المحاضرة الثانية عشر 12 : علم اللغة الآلي الحاسوبي

ولد علم اللغة الحاسوبي (أو اللغويات الحاسوبية) بعد عدد من المحاولات الفاشلة في الترجمة الآلية. وفي أواخر الأربعينات من القرن الماضي، رأت حكومة الولايات المتحدة الحاجة إلى الترجمة الآلية للمعالجة للنصوص، خاصة النصوص الروسية. كان من المفترض، نظرا للسهولة التي تم بها تكييف أجهزة الحاسوب للتعامل مع المشاكل الرياضية، أن تعليمها لترجمة اللغة سيكون مجرد مسألة بسيطة. ولكن سرعان ما أصبح واضحا أن المشكلة كانت أكثر صعوبة، وبعد ذلك، تم إنشاء تخصص علم اللغة الحاسوبي.

في الأيام الأولى من اللغويات الحاسوبية، هيمن علماء الحاسوب على الحقل، ولكن منذ سبعينات القرن العشرين، أصبح واضحًا مدى تعقيد اللغة في الواقع، ولذلك نجد أن علم اللغة الحاسوبي المعاصر يستفيد من الخبراء في مجالات متنوعة.

لطالما كانت الترجمة الآلية هدفًا رئيسيًا في اللغويات الحاسوبية، وهو مجال حقق فيه المجال خطوات هائلة. المهمة معقدة للغاية، وتتطلب تحديد أجزاء الكلام، وفهم قواعد اللغة، ومفردات واسعة، وآليات للتعامل مع العامية واللغة الدارجة. الترجمة الآلية بعيدة كل البعد عن الكمال، ولكن مع كل سنة تصبح الترجمات أكثر دقة وأقل تكلفا.

التعرف على الكلام هو مجال آخر من مجالات علم اللغة الحاسوبي الذي لاقى اهتمام كبيرا. فبعد محاولات قليلة فاشلة في برامج التعرف على الكلام السائدة في منتصف التسعينيات، سارت الأمور بصمت لبعض الوقت. ولكن في أوائل القرن الحادي والعشرين، وصل عدد من حزم برامج التعرف على الكلام الجديدة إلى السوق، التي تميزت بأنظمة تعلم مكثفة ونسب عالية من الدقة. وقد أدى ذلك إلى اهتمام متجدد ببرامج التعرف على الكلام من قبل الجمهور العام وزيادة مصاحبة في التمويل والأبحاث.

توليد الكلام (Speech generation) هو مجال مرتبط باللغويات الحاسوبية وشهد تطورا مطردا منذ الثمانينات. إن الوصول إلى قراءة نصية طبيعية للنص المكتوب يمثل مشكلة صعبة للغاية، ولكنها تنطوي على فوائد محتملة هائلة. بالنسبة إلى المستخدمين العمي ، يمكن أن تكون برامج توليد الكلام حاسمة في الاستمتاع بثمار العصر الرقمي.

يلعب علم اللغة الحاسوبي أيضًا دورًا كبيرًا في أنظمة تصحيح القواعد النحوية الآلية، مثل تلك المدمجة في معظم معالجات النصوص الشائعة. يتطلب المدقق النحوي (grammar checkers) الدقيق قدرة متطورة لتحديد أجزاء الكلام وقائمة شاملة من القواعد النحوية والاستثناءات. في حين أنه لا تزال هناك الكثير من المشاكل في معظم المدققات النحوية الرئيسية، إلا أنها أصبحت لا غنى عنها بالفعل للكثيرين من الجيل الجديد.

علم اللغة الحاسوبي هو مجال مثير يبني على مجموعة واسعة من التخصصات. المشاكل التي يجب معالجتها كثيرة، وكلها ليست بسيطة. ومع ذلك، فإن الرؤى المستقبلية التي يتطلع إليها تجعله نضالًا يستحق المحاولة. فمن حلم المترجم العالمي إلى التعرف على الكلام بشكل مثالي، لا يمكن لأهداف علم اللغة الحاسوبي إلا أن تساعد على إثارة شعوراً بالاندهاش.

