

Crónica

Las aplicaciones de las tecnologías del habla

Applications of speech technology

Mar Mediavilla

Una exposición paralela a la segunda edición del curso sobre Tecnologías de la lengua mostró los trabajos que realizan diferentes empresas y equipos de investigación en este ámbito. Traducción asistida por ordenador, conversión texto-habla, bases de datos orales y textuales, traducción automática o proyectos en el campo de lingüística computacional fueron algunos de los temas que despertaron el interés del público hacia este novedoso campo.

En los sistemas de traducción automática se establecen los *paradigmas lingüísticos* y *no lingüísticos*. Estos últimos son aquellos que traducen palabra a palabra y utilizan técnicas estadísticas para correlacionar palabras en los idiomas de origen y destino y determinan la traducción más probable para las palabras o las secuencias de éstas. Dentro de ellos, algunos contienen una amplia base de datos de textos paralelos en los que se realiza una comparación entre la frase de entrada con las almacenadas en un sistema.

Los sistemas lingüísticos, que utilizan en su tarea un análisis profundo de términos, estructura de frases, rasgos semánticos y contexto gramatical, pueden clasificarse como *interlingua* o *transfer*. En *interlingua*, la entrada se transfiere a una representación artificial, que posteriormente es traducida al idioma de destino. Requieren de análisis semántico y representación abstracta de términos muy profundos, que supone un obstáculo para su utilización a gran escala. Los que emplean la técnica de *transfer* se basan en un módulo central de la traducción, un módulo de transferencia que transforma la entrada suministrada por un módulo de análisis y la entrega a un módulo de generación que proporciona frases gramaticalmente correctas en el módulo de destino.

La tecnología *comprendium* fue el motor de traducción expuesto por la empresa Sail Labs.¹ Según sus promotores es un sistema basado en reglas lingüísticas que realiza un profundo análisis morfológico, gramatical y léxico del texto de entrada construyendo una salida según un modelo lingüístico. Es sensitivo al contexto, esto es, el significado de un término depende de su función sintáctica dentro de la frase, e incluye también un análisis semántico. *Comprendium*, basado en un modelo modificado de *transfer*, realiza la traducción en un proceso de tres fases: análisis, transferencia y generación. En la actualidad, este sistema está disponible, entre otros idiomas para inglés-francés o alemán-español, alemán-francés-español, francés-español, catalán-inglés y ruso-alemán-inglés.

Muchas veces se ha pensado que las máquinas podrían sustituir la tarea humana, pero sin duda antes está el paso previo en el que la máquina, o los servicios que a través de ella se ofrecen, se convierten en herramienta. Un ejemplo es el programa de traducción asistida por ordenador (TAO) denominado *Déjà vu*. Sus distribuidores, la empresa Ampersand,² explican que se trata de algo más que un programa «de memorias de traducción», ya que es un entorno integrado de traducción asistida por ordenador que incluye en un mismo programa componentes de un editor de textos multiformato, una memoria de traducción, un gestor de terminología, un alineador de textos y utilidades complementarias con un importante soporte técnico. Este tipo de programas almacenan todas las traducciones que se realizan en una base de datos denominada *memoria de traducción*. Se trata de un sistema integrado de TAO, es decir, no sólo permite al traductor aprovechar sus traducciones anteriores, sino que además facilita el trabajo adicional vinculado al proceso de traducción. El programa formatea el texto, dando a la traducción la misma forma que el original, permite trabajar en red con acceso a bases de datos de varios usuarios que, al mismo tiempo, pueden enviar nueva información a la base de datos, junto con un programa de gestión de terminología.

Por último, dentro de este primer apartado, en la exposición se presentó un software destinado a humanizar la interacción con el usuario. La base de este software, desarrollado por la empresa Q-

go,³ en colaboración con universidades y centros de investigación, es la tecnología de navegación por preguntas y respuestas basadas en el procesamiento del lenguaje natural. A quien navega por un sitio web ofrece la oportunidad de plantear preguntas modelo con enlaces directos a determinados contenidos. La tecnología lingüística se basa en la comparación entre la pregunta de quien visita la página con las preguntas modelo de la base de datos. Este software está disponible en alemán, holandés, español e inglés, y está trabajando con los modelos en francés, italiano, portugués y catalán. Esta empresa también desarrolla sistemas de recuperación de información en la WWW.

