تقييم نِتاج الترجمة الآلية إلى اللغة العربية بين أداء نظام "الوافي الذهبي" اللغوي ونظام "غوغل" الإحصائي

سهيلة بربارة

جامعة الجزائر2

barbara.translate@gmail.com

تاريخ الاستلام: 2018/04/11 تاريخ القبول: 2018/05/31

الملخّص

شكّلت الترجمة على مرّ العصور، أداة قوية للتواصل بين الناس والاحتكاك بين الحضارات، كما تضاعفت أهميتها بجميع أبعادها وتأجّج الطلب عليها في القرن الحالي الذي يشهد، لا محالة، تزايدا في الطلب على عمل الترجمة في وقت قياسي، ثمّ إنّ ظاهرة المثاقفة (Accultura-tion) التي يعيشها عصرنا الحالي، تجعل من الترجمة عملية أكثر تعقيدا وتعدّدا.

تمخّض عن المطالب المتواصلة لتبادل المعلومات والاتصال بين لغات مختلفة نتيجة التسارع التكنولوجي، والتحاق جميع الجهات الفاعلة بالشبكة العنكبوتية العالمية، ضغط كبير لتبادل المعلومات من لغات مختلفة في 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 لتسريع عملية إنجاز أعمال الترجمة لا سيما في المجالات التقنية وعالم الخدمات التجارية المُتّسم بالنصوص ذات المصطلحات المتكرّرة والخالية من الصور الإبداعية والبلاغية التي يعجز الحاسوب عن تمثيلها.

الكلمات المفاتيح:

الترجمة الآلية - المقاربة اللغوية - المقاربة الإحصائية - المقاربة الهجينة - وحدة نمطية للترجمة الآلية.

Évaluation de la production en traduction automatique vers la langue arabe
Performance du système de traduction linguistique «Golden Al-Wafi» vs le système de traduction statistique de «Google traduction»

Résumé

De tout temps, la traduction a constitué un puissant outil de communication entre les peuples et a joué un rôle clé dans le rapprochement des civilisations. Son importance, dans tous ses aspects, a été renforcée par la forte augmentation de la demande en services de traduction au cours du siècle actuel et nous assistons, inévitablement, à une ruée vers une demande de traduction dans un temps record. De plus, le phénomène d'acculturation caractérisant notre présente ère, a fait de la traduction un processus plus complexe et multiforme.

La demande persistante d'échange d'informations et de communication dans différentes langues induite par la percée technologique et l'adhésion de tous les acteurs actifs au réseau informatique mondial, est la résultante des nombreuses pressions exercées pour l'échange d'informations dans différentes langues et en un temps record afin de faciliter la satisfaction des besoins impérieux dans les différents aspects de la vie humaine (commerciale, militaire, technique culturelle, sociale, etc.)

Si la traduction humaine réussissait il y a quelques temps, à répondre à ces exigences, elle en est désormais incapable face aux performances et à l'efficacité croissantes des ordinateurs connectés à Internet. Les traducteurs ont donc dû recourir à des outils numériques pour accélérer le processus de traduction, en particulier dans les domaines techniques et dans le monde des services aux entreprises, caractérisés par des textes à la terminologie répétitive et lesquels sont exempts de figures créatives et rhétoriques qui ne peuvent, en aucun cas, faire l'objet de représentation automatique.

Mots clés:

Traduction automatique - approche linguistique - approche statistique - approche hybride - modèle de système de traduction automatique.

Evaluation of the output of Machine Translation into Arabic Linguistic- Based approach «Al- Wafi golden» vs «Google» Statistical - based approach

Abstract

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 we are 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 thes e requirements, it could no longer do it, due to the increasing per formance and efficiency of computers connected to the Internet. Therefore, translators had to resort to digital tools to speed up the process of the com pletion of translation work, especially in the technical fields and the world of business services characterized by repeated terminology texts, free of cre ative and rhetorical figures which cannot be subject to automatic represent a tion.

Keywords:

Automatic translation - linguistic approach - statistical approach- hybrid approach - MT system model.

مقدمــة

يتنـاول بحثنـا جانبـا هامـا مـن الرهانـات التـي تواجـه مجـال الترجمـة الآليـة إلى اللغـة العربيـة التـي تعـرف تأخـرا ملموسـا مقارنـة: اللغـات التـي تحتـل الصـدارة في المجـال التكنولوجـي.

نسـتهل دراسـتنا بلمحـة تمهيديـة عـن نشـأة الترجمـة الآليـة وروادهـا ونسترسـل في الحديـث عـن أبـرز الفاعلـين في مجـال تصميـم برمجيـات الترجمـة الآليـة وعـرض العوامل التاريخيـة الداعيـة لحوسـبة اللغـة العربيـة، والعوامـل العمليـة الدوليـة انتهـاءً بالعوامـل الخاصـة المرتبطـة بخصائص هـذه اللغة التي تؤهّلها للتمثيل الحاسـوبي إلى ضوء المبادرات التـي شُرع فيهـا في هـذا المجـال، مـما يسـوق إلى التطـرق للمسـاعي المبذولـة في مجـال حوسـبة اللغـة العربيـة لأغـراض الترجمـة لا سـيما منتجـات شركـة صخـر وشركـة آي تي آي (ATA Software Technology) التـي انفـردت بتصميمهـا لبرمجيـات للترجمـة الآليـة مـن اللغـة العربيـة وإليهـا تعتمـد عـلى المقاربـة اللّغويـة في ظـل توفّـر مقاربـات متباينـة لبنـاء برمجيـات الترجمـة الآليـة.

