اللغة العربية والترجمة الآلية
المشاكل والحلول

أ.د. محمد زكي خضر
الجامعة الأردنية – الأردن

مقدمة

الترجمة هي نقل معاني نص من لغة إلى لغة أخرى مع مراعاة الدقة والأسلوب. ويتطلب ذلك فهم النص الأصلي والتعبير عن المحتوى والأسلوب بلغة أخرى. فالمترجم يجب أن يتقن اللغتين: المترجم منها والمترجم إليها.

هناك طريقتان معروفتان في الترجمة. أولهما تعتمد الترجمة الحرفية والالتزام بمعاني مفردات النص الأصلي ونقلها إلى اللغة الثانية، والطريقة الثانية تعتمد على فهم المعنى العام ثم التعبير عنه باللغة الثانية بأسلوب المترجم نفسه.

المترجم يتعرف الرموز المكتوبة في الترجمة الكتابية، والأصوات المنطوقة في الترجمة الشفوية، أي يقرأ الرموز الكتابية للغة التي يترجم منها إذا كان النص مكتوباً أو يتعرف أصوات اللغة التي يستمع إليها إن كانت ترجمة شفوية. وهنا نذكر أن قراءة اللغة العربية المكتوبة بحروف غير مشكولة يستوجب معرفةً من القارئ لكي يقوم هو بالشكْل الصحيح دون أن تبصر عيناه علامات الشكْل. وهو بذلك يستعمل خبراته التي تعلمها وتدرب عليها في ماضي حياته من قواعد اللغة لكي يستعملها بسرعة، فينصب المفعول ويرفع الفاعل وربما يفعل ذلك بالسليقة وهو لا يعرف من قواعد الإعراب اللغة العربية شيئاً. أما المترجم للكلام

المنطوق فهو يستطيع تمييز الجملة التي يترجمها إن كانت استفهامية أو خبرية أو تعجيبة دون وجود علامات استفهام أو تعجب. بعد ذلك يرجع المترجم إلى الوحدات المعجمية وهي الكلمات والتعبيرات الاصطلاحية ويفهم معانيها في سياقاتها اللغوية والاجتماعية المختلفة، وهنا يجب الإلحاح على السياقات المختلفة، لأن الكلمة الواحدة تعني أشياء كثيرة طبقاً للسياق الذي ترد فيه. هناك إذن مشكلة معاني الكلمات في السياقات المختلفة (أي تعدد المعاني)، إضافة إلى المشتركات اللفظية (الكلمات التي تتشابه في كتابتها أو نطقها وتختلف في معانيها، مثل عين الإنسان وعين الماء[1] وهناك التعبيرات الاصطلاحية. وهذه كما سنرى تمثل مشكلة كبيرة بالنسب للترجمة عامة والترجمة الآلية خاصة.

وهنا ينبغي تحليل المفهوم والوقوف على كنهه وما يتفرع عنه من ظلال في المعنى، وخاصة المجازية منها وغيرها، وذلك حتى تتحدد العلاقة بين هذا المفهوم وفروعه وبين المفاهيم التي يمكن أن تتصل به بشكل أو بآخر، أي العلاقة بينه وبين مجموعات المفاهيم الأخرى، وذلك للوصول في النهاية إلى الجملة أو شبه الجملة. فالمفهوم الذي يرتبط بجسم واحد بعينه، أو يدل عليه، هو مفهوم إفرادي، مثل "المريخ" أما إذا ارتبط المفهوم بعدة أجسام بينها نوع من التجانس الذي يضعها في مجموعة واحدة، فإن المفهوم يكون عاماً (بمعناه الضيق هنا)، مثل "كوكب". فهذا المفهوم يدل على عدد من الأجسام التي تدور حول الشمس. ولتحديد المقصود يمكن أن يكون التعريف مركزاً أو موسعاً. وعلى ذلك فإن دقة التعريف تحدد المفهوم. والتعريف قد يكون شاملاً في بعض الأحيان وفي غالب الأحيان لا يخلو من النقص أو أن يكون تعريفاً بالضد أو ما يعرف بالتعريف السلبي أو أن يكون فضفاضاً أو ضيقاً أو يأخذ بمبدأ الإحلال والتعويض إلى غير ذلك من الصفات التي يتصف بها التعريف. وهذا التعريف قد يكون واضحاً في ذهن المترجم وقد يكون مشوشاً فيتصرف في ترجمته لهذا

[1] د. محمود اسماعيل صالحي (الصيني)، أستاذ اللغويات التطبيقية، مدير مركز الترجمة بجامعة الملك سعود (سابقاً) **الحاسوب في خدمة الترجمة والتعريب**
www.emro.who.int/ahsn/meetings/sep03/day2/Dr.%20Mohammed%20Saleh.doc

المفهوم وقد يعبر عنه بمفهوم يفهمه هو ويكون قريباً أو بعيداً من المفهوم الأصلي.

ومن العمليات التي تساعد على تعُّرف المفهوم، تعرف الوحدات النحوية ووظائفها: أي تحديد الوحدات (الألفاظ والتعبيرات) ومعرفة دور كل منها في الجملة، مثل المسند والمسند إليه والتكملة. ومن أمثلة التعبيرات المضاف والمضاف إليه، والجار والمجرور والصفة والموصوف وغير ذلك. وكل من هذه التعبيرات له وظيفته النحوية، فبعضها يقوم مقام الفاعل أو المفعول به وبعضها يقوم مقام الصفة أو الظرفية إلى غير ذلك. ومن الوحدات النحوية الجمل. ولابد للمترجم أن يفهم معنى كل واحد من هذه العناصر ووظيفته اللغوية والنحوية في النص. فصيغة الأمر قد تعبر عن إلزام السامع أو رجائه أو استعطافه مثلاً. كما أن الصيغة الخبرية قد تأتي للإخبار كما تستعمل للطلب غير المباشر (كما في عبارة "أرى أن لديك قلماً زائداً")، بل وللدعاء كذلك ("رحم الله فلاناً" مثلاً).[1]

ومن الأمور المهمة في الترجمة معرفة حقل النص ومجال تخصصه. مثلاً إذا كان النص في علم الكيمياء فإنك تجد أن بعض الكلمات تختلف معانيها في هذا النص عن معانيها لو وردت في نص الفيزياء أو وردت في الإلكترونيات. أو أن كلمة وردت في حقل تجارة المواد الغذائية أو في حقل من حقول العلوم الزراعية. وهكذا.

الجانب الثاني وهو التعبير عن محتوى النص باللغة الهدف أي اللغة المترجم إليها. ويتطلب هذا معرفة بإنتاج المقابلات الصوتية أو الكتابية والنحوية والمعجمية والبلاغية. فعلى المترجم أن يعرف أن هذه التعبيرة الإسنادية مثلاً تترجم إلى كذا وهو بذلك يجب أن يكون ملماً بقواعد النطق أو الإملاء. وأصول الكتابة: قواعد النطق إن كانت الترجم شفوية في اللغة المترجم إليها، أو قواعد الإملاء وأصول الكتابة في هذه اللغة، إن كانت الترجمة تحريرية. وهو بذلك يقوم باختيار العبارات المناسبة في اللغة التي يترجم إليها التي توافق المفهوم الذي

[1] أ. د. محمد الصرايرة، قسم اللغة الإنكليزية، جامعة اليرموك، **اللغة العربية والترجمة الآلية**، محاضرة في مجمع اللغة العربية الأردني، الموسم الثقافي التاسع عشر 2001م.

توصل إليه في الخطوات السابقة.

وهنا تبرز براعة المترجم في اللغتين حيث يجب أن يتقن ما يسمى بالتعادل أو التقابل المعجمي أو الاصطلاحي، فالتعبير العربي "رجع بخفي حنين" ليس له مقابل حرفي بأي لغة، ولو ترجمت ترجمة حرفية لأصبحت مضحكة. فهو يجب أن يفهم مثل هذه العبارات ويجب أن يعرف العبارات الدارجة التي تقابلها في اللغة الثانية أو على الأقل يؤلّف/يبتكر هو تعبيراً من كلمات تلك اللغة يعبر به عن المفهوم بدقة كافية مستعملاً القواعد النحوية والصرفية في اللغة التي يترجم إليها. فالترجمة كالكتابة أو التأليف تخضع لقواعد الصحة اللغوية نفسها.

أما القواعد الأسلوبية وهي القاعدة البلاغية المعروفة (لكل مقام مقال) فنجد أن الجملة أو العبارة قد تكون صحيحة نحوياً ومعجمياً ولكنها غير مناسبة من حيث المقام. فالمترجم يجب أن يكون ذا معرفة بالقواعد الأسلوبية للغتين المترجم منها والمترجم إليها.[1]

وماذا بعد ذلك؟ لا يزال هنا مشكلة التعامل مع الجمل الطويلة. فالجمل الطويلة هي من أهم مشاكل الترجمة لأنها تحتاج إلى معالجات ذهنية معقدة وقد تشوش ذهن المتلقي بوجهٍ عام. فقد تحوي الجملة الطويلة عدةَ مفاهيم وقد يتجزأ المفهوم الواحد ليتكامل بين بدء الجملة ونهايتها. وقد يتخلل المفهوم الواحد مفاهيم ثانوية. وقد يكون المفهوم المهم مخفياً بين عدد من المفاهيم الأقل أهمية منه. وهكذا فإن الجملة الطويلة قد تستعصي على الفهم والتحليل وكذلك تستعصي على التركيب في اللغة المترجم إليها.

إن معلوماتنا عن الترجمة محدودة إلى الآن، ولا نستطيع إن نقول إن المترجم يتبع هذه الخطوات: 1، 2، 3 بل إن كل مترجم له أسلوبه الذي يتبعه في الترجمة، وأصعب من ذلك تحديد كيفية فهم النص. ماذا نفعل تماماً حينما نحاول فهم نص ما من النصوص؟ إن الطريقة التي يتعامل بها دماغ الإنسان مع النصوص ما زالت غير معروفة على وجه الدقة وتخضع لنظريات غير مؤكدة حتى الآن. كما أنه لم تتوفر حتى يومنا هذا أساليب ناجحة للتعامل مع

[1] المصدر السابق.

توفير عناصر الحس والخبرة الذاتية والمعلومات وغيرها التي يمكن أن ننقلها بشكل من الأشكال إلى الحاسوب في الترجمة الآلية التي سنتناولها فيما بعد. لذلك يمثل "فهم النص" مشكلة كبرى في الترجمة.

لماذا الترجمة الآلية؟

السبب الأول والأهم هو وجود كمية هائلة مما يجب ترجمته، بحيث لا يكفي عدد المترجمين من البشر للقيام بجزء يسير منه، خاصة بعد تفجر ثورة المعلومات وتنوع اللغات التي تنتج المعارف اليوم وينبغي أن يعرفه مَن لا يتكلمون تلك اللغة. يزيد عدد اللغات الحية اليوم على 4000 لغة في كافة أنحاء العالم. وتأتي اللغة العربية ضمن اللغات العشر الأولى منها إذا ما أخذنا بعين الاعتبار عدد الناطقين بها. وبرغم التأريخ الحافل للغة العربية فإنها اليوم ليست المصدر الأهم في العلوم والتقنية الحديثة. لذلك فإن عملية الترجمة من اللغات الأخرى إلى اللغة العربية ذات أهمية بالغة بالنسبة للناطقين بالعربية. فإذا ما أرادوا الاطلاع على آخر ما توصل إليه العلم فإما أن يتعلموا لغة (أو لغات) أخرى بجانب العربية، أو أن يترجم لهم ما يصدر من علوم في اللغات الأخرى. وما نشهده اليوم من إقبال على تعلم اللغة الإنكليزية والتدريس بها في الجامعات والمدارس في الوطن العربي ما هو إلا نتيجة للتخلف في الترجمة إلى اللغة العربية.

