الباب الثانى عشر

الحاسبات وهندسة اللغة

١٢-١ مقدمة عامة

١٢-٢ الإطار العام لهندسة اللغة

١٢-٣ الترجمة الآلية

١٢-٤ التعامل مع الكلام

الباب الثانى عشر

الحاسبات وهندسة اللغة

١٢-١ مقدمة عامة

مع انتشار الحاسبات أصبح من الأفضل التعامل معها باللغات الطبيعية سواء كانت مكتوبة أو منطوقة . كذلك أتاحت شبكة الإنترنت إطارا عالميا للتفاعل بين الأشخاص ، وأزالت الحواجز الجغرافية والسياسية المختلفة ، ولكن مازال حاجز اللغة يمثل صعوبة عند التعامل فى إطار هذا الفضاء المعرفى الواسع . ويظهر ذلك من توزيع المعلومات والمعارف على اللغات المختلفة الذى يشير إليه الشكل (١٢-١) ، والذى يوضح العشرة لغات الأكثر استخداما على الإنترنت [Fox, 1997] .

اللغة	الإنجليزية	الألمانية	اليابانية	الفرنسية	الإسبانية	الإيطالية	البرتغالية	السويدية	الهولندية	النرويجية
نسبة الاستخدام	٨٢,٣	٤,٠	١,٦	١,٥	١,١	٠,٨	٠,٧	٠,٦	٠,٤	٠,٣

شكل (١٢-١) : أكثر ١٠ لغات استخداما على الإنترنت

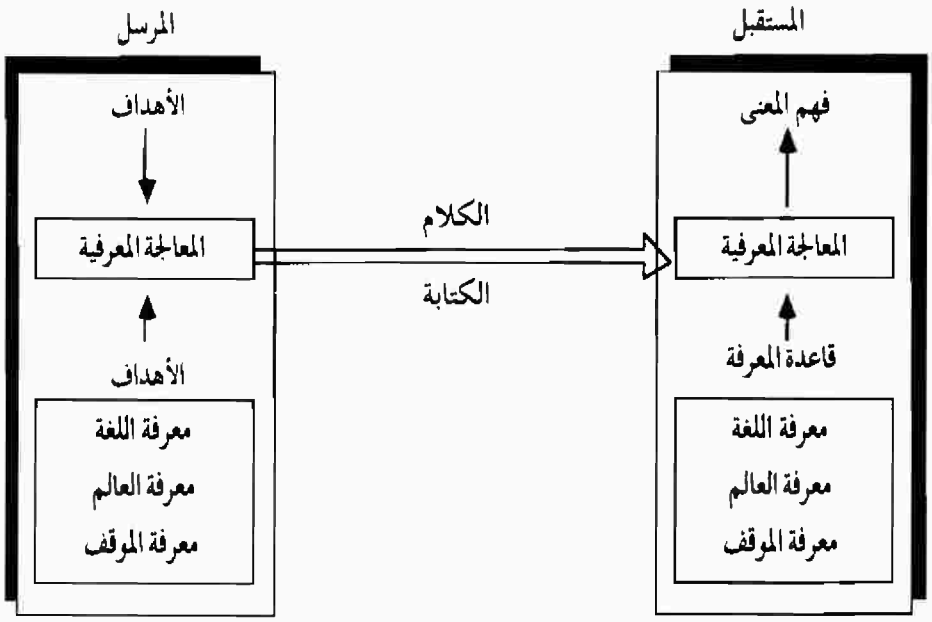
ومع زيادة كم المعلومات المتاحة ازدادت صعوبة التمييز بين المعلومات واختيار أنسبها بالنسبة للمستخدم . لذلك ظهرت حاجة ماسة لتصميم وحدات للربط تقوم بترجمة وعرض المعلومات باللغة الملائمة ، سواء كانت مكتوبة أو منطوقة مع الأخذ فى الإعتبار الخلفية الثقافية للمستخدمين . كل ذلك يتطلب تطوير مكونات نظم الحاسبات من خلال تصميم وتنظيم المنتجات التى تحقق هذا الهدف والتغلب على المشاكل الفنية المتعددة . وقد أدى كل ذلك إلى ظهور مجال مهم فى إطار تكنولوجيا المعلومات يركز على «هندسة اللغة» (Language Engineering) . وقبل عرض الإطار العام لهندسة اللغة وتطبيقاتها المختلفة ، سنقدم عرضاً موجزا لأحد نماذج اللغات الطبيعية الذى ينظر إلى اللغة على أساس أنها وسيلة للاتصال بين الأشخاص وتبادل الأفكار ، وذلك من منظور معرفى .