غايتنـا مـن هـذه الدراسـة هـي التعريـف بأبـرز المقاربـات المتاحـة في هـذا المجـال واقـتراح المقاربـة الأنسـب، إمّـا لتصميـم برمجيـة آليـة للترجمـة مـن اللّغـة العربيـة، لا سـيما إليهـا أو اقـتراح تحسـينات الوحـدة النمطيـة التـي تُوفّرهـا شركـة «آي تي آي»، في ضـوء النتائـج التـي حصلنـا عليهـا عنـد تقييـم نتـاج برمجيتهـا» ذات المقاربـة اللغويـة وبرمجيـة «غوغـل» (Google) ذات المقاربـة الإحصائيـة.

يتعـين التذكـير بـأنّ موضـوع الدراسـة ينحـصر في البرمجيـات المُصمَّمـة لترجمـة النصـوص والشـاملة للغـة العربيـة عـلى وجـه الخصـوص، وننطلـق مـن فكـرة أنّ هـذه المنظومـات أدوات مسـاعدة عـلى الترجمـة (CAT tools) بحاجـة دائمـة لتقيـح بـشري لاحـق (Post- editing).

عـلاوة عـلى ذلـك، تجـدر الإشـارة إلى أن هـذا الجانـب التحليـلي مـن هـذه الدراسـة هـو عـرض لعيّنـة فقـط مـن النـماذج التـي شـكّلت موضـوع دراسـة تحليليـة شـاملة

وواسعة (بربارة، 2017).

يعقب وصفنا لمقاربات الترجمة الآلية، لا سيما اللغوية والإحصائية منها، عرض حالة تطبيقية لترجمة نموذج نصي في الاتجاه الأحادي (انجليزي ـ عربي) بواسطة منظومة «الوافي الذهبي» اللّغوية ومنظومة «غوغل» الإحصائية مع تحديد مواطن ضعف وقوة كلتا المنظومتين وفقًا للطريقة الموصوفة في العرض التطبيقي، ولقد مثّلنا النتائج المستخلصة 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 لولادة مشروع آلة تُترجمْ سواء على يد المهندس الروسي، بيتر سميرنوف ترويانسكي (Peter Smirnov Troyanskij)، والمهندس الروسي المنشأ، جورج أرترستروني (Georges Artsruni)، أو من خلال النظرية التي طوّرها وارن ويفر (Warren Weaver[1]) عن الترجمة الآلية في عام 1949، والتي نقل عنها الخبير الأمريكي جون هتشينز (Hutshins, 1986) 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 والتحويلي التي مكنت بيتر توما (Peter Tom [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 الحادي والعشرين، أصبحت الضرورة ملحة للارتقاء باللغة العربية من أجل تصميم وحدة ترجمة تشارك نظيراتها الأجنبية في المقاربات المتاحة من المنهج اللغوي والإحصائي والهجين في غياب منظومة مثلى لها، و طرح 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:

العامل التاريخي لحوسبة اللغة العربية

في صدد تناول العوامل التاريخية التي تشكّل الدوافع الداعية لبعث اللغة العربية، لقد حصر الراحل عبد الرحمن الحاج صالح 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 تخاطبٍ كبقايا اللّهجات[3]... إلخ.

يكفي أن نشير، أيضا، إلى ما ذكره الباحث الجزائري، محمد بطّاز، مدير شبكات وأنظمة المعلومات والاتّصال الجامعيّة (في محاضرة ألقاها 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... 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

إن 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 ـ اللغة العربية هي لغة مهمة في جميع الجوانب. يختلف تصنيفها من مصدر إلى آخر حيث هناك من يضعها في المرتبة الثامنة[4] في

العالم من حيث عدد المتحدثين بها في حين صنّفها الباحث الأمريكي جورج فيبر (George Weber) في ورقة بحثية له[5] أعدّها عام 1997 في المرتبة الخامسة مستندا إلى معايير خاصة، ويرى هذا الأخير أن هذا التصنيف بقي ساريا إلى غاية 2008.

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 من سوق 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، أن العوامل الخاصة المتعلّقة بخصائص اللغة العربية، هي التي تؤهّلها وتحفّز 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:

- تميّز 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 قياسًا 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 الخنجرية وهمزة الوصل والهمزة العربية وتمّ 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 ليتناول 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

الأمـر.

بالنسـبة لـلأدوات التكنولوجيـة المسـاعدة عـلى الترجمـة، فقـد كانـت شركـة «صخـر» (SAKHR) التـي أنشـئت عـام 1982 مـن بـين الـشركات الرائـدة في مجـال تطويـر العديد مـن التقنيـات المتقدمـة التـي تركـت علامـات بـارزة في صناعـة تقنيـة المعلومـات. فقـد قامـت الشركـة بتطويـر جيـلٍ جديـدٍ مـن تقنيـات المعالجـة الطبيعيـة للغـة العربيـة (NLP)، كـما وضعـت نظامـا للترجمـة أطلقـت عليـه تسـمية «تَرجِـم» (Tarjim) وهـو ثنـائي الاتجـاه أي إنجليـزي / عـربي / انجليـزي[6].

