

## LINGÜÍSTICA

# La Universidad española, vivero de empresas de las tecnologías de la lengua

Thera, Eleka y E-genio comercializan aplicaciones desde 2003. Más de 30 grupos universitarios investigan cómo mejorar los sistemas de búsqueda de información en documentos digitales

## LAIA REVENTÓS

España cuenta con una "próspera y amplia comunidad científica de las tecnologías de la lengua", aquellas que mediante voz o texto permiten programar ordenadores de manera que se comporten como si entendiesen la lengua humana. Las palabras de Donia Scott, presidenta del Congreso de la Asociación Internacional de Lingüística Computacional (ACL), que celebró el pasado julio su 42ª reunión anual en el recinto del Fórum, reflejan el estado de la cuestión en nuestro país de una disciplina científica muy ligada a la inteligencia artificial y que reúne las inquietudes de las humanidades con las de la ciencia y la tecnología.

En España hay más de 30 grupos de investigación, repartidos por las universidades, que trabajan en reconocimiento de voz, procesamiento del lenguaje natural, traducción de texto a texto y síntesis de voz, los cuatro procesos básicos de estas tecnologías. La Sociedad Española para el Procesamiento del Lenguaje Natural, fundada en 1984, agrupa a más de 300, entre socios y empresas. Manuel Palomar, su presidente y catedrático de Lenguajes y Sistemas Informáticos de la Universidad de Alicante, asegura: "Ahora podemos transferir aplicaciones que beneficien a la sociedad. Las de texto son las más avanzadas. Las de voz son más complejas porque es difícil detectar los distintos tonos".

El grupo de Palomar, en colaboración con las universidades de Jaén, la Politécnica de Valencia y la UNED, trabaja para mejorar los actuales sistemas de búsqueda de información, uno de los retos de las tecnologías de la lengua. "Se trata de hacer búsquedas concretas en documentos digitalizados. Es decir que si queremos saber qué mide la torre Eiffel, obtengamos una respuesta concreta ya que la máquina se encarga de la criba de datos". El buscador Tabarca, funciona desde julio con esta tecnología.

## 'Desambiguación'

El problema de estos sistemas "es el tiempo de respuesta. Una de sus soluciones pasa por la *desambiguación* del significado, una técnica que pone cada palabra en su contexto pero que todavía no está resuelta", dice Alfonso Ureña, de la Universidad de Jaén. Su grupo (entre otros) trabaja en sistemas de recuperación de información multilingüe que incorporen técnicas del procesamiento del lenguaje natural; "es decir que incorporan sinónimos, tiempos verbales y realizan análisis sintáctico y semántico tanto en el idioma de la consulta como en otras lenguas".

Otro campo son los sistemas de extracción de información en documentos digitales. "XNotarial es una aplicación que extrae de las escrituras de compraventa los nombres del vendedor, comprador y la finca automáticamente", explica Palomar.

El Centro de Tecnologías y Aplicaciones del Lenguaje y del



Participantes en el diálogo de lingüística computacional.

TEJEDERAS

## El concurso CLEF

L.R.

En España, la investigación en tecnologías de la lengua (o lingüística computacional) no sería lo que es sin la labor de Felisa Verdejo, catedrática de Lenguajes y Sistemas Informáticos en la Universidad Nacional de Educación a Distancia (UNED) desde 1991. Una opinión compartida unánimemente por la mayoría de investigadores que asistían al Congreso Internacional de Lingüística Computacional en Barcelona, a muchos de los cuales les ha dirigido la tesis doctoral:

Fundadora y directiva de las asociaciones españolas para la Inteligencia Artificial y el Procesamiento del Lenguaje Natural (SE-PLN), autora de más de 150 publicaciones, colaboradora de investigadores de las universidades de Cambridge, Amsterdam, Duisburgo, Berlín, Osaka...

Hoy, además de dirigir el grupo de Procesamiento del Lenguaje Natural de la UNED que desarrolla sistemas de recuperación de información y consultas multilingües y multimedia, se encarga de evaluar



Felisa Verdejo.

TEJEDERAS

algunas de las tareas del concurso CLEF (Cross Language Evaluation Forum, en inglés). "Se trata de una competición a través de Inter-

net que promueve la investigación de los sistemas de recuperación de información multilingües y multimodal". CLEF: <http://clef.iei.pi.cnr.it/>

Habla (TALP), de la UPC, está especializado en el del tratamiento automático del lenguaje natural oral y escrito. "TALP participa con varios grupos europeos en el desarrollo de una agenda electrónica capaz de traducir conversaciones completas del inglés al castellano o catalán", explica Horacio Rodríguez.