هناك بعض المدارس الفكرية فى دراسة اللغات تنظر إليها من زاوية معرفية . وعلى ذلك فإن الأسئلة الأساسية التى يبحثون عن إجاباتها يمكن أن تكون :

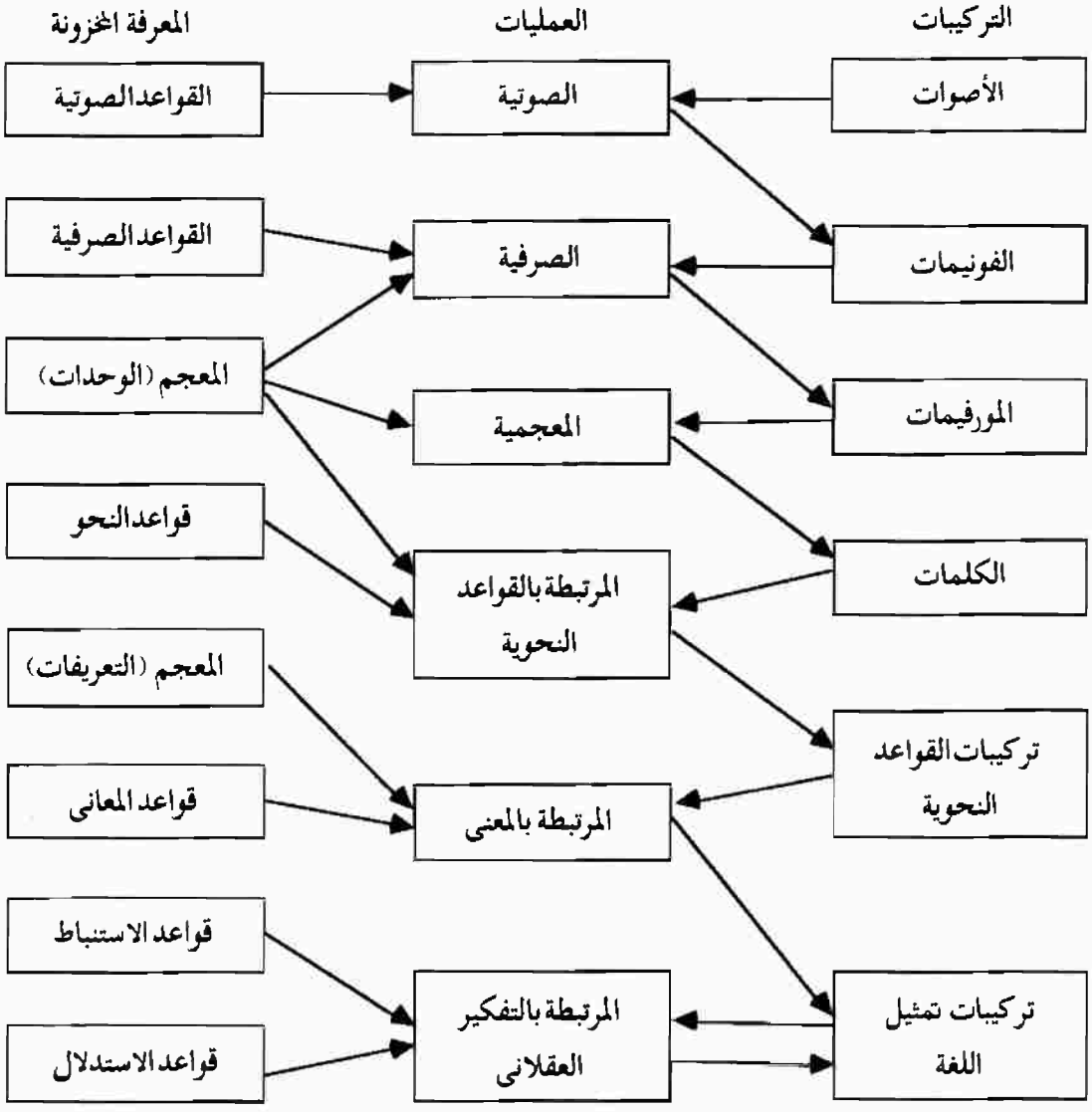
ما المعرفة التى يجب أن تكون عند الإنسان حتى يمكنه تكلم وفهم لغة معينة ؟ كيف يرتب العقل نفسه بحيث يستطيع استخدام هذه المعرفة فى عملية الاتصال مع الآخرين ؟ معنى ذلك أن هذا الاتجاه يعتبر أن اللغة عملية اتصالية مبنية على المعرفة . وعلى ذلك يمكن تعريف اللغة على أنها عملية اتصال بين وحدات ذكية نشطة ؛ حيث يقوم كل من المرسل والمستقبل بإجراء عمليات معرفية معقدة . فالمرسل يبدأ بتحديد هدف الاتصال والأثر أو الآثار المرجوة من هذا الاتصال ، والمعلومات المطلوب توصيلها والحالة المطلوبة التعبير عنها [Dik, 1989] .