مـع ذلـك، تـأتي شركـة «آي تي آي» (ATA Software Technology) التـي تأسسـت عـام 1992 مـن قبـل متخصصـين ومبرمجـين تعـود خبرتهـم إلى الثمانينـات ويقـع مقرهـا الرئيـسي في لنـدن في صـدارة الـشركات المنتجـة لبرمجيـاتٍ تسـتهدف اللغـة العربيـة واللغـة الإنجليزيـة في مجـال الترجمـة الآليـة.

وفقـا للمعلومـات المسـتقاة مـن موقـع هـذه المؤسسـة، فقـد أصـدرت الشركـة في عـام 1995 برنامجهـا للتعريـب الـذي أطلقـت عليـه تسـمية «المترجـم العـربي» باعتبـاره أول برنامـج للترجمـة الآليـة مـن اللغـة الإنجليزيـة إلى العربيـة عـلى الحواسـيب الشـخصية والماكنتـوش وطـوّرت الشركـة حزمـة برامـج لتعريـب النصـوص الإنجليزيـة أطلقـت عليهـا اسـم «الـوافي» التـي لقيـت إقبـالا كبـيرا مـن جانـب المسـتخدمين في العـالم العـربي.

يذكـر الصحفـي مـيرزا الخويلـدي في مقالـه الصـادر عـلى الموقـع الالكـتروني[7] لجريـدة الـشرق الأوسـط بتاريـخ 20 جـوان 2002 والـذي خصّـه لتقديـم منتجـات شركـة آي تي آي وأهـم إسـهامات هـذه الأخـيرة في مجـال تطويـر وتعريـب البرمجيـات، مشـيرا إلى إصـدار الشركـة لثلاثـة برامـج متنوعـة مـن برنامـج الـوافي: الأول: هـو برنامـج الـوافي السريـع وهو قامـوس عربي/إنجليـزي، وإنجليزي/عـربي شـامل يعمـل داخـل أغلـب البرامـج المتوافقـة مـع مايكروسـوفت ونـدوز (Microsoft Windows). هـو أكـثر البرامـج انتشـارا، وسـهل الاسـتخدام ومتوفـر بسـعر زهيـد.

البرنامـج الثـاني هـو «الـوافي الجديـد»، وهـو برنامـج متقـدم لترجمـة النصـوص ترجمـة

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 يشمل أكثر من مليون ونصف 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 من اللغة الانجليزية إلى اللغة العربية.

نظرا لأن هذه الشركة المؤسسة الرائدة في وضع وحدة نمطية لترجمة النصوص ترجمة آلية ثنائية الاتجاه انجليزي / عربي / انجليزي، تعتمد المقاربة اللغوية في تصميمها و تقدّم 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 «سستران» (Systran) هي أول برمجية صُمّمت باعتماد هذه المقاربة قبل تطوير الشركة للمقاربة الهجينة التي سنعرضها لاحقا.

مزايا مقاربة الترجمة الآلية القائمة على قواعد اللغة

- لها الفضل في الجهود الحثيثة والإنجازات التي حُقّقت في مجال تصميم القواميس الإلكترونية 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 الباهضة من الوقت والموارد المالية الناجمة عن تحليل اللغات المعنية بالترجمة الآلية.

- عدم تعميم استعمالها أنموذجًا للترجمة الآلية في 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 الإلكترونية.

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[8] (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 ولقد عملت اللسانيات الإحضار (Cor-puslinguistics) على تطوير مقاربتين ارتأينا تصنيفهما ضمن هذا المنهاج كتطبيقين يقومان على ما يطلق عليه بتسمية ذاكرة الترجمة[9] (Translation's memory).

الترجمة الآلية القائمة على المثال (Example- based MT)

مقاربة الترجمة الآلية القائمة على المثال (EBMT) هي أسلوب الترجمة الآلية يعتمد على متون تشمل نصوص 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 (Shannon, 1964) 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 (Calison-Burch, 2007) عناصر النص المُقرّر تقريره مع استهداف المفردة المعنية بالترجمة و انتقاء 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 الباهضة لإعداد المدوّنات.

- عدم التنبؤ مسبقا بالأخطاء التي يمكن أن ترد و بالتّالي 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 تحض 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

والبرامـج الُمعتمـدة في تصميمهـا.

مزايا المقاربة الإحصائية للترجمة الآلية

- الاستعمال الناجع للموارد البشرية والبيانية.

- ضخامـة المدونـات الموازيـة في شـكل مقـروء آليـا فضـلا عـن البيانـات الأحاديـة اللغـة.

- تعميـم تطبيـق هـذا النمـط عـلى جميـع اللغـات وعـدم تصميمـه لـزوج معـيّن مـن اللغات.