أبرزَ تقرير التنمية الإنسانية العربية الأول لعام 2002 ضعف حركة الترجمة في الوطن العربي، بل ضمورها الشديد، مما يكاد يقضي على دورها المأمول في نقل المعرفة وتوطينها باللغة العربية. وقد كشفت الإحصاءات في هذا الصدد أن ما يترجمه العالم العربي من كتب قد لا يزيد على خُمس ما يترجمه بلد أوربي صغير مثل اليونان الذي يقل عدد سكانه عن 5% من عدد سكان الوطن العربي. وتتعاظم أهمية الترجمة العلمية يوماً بعد يوم نتيجة الانفجار المعرفي وتزداد هذه الأهمية بالنسبة لعالمنا العربي لكونه – أساساً – متلقياً للمعرفة أكثر منه

منتجاً لها. وتعثر الترجمة يعد سبباً رئيسياً وراء تعثر تعريب التعليم الجامعي.[1]

لمقدم هذا البحث تجربة في ترجمة أحد الكتب العلمية التي تدرس في أقسام الهندسة الكهربائية في كافة أنحاء العالم منذ السبعينيات في القرن الماضي. فبعد قضاء أكثر من عام في ترجمة الطبعة الثانية من الكتاب ظهرت الطبعة الثالثة فأعيدت الترجمة من البداية وأدى ذلك إلى تأخر ظهور الترجمة العربية أكثر من عامين آخرين. منذ ذلك الحين وحتى اليوم وصلت طبعات الكتاب إلى الطبعة السابعة وهو يحتل مركز الصدارة في الكتب المقررة في الجامعات العالمية لتلك المادة في أقسام الهندسة الكهربائية منذ ما يقرب من أربعة عقود، لكن لم تصدر ترجمة عربية لأي طبعة لاحقة حتى الآن، ولم تستعمل الترجمة العربية إلا في العراق. ولو قدر لهذا الكتاب وحده أن تستمر ترجمات طبعاته وأن تدرس مادته باللغة العربية في الجامعات العربية، لاستعملت منه مئات الآلاف من النسخ على مستوى العالم العربي ولكان أساساً مهماً لتعريب التعليم في أقسام الهندسة الكهربائية. وهذا يدل على ضرورة العمل المؤسسي المستمر للترجمة بوجه عام وضرورة الترجمة الآلية بوجهٍ خاص، والحاجة إلى تكامل الجهود بين الجامعات العربية في مجال الترجمة.

والسبب الثاني في أهمية الترجمة الآلية هو أن عملية الترجمة للمترجمين من البشر عملية مملة وبطيئة والمترجم يحاول التغلب على مله بتغيير أسلوبه تارة وبالراحة تارة. وهو كطبيعة البشر ينام ويلهو ويمرض ويغير عمله ويتقاعد. كل ذلك يحدد عمل المترجمين ويجعلهم بضاعة نادرة في عصر العولمة. وإذا علمنا أن المترجم غالباً ما يتقن لغة واحدة مع اللغة الأم التي يترجم منها أو إليها فيعني ذلك ندرة شديدة فيمن يتقن لغة كاللغة الكورية أو اليابانية مع العربية. وهذا يعطي أهمية إضافية للترجمة الآلية، فالحاسوب يمكن أن يعمل 24 ساعة في اليوم لا يأخذ إجازة نهاية الأسبوع ويمكن إبداله بما هو أفضل منه وتحسين أدائه وسرعته مع التقدم التقني الجاري اليوم. وهناك أسباب أخرى تقف إلى جانب الترجمة الآلية

[1] د. نبيل علي، **الفجوة الرقمية**، عالم المعرفة 318، أغسطس 2005.

منها حاجة القطاع التجاري إلى ترجمة تعطي فكرة لا بأس بها عن المنتجات الصناعية والتجارية دون أن تكون على درجة عالية من الرصانة، مع ازدياد حجم مثل هذه المعلومات المطلوب ترجمتها وتوسعها لتشمل لغات جديدة في عصر العولمة.

إن الحاسوب لديه قدرة جزئية في هذا الجانب لا شك. فكما هو معروف أن الحاسوب يعطيك ما تعطيه أولاً، حيث إنه إنما يعطيك ما تسجله في ذاكرته أو تعلمه على فعله. ومن ثم كلما كان التحليل اللغوي وكانت القواعد اللغوية المقدمة للحاسوب والمخزونة في ذاكرته دقيقة أدى ذلك إلى ترجمات أقرب إلى الصواب.[1]

قابلية الحاسوب في خدمة الترجمة

التعادل أو التقابل المعجمي والاصطلاحي يعتمد على ما نخزن في ذاكرة الحاسوب. وكذلك معرفة القواعد النحوية تعتمد على المجهود البشري، كما تعتمد معرفة القواعد الأسلوبية على الجانب البشري، أي على المعلومات والبرامج والإمكانات التي يوفرها مصممو البرنامج الحاسوبي، فالحاسوب لديه القدرة الجيدة على التعامل مع هذه الأمور بسهولة كبيرة.

وهنا ينبغي أن نناقش إمكان الآلة القيامَ ببعض العمليات التي يقوم بها المترجم من البشر، وسنعرض فيما يأتي لبعض منها:

1. معرفة الموضوع العام للمادة المطلوب ترجمتها: هناك الكثير من الأبحاث التي أدت إلى تطوير في برامج تعرف الموضوع العام للمادة المراد ترجمتها أو أخذ فكرة عنها أو اختصارها. فمثلاً يمكن للآلة استعمال دلائل تؤدي إلى إن نعرف: هل المادة سياسية أو اقتصادية أو علمية، وإن كانت علمية فهل هي في موضوع رياضي أو زراعي أو كيميائي مثلاً. إن تحديد موضوع المادة يساعد على تحديد المصطلحات المستعملة في ذلك العلم

[1] د. محمود اسماعيل صالحي (الصيني)، أستاذ اللغويات التطبيقية، مدير مركز الترجمة بجامعة الملك سعود (سابقاً) **الحاسوب في خدمة الترجمة والتعريب**
www.emro.who.int/ahsn/meetings/sep03/day2/Dr.%20Mohammed%20Saleh.doc

ومن ثَم الوصول إلى ترجمة أكثر دقة وتحديداً. وبالطبع فإن دقة تعريف موضوع المقال أو البحث أو الفقرة المراد ترجمتها يعتمد على المصطلحات أو ما يسمى بالكلمات المفتاحية التي تحويها، كما يعتمد على كيفية التعامل معها أو ما يسمى (بمصطلح الحاسوب) الخوارزميات المستعملة في تحديد الموضوع. وعلى ذلك فإن برنامجاً قد يكون أكفأ من برنامج آخر وفق الأسس التي اعتمدها واضعو ذلك البرنامج بالمقارنة بالبرنامج الآخر.

2. استعمال المعاجم اللغوية بمقابلة كلمة مقابل كلمة ضمن الموضوع العام للمادة المطلوب ترجمتها. إن المعاجم قد تكون معاجم بلغة واحدة أو بأكثر من لغة. فالمعاجم بلغة واحدة تحدد معنى الكلمة في نطاق معناها بتلك اللغة وحقول استعمالاتها للوصول إلى المفهوم المستعملة فيه تلك الكلمة. أما المعاجم بلغتين أو أكثر فهي تعطي ما يقابل الكلمة من كلمات في اللغة الثانية أو اللغات الأخرى. ويلاحظ هنا كثيراً من التعقيد حيث هناك الكلمة يقابلها كلمة والكلمة مقابلها كلمتان أو أكثر وقد يكون المفهوم المتكون من أكثر من كلمة يقابله كلمة واحدة في اللغة الأخرى. وهناك الكلمة التي يصعب إيجاد مقابل لها في اللغة الأخرى. كل ذلك يبيّن محدودية قابليات المعاجم المستعملة في خدمة الترجمة.
3. تفكيك الكلمة إلى مكوناتها فمثلاً في اللغة العربية يمكن تفكيك الكلمة إلى مكوناتها من سوابق ولواحق كأحرف العطف وحروف الجر وعلامات الإعراب والضمائر المتصلة وغيرها وتطبيق قواعد الصرف على الكلمات ومعرفة وزنها الصرفي وفيما إذا كانت مصدراً أو اسم آلة أو غيرها. ولكن غياب الضبط بالشكل يجعل المهمة أصعب، ووجود أكثر من احتمال للتفكيك أو الشكْل يجعل المهمة أعقد. ففي بعض الأحيان يصعب التمييز بين أن يكون الحرف في الكلمة حرفاً زائداً كحرف جر أو حرف عطف أو أن يكون حرفاً أصلياً فمثلاً كلمة "أهلك" هل هي فعل ماض من الهلاك أم هي للمخاطب من الأهل؟
4. تحديد المفهوم العام لمعنى الجملة المراد ترجمتها ومحاولة إعادة صياغته لكي يكون مناسباً للغة المراد الترجمة إليها. فقد أجريت أبحاث كثيرة لتحديد المفاهيم العامة في الحياة على مجموعات تتشعب منها مجموعات ثانوية وأخرى تتشعب عنها وفق تسلسل منطقي

لتصنيف الأشياء الملموسة أو المعنوية. فمثلاً التفاحة هي ثمرة والثمرة هي جزء من شجرة والشجرة نوع من النبات. وهكذا تصنف الأشياء تصنيفاً وصفياً دقيقاً. وقد استعملت مثل هذه المفاهيم في الترجمة بين اللغات بحيث يكون هناك لغة وسيطة تترجم إليها المفاهيم من أية لغة ثم تترجم تلك المفاهيم من اللغة الوسيطة إلى اللغة المراد الترجمة إليها. وقد توسعت هذه الأفكار وخاصة في جامعة الأمم المتحدة في اليابان. إلا أن نجاح هذه الطريقة لا يزال محدوداً لصعوبة تمثيل المفاهيم بدقة خاصة في الأمور المعنوية والأدبية.

5. تركيب الجملة باللغة المراد الترجمة إليها وفق قواعد تلك اللغة. وهذه العملية تستعمل قوانين النحو والصرف والمعاجم لتركيب الجملة المراد تكوينها وفق المفهوم أو المعنى المستنتج مع خطوات سابقة وتخضع هذه الخطوات لاجتهادات اللغويين وواضعي البرامج الخاصة بذلك.

الحاسوب واستخداماته المختلفة في الترجمة

الحاسوب باستعمالاته اليوم وبالمستوى الذي وصلت إليه عملية الترجمة الآلية لا يعوض عن المترجمين من البشر. وسوف تمضي بضعة عقود قبل أن تصل الترجمة الآلية إلى المستوى الذي يضاهي المترجمين الجيدين. لكن الحاسوب اليوم بإمكانه أن يقدم عوناً لا بأس به للمترجم. ومن هذه المعونة ما يأتي:

آ. **الترجمة الآلية مع تحرير لاحق**، أي مراجعة بشرية بعد الترجمة الآلية. فالترجمة الآلية البدائية التي تعتمد على ترجمة معاني الكلمات وصياغة الترجمة بلغة ركيكة يمكن أن تكون بداية للمترجم لكي يقوم بإعادة صياغة الجمل وتنقيح المعاني ووضع الترجمة على وجهٍ مقبول ومفهوم. وتجدر الإشارة إلى أنه إذا كانت الترجمة الآلية سيئة جداً فربما يكون من الأسهل على المترجم أن يقوم بالترجمة بنفسه من جديد بدل تنقيح نص سيِّئ جداً.

