجزائر 2.

3) ــ بربارة، سهيلة (2017)، الترجمة بمساعدة الحاسوب من الانجليزية إلى العربية: بين مقاربة برمجية "الوافي الذهبي" اللّغوية وبرمجية "غوغل" Google الإحصائية: دراسة تاريخية تعاقبية وتزامنية مقارنة لنماذج نصية مترجمة آليا. الجزائر: رسالة دكتوراه، جامعة الجزائر 2، معهد الترجمة.

4) ــ خطاب، شريف (2011): "المعالجة الآلية للغة العربية، أساسيات الحاسوب". القاهرة: بحث، قسم علوم الحاسب، كلية الحاسبات والمعلومات، جامعة القاهرة.

5) ــ. نبيل علي (1998): اللغة العربية والحاسوب. الكويت: مؤسسة تعريب الكويت

مواقع ويب:

1) موقع شركة "آي تي آي" ATA: https://site2corp.com/uk/ata-software-technology-limited

2) موقع شركة "صخر": http://www.sakhr.com/index.php/en/

باللغة الأجنبية

1) Alaouneh, Mouiad Fadiel (2011) . *Rule-Based and Example-Based Machine Translation from English to Arabic. Malaysia:* Fac. of Inf. Sci. & T.echnol., Nat. Univ. of Malaysia, Bangi, Malaysia. Published by EEE Conference Publications.

2) Callison-Burch, Chris (2007) . *Machine translation: Word-based models and the EM algorithm*.Slides borrowed from Philipp Koehn: John Hopkins University December 3 http://www.sciencedirect.com/science/article/pii/S0885230814001077

3) Costa-Juss , Marta R. (2012) . *Study and comparison of rule-based and statistical Catalan- Spanish Machine Translation Systems, Computing and Informatics*, Vol. 31, P 248.

4) Costa-juss, Marta R. (2015). *Latest trends in hybrid machine translation and its applications*. Computer speech and language, volume 32, issue 1, , PP 3- 10. Paper on ScienceDirect website. http://www.sciencedirect.com/science/article/pii/S0885230814001077

5) Costa-juss, Marta R. (and others) (2016). *Hybrid Approaches to Machine Translation*. eBook ISBN : 978-3-319-21311-8, pp 67-

75.

6) Grajales, Carlos Alberto Gómez (2015). "*The statistics behind Google translate".* Statistics Views, available from: http://www.statisticsviews.com/details/feature/8065581/The-statistics-behind-Google-

7) Hutchins, W.J.(1987). *Machine Translation: Past, Present, Future*: Ellis Horword Limited, West Sussex <http:ourworld.compuserve.com/hompages/WJ Hutchins/PPF_3.pdf) (assessed 29 December 2005)

8) Limarez, Jesus Angel Gimenez (2008) . *Empirical Machine Translation and its Evaluation. Barcelona:* Tesi Doctoral per a optar al grau deDoctor en Inform`atica. Programa de Doctorat en Intellig`encia Artificial, Departament de Llenguatges i Sistemes Inform`atics, Universitat Polit`ecnica de Catalunya, Barcelona, , PP 7-9.

9) Richens, R.H. and Booth, A.D. (1955).*'Some methods of mechanized translation.'* In: Locke, W.N. and Booth, A.D. (eds.) Machine translation: fourteen essays (Cambridge, Mass.: Technology Press of the Massachusetts Institute of Technology), pp. 24-46.

10) Shannon, C. E. and Weaver, W. (1949). *The mathematical theory of communication.* University of Urbana

11) Soudi, Abdelhadi. Farghaly, Ali. Günter Neumann. Zbib, Rabih (2012). *Challenges for Arabic Machine Translation* : .John Benjamins Publishing, pp343 – 347. Translate.html.

الهوامش

1 July 17, 1894 – November 24, 1978) was an American scientist, mathematician, and science administrator.

2 A computer scientist and a researcher in Hungarian language.

[3] رئيس المجمع الجزائري للغة العربية.

[4] من مقال: "إعادة الاعتبار للغة العربية في المجتمع العربي" نشر في العدد الخاص من كتاب المجلس الأعلى للغة العربية الموسوم "العربية الراهن والمأمول"، 2009، ص 65.

[5] Summer Institute for Linguistics (SIL) Ethnologue Survey (1999) available from : http://populationsdumonde.com/classements/les-10-langues-les-plus-parlees-dans-le-monde

6 George Weber," the Word's10 Most Influential Languages" .In: Language Today Vol. 2, Dec 1997

available from: http://french.server276.com/bulletin/articles/promote/advocacy/useful/toplanguages.pdf

[7] : برمجية متوفّرة على الانترنت للتحميل من أجل ترتيب المدوّنات الموازية ترتيبا خطيا و هو يفيد في إعداد القواميس الثنائية اللّغة. تستخدم شركة آي بي أم هذه المنظومة على نطاق واسع. [7]GIZA++

[8] يرتبط استعمال ذاكرة الترجمة بأسلوب الترجمة بمساعدة الحاسوب خلافا للترجمة الآلية الصرفة التي تعتمد على قواميس الكترونية ضخمة وخوارزميات رياضية. تتمثّل ذاكرة الترجمة في تخزين ترجمات سابقة في شكل نصوص متوازية ثنائية اللغة أو متعددة اللغة يتم إحضارها وتظهر في شكل خيارات ينتقي منها المترجم الترجمة الأنسب والأقرب من السياق الذي هو بصدد الترجمة فيه (شرح الباحثة). تعرّف مجموعة EAGLES (*Expert Advisory Group on Language Engineering Standards*) ــ مجموعة الخبراء المستشارين حول معايير هندسة اللغة ــ ذاكرة الترجمة على أنّها " "أرشيف من نصوص متعددة اللغات (مجزأة، ومصفّفة، ومحلّلة ومصنّفة) قابلة للتّخزين ولاسترجاعها أو مقاطع منها في لغات متعدّدة وفقا لشروط بحث مختلفة".

[9] بربارة، سهيلة (2016)، *الترجمة بمساعدة الحاسوب من الانجليزية إلى العربية:* بين مقاربة برمجية "الوافي الذهبي" اللّغوية وبرمجية "غوغل" Google الإحصائية دراسة تاريخية تعاقبية وتزامنية مقارنة لنماذج نصية مترجمة آليا، جامعة الجزائر 2، معهد الترجمة، ص 238ـــــ. تم استقاء المدوّنة من الوقع التالي: //www/ :
(http exampleessays.com/categories/a.html)

[10] المصدر السابق، ص 238.

[11] المصدر السابق، ص 249

[12] المصدر السابق، جدول 21 نموذج 5 ، ص 283.

[13] المصدر السابق، رسم بياني دائري 19 نموذج 5، ص 283..

[14] المصدر السابق، ص284.

[15] المصدر الصادق، ص 284.

[16] المصدر السابق، جدول 22 نموذج 5 : ص 285.

[17] المصدر السابق، رسم بياني دائري 20 نموذج 5:ص 286.

[18] المصدر السابق، رسم بياني 21 نموذج 5: ص 286

[19] المصدر السابق، جدول 25: ص 298.