- الحصـول عـلى ترجمـة أكـثر طلاقـة وسلاسـة بفضـل اسـتعمال نمـوذج اللّغـة القائـم عـلى نظريـة توزيـع الاحتـمالات بحيـث أنّـه يوفّـر السـياق للتّمييـز بـين الكلـمات والعبـارات التـي تبـدو مشـابهة. تسـتخدم نماذج اللّغـة في اسـتخراج المعلومات باسـتخدام نمـوذج الاحتـمالات المطلوبـة بصفـة أكـثر.

باسـتثناء برمجيـة «غوغـل» الموصوفـة أعـلاه لأغـراض هـذه الدراسـة، هنـاك برمجيـة أخـرى تعتمـد الترجمـة القائمـة عـلى المقاربـة الإحصائيـة وتتمثّـل في نظـام بينـغ (Bing) للترجمـة وهـو محـرّك بحـث حديـث النشـأة والمعـروف سـابقا بتسـميات مثـل لايـف سـورتش (Live Search) وويـنـدووز لايـف سـورتش (Windows Live Search) وأم أس أن سـورتش (MSN Search) هـو نظـام طوّرتـه شركـة مايكروسـفت (Microsoft) وعرضتـه عـلى الاسـتعمال العـام في الفاتـح جـوان مـن عـام 2009. إنـه نظـام يعتمـد عـلى المقاربـة الإحصائيـة ويسـعى لمنافسـة نظـام «غوغـل».

بـين هـذه المقاربـة وتلـك، تظـل الترجمـة الآليـة مـن اللغـة العربيـة وإليهـا تعـاني ضعـف النتـاج بالمقارنـة مـع اللغـات الأخـرى ومـن الأرجـح أن ذلـك راجـع أساسـا إلى محـاولات تطبيـق المقاربـات المسـتهدِفة للغـات الغربيـة عـلى اللغـة العربيـة. بالتـالي، مـا هـي المقاربـة الأنسـب التـي يمكـن تطويرهـا لتصميـم وحـدة نمطيـة للترجمـة الآليـة العربيـة؟

لـكي تتّضـح الرؤيـة حـول هـذه القضيـة، ارتأينـا تقديـم النتائـج التـي خلصنـا إليهـا في

إطار تقييم ومقارنة أداء ترجمة منظومة «الوافي الذهبي» القائمة على قواعد اللغة، مما يتبيّن من اعتماده على نمطية ترجمة آلية تشمل أدوات تحليل لغوية عميقة وقواميس متعددة المجالات، وأداء ترجمة «غوغل» القائمة على المقاربة الإحصائية[10] 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

الإملائيـة والنّحويـة ظاهـرة مُنتـشرة بكـثرة أثنـاء الكتابـة باللغـة العربيـة دون غيرهـا مـن اللغـات الأخرى.

لقـد عمدنـا إلى تحديـد أنـواع الأخطـاء الـواردة في نتـاج ترجمـة برمجيتـي «الـوافي الذهبـي» و«غوغـل» مـن خـلال تحليـل عميـق لاسـتخلاص نسـب النجـاح المئويـة للبرمجيتـين والتّرجيـح لإحداهـما مـع تمثيـل النّسـب المحصـل عليهـا في جـداول ورسـوم بيانيـة.

النموذج المُقترح: النص الانجليزي الأصلي[11]

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

الإستخبارات الاصطناعية

الإسـتخبارات الاصطناعيـة (AI) منطقـة علـم الحاسـبات تركّـز عـلى خلـق المكائـن التـي يمكـن أن تشـغل عـلى السـلوك ذلـك البـشر يعتـبر ذكي.

القـدرة لخلـق مكائـن ذكيـة فتنـت بـشرا منـذ أوقـات قديمـة و اليـوم بوصـول الحاسـوب و50 سـنة مـن البحـث إلى تقنيـات برمجـة الإسـتخبارات الاصطناعيـة، حلـم المكائـن الذكيـة **يصبـح** a حقيقـة.

تحديد أخطاء الترجمة الآلية بمنظومة الوافي الذهبي لعام[12] 2015

الأخطاء التركيبية واللّغوية والإملائية

ـ الخطـأ الأول: العنـوان: ... الإسـتخبارات الاصطناعيـة: ... كتابـة خاطئـة للمبتـدأ «الإسـتخبارات» باسـتخدام همـزة القطـع. يتعـيّن كتابتهـا بهمـزة الوصـل.

ـ الخطـأ الثـاني: الفقـرة 01 الجملـة 01:.... الإسـتخبارات الاصطناعيـة...: ورود نفـس

الخطأ الإملائي السابق.

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

ـ الخطأ الرابع : الفقرة 02 الجملة 01:...**القدرة لخلق**. حرف الجر «اللام» غير مناسب والصواب تعويضه بالحرف «على».

ـ الخطأ الخامس: الفقرة 02 الجملة 01:... **البحث إلى تقنيات**. حرف الجر «إلى» غير مناسب والصواب تعويضه بحرف الجر «في».

ـ الخطأ السادس: الفقرة 02 الجملة 02: ... **الإستخبارات** ....تكرار نفس الخطأ الإملائي.

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

الأخطاء النحوية

ـ الخطأ الوحيد: الفقرة 01 الجملة 01:...**خلق المكائن** ...: الصواب في التنكير بدل التعريف.