ب. **الترجمة مع التحرير السابق**، بمعنى أن الإنسان يحرر النص المراد ترجمته. مثلاً يبسط الجمل المعقدة. والكلمات التي لها معان كثيرة يحدد معناها المطلوب وهكذا، أي إننا نعدل

النص بحيث يستطيع أن "يفهمه" الحاسوب، وتسمى هذه اللغة المقبولة للآلة Machine Acceptable Language. ويشبه ذلك التحاور مع الحاسوب بلغات البرمجة التي تتضمن كلمات محدودة بصيغ محددة لا يجوز التصدي لها.

ج. هناك نوع ثالث يسمى **بالترجمة التحاورية** interactive وهي مثال للتعاون بين الحاسوب والمترجم البشري، وذلك بأن يكون برنامج الترجمة ذا إمكانية حوارية بأن يعطي الترجمة جملة جملة ويتوقع من المترجم أن يوافق أو يعدل بعض أجزائها لكي يصل إلى الترجمة المقبولة.

د. **الترجمة البشرية بمعاونة الآلة**، أي إن الإنسان يترجم والآلة تعاونه في هذه العملية. وهذا عكس الترجمة التحاورية. هنا نجد أن الإنسان يترجم والآلة تبحث له في المعجم عن الكلمات وتعطيه معاني الكلمات كما تعطيه المرادفات من ذاكرتها.

هـ. **الخدمات الحاسوبية الأخرى للترجمة** مثل بنوك المصطلحات الآلية بشكل نصوص لمصطلحات متسلسلة أو البحث عن مصطلح فيها أو بإدخال مصطلح لمعرفة مرادفاته أو مقابله بلغة أخرى أو بالحصول على جميع المصطلحات في حقل معين إلى غير ذلك من الخدمات.[1]

تاريخ الترجمة الآلية

أول من استخدم الحاسوب في الترجمة بوضوح هو "وارن ويفر" عام 1947. ومنذ عام 1949 سارت بحوث الترجمة الآلية في الولايات المتحدة قدماً في جامعات كاليفورنيا ولوس أنجلوس وتكساس وغيرها. وفي جورج تاون أجريت بنجاح أول ترجمة من اللغة الروسية إلى الإنكليزية عام 1954. وفي عام 1955 أجريت في الاتحاد السوفيتي أول تجربة في الترجمة

[1] د. محمود اسماعيل صالحي (الصيني)، أستاذ اللغويات التطبيقية، مدير مركز الترجمة بجامعة الملك سعود (سابقاً) **الحاسوب في خدمة الترجمة والتعريب**
www.emro.who.int/ahsn/meetings/sep03/day2/Dr.%20Mohammed%20Saleh.doc

الآلية من الإنكليزية إلى الروسية في الرياضيات (على أساس قاموس يحوي 2300 كلمة). وقد استمرت الأبحاث في هذه الحقبة باستعمال ما يسمى الجيل الأول من برامج الترجمة الآلية حتى عام 1966.[1] وقد تبين خلال تلك المدة حجم الصعوبات الهائلة التي تعترض الترجمة الآلية مما أدى إلى تباطؤ في أبحاث الترجمة الآلية وربما إهمال الموضوع حتى عام 1975.

شهد الاهتمام بالترجمة الآلية في أوربا وكندا بين عامي 1975 و1985 عودة إلى الأبحاث في هذا الحقل على مدى عقد من الزمان تطورت فيه آلية استعمال الأنظمة الخبيرة وتقعيد معالجة اللغات الطبيعية واستحداث ما يمكن اعتباره الجيل الثاني من برامج الترجمة الآلية. وكانت نتيجة ذلك ظهور بعض برامج الترجمة الآلية التجارية في الأسواق التي تستعمل الحواسيب المكروية. لقد تطور في هذه البرهة البحث العلمي في معالجة اللغات الطبيعية وخاصة اللغات الأوربية واليابانية. وهذا التطور شمل البحوث المعجمية والنحو والصرف والدلالة.

وفي هذه المرحلة تطورت كذلك أساليب الذكاء الإصطناعي الحديثة التي تستند إلى استعمال أنماط رياضية وحاسوبية تحاكي عمل الإنسان أو الكائنات الحية فأساليب الشبكات العصبية والنظم الخبيرة والمنطق المشوش والخوارزميات الجينية كلها تحوي أساليب رياضية تقلد ما يجري في دماغ الإنسان أو تقلد أحياء خلقها الله وعلمها كيف تعيش في مجتمعات وكيف تتصرف كما هو الحال في محاكاة خلايا النمل وأنماط طيران الوز. وهذه الأساليب ما زالت في تطور مستمر إلى يومنا هذا؟

وفي عام 1989 بدأ عهد جديد في أسس الترجمة الآلية بالاستناد إلى المعلومات الإحصائية حينما قامت شركة A.B.M بمشروع كاردید الذي استند إلى الترجمة بالأمثلة والترجمة المحدودة الموضوع وتعدد اللغات المترجم منها والمترجم إليها. وأعقب ذلك تزايد الاهتمام

---

[1] **الترجمة الآلية... تحديات وآمال!** مجلة الحرس الوطني

haras.naseej.com/Detail.asp?InNewsltemID=62556 – 33k.

بالترجمة الآلية إلى حدٍّ لم يسبق له مثيل. وعلى ذلك يمكن اعتبار تسعينيات القرن العشرين مرحلة ظهور الجيل الثالث من برامج الترجمة الآلية المستندة إلى الذخيرة اللغوية Cirpus Based MT والتي لا تزال في تطور حتى اليوم مع بعض التكامل مع الأساليب الأخرى للترجمة الآلية.

واليوم تأتي الولايات المتحدة واليابان وروسيا والصين في طليعة البلدان التي توظف الترجمة الآلية لخدمة متطلباتها الاجتماعية والاقتصادية والعلمية والتقنية.

واقع الترجمة الآلية

هناك الآن ما يقرب من 1000 برنامج ترجمة آلية (خاصة للغات الأوربية) في السوق مع أن نوعيتها ليست جيدة بوجهٍ عام لكن الطلب عليها عال جداً. وقد زادت الشابكة (الإنترنت) من الحاجة إلى الترجمة الآلية وهي كذلك وسيلة سهلة لتسليم المادة المترجمة إلى من يحتاجها. وسيشهد المستقبل تكاملاً بين عمل المترجمين من البشر والترجمة الآلية حيث تحتاج الترجمة الآلية إلى مترجمين أكفياء لتطويرها ومتابعة عملها وإدخالها في مجالات لم تدخلها بعد.[1]

إن استعمال أنظمة الترجمة الآلية القليلة الدقة سيدفع إلى الحاجة إلى تحسين النوعية باطراد. وسينقلب التوتر السابق بين المترجمين من البشر ومؤيدي الترجمة الآلية إلى تعاون بعد إدراك الجميع أن الترجمة الآلية لن تستطيع إلغاء دور المترجمين من البشر.

تبلغ مبيعات برامج الترجمة الآلية في اليابان ملايين النسخ وبتكلفة تبلغ عشرات الملايين من الدولارات. ويتوقع أن يتضاعف ذلك مئات المرات خلال العقدين القادمين مع التطور والتحسين المتوقع في الترجمة الآلية مستقبلاً.[2]

---

[1] John Hutchins, **The state of machine translation in Europe and future prospects**. Aymara.org/biblio/mtranslation.pdf

[2] Setsuo Yamada Syuuji Kodama Taeko Matsouka, Hirosgi Araki Yoshiaki Mirakami Osamu Takano Yoshiyuki Sakamoto, **A Report in the Machine Tranlation Market in Japan**. www.mt-archive.info/MTS-2005-Yanada.pdf.

أساليب الترجمة الآلية – محاكاة الآلة لعملية الترجمة البشرية

يمكن تصنيف برامج الترجمة الآلية بحسب مستوياتها التي تُبرز مدى تعقيدها ومدى كفاءتها في الترجمة إلى الأصناف التالية على نحوٍ تقريبي:[1]

المستوى الأدنى

وهو بوضع كلمة محل كلمة مكافئة لها. وهذه تحتاج إلى معجم ثنائي اللغة ضخم. وهنا ينبغي الأخذ بعين الاعتبار أن بعض الكلمات لا مقابل لها وبعضها يحتاج أكثر من كلمة وبعضها لها أكثر من مقابل وهكذا.

المستوى الثالث الأعلى من المستوى الأدنى

وهو القيام بعمليات معالجة صرفية للوصول إلى عبارات قياسية بغرض تقليص حجم المعجم المطلوب. وعلى ذلك فهناك حاجة إلى التعامل مع الكلمات بمستوى مخفي بحيث توصف الكلمة بعدة مواصفات صرفية ونحوية ويقع تحت هذا المستوى طريقة الترجمة بالأمثلة المباشرة، حيث يكون هناك ذخيرة لغوية متوازية جملة جملة. وتكون العبارات المتوازية في الذخيرة مكونة من كلمتين فأكثر لكن المشكلة في هذه العبارات هي في مطابقة الجمل المراد ترجمتها مع ما هو موجود في الذخيرة.

المستوى المتوسط الأول

لكي يمكن الحصول على الجملة المصدر بوجهٍ صحيح يجب تكوين شجرة إعراب للجملة الأصلية ثم بعد ذلك إسقاط ذلك على اللغة المراد الترجمة إليها. وعلى ذلك فهناك حاجة إلى محلل نحوي إضافةً إلى المعجم الثنائي اللغة.

المستوى المتوسط الثاني

هناك الكثير من الظواهر اللغوية التي لا يمكن نقلها من لغة إلى لغة بمجرد التحليل

[1] Thepchai Supnithi, Virach Sornlertanvanich, Thatsanee Charoenporn, **A Cross System Machine Translation COLING-02 on Machine Translation in Asia** – Vol 16, Sep. 2002.

النحوي والصرفي. ففي العربية إذا قلنا "إن رأسي يؤلمني" فترجمتها بالإنكليزية هي "لدي صداع". وعلى هذا من الضروري فهم المعنى وتمثيل المعنى تمثيلاً سليماً لكي يمكن إيجاد مرادفه في اللغة الأخرى. وعلى هذا يجب تطوير المعجم لكي يحوي ترجمة لمثل هذه المعاني المتقابلة. ولهذا فإن معظم برامج الترجمة الآلية اليوم (مثل برنامج سيستران ولوغوس ويوروترا) تحوي محللاً صرفياً ونحوياً ونوعاً من التمثيل الدلالي لمثل هذه الحالات.

المستوى الأعلى

في هذا المستوى يجب أخذ الأساليب البلاغية العميقة في اللغة المصدر واللغة المترجم إليها وما زال البحث في هذا المستوى يأخذ أبعاداً مختلفة ويحتاج إلى دراسات لغوية وتمثيل حاسوبي عميق ويمثل قصوراً واضحاً في برامج الترجمة الآلية المتوفرة اليوم.