Otra consecuencia de la explosión investigadora son las empresas salidas de las universidades. Todas, en 2003. Como Thera, del grupo Clic de la Universidad de Barcelona, que comercializa Ontology, un programa para clasificar documentos de cualquier fuente electrónica a gran velocidad (65.000 palabras por segundo). Además, tiene analiza-

dores morfológicos y sintácticos del catalán, castellano e inglés.

Otro ejemplo es Eleka, surgida del Grupo IXA de la Universidad del País Vasco, que vende un *software* que reconoce textos en euskera mientras están siendo escaneados de un libro. "Eleka utiliza *lematizadores* (programas que detectan la raíz de una palabra); WordNet, diccionario multilingüe donde las palabras están organizadas por campos semánticos o un corrector ortográfico que se adapta a los procesadores de texto", dice Arantxa Díaz de Ilarraza, de IXA.

E-genio, salida del laboratorio de bases de datos de la Universidad de A Coruña, ha incorporado a la Biblioteca Virtual

Galega un sistema que permite buscar cualquier palabra, frase o conjunto de caracteres en cualquiera de las obras almacenadas allí. Nieves R. Brisaboa, directora del laboratorio coruñés, explica que "E-genio también ha digitalizado todos los fondos documentales de la Real Academia Gallega. En breve, su página incorporará la *hemeroteca virtual resultante*".

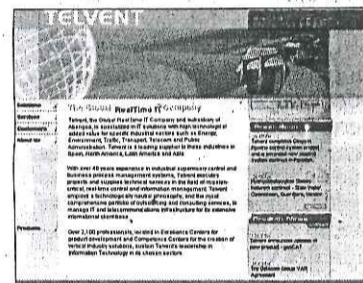
SEPLN: [www.sepln.org](http://www.sepln.org)  
 UA: <http://gplsi.dlsi.ua.es>  
 TALP: <http://www.talp.upc.es/>  
 UJAEN: [www.wdi.ujaen.es/~gsi/](http://www.wdi.ujaen.es/~gsi/)  
 UPValencia: [www.upv.es](http://www.upv.es)  
 ACL: [www.aclweb.org](http://www.aclweb.org)  
 THERA: [www.thera-clic.com](http://www.thera-clic.com)  
 ELEKA: [www.eleka.es](http://www.eleka.es)  
 BVG: <http://bvg.udc.es/index.jsp>

## Telvent desarrolla un sistema para acceder al metro o al autobús con la voz

L.R.

Tecnologías de la lengua para tomar los transportes públicos sin intervención humana. Telvent, filial tecnológica de la multinacional andaluza Abengoa, está desarrollando un sistema que, mediante la voz, permitirá a cualquiera comprar billetes de autobús, metro o tren en las máquinas expendedoras automáticas y, en especial a los ciegos, que serán capaces de adquirir los boletos sin recibir ayuda de otras personas. En España hay 50.000 invidentes o personas con dificultades visuales, según el Instituto Nacional de Estadística (INE).

La tecnología, llamada Genio, emplea simultáneamente sistemas de reconocimiento de voz y visualización de las personas (cámara) para que la máquina pueda comprobar que realmente tiene a alguien que solicita su ayuda y no se trata, por ejemplo, del sistema de megafonía de la estación. De momento, la empresa está realizando pruebas piloto en el metro de Bilbao, ciudad a la que Telvent ya suministra los sistemas de gestión de accesos y billetes.



Telvent.

Si todo va bien Genio estará disponible a finales de 2005. Antes, deben resolver varias dificultades: "La diversidad de voces y entonaciones que la máquina debe ser capaz de reconocer y el ruido ambiental, que es muy difícil de aislar. Los niveles de ruido dependen de cada estación y de dónde están situadas las máquinas expendedoras de billetes", explica Carlos Gil, director de tecnologías de sistemas de pago de la compañía.

En el desarrollo de este proyecto piloto de tres años de duración participa un equipo de 50 ingenieros españoles del Centro Tecnológico de Sistemas de Pago de Telvent, Fagor Electrodomeísticos, Ikerlan y la Universidad del País Vasco. Cuenta con una inversión de unos 1,2 millones de euros.

## Hablar a la lavadora

El grupo también está trabajando para que Genio sea incorporado en las lavadoras y hornos de Fagor, la empresa de electrodomésticos. "Queremos que el usuario pueda dirigirse de forma natural a la lavadora; es decir con su voz", explica Gil. O sea que el usuario le diga 'he metido en el bomo ropa mezclada muy sucia' y la máquina decida qué programa debe utilizar y, en el caso del horno, éste decida la temperatura y tiempo de cocción cuando la persona diga, con la voz, 'quiero hacer pollo asado'.

TELVENT: [www.telvent.com](http://www.telvent.com)  
 FAGOR: [www.fagor.com](http://www.fagor.com)  
 IKERLAN: [www.ikerlan.es](http://www.ikerlan.es)  
 UPV: [www.ehu.es](http://www.ehu.es)