بعد ذلك يعمل المرسل على تحويل هذه الأهداف إلى إشارات صوتية أو أشكال أو كتابة . وعند وصول هذه إلى المستقبل ، فإنه يقوم بإجراء عملية عكسية للاستدلال على الرسالة الأصلية التى قصد المرسل إبلاغها له . ويوضح الشكل (١٢-٢) النموذج الأساسى للاتصال عن طريق اللغة . ولتسهيل دراسة اللغة فقد تم تقسيمها إلى عدد من المستويات أو الطبقات ، يعتبر كل منها مستقلا عن الآخر عند الدراسة ، ولكنها لابد أن تتكامل فى النهاية لتشكل منظومة لغوية متكاملة . ويوضح الشكل (١٢-٣) النموذج الطبقى لفهم أو استقبال اللغة . وتبعا لهذا النموذج فإن المعرفة عن لغة معينة تتكون من القواعد الخاصة بالتعامل مع كل مستوى من البناء الطبقى . فمثلا عندما يستمع شخص إلى جملة معينة ، فإن تتابع الأصوات التى يسمعها تشكل المستوى الأول . وتتولى العملية الصوتية تحليل هذه الأصوات وتحويلها إلى تتابع من «الفونيمات» (Phonemes) .

ثم تتولى العملية الصرفية تحويل الفونيمات إلى مجموعة من «المورفيمات» (Morphemes) وتتولى العملية المعجمية إيجاد العلاقة بين المورفيمات والكلمات ثم تبدأ العمليات المرتبطة بالقواعد النحوية بتحليل الجمل وهكذا حتى نصل إلى المستوى المرتبط بالتفكير العقلانى . وبالطبع فإن المرسل يستخدم المستويات نفسها، ولكنه يبدأ من أسفل إلى أعلى [Winograd, 1983] [عـلـى ، ١٩٨٨] [تشومسكى ، ١٩٩٣]



شكل (١٢-٢) : النموذج الأساسى للاتصال عن طريق اللغة .



شكل (١٢-٣) : النموذج الطبقى لفهم أو استقبال اللغة .

١٢-٢ الإطار العام لهندسة اللغة

تتعلق هندسة اللغة بتصميم وتنفيذ المنظومات المطلوبة للتعامل بفعالية مع المعلومات والمعارف عبر اللغات المختلفة ؛ فتعرف الكلام المنطوق بلغات ولهجات متعددة وترجمته إلى لغة ولهجة المستخدم سيتيح له التفاعل بشكل أكثر إيجابية مع المعلومات المنطوقة . كذلك يمكن أن يتم توليد اللغة المسموعة من النص المكتوب أو توليد وعرض المعلومات بلغات متعددة يتم إختيارها من جانب المستخدم . ويتطلب تنفيذ ذلك تكامل عدة مجالات فى منظومة واحدة تشتمل على المكون اللغوى والثقافى والجامد واللين والشبكى وغيرها .