اللغة الوسيطة

هناك اتجاه إلى استعمال لغة وسيطة بين اللغات. وقد لاقت فكرة اللغة العالمية الوسيطة صعوبات في أن تحل كل مشاكل الترجمة الآلية، واقترح أن تكون اللغة الإنكليزية بوضعها الحالي هي اللغة الوسيطة بين اللغات العالمية جميعاً. وقد يكون من المناسب أن تكون هي اللغة الوسيطة بين اللغات الأوربية الحديثة كوارثٍ للغة اللاتينية التي انبثقت منها معظم اللغات الأوربية. ولكن أليست اللغة العربية جديرة بأن تكون هي اللغة الوسيطة بين اللغات التي يتكلم بها المسلمون اليوم كالتركية والفارسية والأوردية والسواحيلية والملاوية والبنغالية التي قد اشتُق كثير من كلماتها من العربية؟ هذا السؤال على العرب اليوم أن يجيبوا عنه.

وفي كل الأحوال على برنامج الترجمة الآلية أن يقوم بعمليتين رئيسيتين:

آ. **النقل:** أي عملية إيجاد المقابلات المعجمية والنحوية والأسلوبية لأجزاء النص المترجم، مثل إيجاد الكلمات والتعبيرات المقابلة في المعنى والوظيفة للكلمات والتعبيرات الواردة في النص الأصل، وكذلك إيجاد التراكيب النحوية المقابلة للتراكيب الواردة في النص الأصل.

ب. أما العملية الفرعية الثانية فهي **التأليف أو التوليف**، أي صياغة الجمل الناتجة من عملية

النقل السابقة صياغةً صحيحة صرفياً ونحوياً وأسلوبياً، مثل صوغ (رجل + جمع) في صورة (رجال) و(مسلم + جمع + حالة النصب أو الجر) في صورة (مسلمين)، وكذلك وضع الصفة في العربية بعد الموصوف ومراعاة قواعد المطابقة اللازمة.[1]

الترجمة الآلية المستندة إلى الإحصاء

تستند هذه الطريقة الحديثة إلى جمع أكبر ما يمكن من ذخيرة لغوية (corpus) وإعمال أكبر ما يمكن من جهد إحصائي عليها لكي تهيأ للاستخدام في الترجمة الآلية. وهذه الذخيرة المترجمة هي بالأساس مترجمة من قبل مترجمين من البشر فهي تستخلص خبرات البشر للإفادة منها في الترجمة الآلية.

الفلسفة وراء هذه الطريقة هي أن الذخيرة اللغوية إذا كانت كبيرة الحجم بقدْرٍ كافٍ فهي تجمع بين دفاتها معظم الكلمات الشائعة في اللغة ومعظم التعابير اللغوية ومعظم التراكيب النحوية والصرفية فيها. وعلى ذلك فإن أي عملية إحصائية على هذه الذخيرة تعطي نتائج قريبة من واقع اللغة الفعلي.

ويحق لنا أن نتساءل عن مدى صحة هذه الفرضية نظراً لأهميتها واعتماد نظرية الترجمة الآلية عليها اعتماداً كبيراً.

وللإجابة عن هذا التساؤل يجب أن نحدد أولاً عن أي ذخيرة لغوية نتحدث؟ ومن أين جمعت؟ وكيف جمعت؟ وأي المواضيع اللغوية تشمل؟ ولأي نوع من الترجمة ستستعمل؟ وهل هي بلغة واحدة أم متعددة اللغات؟ وهل هي جاهزة للتقابل اللغوي؟

وما يحدث اليوم في اللغات العالمية الأخرى غير العربية هو افتراض أن اللغة المتوفر بها كتابات كالصحف والمجلات والكتب المطبوعة والمقالات المنشورة على الشابكة والإعلانات

---

[1] د. محمود اسماعيل صالحي (الصيني)، أستاذ اللغويات التطبيقية، مدير مركز الترجمة بجامعة الملك سعود (سابقاً) **الحاسوب في خدمة الترجمة والتعريب**
www.emro.who.int/ahsn/meetings/sep03/day2/Dr.%20Mohammed%20Saleh.doc

ونشرات الشركات التجارية تمثل اللغة. ومن ثم استناد الترجمة الآلية إلى مثل هذه الذخيرة يمثل هذه اللغة أو تلك. ويعتبر حجم الذخيرة الذي يبلغ مئات الملايين من الحروف (مئات الميكابايت) مقداراً يمثل الحدود الدنيا لحجم الذخيرة التي يمكن الاعتماد عليها للحصول على ترجمة معقولة ومقبولة. وقد سبق أن ذكرنا الفقر الشديد الذي تعاني منه الترجمة من العربية وإليها، وعلى ذلك فالذخيرة اللغوية الممكن توفرها باللغة العربية لا تزال محدودة.

ويقع تحت هذا الأسلوب من الترجمة عدة طرائق منها: الترجمة الإحصائية المباشرة والترجمة بالأمثلة والترجمة بالتجاور، ففي الأسلوب المباشر يكون هناك معجم عبارات متقابلة بين اللغات المراد الترجمة منها وإليها. وفي الأسلوب غير المباشر تُجرى تحويلات لغوية متعددة للإفادة من الذخيرة كالتحويلات الصرفية والنحوية والدلالية واستعمال الإحصاء في الأنظمة الخبيرة.

هذه الطريقة تحتاج إلى إعراب الجملة نحوياً بلغتها الأصلية ثم تطبيق جملة من التحويلات على شجرة الإعراب المستحصلة، وذلك بإعادة تسلسل السلسلة الظاهرية من جهة اللغة الأصلية، للوصول إلى تسلسل أفضل لكلمات الجملة، بحيث تكون أقرب إلى اللغة المراد الترجمة إليها وهذه العملية تطبق على الذخيرة المستعملة للتدريب كما تطبق على النصوص المستعملة في الترجمة.

كيف تجري العملية الإحصائية على الذخيرة؟

يقوم الحاسوب أولاً باستخراج الكلمات غير المكررة في الذخيرة ليعمل منها قائمة بكلمات الذخيرة. ثم بعد ذلك يقوم بعمل قائمة بكل كلمتين متعاقبتين فيها ليكون من ذلك قائمة بالعبارات المكونة من كلمتين فيحدد مرات ورود كل منها، ثم يقوم بعمل قائمة بثلاث كلمات متعاقبة ويعمل منها قائمة تحوي عدد مرات ورود هذه العبارات المكونة من ثلاث كلمات وهكذا يزيد في أعداد الكلمات إلى أن لا يبقى هناك كلمات متعاقبة مكررة أخرى أو يتوقف عند عدد محدد من الكلمات.

هذه العملية الإحصائية هي واحدة من العمليات الإحصائية المهمة التي تستند إليها كثير من أبحاث المعالجة الآلية للغات الطبيعية اليوم. لكن الترجمة الآلية تحتاج إلى نصوص متوفرة بلغتين أو أكثر. فالذخيرة اللغوية يجب أن تكون متوفرة بلغتين يجب أن تكون مصوغة بشكل متقابل.

تستند هذه الطريقة إلى عمل جداول بالعبارات المكررة بين اللغتين والمكونة من كلمتين أو أكثر. فترجمة عبارة معينة بأكثر من موقع واحد في الذخيرة بالترجمة نفسها وربما من قبل مترجمين مختلفين يزود برنامج الترجمة الآلية بترجمة مفضلة لتلك العبارة. أما إذا اختلفت الترجمة بين مصدر من مصادر الذخيرة وآخر أو في المصدر نفسه، فتتبع الأساليب الإحصائية باختيار الترجمة الأكثر تكراراً أو التي يقع الموضوع الذي وردت فيه ضمن موضوع النص المراد ترجمته. هذا وإن استخلاص العبارات المحتوية على كلمتين أو أكثر يجري ببرامج حاسوبية متوفرة اليوم.

هذه الطريقة بأسلوبها البسيط تُعطي ترجمة غير دقيقة بالتأكيد، لكن إلحاق معلومات نحوية وصرفية بالمعجم المستخلص من الذخيرة سوف يزيد من دقة الترجمة. وقد استعملت هذه الطريقة في ترجمة محاضر البرلمان الكندي بين اللغتين الفرنسية والإنكليزية فوجدت ناجحة إلى حدٍّ بعيد وذلك بسبب أن موضوع التحاور في أروقة البرلمان ذو نمط معين وبأساليب متعارفة بين المتحاورين. ويعني ذلك أن هناك أسلوبين للإفادة من الذخيرة اللغوية وهما الأسلوب المباشر والأسلوب غير المباشر.

الترجمة بالأمثلة

تستند هذه الطريقة إلى إجراء أبحاث على الذخيرة اللغوية لاستخلاص الأمثلة والعبارات الشائعة المتقابلة بين اللغتين المراد الترجمة بينهما. كما أن استخلاص قوالب نحوية بين اللغتين وتعويض الكلمات المقابلة بين اللغتين هو أحد الأساليب التي تقع ضمن هذا الأسلوب. هذه الطريقة عادة فعالة للترجمة بين اللغات التي تعود للعائلة نفسها كالإنكليزية والفرنسية أو

العربية والسواحلية بسبب سهولة تكوين قوالب متقابلة بين اللغتين العائدتين للعائلة نفسها لكن الأمر يزداد صعوبة بين اللغات التي تعود إلى عوائل مختلفة.

لنأخذ العبارات القرآنية المتشابهة التركيب اللغوي مع اختلاف كلمة فيها:[1]

فـ + اعلموا أن الله + ... 3 آيات

و + اعلموا أن الله + ... 12 آية

والله (بصير، خبير، عليم) بما (تعملون، يفعلون، يصنعون، يعملون)

والله بما (تـ ، يـ) ـعملون (محيط، بصير، خبير، عليم)

كأنهم أعجاز نخل (خاوية، منقعر)

مثل هذه العبارات القرآنية يمكن أن تشكل قوالب للترجمة بالأمثلة بحيث إذا تغيرت كلمة فيها فإن القالب يكون نفسه باستبدال ترجمة الكلمة الواحدة فقط.

الترجمة بالتجاور Connectionist

تستند هذه الطريقة إلى استعمال وسائل الذكاء الاصطناعي كالشبكات العصبية، وتحليل الجملة إلى شبكة تستعمل كمعلومات لما يرد من جمل لاحقة، بحيث يزداد البرنامج خبرة كلما استعمل مرة بعد مرة، فهو يخزن خبراته لكي يستعملها مستقبلاً. فإذا ما قام المستخدم بتعديل معنى معين سبق أن أخطأ فيه البرنامج فإنه يتعلم من أخطائه ولا يكرر الخطأ بل يسلك الأسلوب الجديد بدل ذلك. كما سيأتي ذكره في برنامج غوغل.

المشاكل التي تعاني منها الترجمة الإحصائية

لا تزال الترجمة الإحصائية تعاني من كثير من الصعوبات. من هذه المشاكل أن الدقة في الترجمة الإحصائية تعتمد على نوع ودقة الذخيرة المستعملة وعلى البرمجيات المستعملة فيها.

---

[1] أ. د. محمد زكي خضر، **الطريقة الإجمالية في ترجمة معاني القرآن الكريم**، المؤتمر الدولي الثاني في اللغة والترجمة، مركز أطلس العالمي للدراسات والأبحاث – عمان – الأردن، 14-2002/12/15.