Automatic Trans,⁴ empresa integrada por especialistas en el lenguaje y la tecnología del conocimiento, presentó los traductores denominados *Automatic Translation Server*, que actualmente se comercializan para traducir del castellano al catalán y viceversa, además de castellano a portugués de Brasil y su próximo lanzamiento del traductor automático de inglés a castellano. Una de las principales virtudes de este producto es su capacidad de aprendizaje, ya que el sistema recoge las dudas y errores durante la traducción, y esa información segmentada y agrupada se presenta al responsable del sistema de traducción, quien se encargará de añadir entradas nuevas a cualquiera de los cuatro niveles existentes: palabras nuevas, excepciones, expresiones o frases hechas y reglas de traducción. Desde la empresa apuntan que con un mantenimiento regular, Automatic Translation Server puede alcanzar niveles del 99,99 % de calidad en la traducción automática, con un lenguaje normalizado en estilo correcto y con expresiones perfectamente construidas y de total comprensión para el lector.

Investigación y desarrollo

En la exposición se presentaron también las principales líneas hacia las que se dirige la investigación en el campo de las tecnologías del habla. Por ejemplo, Telefónica I+D (TID),⁵ según explica Juan María Garrido, destaca entre sus ámbitos el reconocimiento de habla, la conversión texto-habla y el procesamiento del lenguaje natural. En el campo del reconocimiento del habla, la tecnología desarrollada por TID permite reconocer frases pronunciadas de forma continua, sin pausas entre las palabras, y con independencia del locutor que las pronuncie. Esta tecnología está integrada en el llamado *Reconocedor de Lenguaje Natural*. Frente a otros sistemas que sólo permiten el reconocimiento de palabras o enunciados cortos pronunciados aisladamente, el Reconocedor de Lenguaje Natural, comenta Garrido, permite el diálogo hombre-máquina de forma más natural y con menos restricciones. Además, incorpora procedimientos que permiten minimizar la influencia del ruido de fondo y de las diferencias debidas al canal (aparato telefónico empleado o red de transmisión). En el terreno de la conversión texto-habla, el *Conversor Multilingüe* de TID permite la lectura de textos con diferentes «locutores» (voces sintéticas) en distintos idiomas: español peninsular y de América, catalán, euskera, gallego, portugués europeo y portugués brasileño. La arquitectura del sistema permite emplear el Conversor a través de diferentes canales telefónicos, utilizando «locutores» distintos, incluso de idiomas diferentes de forma simultánea. La investigación actual en este campo se orienta hacia el desarrollo de nuevos «locutores» de forma cada vez más automatizada y rápida, la creación de «locutores» sintéticos que reproduzcan lo más fielmente posible las especificidades fónicas de los locutores humanos de referencia, y la corrección automática del texto de entrada del Conversor.

En un último apartado, el trabajo en procesamiento del lenguaje natural se centra, por un lado, en el desarrollo de modelos de lenguaje (gramáticas contextuales que definen las combinaciones posibles de palabras en una situación determinada), que se emplean como apoyo al Reconocedor y, por otro, en el desarrollo de analizadores semánticos que permitan la interpretación semántica de los enunciados reconocidos. Una tecnología suficientemente madura en estos tres campos ha permitido abordar el desarrollo de los denominados *sistemas de diálogo* capaces de reconocer la demanda de un usuario por medio de la voz, interpretar y generar una respuesta (fig.1).

Una de las aplicaciones con más proyección son los «portales de voz», sistemas que permiten acceder por medio de la voz, es decir a través del teléfono sin intervención de operador humano, a los contenidos y servicios disponibles en Internet. Éstos permitirán lectura de correo electrónico, accesos a prensa electrónica, directorios, bases de datos y comercio electrónico, entre otras funciones.