كما تعتبر هندسة اللغة عنصرا أساسيًا فى دعم التجارة الإلكترونية العالمية . فنجاح المؤسسات المختلفة فى هذا المجال يعتمد على جودة المعلومات المقدمة إلى الزبائن عن المنتجات المعروضة . وترجمة المعلومات من الناحية اللغوية فقط لا يكفى بل يستلزم الأمر أيضا ترجمة «ثقافية» للمحيط الذى يوجد فيه الزبائن . ولذلك يجب أن تهتم الشركات بالنقاط التالية [Ghonaimy, 1998,2] .

١- تحديد «المكونات اللغوية» فى المنتجات ؛ حتى يمكن معرفتها بسهولة واتخاذ اللازم نحو «عولمتها» .

٢- سهولة ترجمة بيانات المنتجات إلى لغات متعددة .

٣- إضافة البعد الثقافى إلى لغات الهدف .

وكل هذه النقاط تتطلب التطوير الفنى لنظم البرمجيات والمعلومات لضمان تحقيق التعددية اللغوية المطلوبة [Ghonaimy, 1998,3] .

وقد وضعت بعض التجمعات الدولية إطارا للبحث والتطوير فى هندسة اللغة يتضمن الموضوعات التالية :

الحسابات المعجمية (Computational Lexicons) .

الصياغات النحوية .

معالجة اللغة المنطوقة .

منهجيات التقييم سواء للمصادر اللغوية أو الأدوات أو المنتجات النهائية .

نظم الترجمة الآلية بمساعدة الحاسبات .

إنشاء مجموعة النصوص الكبيرة (Large Corpora) ومراجعتها وإتاحتها على الوسائط الإلكترونية أو الشبكات .

وسنقدم فيما يلى موجزا لموضوع الترجمة الآلية ، وكذلك موضوع التعامل مع الكلام .

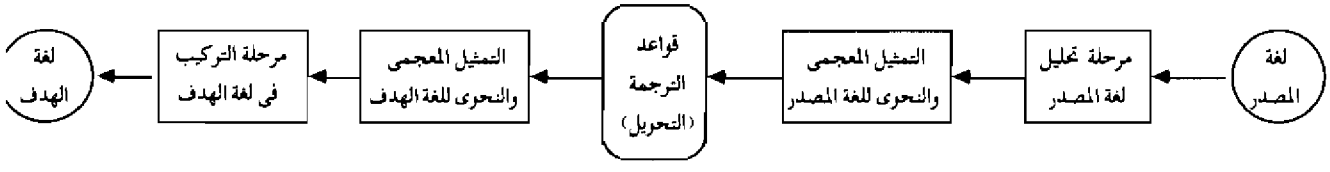
٣-١٢ الترجمة الآلية

هناك مجهودات كبيرة تبذل فى مجال الترجمة الآلية بين اللغات المختلفة ، وتستخدم لذلك طرق متعددة . بعضها يعتمد على المعرفة الخاصة بمجال الترجمة وتسمى «الترجمة الآلية المبنية على المعرفة» (Knowledge-based Machine Translation) [Nirenburg, 1992] . وبعضها يعتمد على المعاجم (Lexicons) [Guthrie, 1996] . كما يعتمد بعضها على الطرق الإحصائية [Brown, 1993] والتى تتطلب وجود مجموعات نصوص كبيرة بالنسبة لكل لغة على حدة أو فى حالة وجود نصوص كبيرة بلغتين مثل مجموعات نصوص (Hansard) الخاصة

بالبرلمان الكندى والتـــى تحتوى على أكثر من ١٠٠ مليون كلمة باللغتين الإنجليزية والفرنسية ، يمكن الاستفادة مـــن ذلك فى عملية الترجمة مثل مشروع «شمبليون» (Champollion) [Smadja, 1996] .