هدف علم اللغة الحاسوبي

إن هدف علماء علم اللغة الحاسوبي هو كتابة البرامج التي يمكنها التعامل مع (فهم أو توليد) أكبر قدر ممكن من مواد اللغة. هذه البرامج جيدة ولكنها حلول تقريبية. لا يمكنها التعامل مع جميع الجمل في اللغة الطبيعية، على الرغم من أنها تتعامل مع التراكيب المثيرة والأكثر شيوعًا. هذه الحقيقة، التي يتم قبولها بشكل عام من قبل اللغويين الحاسوبيين، ستكون غير مقبولة للغويين النظريين، لأنها جزء من هدفهم أن تفسر نظريتهم النحوية كل جمل اللغة الصحيحة نحويا.

العمل في اللغويات النظرية له علاقة بعلم اللغة الحاسوبي، لأن جميع جهود التعميم (على سبيل المثال، عن طريق اختزال مجموعة كبيرة من الجمل في مجموعة صغيرة من القواعد والقيود المفروضة على تلك القواعد) التي تم إجراؤها في السابق لوصف قواعد لغة طبيعية معينة، تعتبر ضرورية لعلم اللغة الحاسوبي ، خاصة في عملية تحليل الجملة.

كما أن العمل في اللغويات الحاسوبية له صلة باللسانيات النظرية. إن النظرية التي ترتكز على العمل الأكثر واقعية في اللغويات الحاسوبية ، تنظر إلى فهم اللغة وتوليدها كعمليات للتلاعب بالرموز بطريقة تحكمها القواعد. إن اهتمام اللغويات النظرية بكل جوانب القدرات اللغوية (المعرفة النظرية للغة وكيف يتم استخدامها)، يجب أن يكون قابلاً للاختبار من خلال الأنظمة التي يصممها اللغويين الحسابيين(Grishman 1986 ، .Winograd 1983).

مهمة بناء الأنظمة التي تفهم أو تولد لغة طبيعية هي مهمة معقدة. فهي تتطلب دمج العديد من أنواع البيانات اللغوية (صرفية، بنائية، ودلالية) وغير اللغوية (معرفة مجال الخطاب). كما تتطلب أيضًا الاستخدام الفعال لكل البيانات. من هذا المنطلق، يُعد تصميم وبناء تطبيقات اللغة الطبيعية مهمة هندسية. إحدى الاستراتيجيات العامة لجعل وظائف البناء أسهل هو التقسيم (أي تقسيم المشكلة إلى

مشاكل فرعية أصغر)، وهذا المفهوم ليس غريباً عن علم اللغة، فعادة ما يتم تمثيل القدرة اللغوية ودراستها باعتبارها تعمل في مستويات هيكلية: أصوات وكلمات وجمل. يدرس اللغويون علم الصوتيات، علم الأصوات، علم الصرف، علم الدلالة ، وبنية اللغة ويفترضون وجود مستويات أو وحدات (modules) في القدرة البشرية (على سبيل المثال ، مبدأ بناء الجمل المستقل لتشومسكي). وحتى إن كانت الرؤية المعيارية تبسيط للقدرة اللغوية، فإنها تجعل أنظمة اللغة الطبيعية مرنة ويسهل توسيعها.

يعتمد مقدار المعرفة المستخدمة في عملية الفهم أو التوليد على الغرض من التطبيق. بالنسبة للعديد من التطبيقات، فإن المهمة الأساسية هي تحليل الجمل، (أي تحديد ما تعنيه الجمل). تتطلب بعض التطبيقات أيضًا تحليلًا للوحدات فوق الجمل، مثل الخطاب والحوار.

أصول علم اللغة الحاسوبي

يقول مارتن كاي (Martin Kay, 2003) أنه ربما بدأ علم اللغة الحاسوبي في عام 1949 فيمل يتعلق بالترجمة الآلية. عقد أول مؤتمر حول الترجمة الآلية في عام 1952. وصدرت أول مجلة باسم 'الترجمة الميكانيكية' في عام 1954. ولكن مصطلح 'علم اللغة الحاسوبي' بدأ في الظهور في عام 1965 عندما ظهر كعنوان فرعي للمجلة، 'الترجمة الميكانيكية وعلم اللغة الحاسوبي' حيث كان مكتوبا بخط صغير جدا. وفي عام 1974، تم تغيير اسم مجلة "'الترجمة الميكانيكية وعلم اللغة الحاسوبي' إلى 'المجلة الأمريكية لعلم اللغة الحاسوبية .(American Journal of Computational Linguistics) 'وفي عام 1980، أصبح اسمها "علم اللغة الحاسوبي (Computational Linguistics) 'حيث لا تزال تصدر تحت هذا الاسم.