كما أن تغير اللغة والمفردات المستعملة مع مضي الزمن وضخامة الذخيرة المطلوبة يجعل هناك ضرورة لتحديد الزمن التاريخي للنص. فكلمة "السيارة" حين ترد في نص مكتوب قبل قرون تعني غير "السيارة" المكتوبة في نص مستعمل اليوم. كما أن هناك مشكلة في تحديد الأسماء فكلمة "أحمد" في جملة "أحمد الله على نعمائه" هي ليست اسم عَلَم. ومن الصعوبة بمكان على الآلة أنْ تُحدِّد هل الكلمة ذات معنى أم اسم علم. وتحتاج إلى الكثير من الضوابط البرمجية لكي تميز الآلة بين الكلمة كاسم علم أو غيره.

حل الغموض بواسطة المعالجة المتوازية

تتلخص طريقة الترجمة الآلية بالتوازي بأن تستعمل أكثر من طريقة واحدة لترجمة نص معين في وقت واحد، ثم بعد ذلك يجري تقييم هذه الترجمات والمفاضلة بينها لاختيار أفضلها وربما يجري التزاوج بين بعضها للوصول إلى ترجمة تجمع محاسن أكثر من طريقة واحدة بوقت واحد.

بعد إجراء هذه التحليلات نصل إلى احتمالات متعددة للترجمة الوسيطة يجري تقييمها وإسقاط ما هو أقل احتمالاً منها، ومن ثَم الوصول إلى الترجمة الفضلى بحيث يعاد تركيب الجملة في اللغة المراد الترجمة إليها بخطوات معاكسة.

الترجمة الآلية واللغة العربية

تتشابه المشاكل التي تعاني منها الترجمة من اللغة العربية وإليها مع المشاكل التي تعاني منها اللغات الأخرى، فإذا ما استثنينا الشعر والنثر الأدبي الراقي والمحتوي على دلالات بلاغية عالية فإن هذا ينطبق على اللغة العربية أيضاً. وفيما يأتي بعض خصائص اللغة العربية ذات العلاقة بالترجمة، فليس المجال هنا للكلام عن خصائص اللغة العربية بوجهٍ عام، ولكن هناك بعض الخصائص التي تؤثر في التقدم في حقل الترجمة الآلية بوجهٍ خاص وهو موضوعنا هنا. فمن الخصائص المهمة للغة العربية المستعملة بكثرة اليوم هو غياب الضبط

بالشكْل (بالحركات). ومع أن هناك محاولات لكتابة برامج للشكْل الآلي للنص غير المشكول، فإن أقصى دقة قد تصل إليها مثل هذه البرامج اليوم لا تتعدى نسبة 95%. ويزيد الخطأ عن 5% بكثير بالنسبة لشكْل أواخر الكلمات. إن غياب الشكل يجعل القارئ يفترض من عنده شكلاً افتراضياً بما يملكه من ثروة لغوية وكذلك يقوم المترجم. أما الآلة فيجب تعليمها ذلك.[1]

كما أن ندرة استعمال علامات الوقف والفواصل في النصوص العربية يضيف تعقيداً آخر للنص العربي. أما من ناحية الصرف والنحو وطريقة الكتابة والعمليات التي تجري على الكلمة من دمج للسوابق واللواحق وإدغام وإقلاب وغيرها فهو ما تختص به العربية. ولكل لغة خصوصيتها.

وكثيراً ما تكون الجملة العربية طويلة، حيث يبلغ متوسط عدد كلمات الجملة العربية ما يزيد على 35 كلمة.[2] وهذا يؤدي إلى بطء المعالجة الحاسوبية للجملة العربية. ويمكن التغلب على هذا الطول بتقسيم الجملة إلى عبارات... مثلاً المضاف والمضاف إليه وأشباه الجمل من جار ومجرور.

لكنْ هناك بعد آخر في اللغة العربية ينبغي الانتباه إليه هو البعد التأريخي. فالنصوص الإنكليزية مثلاً التي تتضمنها الذخيرة لا يتعدى تأريخها القرن أو القرنين من الزمان وهي مع هذا قد يلاحظ فيها التغير بين كتابات القرن الماضي والقرن الحالي. وهذا الفرق قد يوازي الفرق بين اللغة العربية في صدر الإسلام (التي كتبت بها معظم المصادر الدينية والتاريخية ولم تتغير إلا قليلاً) واللغة العربية اليوم.

---

[1] Moustafa Elshafei, Husin Al-Muhtaseb, and Mansour Alghamdi, **Machine Generation of Arabic Diacritical Marks, The 2006 Inernational Conference on Machine Learning**; Models Technologoes & Applications (MLMTA'06).

[2] Nizar Habash, Bonnie Dorr, Christof Monz, **Challenges in Bilding an Arabic-English GHMT system with SMT Componentsm Center for Computational Learning Systems**. Columbia University. www.nizarhabash.com/publications.html.

المعالجة الابتدائية للغة العربية قبل الترجمة الإحصائية

تتضمن العمليات التمهيدية للغة العربية قبل الترجمة الآلية إجراء عمليات عديدة منها العمليات الآتية وهذه على سبيل المثال لا الحصر:[1]

1. المعالجة الكتابية للنص فيما يتعلق بالهمزات والهاء آخر الكلمة والتاء المربوطة والشكْل.
2. تجزئة الكلمة إلى سوابق ولواحق وتحديد أجزاء كل منها إن وجدت كألف لام التعريف وحروف العطف وحروف الجر المتصلة والضمائر وعلامات الإعراب وتحديد جذر الكلمة ووزنها الصرفي.
3. الرجوع إلى قواعد تغيير كتابة الكلمة بعد إلحاق بعض الضمائر المتصلة مثل كلمة مكتبة – مكتبتهم وأعلى – أعلاه..
4. تحديد فيما إذا كان العدد المتضمَّن ضمن الكلمة مفرداً أو مثنى أو جمعاً وتحديد علامة ذلك كالألف والنون أو الواو والنون وغيرها.
5. حل بعض المشاكل المتوقعة لوجود أكثر من احتمال في تحديد الكلمة المراد ترجمة معناها مثل الكلمات التي تعطي أكثر من معنى باختلاف الشكل، كأن يكون الفعل مبنياً للمجهول أو للمعلوم أو أن يكون اسماً أو فعلاً مثل كَتَب – كُتِب – كُتُب.

هذه بعض العمليات التي يجب القيام بها أو ببعضها بحسب البرامج التي تحتاجها الترجمة الآلية والتي يجب القيام بها، سواء على الذخيرة اللغوية أو على النص المراد ترجمته من العربية، أو النص المراد توليده باللغة العربية إن كان مترجماً من لغة أخرى. ولنأخذ العبارات القرآنية الآتية مثالاً على العبارات المكونة من كلمتين والواردة في القرآن الكريم بكل أشكالها.

وَيُقِيمُونَ الصَّلاة، يُقِيمُونَ الصَّلاة، لِيُقِيمُوا الصَّلاة، ويُقِيمُوا الصَّلاة، يُقِيمُوا الصَّلاة، وأَقام الصَّلاة، وَأقامُوا الصَّلاة، أقمْتُمُ الصَّلاة، أقِمِ الصَّلاة، أقِيموا الصَّلاة، فأقِيمُوا الصَّلاة، وَأقِمِ

[1] Fatiha Dadat, Nizar Habash, **Combination of Arabic Preprocessing Schemes for Statistical Machine Translation, Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL**, pages 1-8, Sydney, July 2006.

الصَّلاة، وَأَقِمْنَ الصَّلاة، وَأقِيمُوا الصَّلاة، وَإقامَ الصَّلاةِ، وَإقامِ الصَّلاةِ، قائِمٌ يُصَلي، مُقِيمَ الصَّلاةِ، وَالمُقِيمِي الصَّلاةِ، وَالمُقِيمِين الصَّلاة، فأَقَمْتَ لهُمُ الصَّلاةِ.

فهذه العبارات كل منها يحتوي على كلمة تتعلق بإقام الصلاة مع بعض الإضافات من سوابق أو لواحق أو تغيير في الكلمة الأولى بين الفعل والاسم. فإذا ما حددت الإضافة من ناحية مفعوليتها أمكن ترجمة ذلك إلى اللغة الثانية. فإذا عرفنا من ترجمة معينة ما معنى "إقام الصلاة" بتلك اللغة، أمكن ترجمة كل العبارات السابقة بالاستناد إلى ذلك المعنى بعد إلحاقه وتغيير تركيبته وفق قواعد محددة تترجم الإضافات الموجودة في اللغة الأصلية. ويلاحظ أن بين العبارات السابقة العبارة الأخيرة المكونة من ثلاث كلمات: **فأَقَمْتَ لهُمُ الصَّلاةِ** حيث إن الكلمتين المعنيتين دخل بينهما جار ومجرور "لهم" ومثل هذه العبارات يجب البحث عنها أيضاً لتلحق بمثل هذه المجاميع من العبارات: وكمثال على ثلاث كلمات بينها ضمائر أو ما يقابل تلك الضمائر فتصبح أربع كلمات:

أَنفِقُوا مِمَّا رَزَقْنَاكُم، أَنفِقُوا مِمَّا رَزَقَكَ، وأَنفِقُوا مِمَّارَزَقْنَاهُم، وأَنفِقُوا مِمَّا رَزَقَهُمْ، وَيُنفِقُوا مِمَّا رَزَقْنَاهُمْ، أَنفِقُوا مِمَّا رَزَقَهُمُ اللهُ، أَنفِقُوا مِمَّا رَزَقَكُمُ اللهُ، وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ.

ويلاحظ هنا تغير تسلسل الكلمات بين عبارة وأخرى. على ذلك فإن الترجمة يجب أن تأخذ بعين الاعتبار ما هو متعارف باللغة المترجم إليها بعد معرفة مفهوم النص وهو "الإنفاق من رزق الله". وكذلك العكس عند الترجمة إلى اللغة العربية.

وصف الكلمة وفق مشروع مداد البيان[1]

ومن الأمثلة السابقة يتبين أن المعالجة اللغوية للنصوص الأصلية قبل أن تكون جاهزة للترجمة الآلية ضرورية لتمام الإفادة من الذخيرة اللغوية، التي يجب أن تكون بحجم كافٍ لكي تشمل اللغة أو معظم الأساليب الواردة فيها على الأقل. وقد احتوى مشروع مداد البيان الخاص

[1] أ. د. محمد زكي خضر، **مداد البيان**، ندوة حول مشروع قاعدة بيانات حاسوبية للقرآن الكريم، مجمع اللغة العربية الأردني، 20 ذي الحجة 1427هـ/9 كانون الثاني 2007.

بوصف كلمات القرآن الكريم في قواعد بيانات تفصيلية على معالجات للكلمات القرآنية يمكن أن يخدم اللغة العربية بوجهٍ عام. وكمثال على بعض المعالجات التي ينبغي القيام بها في اللغة العربية عند تحليل الأمثلة أو النص المراد ترجمته أو عند توليد نص مترجم المعالجات الآتية، نورد ما يأتي:

1. الإدغام: مثل **مُضَارٌ وَازَّيَّنَتْ**
2. الإبدال: مثل **وَاصْطبرْ**
3. الإعلال بالقلب: مثل: **قِيلَ** أو بالحذف مثل: **قلْتُ** أو بالنقل مثل: **وَيَقُولُونَ** أو بالحذف الصوتي مثل: **نَبْغِ**
4. حذف الهمزة: مثل **كُلُوا** وتسهيل الهمزة مثل: **بَادِيَ**
5. حذف ياء المنقوص: مثل **بَاغٍ** ولا **عَادٍ**
6. حذف النون للإضافة: مثل **وَالمُقِيمِي**
7. حذف حرف العلة للحزم: **وَلا تَمْشِ**
8. حذف نون الأفعال الخمسة: **ولا تَقُولوا**

وعلى ذلك فإن التدريب الذي يجب أن يتم يحتاج كمية كبيرة من الأمثلة التي تحوي هذه الحالات لكي يكون كافياً للقيام بعملية الترجمة على وجهٍ مقبول.