وهناك مشروعات أخرى تم تنفيذها مثل مشروع (PRINCITRAN) والذى يقـــوم بالترجمة بين اللغات الآتية : العـــربية والإنجليـــزية والكورية والإسبانية [Dorr, 1995] . وقد استخدم هذا المشروع أحد القواميس الثنائية الإلكترونية يسمى (Alpnet) للترجمة من اللغة العربية إلى اللغة الإنجليزية بالإضافة إلى إستخدام الأكواد اللغوية الموجودة فى أحد القواميس الإلكترونية الذى يسمى (Longman's Dictionary Of Contemporary English) (LDOCE) . هذا بالإضافة إلى المجهودات التى تقوم بها بعض الشركات للترجمة من اللغة الإنجليزية إلى اللغة العربية [جهاد عبد الله ، ١٩٩٦] ولكنها تحتاج إلى عملية تقييم مكثفة من جانب المستخدمين ؛ نظرا لأن عملية تقييم نظم معالجة اللغات الطبيعية تشكل عنصرا أساسياً فى تطويرها [King, 1996] .

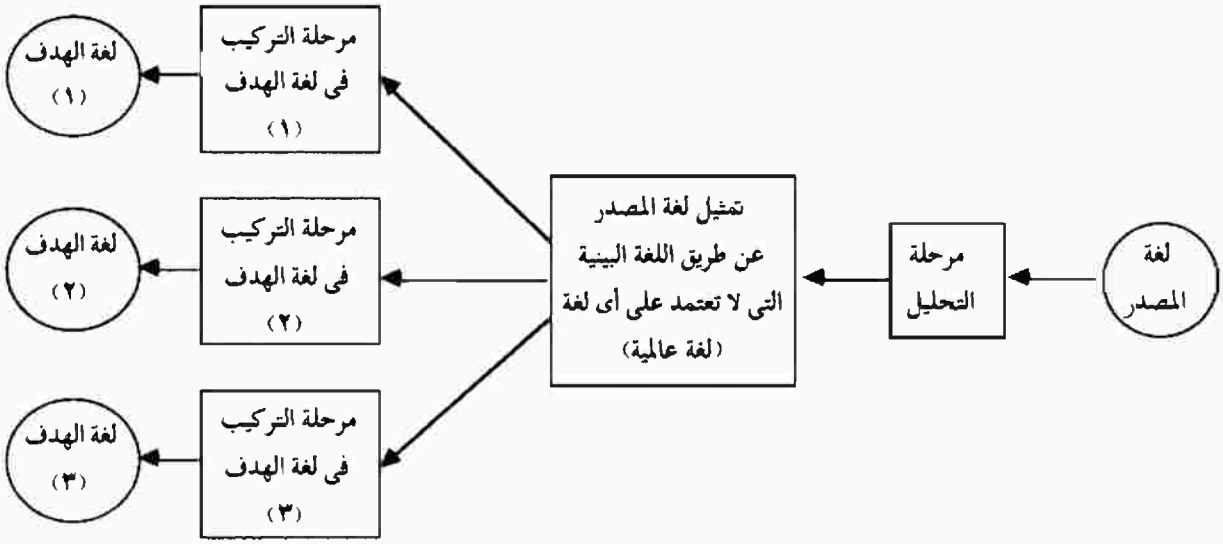
وبوجه عام هناك طريقتان أساسيتان للترجمة الآلية إحداهما ، تعتمد على «النموذج التحويلى» (Transfer model) والأخرى تعتمد على نموذج «اللغة البينية» (Interlingua) . فى النموذج التحويلى يتم أولا «إعراب» (Parsing) جملة معينة فى «لغة المصدر» (Source Language) إلى صورة داخلية مجردة من خلال مرحلة التحليل ؛ حيث يتم تحديد الخصائص المعجمية والنحوية للجملة . بعد ذلك تتم عملية «التحويل» إلى «لغة الهدف» (Target Language) من خلال إيجاد التمثيل المعجمى والنحوى لها باستخدام قواعد التحويل ، ثم الترجمة النهائية من خلال مرحلة التركيب (Synthesis) كما هو موضح فى الشكل (١٢-٤) .



شكل (١٢-٤) : منظومة الترجمة الآلية باستخدام النموذج التحويلى .