الأخطاء الصرفية

ـ الخطأ الوحيد: الفقرة 01 الجملة 01: ... ـ **يصبح** ...: -الصواب تصريف الفعل في الماضي-.

الأخطاء المعجمية والدلالية

ـ الخطأ الأول: العنوان: ... **الإستخبارات الاصطناعية** ...: وردت الترجمة خاطئة سياقيا ودلاليا والصواب هو اختيار العبارة «الذكاء الاصطناعي».

ـ الخطأ الثاني: الفقرة 01 الجملة 01: ... **الإستخبارات الاصطناعية ...:** تكرار نفس الخطأ السابق.

ـ الخطأ الثالث: الفقرة 01 الجملة 01: ... **منطقة** علم الحاسبات ...: الكلمة «منطقة» غير مناسبة سياقيا ودلاليا ويتعيّن استبدالها بكلمة «مجال».

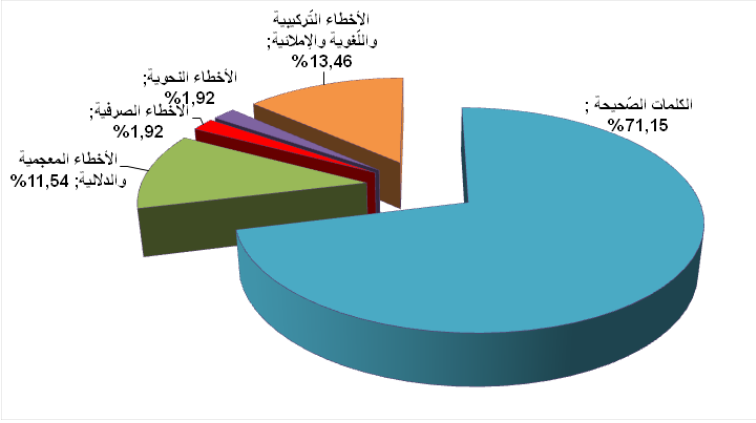
ـ الخطأ الرابع: الفقرة 01 الجملة 01: ... **خلق** المكائن ...: الاسم «خلق» غير مناسب سياقيا ودلاليا ويتعيّن استبداله بكلمة «ابتكار».

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 و دلاليا و يتعيّن استبدالها بكلمة «بظهور».

الجدول رقم 1: تقييم أداء منظومة الوافي الذهبي[13] 2015

	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



المخطط رقم 1: تمثيل نسب أخطاء الوافي الذهبي لعام[14] 2015

ترجمة «غوغل» لعام 2015 [15]

الذكاء الاصطناعي

الذكاء الاصطناعي (AI) هو مجال علوم الكمبيوتر **مع التركيز** على الأجهزة **التي يمكنها الانخراط** في السلوكيات **بأن (تنظر البشر ذكاء).**

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

الأخطاء التركيبية واللّغوية والإملائية[16]

ـ الخطأ الأول: الفقرة 01 الجملة 01: .... **مع التركيز** الصواب استعمال تقنية استبدال الاسم بالفعل على النحو التالي: التي تركّز.

ـ الخطأ الثاني: الفقرة 01 الجملة01: ... **التي يمكن أن خلق الانخراط ...:** العبارة بحاجة إلى إعادة صياغة.

ـ الخطأ الثالث: الفقرة 01 الجملة 02: ... **بأن تنظر البشر ذكاء ...:** العبارة بحاجة إلى إعادة صياغة.

ـ الخطأ الرابع: الفقرة 02 الجملة 01: ... **لها البشر مفتون** ...: العبارة بحاجة إلى إعادة صياغة.

الأخطاء النحوية

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

الأخطاء الصرفية

الخطأ الوحيد: الفقرة 02 الجملة 03: **أصبحت** ....: يجدر التذكير بدل التأنيث.

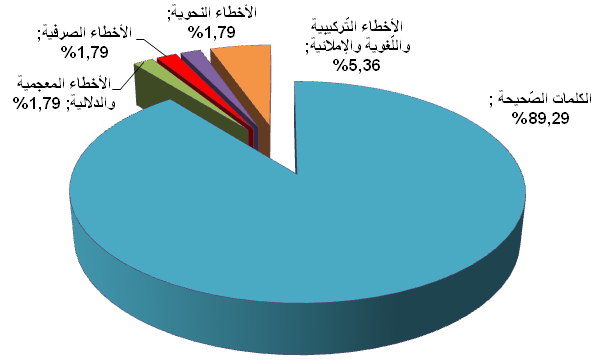
الأخطاء المعجمية والدلالية

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

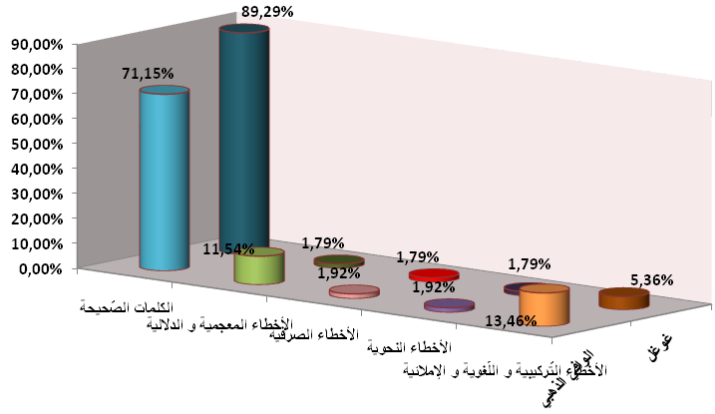
الجدول رقم 2: تقييم أداء منظومة غوغل[17] 2015