Representación universitaria

En la Unidad de Investigación en Lingüística Computacional (UILC), del Instituto Universitario de Lingüística Aplicada de la Universitat Pompeu Fabra (UPF, Barcelona),⁶ se desarrolla un sistema de procesamiento de texto no restringido para el catalán, con el fin de anotar textos de manera automática con información lingüística. Según explican desde la UILC, la información facilitará la explotación de los textos en tareas relacionadas con el campo de la Lingüística Computacional, en el sentido amplio del término: desde corrección gramatical hasta extracción y recuperación de información, pasando por traducción automática (fig. 2).

Según el diseño de la figura 2, la primera etapa por la que pasa el texto, en formato ANSI, es el *preproceso*, en el que se identifica cada palabra, oración y párrafo, se detectan unidades tales como nombres propios y fechas y se da al texto un formato adecuado para el siguiente paso: la *proyección morfológica*. Cada palabra del texto se etiqueta con todos los posibles lemas y categorías morfológicas que puede representar. Por ejemplo, tanto si se procesa *la finestra és petita* [la ventana es pequeña] como si se procesa *me la menjaré* [me la comeré], *la* recibirá etiqueta de artículo y de pronombre, que son las dos categorías que puede tener este vocablo. En la siguiente fase, la de *desambiguación morfológica*, se descartan las etiquetas inadecuadas en función del contexto lingüístico: en el primer ejemplo, se puede decidir que *la* es un artículo porque a continuación tiene un nombre (*finestra*), mientras que en el segundo ejemplo se puede decidir que es un pronombre porque a continuación tiene un verbo (*menjaré*). En un paso posterior, en la fase de *proyección sintáctica* se asignan las etiquetas sintácticas correspondientes a la función que cada palabra puede realizar, en función de dos variables: la categoría morfológica de la palabra (por ejemplo, los nombres pueden ser sujeto, pero los adverbios no), que se ha determinado en el paso previo, y el contexto lingüístico (*finestra* en el primer ejemplo no puede ser complemento directo, porque en la oración no hay ningún verbo transitivo). Mediante este mecanismo se asigna posiblemente más de una etiqueta a cada palabra, por lo que es necesario un último paso, el de *desambiguación sintáctica* (que funciona como el de desambiguación morfológica), para tener un texto completa y --en principio-- correctamente etiquetado.

El proceso descrito hasta aquí es completamente automático: no se corrige el texto manualmente en ningún momento del proceso. En cuanto a las herramientas que se utilizan, son básicamente dos: el lenguaje de programación *Perl* y el formalismo *Constraint Grammar*, desarrollado en la Universidad de Helsinki (escrito en C++). *Perl* se utiliza para el preproceso y la proyección morfológica. La proyección se hace a partir de un formulario o tabla de formas con lemas y categorías morfológicas asociadas, que se actualiza periódicamente a partir de una herramienta llamada *CATMORF*, desarrollada por la propia UILC, cuyos datos léxicos fueron obtenidos de manera semiautomática a partir de diccionarios electrónicos. En cuanto al formalismo de la *Constraint Grammar*, se ha usado para la creación de tres gramáticas (archivos de reglas), mediante las cuales se realizan las tareas de desambiguación morfológica y de proyección y desambiguación sintáctica respectivamente. En cuanto al estado del proyecto se prevé que una primera versión de todo el sistema esté operativa a finales de 2001.

Por otro lado, el Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP) es un centro específico de investigación de la Universidad Politécnica de Catalunya (UPC) de ámbito interdepartamental.⁷ Está formado por los dos grupos de investigación de la UPC que trabajan en tecnologías del lenguaje, oral y escrito: el Grupo de Tratamiento del Habla del Departamento de Teoría de Señal y Comunicaciones⁸ y el Grupo de Tratamiento del Lenguaje Natural del Departamento de Lenguajes y Sistemas Informáticos.⁹

El TALP, como centro de I+D, fomenta la transferencia de conocimientos, experiencia y tecnología, en particular mediante la cooperación con instituciones públicas y empresas con proyectos de investigación aplicada y desarrollo. Actualmente, explican Climent Nadeu y Jaume Pedrell, dispone de toda una serie de recursos y productos propios básicamente en castellano y catalán. Entre otros, reconocimiento del habla, conversión de texto a voz, análisis morfológico, desambiguación, además de bases de datos orales y textuales, red léxico-conceptual *Euro WordNet*.