وفى النموذج الذى يعتمد على اللغة البينية كما هو موضح فى شكل (١٢-٥) ، يتم أولا تحويل لغة المصدر من خلال مرحلة التحليل إلى اللغة البينية وهى لغة عالمية لا تعتمد على أية لغة ، وبعد ذلك تتم عملية التركيب بالنسبة لكل لغة من لغات الهدف . ويجب ملاحظة أن نموذج اللغة البينية يصلح بشكل أفضل عند الترجمة بين عدد كبير من اللغات مثل ترجمة الوثائق والمعلومات بين دول

الاتحاد الأوروبى على سبيل المثال . فباستخدام هذا النظام يتم توفير مجهود كبير فى عملية تنفيذ برامج الترجمة [Nirenburg, 1987] .



شكل (١٢ - ٥) : منظومة الترجمة الآلية باستخدام نموذج اللغة البينية .

ويجب ملاحظة أن عملية الترجمة الآلية قد تحتاج فى معظم الأحيان إلى تدخل «المترجم البشرى» ، إما فى تجهيز النص لعملية الترجمة وتسمى عملية «التحرير القَبْلى» (Pre-editing) أو فى تنقيح ناتج الترجمة من خلال عملية «التحرير البَعْدى» (Post-editing) وذلك نظرا لصعوبة الترجمة الآلية الكاملة فى الوقت الحالى . ويمكن بالطبع لنظم الحاسبات مساعدة المترجم البشرى أيضا فى هذه العمليات .

١٢-٤ التعامل مع الكلام

يعتبر الكلام أكثر الوسائل التى تلائم الاتصال بين الأشخاص . وقد كانت اللغة المنطوقة أسبق فى الظهور من اللغة المكتوبة . لذلك بدأ الاهتمام باستخدام نظم الحاسبات فى معالجة اللغة المنطوقة حيث إن هناك تطبيقات متعددة فى هذا المجال . وتشتمل تطبيقات معالجة الكلام على سبيل المثال على مايلى [Rabiner,1994] :

١- تكويد الكلام (Speech Coding) ويتضمن ذلك ضغط الإشارات الكلامية بحيث يمكن تخزينها على الوسائط الإلكترونية أو نقلها عبر الشبكات بصورة أفضل .

٢- تركيب الكلام (Speech Synthesis) ويتم ذلك عادة من خلال تحويل نص مكتوب إلى نص منطوق ، ولذلك يسمى أيضا «ترجمة النص المكتوب إلى كلام» (Text-to-Speech Translation) .

٣- تعرف الكلام (Speech recognition) ويشتمل ذلك عادة على معرفة

الكلمات أو التعبيرات القصيرة المنفصلة ، والتى قد تستخدم على سبيل المثال فى تشغيل بعض الأجهزة أو إعطاء الأوامر البسيطة للحاسبات .

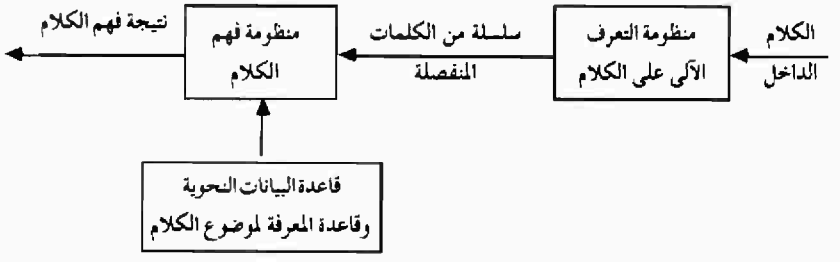
٤- فهم الكلام (Speech understanding) ويشتمل على فهم الكلام المتصل ويتطلب ذلك وجود قاعدة بيانات نحوية للغة معينة ، بالإضافة إلى قاعدة معرفية خاصة بموضوع أو مجال الكلام .

٥- تعرف المتكلم (Speaker recognition) ويتعلق بتمييز المتكلم والتحقق من هويته ويمكن الاستفادة من ذلك فى أنظمة تأمين الحاسبات والشبكات . كما يمكن أيضا استخدامه فى تمييز المتحدثين عند وجود تسجيلات صوتية لأكثر من متكلم فى أثناء المحادثات أو المؤتمرات .