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

المخطط رقم 2: تمثيل نسب الأخطاء الواردة في ترجمة غوغل لعام 2015[18]



المخطط رقم 3: تمثيل مقارن للنسب المئوية لأداء منظومة الوافي لعام 2015 ومنظومة غوغل[19] **2015**



الترجمة النموذجية للنموذج الخامس

الذكاء الإصطناعي

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 اللغة مِن الإنجليزية 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 تفوق 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

بحصرهـا في الجـدول أدنـاه:

الجدول رقم 3: مقارنة بين المقاربة اللغوية و المقاربة الإحصائية[20]

المقاربـة اللغوية (الوافي الذهبي)	المقاربـة الإحصائية (غوغل)
هي مقاربة قائمة على قواعد اللغة	هـي مقاربـة قائمـة عـلى المتـون الموازيـة أي ترجـمات بشريـة موازيـة وتتجاهـل القواعـد اللغويـة.
تطبـق قواعـد اللغـة عـلى مراحـل التحليـل والتحويـل والتوليد.	تعتمد قوانين الاحتمالات.
هـي بالتـالي تولّـد نصوصـا تقـل فيهـا الأخطـاء اللغويـة عـلى العمـوم ولكنهـا تتضمـن مشـاكل دلاليـة وسـياقية.	توفر أداء ذا جودة من حيث قلة الأخطاء الدلالية.
يسـتلزم بنـاء منظومـة قائمـة عـلى قواعـد اللغـة تمثيـل قواعـد لغويـة واسـعة النطـاق و جهـد بـشري، خاصـة عندمـا لا تنتمـي اللغتـان، المصـدر و الهـدف إلى نفـس عائلـة اللغـات.	لا يسـتلزم بنـاء منظومـة إحصائيـة عـلى قواعـد اللغـة ولا عـلى الجهـد البـشري لتمثيـل تلـك القواعـد وتحديدهـا.
الاعتـماد عـلى قواميـس ثنائيـة اللغـة و عـلى الأدوات اللغويـة (المحلـل الـصرفي، المدقـق الإمـلائي، المحلـل الإعرابي...الـخ).	لا تُسـتخدم القواميـس ولا الأدوات اللغويـة في هـذه المقاربـة.
لا تعتمـد هـذه المقاربـة عـلى ترجـمات سـابقة ولا عـلى متـون مُخزّنـة.	تعتمـد هـذه المقاربـة عـلى ترجـمات سـابقة مـن خـلال متـون كبـيرة الحجـم.
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.
تواجـه المقاربـة معضـلات اللبـس عـلى اختـلاف أنواعـه.	لا تواجه المقاربة معضلات اللبس في أغلب الأحيان.
يكـون النتـاج ذا جـودة كلـما كانـت الجمـل قصـيرة في النـص.	يكـون النتـاج ذا جـودة كلـما كانـت الجمـل قصـيرة في النـص.
يبقى الأداء ضعيفا مقارنة بالترجمة البشرية.	يبقـى الأداء ضعيفـا في غيـاب المتـون ذات الصلـة بالنـص المترجـم.

عـلى الرغـم مـن مواطـن القصـور التـي تمـت معاينتهـا في برمجيـات ترجمـة النصـوص عـلى اختـلاف مقارباتهـا، إلا أنّهـا أدوات لا يمكـن الاسـتغناء عنهـا في المجـالات التبليغيـة وفي الترجمـة التفسـيرية، كـما أنهـا تشـكل أداة ثمينـة لربـح الوقـت وتوفـير العنـاء عـلى

المترجمين المحترفين في البحث في القواميس الورقية في 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 لما يتعلّق الأمر بالتعامل مع اللغة العربية لُغةً مُترجمًا 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.

خـــاتمة

لمّا كانت جودة أي منظومة آلية ترتكز على تحليل الأخطاء الناتجة عن تلك المعالجة التي توحي بالأساليب المعتمدة لتصميم برمجيات الترجمة، أظهرت دراستنا التطبيقية بأنّ تباين النتاج من برمجية إلى أخرى يبقى مرتبط 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 تؤديه في الحصول على التجانس بين أجزاء الجمل والعبارات مع احترام القواعد البنوية الخاصة بكل لغة. بالتالّي، 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 (Hybride) 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 والإستراتيجية وذلك، في مرحلة التصميم.

- البرمجية العربية بحاجة إلى عناصر بشرية لتقييم أدائها بعد مرحلة التصميم.

في هـذه الحالـة، دور المترجـم المحـترف حاسـم في تقييـم الأداء مـن منطلـق جـودة النتـاج وسرعـة الأداء وكلفتـه، وتوجيـه المصمّمـين فيـما يتعلـق بالغرض المنشـود مـن البرمجية وفي تجميـع المتـون اللازمـة لتعبئتهـا بها، في حين يتـولى عالم اللغة واللسـانيات تحديـد القواعد اللغويـة والمسـتويات الأساسـية التـي يجـب تمثيلهـا بعـد القيـام بدراسـة عمليـة مسـبقة.