En la exposición paralela a los cursos celebrados en la UIMP de Barcelona se presentaron diversos

productos y herramientas del TALP. Entre éstos, había programas de demostración de reconocimiento del habla y de conversión de texto a voz y demostraciones del uso de las herramientas de procesado natural del lenguaje que utilizan como fuente de información la red Euro WordNet en castellano y catalán. Las demostraciones de reconocimiento permitían al visitante escribir una lista de palabras que inmediatamente podía empezar a pronunciar para que fueran reconocidas por el sistema y mostradas por pantalla, sin ninguna sesión previa de adiestramiento. Este sistema presenta una alta tasa de acierto, afirman Nadeu y Pedrell, entre el 80 % y el 97 % de las palabras de la lista confeccionada por el usuario e indica si la palabra no está en la lista, siendo útil en aplicaciones donde se requiere el control de un ordenador o un aparato mediante órdenes orales.

En una segunda representación se mostraba la tecnología desarrollada en el Centro TALP para realizar reconocimiento del habla en un entorno de diálogo persona-máquina. En éstos casos se espera que el ordenador sea capaz de reconocer frases enteras, donde el usuario puede utilizar las palabras del vocabulario en diferentes combinaciones. En concreto, la demostración consistió en el sistema de reconocimiento desarrollado para suministrar información sobre trenes, donde el sistema es capaz de transcribir al texto preguntas orales del usuario y donde un solo ordenador puede atender a decenas de clientes simultáneamente.

La demostración de la tecnología de conversión de texto a voz consistía en un programa que lee cualquier texto tecleado en el ordenador por el visitante, bien en castellano o catalán. Una demostración que también puede realizarse a través de la página web del Centro TALP. Este tipo de producto, además de complementar los sistemas de reconocimiento en las aplicaciones telefónicas, tiene una aplicación inmediata para personas invidentes o con discapacidades orales.

Por último, de entre las herramientas de procesamiento del lenguaje natural desarrolladas en el Centro, las que llamaron más la atención fueron las de desambiguación semántica y de análisis sintáctico. En la primera, el visitante podía teclear en el ordenador cualquier frase como, por ejemplo, «*El hombre bajo bajó a tocar el bajo*» y el ordenador le indicaba que el primer «*bajo*» es una característica del hombre, el segundo «*bajo*» una acción y el tercer «*bajo*» un instrumento musical. La segunda herramienta proporciona de forma automática el árbol sintáctico que todo alumno de enseñanza básica ha tenido que hacer alguna vez, señalan desde el citado centro. Entre las aplicaciones de estos tipos de herramientas se tienen, por ejemplo, la extracción automática de información de un texto, la producción de resúmenes o la traducción de un texto de un idioma a otro.

Notas

¹ Sail Labs: <http://www.sail-labs.es>

² Ampersand: <http://www.ampersand.es>

³ Q-go: <http://www.q-go.com>

⁴ Automatic Trans: <http://www.automatictrans.es>

⁵ Telefónica I+D (TID): <http://www.telefonica.es/index/tid.html>

⁶ Unidad de Investigación en Lingüística Computacional (UILC), del Instituto Universitario de Lingüística Aplicada de la Universitat Pompeu Fabra (UPF, Barcelona): <http://www.iula.upf.es>

⁷ Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP), de la Universidad Politécnica de Catalunya (UPC): <http://www.talp.upc.es>

⁸ Grupo de Tratamiento del Habla del Departamento de Teoría de Señal y Comunicaciones, de la UPC: <http://gps-tsc.upc.es/veu/veu.html>

⁹ Grupo de Tratamiento del Lenguaje Natural del Departamento de Lenguajes y Sistemas Informáticos, de la UPC: <http://www.lsi.upc.es/~acquirex/nlrg.html>

Figura 1 Independientemente de las características de cada portal, el esquema general de funcionamiento es el mismo: un usuario llama a un número de teléfono determinado, y entabla una

«conversación» con un sistema de diálogo, que le permite «navegar» a través del sistema por medio de la voz, normalmente por medio de menús, hasta acceder a la información o el servicio deseados. Todo de forma totalmente automática, sin intervención alguna de un operador humano



Figura 2 La arquitectura del sistema y las herramientas que se utilizan