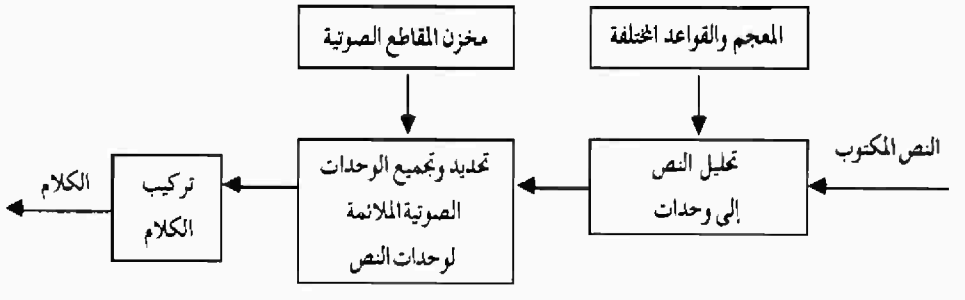
٦- الترجمة الآلية للغة المنطوقة (Automatic Translation of Spoken Lan-guage) وهناك نشاط كبير فى هذا المجال لاستخدامه على وجه خاص فى المحادثات التليفونية عبر اللغات المختلفة .

٧- تمييز اللغة المنطوقة (Spoken Language Identification) ويتعلق ذلك بتمييز اللغة التى يتحدث بها شخص معين من خلال التسجيلات الصوتية .

ويوضح الشكل (١٢-٦) مكونات نظام فهم الكلام [Juang, 2000] كما يوضح الشكل (١٢-٧) مكونات نظام لتحويل النصوص إلى كلام [Cox, 2000] .



شكل (١٢-٦) : مكونات نظام فهم الكلام .



شكل (١٢-٧) : الإطار العام لنظام تحويل النص المكتوب إلى كلام .

وهناك مشروعات متعددة على المستوى العالمى لدعم البحوث فى هذا المجال ، أحدها يسمى «منظومة الكلام المنطوق» (Spoken Language System) (SLS) «وكالة مشروعات البحوث المتقدمة فى مجال الدفاع» (Defense Advanced Research Projects Agency) (DARPA) . وقد إبتدأ هذا المشروع فى نهاية الثمانينيات من القرن العشرين ويركز أساسا على اللغة الإنجليزية . كذلك هناك مشروع الإتحاد الأوروبى المسمى «فهم الكلام والمحادثات» (Speech Understanding and Dialogue) SUNDIAL والذى يركز على أربع لغات هى الإنجليزية والفرنسية والألمانية والإيطالية . كما أن هناك مشروعات أخرى تتم فى الجامعات المختلفة ، وتتعامل مع عدد أكبر من اللغات مثل مشروع «التليفون العالمى» (Global Phone) والذى يتم تنفيذه فى «معمل المنظومات التفاعلية» (Interactive Systems Laboratory) بجامعة «كالسروه» (Karlsruhe) بألمانيا ، ويتعرف الكلام بالنسبة لعدد كبير من اللغات ، يصل عددها إلى ١٢ لغة من بينها اللغة العربية [Waibel , 2000] .

كما أن هناك عديداً من التطبيقات التى تهتم بها بوجه خاص شركات تقديم الخدمة على شبكة الانترنت ، توجز بعضاً منها فيما يلى [Cox, 2000] :

١- تسهيل طرق الاتصال الصوتى بالأدلة (Directories) والخدمات المختلفة المتاحة على الشبكة .

٢- تمييز الإتصالات التليفونية الواردة لشخص ما .

٣- التحويل بين الوسائط المختلفة مثل تحويل البريد الإلكترونى إلى كلام أو الرسائل الصوتية إلى نصوص تعرض على الشاشة .

٤- تدوين الملاحظات آليا أثناء المحادثات .

٥- الترجمة الآلية فى الزمن الحقيقى .

٦- مساعدة ذوى الاحتياجات الخاصة بالنسبة للسمع والرؤية والكلام .