لقد كان هناك دافعين أساسين وراء أنشطة علم اللغة الحاسوبي:

.1نظري: جاء من التفكير أن اعتماد الأهداف الحاسوبية من شأنه أن يسبب تقدما هاما في علم اللغة.

.2تكنولوجي: جاء من الرغبة في إنتاج التكنولوجيا لخدمة الاحتياجات العملية للترجمة، واستخراج المعلومات، والتدقيق النحوي، الخ.

لا يمكن تحقيق أي من هذه المشاريع من خلال اعتماد الطرق اللغوية وحدها (المرجع السابق نفسه.(

إن علم اللغة الحاسوبي، كحقل، يسبق الذكاء الاصطناعي، وهو الحقل الذي يندرج تحته في الغالب. نشأت علم اللغة الحاسوبي مع الجهود المبذولة في الولايات المتحدة في الخمسينات من القرن العشرين لاستخدام أجهزة الكمبيوتر لترجمة النصوص تلقائيًا من اللغات الأجنبية، وخاصة المجلات العلمية الروسية، إلى الإنجليزية. نظرًا لأن أجهزة الحاسوب يمكنها إجراء حسابات رياضية بشكل أسرع وأكثر دقة من البشر، فقد كان يُعتقد أنها مجرد مسألة وقت قصير قبل أن يتم الاهتمام بالتفاصيل التقنية التي من شأنها أن تسمح لها بنفس القدرة الرائعة على معالجة اللغة.

عندما فشلت الترجمة الآلية (المعروفة أيضًا بالترجمة الميكانيكية) في الحصول على ترجمات دقيقة فورية، عُرف أن المعالجة الآلية للغات البشرية أكثر تعقيدًا بكثير مما كان متوقعًا في الأصل. ولد علم اللغة الحاسوبي كاسم لحقل الدراسة الجديد المكرس لتطوير الخوارزميات والبرمجيات الخاصة بالمعالجة الذكية لبيانات اللغة. عندما نشأ الذكاء الاصطناعي في الستينات، أصبح مجال علم اللغة الحاسوبي قسما فرعيا من الذكاء الاصطناعي محصص للتعامل مع الفهم البشري وإنتاج اللغات الطبيعية.

من أجل ترجمة لغة إلى لغة أخرى، لوحظ أنه على المرء أن يفهم القواعد اللغوية لكلتا اللغتين، بما في ذلك كل من الصرف (قواعد تشكيل الكلمات) والنحو (قواعد بناء الجمل). ولفهم التركيب اللغوي، كان على المرء أيضا أن يفهم الدلالات والمعجم (أو المفردات)، بل وحتى فهم شيء من براغماتية استخدام اللغة. وهكذا، تحول ما بدأ كمجهود للترجمة بين اللغات إلى نظام كامل مخصص لفهم كيفية تمثيل ومعالجة اللغات الطبيعية باستخدام أجهزة الحاسوب.

يتم في الوقت الحاضر إجراء الأبحاث في نطاق علم اللغة الحاسوبي في أقسام ومختبرات علم اللغة الحاسوبي.

الحقول الفرعية لعم اللغة الحاسوبي

يمكن تقسيم علم اللغة الحاسوبي إلى مناطق رئيسية حسب طبيعة اللغة التي تتم معالجتها، سواء كانت منطوقة أو نصية. ووفقا للمهمة التي يجري تنفيذها، سواء تحليل اللغة (التعرف) أو توليف اللغة (التوليد.(

يتعامل التعرف على الكلام وتوليف الكلام مع كيفية فهم اللغة المنطوقة أو إنشاؤها باستخدام أجهزة الحاسوب. التحليل والتوليد هما قسمين فرعيين لعلم اللغة الحاسوبي يتعاملان على التوالي مع تباعد اللغة ووضعها معًا. تظل الترجمة الآلية القسم الفرعي لعلم اللغة الحاسوبي الذي يتعامل مع استغلال أجهزة الحاسوب في الترجمة بين اللغات.