الذخيرة اللغوية العربية

تعتبر الذخيرة اللغوية لأية لغة مرتكزاً أساسياً اليوم للباحثين في اللغات الطبيعية. فالخزن على الحواسيب جعل جمع وتصنيف وتحليل الذخيرة اللغوية ميسوراً. وهذه الذخيرة تمكن الباحثين من إطلاق الوصف على خصائص اللغة وعلى النحو فيها ودراستها تأريخياً وتغيرها مع الزمن.[1]

[1] Ahmad Abdelali, Jim Cowie and Hamdy S. Soliman, **Building A Modern Syandard Arabic Corpus, Workshop on computational Modeling of Lexical acquisition.** The split meeting Croatia 25^{th} – 28^{th} of July 2005.

كانت أولى محاولات تكوين ذخيرة لغوية عربية صغيرة في عام 1994. ثُم بدأت محاولات متعاقبة لجمع ذخيرة لغوية عربية ذات شأن. ومن هذه المحاولات جمع نصوص جريدة الحياة اللندنية ما بين 1994 و2000 والتي احتوت على ما يقارب من 76 مليون كلمة منها 666 ألف كلمة غير مكررة. كما حدثت محاولات عديدة أخرى لتكوين ذخيرة لغوية من مستندات الأمم المتحدة المتعددة اللغات منها محاولة لعدد من الوثائق بلغ 38 ألف وثيقة تحوي ما يزيد على 50 مليون كلمة.

يزداد استعمال اللغة العربية على الشابكة بسرعة هائلة. ففي الوقت الذي ازداد استعمال اللغة الإنكليزية بين عامي 2000 و2008 بنسبة 200% ازداد استعمال اللغة العربية خلال المدة نفسها بنسبة 2000% على حين كانت الزيادة في استعمال اللغة الصينية بنسبة 600% خلال المدة نفسها.

يبلغ عدد مستخدمي الشابكة باللغة العربية اليوم نحو 60 مليون مستخدم أي بنسبة 4.2% من عدد المستخدمين في العالم وهو ما يقرب من نسبة عدد السكان العرب بالنسبة لمجمل سكان العالم والبالغ قُرابة 4.7%،[1] ويعني ذلك أن نمو استخدام الشابكة في العالم العربي قد وصل إلى ما يقرب من المعدل العالمي وهو بازدياد. ولكن هذا بالطبع لا يُظهر مستوى المحتويات العربية حيث لا تزال أقل من 1% من محتوى الشابكة حسب توزيع اللغات.

بعد اعتماد الرمز الدولي الموحد المعروف باليونيكود في نهاية القرن الماضي، احتلت الحروف العربية مواقع محدودة في جدول الرموز، مما يعني أن أي وثيقة على الشابكة تستعمل هذه الرموز يمكن مباشرة معرفة فيما إذا كانت هذه الوثيقة تستعمل الحروف العربية أم لا. ومن ثم يمكن معرفة لغتها غالباً.

باستعمال ذخيرة محدودة يمكن توسيع تطبيقها على أمثلة أوسع بكثير منها بتحديد شروط معينة لما يمكن أن يقاس على هذه الذخيرة، وإضافة كلمات مركبة بدل الكلمات الواردة في

[1] Top ten Internet Languages – Internet Statistics. http://www.internetworldstats.com/stats7.htm.

الذخيرة وجداول احتمالية لما يمكن أن يضاف في المستقبل. فمثلاً إذا وضعت ذخيرة تشمل كلمات القرآن الكريم فإنه ليس من الصعوبة بمكان تطوير الذخيرة إلى الحديث النبوي مثلاً.

خطوات تكوين ذخيرة لغوية[1]

1. الحصول على نص مكتوب متوفر إلكترونياً وليس عليه حقوق ملكية فكرية. وتعتبر الشابكة أيسر المصادر المتوفرة. يمكن الوصول إلى الصفحة الموضوعة على الشابكة عن طريق التنقيب باستعمال ما يسمى بالبرامج العنكبوتية بمساعدة رموز HTML. وبعد تنقية النصوص من هذه الرموز يمكن الوصول إلى النصوص الخام. وقد تحتاج هذه التنقية في أغلب الأحيان إلى تدخل يدوي. ثم توضع رموز خاصة على هذه النصوص للإشارة إلى مصادرها ومعلومات أخرى عنها. وحيث أن طريقة ترميز النصوص العربية مختلفة بين وثيقة وأخرى بحسب مواصفات الترميز العربية المختلفة، ينبغي أن توحد طريقة ترميزها ليسهل التعامل معها. ويعتبر اليوم الترميز الدولي الموحد (اليونيكود)، على نقائصه، أفضل ترميز لأغراض توحيد النصوص.

بعد إجراء عمليات عديدة لتنقية النصوص من بعض الرموز كالمدة والشكْل (إن وجد نظراً لأن الغالبية العظمى من النصوص هي غير مشكولة) وأنواع الهمزة والتاء المربوطة والهاء آخر الكلمة وغيرها مما يكثر اللبس فيها، يلاحظ أن مثل هذا التوحيد يزيد من الغموض واحتمالات الخطأ كما يخفي في طياته بعض المؤشرات إلى مصادر النصوص وأماكن تداولها وغير ذلك من المعلومات.

ما الذخيرة إلا خزان ضخم للغة المستعملة. والبرامج الخاصة بالذخيرة هي وسائل لإعادة ترتيب هذا الخزان لكي يكون بالإمكان استخلاص الملاحظات عن اللغة بالاستفادة منها. وتشير الأبحاث إلى أن الشابكة العنكبوتية اليوم مؤهلة تماماً للحصول على ذخيرة لغوية

[1] Chris Callison-Burch Philipp Koehn Miles Osborne, **Improved Statistical Machine Translation Using Paraphrases In Proceedings** NAACL-2006.

مناسبة للغة العربية من نصوص الصحف اليومية والصفحات التجارية والشخصية والكتب القديمة والحديثة المتوفرة على الشابكة. إلا أن النصوص المترجمة بلغة أخرى توازي اللغة العربية، فيما عدا وثائق الأمم المتحدة، لا تزال قليلة لا تكفي لتكوين أساس لذخيرة لغوية متعددة اللغات يمكن الاعتماد عليها بكفاءة.

2. عملية التوافق بين النصوص في الذخيرة المكونة لأكثر من لغة واحدة. يجب مقابلة النصوص المترجمة بعضها ببعض بدقة، فقرة فقرة. وإذا أمكن وضعها جملة جملة أو عبارة عبارة فذلك أفضل بالطبع ولكن في العادة لا يمكن ذلك تماماً. وهنا يجب حل بعض جوانب الغموض. والجدول الآتي يبين مثالاً لذلك:

عن عمر بن الخطاب قال: سمعت رسول الله صلى الله عليه وسلم يقول:	Narrated "Imar bin Al-Khattab": I heard Allah's Apostle peace be upon him saying:
إنما الأعمال بالنيات	The rewaed of deeds depends upon the intertions
وإنما لكل امرئ ما نوى	And every person will get the rewaed according to what he has intended
فمن كانت هجرته إلى دنيا يصيبها، أو إلى امرأة ينكحها، فهجرته إلى ما هاجر إليه	So whoever emirgrated for worldly benefits or for a woman to marry, his emigration was for what he emigrated for.

3. يلاحظ في أغلب الأحيان أن الكلمات ومعانيها المقابلة لها بشكل منفرد هي الغالب، أي إن الغالبية العظمى من كلمات الذخيرة لها مقابل واحد أي كلمة مقابل كلمة. وكذلك عبارة مقابل عبارة، حتى لو كان هناك في بعض الأحيان قلب لتسلسل الكلمات ضمن العبارة الواحدة.

4. بعد ذلك يمكن جمع الإحصائيات عن تكرار الكلمات والعبارات والتركيبات الهيكلية في النصوص وهناك برامج متخصصة لاستخلاص المعلومات الإحصائية عن الذخيرة. ومن

هذه الإحصائيات يمكن استخلاص معاني الكلمات والعبارات بحسب أعلى تكرار لها وحل بعض مشاكل الغموض الذي يحدث.

5. يلاحظ في النصوص وجود نقاط ارتكاز لمقابلة التراجم بعضها ببعض، مثل وجود الأرقام في النصوص والتواريخ وأسماء الأعلام وعناوين الفقرات. ومثل نقاط الارتكاز هذه، إضافةً إلى النقاط الفوارز بين الجمل، يمكن استعمالها لترتيب الذخيرة ترتيباً متوازياً. وينتج عن ذلك ثلاثة أنواع من التقابل بين النصوص:

آ. **تقابل قوي** حينما تحوي النصوص وترجمتها عدداً متساوياً من الكلمات المتقابلة وهي حالات نادرة.

ب. **تقابل متقارب** حينما يكون عدد الكلمات نفسه لكن هناك اختلاف في تسلسل الكلمات.

ج. **تقابل ضعيف** حينما يكون تسلسل الكلمات وعددها مختلفاً ولكن المعنى متوافق وفق المعجم، وعلى العموم فإن:

- هناك علاقة معنى بين فصلين مترجمين متقابلين.
- هناك علاقة معنى بين كل فقرتين مترجمتين متقابلتين.
- هناك علاقة معنى بين كل جملتين مترجمتين متقابلتين.
- هناك علاقة معنى بين كل كلمتين مترجمتين متقابلتين إذا كان أحد معاني الكلمات وارداً في المعجم.
- يعتبر ورود الأرقام والتواريخ نقاط ارتكاز للتقابل بين النصوص المترجمة
- تستعمل نظرية الاحتمالات لحل بعض الغموض الذي يحيط بالتقابل بين النصوص.

الأبحاث الغربية الأخيرة في الترجمة من اللغة العربية وإليها

بعد الحملة الشرسة على الأمة بحجة أحداث 11 أيلول برزت الحاجة لدى الدوائر العسكرية والمخابراتية والإعلامية الغربية إلى الترجمة من اللغة العربية خاصة. فبعد الاحتلال الأمريكي للعراق جُهز بعض الجنود بأجهزة تشبه الهاتف النقال بحيث يتكلم الجندي بكلمات أو عبارات

باللغة الإنكليزية فتنطق ترجمتها بالعربية لكي يسمعها العراقي وبالعكس. كانت الكلمات والجمل محدودة وقد سجلت باللهجة العامية العراقية. إلا أن كل ذلك توقف خلال أسابيع عند بدء المقاومة العراقية وتوجس جنود الاحتلال بعد ذلك من التحاور مع السكان المدنيين خوفاً على حياتهم من أن يكون من يقترب منهم هو من أفراد المقاومة.[1]

وقد شهد كثير من الجامعات الأمريكية والأوربية جهوداً حثيثة للبحث عن معالجة اللغة العربية حاسوبياً والترجمة منها وإليها. وقد شملت هذه الأبحاث النحو والصرف والدلالة والترجمة، وجمع الذخيرة اللغوية وكيفية تكوين المعجم المحوسب والاستفسار باللغة العربية وتكوين الخلاصات وغير ذلك من الأبحاث المتعلقة باللغة العربية.