- البرمجيـة النموذجيـة العربيـة تسـتلزم حسـن تحديـد الأغـراض وانتقاء المـادة اللغوية والإحصائيـة وتجنيـد المتدخلـين المؤهلـين في تصميمهـا مـن أجـل تحقيـق الأهـداف المرتقبـة.

إنّ اقـتراح برمجيـة عربيـة للترجمـة مـشروع عمـلي يمكـن تجسـيده عـلى أرض الواقـع لتوفـر البـلاد العربيـة عـلى الكفـاءات والمهـارات في مجـال المعلوماتيـة وغيرهـا مـن الشُّـعَب التـي هـي بحاجـة إلى أن تُقيـم بشـأنها المؤسسـات الجامعيـة جـسرا للتواصـل فيـما بينهـا، كـما يجـب حـث القطـاع الخـاص عـلى بـذل الجهـد والمـال في مثـل هـذه المشـاريع. ولعـلّ أبسـط المشـاريع التـي تنطـوي عـلى أهميـة قصـوى ولا تسـتدعي تجنيد مـوارد ماليـة ضخمـة، هـي تشـجيع تصميـم قواميـس ومعاجـم الكترونيـة مفتوحـة المصـدر ومتخصصـة المجـالات، بحيـث تسـعى كل أسرة جامعيـة إلى بنـاء مـسرد رقمـي خـاص بمختلـف تخصّصاتهـا لتعريـب العلـوم والمعـارف تقـوم مقـام بنـوك معطيـات.

تكمـن فائـدة القواميـس الإلكترونيـة المفتوحـة المصـدر في توفـير المـادة الأساسـية للمجـال المُختـص وفتـح المجـال للمختصـين وغيرهـم في ميـدان معـيّن لتعبئـة القامـوس بصفـة مسـتمرة وتحديثـه والرجـوع إليـه عنـد الحاجـة وخـلال إعـداد البحـوث باللغـة العربيـة، كـما يمكـن اسـتغلال محتـوى القامـوس كـمادة لغويـة لتعبئـة برمجيـات الترجمـة.

تجـدر الإشـارة إلى أنّ الاتجـاه إلى اعتـماد مقاربـات هجينـة، أصبـح محـلّ دراسـة لـدى الـشركات العملاقـة المُصمّمـة لبرامـج الحاسـوب وبرمجيـات الترجمـة. في هـذا المجـال، نذكـر جهـود شركـة «غوغـل» في رعايـة بحـوث أكاديميـة تُقـدّم اقتراحـات حـول تحسـين النظـم الإحصائيـة وإرفاقهـا بوحـدات نمطيـة لغويـة. في هـذا الشـأن، نحـن ننـوّه بجهـود

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[21] باللغة الإنجليزية 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 الإنجليزية 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.

الإحـالات

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- مـن مقـال: «إعـادة الاعتبـار للغـة العربيـة في المجتمـع العـربي»، نـشر في العـدد الخـاص لكتـاب المجلـس الأعـلى للغـة العربيـة الموسـوم «العربيـة الراهـن والمأمـول»، 2009، ص 65.

4- Summer Institute for Linguistics (SIL) Ethnologue Survey (1999) availablefrom : http://populationsdumonde.com/classements/les-10-langues-les-plus-parlees-dans-le-monde

5- George Weber, (1999). «the Word's 10 Most Influential Languages». AATF National Bulletin, Vol. 24, N°3.

available from: http: //french.server276.com/bulletin/articles/promote/advocacy/useful/toplanguages.pdf. accessed July25th, 2017.

6- http://ieeexplore.ieee.org/document/4631625/?reload=true. Accessed August 18th, 2017.

7- https://aawsat.com/home/writer/mirza-alkhuwaldi. accessed May 14th, 2012.

8- برمجيـة متوفّـرة عـلى الإنترنـت للتحميـل مـن أجـل ترتيـب المدوّنـات الموازيـة ترتيبـا خطيـا وهـو يفيـد في إعـداد القواميـس الثنائيـة اللّغـة. تسـتخدم شركـة آي بي أم هـذه المنظومـة عـلى نطـاق واسـع. [8] GIZA++

9- يرتبـط اسـتعمال ذاكـرة الترجمـة بأسـلوب الترجمـة بمسـاعدة الحاسـوب خلافـا للترجمـة الآليـة الصرفـة التـي تعتمـد عـلى قواميـس إلكترونيـة ضخمـة وخوارزميـات رياضيـة. تتمثّـل ذاكـرة الترجمـة في تخزيـن ترجـمات سـابقة في شـكل نصـوص متوازيـة ثنائيـة اللغـة أو متعـددة اللغـة يتـم إحضارهـا وتظهـر في شـكل خيـارات ينتقـي منهـا المترجـم الترجمـة الأنسـب والأقـرب مـن السـياق الـذي هـو بصـدد الترجمـة فيـه (شرح الباحثـة). تعـرّف مجموعـة EAGLES (Expert Advisory Group on Language Engineering Standards) ـ مجموعـة الخـبراء المستشـارين حـول معايـير هندسـة اللغـة ـ ذاكـرة الترجمـة عـلى أنّهـا «أرشـيف مـن نصـوص متعـددة اللغـات (مجـزأة، ومصفّفـة، ومحلّلـة ومصنّفـة) قابلـة للتّخزيـن ولاسـترجاعها أو مقاطـع منهـا في لغـات متعـدّدة وفقـا لـشروط بحـث مختلفـة».