تشمل بعض مجالات البحث التي يدرسها علم اللغة الحاسوبي ما يلي:

- التعقيد الحسابي للغة الطبيعية، الذي تم تصميمه بشكل كبير على نظرية الأوتوماتا[1]، مع تطبيق القواعد النحوية الحساسة للسياق وآلات التغيير ذات الحدود الخطية.
- علم الدلالة الحاسوبي الذي يشمل تعريف المنطق المناسب لتمثيل المعنى اللغوي، وإنشاءه آليًا والتفكير به.
- لغويات المجاميع بمساعدة الحاسوب
- تصميم المعربات أو المجزئات للغات الطبيعية
- تصميم المعلّمات (taggers) مثل معلّمات أجزاء الكلام
- الترجمة الآلية باعتبارها واحدة من أقدم التطبيقات اللغوية الحاسوبية وأكثرها صرامة، وهو تخصص يعتمد على العديد من الحقول الفرعية.

يمتد علم اللغة الحاسوبي بين العلوم المعرفية والذكاء الاصطناعي. كما أن لديه مكونات نظرية وتطبيقية. يعتمد المكون النظري على علم اللغة النظري والعلوم المعرفية. وبما أن النظريات اللغوية أصبحت اليوم معقدة للغاية، فقد قرر علماء اللغة تبسيطها من خلال استخدام أجهزة الحاسوب. وهكذا، بدأوا التعاون مع المبرمجين الحاسوبيين بمساعدة علم النفس المعرفي من أجل تطوير نماذج حسابية للنظريات اللغوية الرسمية. يطور المكون التطبيقي نماذج عملية للغة البشرية اعتمادًا على الذكاء الاصطناعي. يمكن وضع هذا أيضا تحت عنوان 'هندسة اللغة' أو 'تقنية اللغة البشرية'. إن وظيفة علم اللغة الحاسوبي التطبيقي هي إيجاد برنامج يمكنه تحسين التفاعل بين الإنسان والآلة حتى يتمكن الإنسان والحاسوب من التواصل بسهولة.(Uszkoreit, 2000)

فهم اللغة الطبيعية

يشتكي اللغويون من الافتقار للذكاء الاصطناعي لدرجة أنهم مجبرون على البحث عن بديل يُعرف باسم 'معالجة اللغات الطبيعية'. وبما أن فهم اللغة الطبيعية هو مرادف لعلم اللغة الحاسوبي، فمن المهم معرفة أهدافه الأساسية التي تشمل الآتي:

- توليد وإنتاج المحتويات بأي لغة طبيعية في أي مجال.
- دعم الخدمات متعددة اللغات.

هناك مسارين من البحث المتعلق بفهم اللغة الطبيعية. أحدها موجه نحو التطبيقات القائمة على النص والآخر نحو التطبيقات القائمة على الحوار. تتضمن التطبيقات القائمة على النصوص معالجة جميع أنواع النصوص المكتوبة، مثل الكتب والمقالات والرسائل والمجلات وغيرها، بطريقة تُمكن المستخدمين من قراءتها بسهولة. وهكذا، فإن الباحثون مستمرين في تطوير وسائل مختلفة للوصول إلى مثل هذه النصوص. تتمثل بعض هذه التطبيقات في:

- العثور على الوثائق الملائمة حول الموضوع المطلوب من قاعدة بيانات النصوص، مثل العثور على الكتب المناسبة في المكتبة.
- استخراج المعلومات من نوع ما من النصوص حول موضوع معين، مثل بناء قاعدة بيانات لجميع المعلومات الموجودة على الانترنت المكتوبة في الأخبار في يوم معين.
- ترجمة المستندات من لغة إلى أخرى.
- تغيير النصوص الطويلة إلى ملخصات قصيرة.

أما التطبيقات القائمة على الحوار فتشمل:

- أنظمة الرد على الأسئلة.
- خدمة العملاء الآلية عبر الهاتف.
- نُظم التعليم التي يمكن للطالب التفاعل فيها مع الجهاز.(Bolshakov; 2004)

معالجة اللغة الطبيعية ومشاكلها:

أحد أهم المشاكل التي يعاني منها المحققون هي سوء فهم الآلات للغات الطبيعية. معالجة اللغة الطبيعية معقدة لأنه سيكون هناك سوء فهم في تطبيق البرامج الحاسوبية المتعلقة بما يلي:

.1علم الأصوات والصوتيات الذي يهتم بالنطق. مشكلة البرامج الحاسوبية المتعلقة بهذا المجال هي أن بعض الكلمات لها نفس النطق ولكن بمعان مختلفة مثل كلمتي *'weak'* و *'week'* الإنجليزية، حيث لا يمكن لأجهزة الحاسوب التفريق بين الكلمتين.