المعجم العربي المحوسب

يكاد أن يتفق الحاسوبيون العرب المهتمون بمعالجة اللغة العربية على أن من أهم ما تحتاجه اللغة العربية اليوم لدخولها عالم تقنية المعلومات بفاعلية، هو تكوين معجم عربي محوسب، يتضمن كل مفردات اللغة العربية على وجهٍ يسهّل التعامل مع مفرداتها في كافة التطبيقات ذات العلاقة. إن السبب في عدم إعداد مثل هذا المعجم حتى الآن هو عدم وجود جهة علمية ذات قدرة مالية على مستوى العالم العربي تأخذ على عاتقها القيام بمثل هذا المشروع الضخم والبالغ الأهمية.

سنحاول هنا أن نبين أهمية المعجم والخطوة الأولى في تكوينه، وهي وضع مواصفات له بحيث يبني عليها من يأتي من بعد.

ما المقصود بالمعجم العربي المحوسب

تتوفر للغة العربية قواميس تعطي معاني كلماتها في لغات أخرى. كما تتوفر قواميس

[1] Jason Riesa, Behrang Mohit, Kevin Knight, Daniel Marcu, **Building an English-Iraqi Arabic Machine Translation System for Spoken Utterances with Limited Resources**. www.isi.edu/natural-language/mt/iraqi_interspeech.pdf.

عربية عربية تعطي معاني الكلمات بكلمات أخرى. ويوضع في هذه القواميس عادة تصريف الكلمة وبعض المدخلات الأخرى. وقد تعددت المعاجم العربية وتنوعت خلال العصور السالفة ولكن القصد منها في كل الأحوال كان واحداً، وهو حراسة القرآن الكريم من أن يقتحمه خطأ في النطق أو الفهم. وحراسة العربية من أن يقتحم حرمَها دخيل لا ترضى عنه العربية، وصيانة هذه الثروة من الضياع.

مرت المعاجم العربية بأطوار مختلفة وتعددت مدارسها المعجمية واللغوية، ومن هذه المدارس: مدرسة الخيل ومدرسة أبي عبيد ومدرسة الجوهري. ومن المعاجم المعاصرة التي سارت على نهج الترتيب الألفبائي "المعجم الوسيط" الذي أصدره مجمع اللغة العربية بالقاهرة.

أما المعجم المحوسب فهو معجم خاص يجب أن يحوي رموزاً خاصة لتصريف الكلمة ومعلومات أخرى عنها تدرج فيه مفردات اللغة بالتفصيل بحيث يكون بالإمكان الإفادة منها حاسوبياً.

أهمية المعجم العربي المحوسب

المعجم المحوسب ذو أهمية بالغة في كافة تطبيقات اللغة حاسوبياً. فالمعجم أساس للشكْل الآلي وللترجمة الآلية وللترجمة الفورية آلياً، ولتوليد الكلام حاسوبياً وللإملاء الآلي ولبرامج فهم الكلام آلياً ولتعليم النطق للأطفال وذوي الاحتياجات الخاصة، ولتحليل النصوص ولفهم دلالة النص ولتلخيص النصوص آلياً وللاستفسار والإجابة آلياً ولكافة تطبيقات توليد الأصوات آلياً، وغير ذلك من التطبيقات التي بدأت بالانتشار بلغات أخرى، وما سيظهر في المستقبل منها كثير وكثير جداً خاصة بعد التوسع في التخاطب مع الآلات بالصوت في مختلف التطبيقات وسيكون لبصمة الصوت مستقبل يشبه بصمة الإصبع والتوقيع.

يقصد بمواصفات المعجم تحديد خصائص الكلمة العربية التي هي وحدة بناء الجملة وينبغي أن توضع لها رموز في المعجم وذلك لكي يمكن للمعجم المساعدة في كل التطبيقات التي ذكرت سابقاً أو المحتمل أن تبرز لها الحاجة مستقبلاً. وينبغي وصف الأسماء من ناحية البناء

والإعراب والتذكير والتأنيث وحالة أخرى من ناحية المقصور والمنقوص والممدود والتعريف والتنكير ومن ناحية العدد مفرداً أو مثنى أو جمعاً وإن كان جمعاً هل هو للمذكر السالم أم المؤنث السالم أو للتكسير وبالنسبة للمجرد أو المزيد هل الجذر ثلاثي أم رباعي أم خماسي وبالنسبة للزيادة هل هي بحرف أم حرفين أم ثلاثة وغير ذلك.

وأما الفعل فيأخذ الصفات الآتية: زمن الفعل ماض أم مضارع أم أمر، ومن ناحية الصحة والاعتلال هل هو صحيح سالم أم مهموز الآخر أم معتل ناقص أم غير ذلك ومن ناحية التمام والنقصان هل هو تام أم ناقص ومن ناحية الجمود والتصرف واللزوم والتعدي هل هو متعد بنفسه أم بحرف جر وهل هو مجرد أم مزيد وعدد حروف الزيادة وهل هو مبني للمعلوم أم للمجهول إلى غير ذلك من الصفات التي تلازم الفعل.

وأما الحرف فيؤخذ نوعه إن كان حرف جر أو حرف عطف أو غير ذلك وتفاصيل إعمال الحروف وهل هي متصلة أم منفصلة كتابة إلى غير ذلك من الصفات.

المعاجم المحوسبة في اللغات الأخرى

وبوجهٍ عام ليس هناك طريقة عالمية موحدة لتركيب المعجم الآلي اليوم. وقد قام بعض الباحثين بمقارنة 55 معجماً آلياً.[1]

تتباين المعاجم في محتوياتها تبايناً كبيراً. وليس هناك نمط عام واحد تلتزم به المعاجم المحوسبة. فهي تختلف بين لغة وأخرى، بل بين الجهات العلمية والتجارية التي قامت بوضع مواصفاتها وتغذيتها بالمعلومات اللغوية.

فهناك المعاجم الأحادية اللغة التي تصف مفردات اللغة وتصريفاتها ومعانيها. وهناك المعاجم الثنائية اللغة أو المتعددة اللغات. وهذه المعاجم تعتمد في إعدادها على المعاجم بكل واحدة من اللغات المكونة لها. لذلك فأي معجم يحوي اللغة العربية مع لغة أو لغات أخرى

[1] Ali Fraghaly and Jean Senellart, Intuitive Coding of the Arabic Lexicon, Proceedings of MT Summit IX, … Jean Senellart, Director of Research and Development, SYSTRAN Software.

يحتاج إلى المعجم العربي المحوسب الأحادي اللغة أساساً.

لقد تأخر ظهور معجم عربي محوسب الآن، وإن المزيد من التأخير سيدفع الشركات الأجنبية إلى أن تبني معاجم قاصرة خاصة بها نتيجة حاجاتها التجارية له. ومما لا شك فيه أن الدوافع التجارية وحدها لن تخدم اللغة العربية كما يجب أن تُخدم. وقد بدأت بعض الشركات فعلاً بجمع الكلام العامي ليكون أساساً لأجهزة تعرف الكلام المنطوق مما يقلل من استعمال اللغة الفصحى ويشيع اللهجات العامية وكتابتها، مما يؤدي إلى ازدهار هذه اللهجات وتطويرها لتكون لغات مختلفة بحيث لا يفهم المتكلم بإحداها الأخرى.

ومن ناحية أخرى فإن خدمة اللغة العربية من قبل الشركات الأجنبية ينتج عنه كثير من الأخطاء اللغوية الجسيمة التي يمكن أن يقع بها أفراد من العرب ممن يقدم الخدمات لمثل هذه الشركات والمؤسسات الأجنبية، وما الخلط بين ترجمة الأربعين النووية والأسلحة النووية عنا ببعيد.

محاولات الترجمة الآلية من اللغة العربية وإليها

قامت محاولات عديدة للبدء بترجمة آلية من اللغة العربية وإليها. وقد أثمر بعضها بتكوين أنظمة ترجمة آلية، على حين أصبح البعض الآخر طي النسيان. وفيما يأتي بعض هذه المحاولات:

1. محاولة الدكتور بشاي الأستاذ السابق بجامعة هارفرد، وذلك منذ أوائل السبعينيات.
2. برنامج ترجمان التونسي والبرامج الأخرى التي تعمل عليها عدة جهات في مصر والأردن.
3. نظام "المترجم العربي" الذي طورته شركة ATA في لندن، وقد طورت الشركة المذكورة برنامجاً مصغراً اسمته "الوافي".
4. نظام "عربترانز"، وقد طورته شركة عربية أيضاً في لندن.
5. نظام "الناقل العربي" الذي طورته شركة سيموس العربية في باريس، ولدى الشركة المذكورة أربعة برامج للترجمة بين الإنكليزية والعربية وبين الفرنسية والعربية – برنامج

لكل اتجاه.

6. نظام شركة أبتك Apptek.
7. نظام سيستران Systran برنامج للترجمة من الإنكليزية إلى اللغة العربية وهناك موقع www.systranet.com وتقوم عليه واحدة من أشهر وأقوى الشركات في هذا المجال وتم تأسيسها عام 1968م.
8. نظام وايدنر Weidner الذي طور أيضاً برنامجاً للترجمة من الإنكليزية إلى العربية.
9. شركة ألبس Alps التي ما زال لديها برامج للترجمة بين عدد من اللغات، وتطبق مبدأ الترجمة التحاورية.
10. في فرنسا لدى جامعة غرينوبل Grenoble.
11. موقع المسبار وهو موقع يهتم أيضاً بالترجمة الآلية من الإنكليزية إلى العربية وبالعكس ويمتاز بالسهولة والمرونة عند استخدامه ويحقق مستوى مقبولاً من الترجمة في مستوياتها البسيطة التي أشرنا لها عند الحديث عن أنماط الترجمة.
12. موقع www.freetranslation.com.
13. موقع www.itranslatoronline.com.
14. لقد وظفت شركة صخر لبرامج الحاسوب محركاً للترجمة الآلية الخاص بها في دعم موقع http://tarjim.com وهي خدمة ترجمة فورية تقوم بترجمة أي صفحة. وقد كانت أولاً ترجمة مجانية على حين يمكن الآن إرسال النص المطلوب ترجمته والحصول على الترجمة بأجور.
15. برنامج شركة غوغل: وهو برنامج مجاني أُعلن إنتاجه أخيراً ويستند إلى الترجمة الإحصائية من ذخيرة لغوية مأخوذة من الإنترنت. والجدير بالذكر أن هذا البرنامج يتعلم من أخطأئه فإذا ما ترجم جملة خاطئة وأخبره المستخدم أن الترجمة خاطئة وأن المفروض أن تكون بصيغة أخرى، فإنه يخزن هذه المعلومات ويستعملها في المستقبل على نحوٍ أصح. ولا تزال الذخيرة اللغوية التي يستعملها محدودة لكن مستقبل هذا النظام يتوقع له

التقدم حيث أن شركة غوغل من أكبر شركات البحث على الشابكة اليوم.