10- http://www.statisticsviews.com /details /feature /8065581 /The-statistics-behind-Google-Translate.html. accessedJuly 25th, 2017.

11- المصدر السابق، ص 238.

12- المصدر السابق، ص 249.

13- المصدر السابق،جدول 21 نموذج 5، ص 283.

14- المصدر السابق، رسم بياني دائري 19 نموذج 5، ص 283.

15- المصدر السابق، ص284.

16- المصدر السابق، ص 284.

17- المصدر السابق، جدول 22 نموذج 5 : ص 285.

18- المصدر السابق، رسم بياني دائري 20 نموذج 5: ص 286.

19- المصدر السابق، رسم بياني 21 نموذج 5: ص 286

20- المصدر السابق، جدول 25: ص 298.

21- https:// pdfs.semanticscholar.org/ 709e/ 70d0b6d1f00ff35caa1a7af8fed58e742636. ccessed April 24th, 2015.

المصادر والمراجع

أ- مراجع باللغة العربية:

- القحطاني، سعد بن هادي. «**تحليل اللغة العربية بواسطة الحاسب الآلي**». مجلة علوم اللغة. القاهرة: دار غريب، م 5، ع 3، ص 822.

- بربارة، سهيلة (6002). الترجمة **بمساعدة الحاسوب من الانجليزية إلى العربية: دراسة تحليلية ونقدية لنصوص مترجمة آليا.** الجزائر: مذكرة لنيل شهادة الماجستير، قسم الترجمة، جامعة الجزائر 2.

- بربارة، سهيلة (7102). **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.

- خطاب، شريف (01102). **المعالجة الآلية للغة العربية، أساسيات الحاسوب.** القاهرة: بحث، قسم علوم الحاسب، كلية الحاسبات والمعلومات، جامعة القاهرة.

- نبيل علي (8991). **اللغة العربية والحاسوب.** الكويت: مؤسسة تعريب الكويت.

مواقع ويب

موقع المسبار لشركة آي- تي - آي: http://www.almisbar.com/

موقع شركة «صخر»: http://www.sakhr.com/index.php/en/

ب- مراجع باللغات الأجنبية:

- Alawneh, M. F., Sembok, T. M. (2011). Rule-based and example-based Machine Translation from English to Arabic. In: Proceedings - 2011 6th International Conference on Bio-Inspired Computing: Theories and Applications, BIC-TA 2011 (pp. 343-347). DOI: 10.1109/BIC-TA.2011.76

- Callison-Burch, Chris (2007) .Machine translation: Word-based models and the EM algorithm.Slides borrowed from Philipp Koehn: John Hopkins

University December 3 https://www.cs.jhu.edu/~jason/665/PowerPoint/lect32a-mt-word-based-models.pdf. Accessed November 14, 2009.

- Costa-Jussa Marta R., Mireia Farrus, Marino Jose B., Fonollosa Jose A. R. (2012) .Study and comparison of rule- based and statistical Catalan- Spanish Machine Translation Systems.Computing and Informatics, Vol. 31, N°2, pp.245-270.
- Costa-jussa, Marta R, Fonollosa José A.R. (2015). Latest trends in hybrid machine translation and its applications. Computer speech and language, volume 32, issue 1, PP 3- 10.
- Costa-jussa,Marta R and al (2016).Hybrid Approaches to Machine Translation. eBook ISBN: 978-3-319-21311-8, pp 67- 75.
- 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-Translate.html.Accessed January 10, 2016.
- Hutchins, W.J.(1986). Machine Translation: Past, Present, Future: Ellis Horwood Limited, West Sussex <http:ourworld.compuserve.com/homepages/WJ Hutchins/PPF_3.pdf) (assessed 29 December 2005).
- Limarez, Jesus Angel Gimenez (2008). Empirical Machine Translation and its Evaluation. Barcelona: Tesi Doctoral per aoptar al graude Doctor en Inform`atica. Programa de Doctorat en Intellig`encia Artificial, Departament de Llenguatges i SistemesInform`atics, UniversitatPolit`ecnica de Catalunya. Available from: http: //nlp.lsi.upc.edu/papers/gimenez08-thesis.pdf. Retrieved November 13, 2012.

- 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.
- Shannon, C. E., Weaver, W. (1964). The mathematical theory of communication.Tenth printing: The University of IllinoisPress, Urbana. http://www.magmamater.cl/MatheComm.pdf. Retrieved June 12, 2015.
- Soudi A., Farghaly A., Neumann G., Zbib R. (2012). Challenges for Arabic Machine Translation. Natural Language Processing 9, 157pp., John Benjamins Publishing Company. DOI: 10.1075/nlp.9