.2المورفولوجيا أو علم الصرف الذي يهتم بالتركيب الداخلي للكلمات في شكلها المكتوب والمنطوق. للصرف وظيفتان أساسيتان هما:

- التصريف :(Inflection) يتعلق الأمر بالوظيفة النحوية للكلمات في نفس الجزء من الكلام ؛ كما في نموذج الفعل 'work' في اللغة الانجليزية:
  - workفي المضارع البسيط ، للمتكلم والمخاطب.
  - worksفي المضارع البسيط ، للمفرد الغائب.
  - workedفي الماضي البسيط والماضي التام.
  - workingفي المضارع المستمر.
- الاشتقاق: يتعلق بإنتاج كلمات جديدة لأجزاء مختلفة من الكلام ؛ كما في المثال التالي:
  -) ——————— Globeاسم(
  -) -——————— Globalصفة(
  -) -——————— Globalizeفعل(

يجب أن يكون المحلل المورفولوجي (الصرفي) ذكيًا بما فيه الكفاية لمعرفة واستخراج النماذج الأساسية من المستندات المدرجة في أجهزة الحاسوب. التطبيقات التي تم إنجازها في هذا الصدد هي:

- الوصل (hyphenation) تقسيم الكلمات إلى أصغر وحداتها.(morphs)
- تصحيح الإملاء.
- التجذير (stemming) الذي يقلل من الكلمات المرتبة قدر الامكان.

مشكلة هذه البرامج الحاسوبية هي المدخلات التي ينبغي أن تكون واسعة جدا. تشمل الأشكال الأخرى للتطبيقات الإعراب وتوليد كلمات اللغة الطبيعية في صيغة مكتوبة أو منطوقة، والترجمة الآلية.(Trost, 2006)

.3بناء الجمل (syntax) الذي يتعلق بتراكيب الجمل. ففي بعض الأحيان يكون ترتيب الكلمات في بعض أنواع التراكيب مضللا كما في الأمثلة التالية:

(1) I saw her *with* a telescope.

قد تكون كلمة *'with'* إما مساعدة أو ملحقة مع الفعل *"saw"* الفعل أو مع الضمير.*"her"*

(2) The article covers the rights of women *and* childhood.

تُفهم أداة الربط *"and"* على أنها تربط الاسمين "women" و ."childhood" لكن قد تُفهم أيضًا على أنها تقرن عبارة *"the rights of women"* بالاسم*"childhood"* ، وهذا غير صحيح. تعتبر هذه واحدة من المشاكل التي تواجه الترجمة بالحاسوب.

.4علم الدلالة (semantics) الذي يتعامل مع معاني الكلمات والعبارات والجمل. ولكن نظرًا لأن الكلمة قد تحتوي على عدة معانٍ مثل كلمة *"cover"* التي تعني *"to hide"*، *"to spread over"* أو*"to deal with"* فإنها ستسبب أيضًا مشكلة في الترجمة بواسطة الحاسوب.

.5البراغماتية (pragmatics) التي تتعامل مع معاني الكلام حسب السياق. في كثير من الأحيان تكون معاني كلمات الجملة واضحة لكن التفسير يعتمد على سياقها. فمثلا:

(3)نحن في الانتظار.

قد تحمل الجملة أيًا من أنواع التفسير المختلفة وفقًا لسياقها: أ. حقيقة عادية ، ب. وعد و ج. تهديد.

لا يمكن أن تميز الترجمة الحاسوبية بين تلك الأنواع من التفسيرات التي تصف ظاهرتي التهكم (قول عكس القصد) أو المجاز (الذي يظهر التعبيرات التي لا توجد حرفياً) ظواهر كما في الأمثلة التالية:

(4)أنت ذكي بما فيه الكفاية لتحقيق ذلك. (ساخر(

(5)ان رأسه كالحائط. (مجازي(

تظهر المشاكل المذكورة أعلاه أنه لا يمكن حلها عن طريق علوم الحاسوب وحدها ولا عن طريق اللغويات وحدها (Wintner, 2004).