ولتبيان القصور الذي تعاني منه المترجمات الآلية المتوفرة، أُعطيت الجملة التالية إلى مترجمي صخر (غريب) وغوغل فكانت الترجمة كالآتي:

الجملة الإنكليزية	**Lawyers for the four Iraqis confirmed the first wuit was registered at the U.S. District Court in Seattle, washington**
مترجم غريب	المحامون للأربعة عراقيين أكدوا أن القضية الأولى سُجلت في محكمة المقاطعة الأمريكية في سياتل، واشنطن
مترجم غوغل	محامون لأربعة عراقيين وأكدت الدعوى الأولى المسجلة في الولايات المتحدة أمام المحكمة المحلية في سياتل، واشنطن
الترجمة الصحيحة	أكد المحامون عن العراقيين الأربعة أن الدعوى الأولى قد سُجلت في الولايات المتحدة لدى المحكمة المحلية في سياتل بواشنطن

وهكذا يتبين أن الترجمة الآلية بوضعها الحالي لا تزال قاصرة ولكن يمكن أن تساعد المترجم الذي يعرف اللغتين أن يراجع الترجمة لوضعها بالصيغة المناسبة الصحيحة. ويبقى أن تثار استفسارات عن دلالة النص الأصلية التي قد يحار فيها المترجم من البشر، هل يشير النص إلى أنّ المحامين الموكلين من قبل العراقيين، قد كان تأكيدهم المذكور موجهاً للعراقيين الأربعة أم هو تأكيد صحفي عام أم غير ذلك. وربما يبرز تساؤل عن دقة ترجمة confirmed بكلمة "أكد" حيث أنها تعني تأييد معلومة معروفة سلفاً بشكل غير يقيني. ويتبين أن الترجمة الآلية حتى لو وصلت إلى مستوى الترجمة البشرية فإنها تعاني من غموض تحتوي اللغة بكل حال من الأحوال.

جوانب أخرى ذات علاقة بالترجمة الآلية

تركز بعض الأبحاث على نطاق محدود من فوائد الترجمة الآلية: من هذه المجالات ترجمة

عبارات الاستفسار بين اللغات. في هذا المجال تؤدي الترجمة الآلية لعبارة الاستفسار إلى لغات أخرى ومن ثم يبحث عن موضوع الاستفسار بشتى اللغات التي تجري الترجمة إليها. ولغرض تقليل المعاني المحيرة عند الترجمة وزيادة كفاءة الترجمة يجب أن تحوي المعاجم إشارة إلى موضوع الكلمات.[1]

ترجمة الكلام المنطوق

ليس هناك ترجمة فورية آلية موثوقة حتى الآن لكن المتوفر لا يتعدى قواميس صوتية أو ترجمة لعبارات بين لغات متعددة. ويمكن أن تجد مثل هذه الأنظمة تطبيقات في حجز الفنادق والحجز على الخطوط الجوية والمشاركة في المؤتمرات وطلب الطعام من المطاعم والاستفسار عن اتجاه السير والحجز لدى العيادات الطبية والمستشفيات واستئجار السيارات وغيرها.[2]

إن مشاكل ترجمة الكلام المنطوق تفوق مشاكل النص المكتوب، نظراً لأن هناك عادة الكثير من الكلام المتداول الذي يحوي أخطاء نحوية أو جملاً ناقصة. هذه المشاكل يمكن حل بعضها بالتحليل التركيبي المحدد. فمثلاً يمكن إهمال ترتيب الكلمات في الجملة أو إهمال مفعول الكلمات المساعدة في الجملة. ويعني ذلك وضع قواعد لغوية مطاطة للترجمة الفورية الآلية.

وهذا يحدد نطاق الكلمات التي يستعملها المتكلم وطريقة نطقها وسرعة النطق وطول الجملة وصيغ التوقف بين الجمل...

وهناك محاولات لاستعمال الرسائل القصيرة على الهاتف النقال للترجمة وتمتاز بأنها

---

[1] Mohammed Aljlayl, Ophir Frieder & David Grossman, **On Arabic-English Cross-Language Information Retrieval: A Machine Translation Approach.** ieeexplore.ieee.org/iel5/7847/21600/01000351.pdf

[2] Ying Zhang, **Survey of Current Speech Translation Research, Language Technologies Institute**, Carnegie Mellon University. citeseer.ist.psu.edu/752690.htm

قصيرة ومحدودة المعجم وبأنها قابلة للتطور والتوسع.[1]

المشاكل والحلول

مما سبق يتبين أن هناك الكثير من المشاكل التي تعاني منها الترجمة الآلية بوجهٍ عام وما يخص اللغة العربية بوجهٍ خاص، ونوجزها فيما يأتي:

1. عدم وجود معجم عربي محوسب.
2. قلة النصوص المترجمة بين اللغة العربية واللغات الأخرى التي يمكن الإفادة منها في تكوين ذخيرة لغوية مفيدة للترجمة الآلية التي تستند إلى أسس إحصائية.
3. قلة الأبحاث اللغوية المتعلقة بالترجمة الآلية من اللغة العربية وإليها وعدم وجود دعم كافٍ للبحث في هذا المجال، ومن ذلك البحث في التحليل الإحصائي وتعرف الأصوات ومشكلة الكلمات المتعددة المعاني ومشكلة التحليل الصرفي المشترك لفظياً ومشكلة فهم المعنى من السياق ومشكلة الإعراب والنحو ومشكلة الضبط بالشكْل (بالحركات).

إن الحلول لهذه المشاكل تتلخص بما يأتي:

1. القيام بحملة توعية للقيادات السياسية والعلمية والجهات الداعمة للبحث العلمي على أهمية البحث العلمي في حوسبة اللغة العربية على نطاق المجامع اللغوية العربية وأقسام الحاسوب واللغة العربية واللسانيات في الجامعات العربية وتوجيه الأبحاث نحو التطبيق العملي.
2. تكوين قيادات بحثية في أقسام اللغة العربية ذات خلفية حاسوبية جيدة وفي أقسام الحاسوب ذات خلفيات جيدة باللغة العربية لكي يكون التواصل والبحث العلمي على أتم وجه.
3. السعي لجعل اللغة العربية لغة وسيطة للترجمة الآلية بين اللغات الشرقية كالتركية

[1] Avinash J. Agrawal, Manoj B. Chandak Mobile Interface for Domain Specific Machine Translation Using Short Messaging Service.
Ieeexplore.ieee.org/iel5/4151644/4151645/04151821.pdf?tp=&isnumber=&arnumber=4151821

والفارسية والأوردية والبنغالية والماليزية والسواحيلية والكردية، والتواصل مع الجهات البحثية في هذه اللغات والإفادة من نتئج أبحاثها دون الإقلال من أهمية البحث العلمي في الترجمة الآلية من اللغات الأوربية إلى اللغة العربية أيضاً.

الخلاصـــة

اللغة العربية أمانة اليوم في أعناق هذا الجيل. فالتحديات التي تواجهها الأمة في النواحي الثقافية والفكرية والتقنية كلها ذات علاقة باللغة. لذلك فإن التقاعس عن خدمة اللغة العربية اليوم يؤثر في مستقبل الأجيال القادمة وفي نهضة الأمة وتقدمها.

الترجمة علم تطبيقي يخدم الأمة، وعلى اللغويين أخذ موقف عملي للحفاظ على اللغة من جهة وتقديم الخدمة لمن يتكلم باللغة بأفضل وجه. لأن الخيار الآخر هو عزوف العامة عن اللغة وإبدالها باللغات الأخرى أو اللهجات المحلية. على اللغويين بذل جهدهم في البحث التطبيقي للغة بما ييسر تعامل اللغة العربية مع الترجمة الآلية، وقد يحتاج ذلك إلى سلوك مسالك جديدة في أبحاث اللغة العربية، فإن التضحية ببعض جوانب اللغة أهون من التضحية باللغة كلها.

حين سُئل الزعيم الصيني ماوتسي تونغ عن رأيه بالثورة الفرنسية بعد مضي 200 عام عليها قال: من المبكر الحكم عليها الآن. ولذلك فمن المبكر الحكم على الترجمة الآلية اليوم بعد مضي نحو 50 سنة على بدء البحث العلمي فيها.[1]

عند بدء البحث في الترجمة الآلية كان الهدف هو الوصول إلى ترجمة آلية لكل أنواع ومحتويات النصوص وبمستوى نوعية جيدة تكافئ المترجم البشري. لكن بعد وقت قصير تبين أن هذا الهدف لا يمكن تحقيقه في المدى المنظور، لكن كذلك تبين أن كثيراً من النصوص المترجمة بدقة غير عالية مفيدة للمترجمين لكي يقوموا بتحسينات عليها، كما أنها تعطي فكرة

[1] Nicola Ueffing, Hermann Ney, **Word-Level Confidence Estimation for Machine Translation**, RWTH Aachen University, portal.acm.org/citation.cfm?id=1220671.

عامة عن الموضوع حينما لا تكون الدقة العالية ضرورية. وحينما يصبح بالإمكان استعمال الترجمة الآلية على الإنترنت والراديو والتلفاز والهاتف ستكون الترجمة الآلية جزءاً من هذه المعدات وسيكون استعمالها من كل البشر.[1]

إن الترجمة الآلية اليوم هي بيد الحاسوبيين واللسانيين واللغويين. لكن بعد عقدين أو ثلاثة من الزمن وفي ضوء التطور الحاصل في قدرات الحواسيب يتوقع أن تتوفر حواسيب ذات قدرات تفوق حواسيب اليوم بمئات أو آلاف المرات وعندها يكون بإمكانها القيام بعمليات تسهل عملية الترجمة الآلية كثيراً. ولكن في ضوء العولمة والحاجة إلى التواصل بين البشر بلغاتهم المختلفة أليس من الأجدى استغلال الإمكانات المتوفرة اليوم؟ وهل استغلت جهود المترجمين من البشر كما يجب لخدمة الترجمة الآلية؟[2]

نظراً لعدم الوصول إلى طريقة مثلى للترجمة حتى الآن وقد ننتظر عقدين أو أكثر من الزمن للوصول إلى طريقة عامة دقيقة للترجمة الآلية، فإن اعتماد أكثر من طريقة واحدة بالتوازي هو ما يحبذ الآن. وهناك أبحاث عديدة تستند إلى اعتماد طرائق متعددة بوقت واحد لترجمة النص نفسه، ثم يُختار منها بعد ذلك. وتعتمد طريقة الترجمة الآلية على نوعية الترجمة المطلوبة من ناحية الدقة والرصانة اللغوية وعلى التكلفة المُرصدة وعلى الزمن المستغرق وفيما إذا كان النظام قابلاً للتحسين في المستقبل أم لا.[3]

---

[1] Gerd Willée, Bernhard Schröder, Hans-Christian Schmitz (eds.) John Hutchins, Computerlinguistik: was geht, was kommt? Computational Linguistics: Achievements and Perspectives. Festschrift fur Winfried Lenders (Sankt Augustin: Gardez! Verlag, 2001), p. 159-162] Machine Translation Today and Tomorrow. www.hutchinsweb.me.uk/Lenders-2002.pdf

[2] Zhuang Xinglai, The Emerging Role of Translation Experts, in the coming MT Era, Translation Journal, Volume 6, No. 4 Oct. 2002

[3] Fuji Ren and Hongchi Shi, Parallel Machine Translation: Principles and Practice. doi.ieeecomputersociety,org/10.1109/ICECCS.2001.930184