

SOCIEDAD

CIENCIA CONTADA EN ESPAÑOL

[Iberoamérica](#) | [Opinión](#) | [Ciencia](#) | [Salud](#) | [Innova](#) | [Tierra](#) | [Tecnología](#) | [Sociedad](#)

El Grupo IXA de la Universidad del País Vasco participa en un nuevo proyecto europeo para dar un salto cualitativo en la traducción automática: QTLeap. El objetivo del proyecto es garantizar la calidad de los resultados de las traducciones automáticas actuales, que se encuentran lejos de la perfección.



Basque Research 27/6/2013 17:00 CEST



El objetivo es ayudar a los traductores automáticos en la sintaxis, gracias al empleo de bancos de árboles. / Conny Beyreuther.

Todos los traductores automáticos tienen sus limitaciones. Las traducciones que se obtienen a través de ellas no son perfectas, ni mucho menos, pero son herramientas de gran ayuda en más de una ocasión. ç

El Grupo IXA de la Universidad del País Vasco (UPV/EHU) pretende superar dichos límites y continuar investigando en el campo de la traducción automática a través del proyecto europeo QTLeap, en colaboración con entidades pioneras de Europa en el campo de la traducción automática: DKFI de Alemania, la Universidad de Lisboa, La Universidad Charles de Praga, La Academia de Ciencias de



Iberoamérica | Opinión | Ciencia | Salud | Inova | Tierra | Tecnología | Sociedad

En el caso del euskera, sin embargo, resulta más complicado, y la calidad de la traducción disminuye. "Tiene tres grandes inconvenientes: por un lado, las enormes diferencias gramaticales y morfológicas que presenta el euskera respecto al resto de las lenguas de su entorno; por otro lado, la elección del equivalente adecuado en las demás lenguas, es decir, hay que elegir la acepción adecuada para cada contexto entre los diferentes significados posibles de una misma palabra; y por último, la carencia de corpus extensos de textos bilingües traducidos".

El euskera tiene una estructura muy distinta respecto al resto de las lenguas, y dar con el orden correcto es prácticamente imposible para un traductor automático. Además de eso, uno de los mayores retos para los traductores de euskera es la obtención de grandes cantidades de textos traducidos al euskera, es decir, poder disponer de corpus extensos.

Más cerca de la perfección

El Grupo IXA de la UPV/EHU trabaja con empeño en el desarrollo de traductores automáticos, y quiere dar un paso más allá en la investigación de dichas herramientas a través del proyecto QTLeap. Intentarán superar los límites que presentan los traductores automáticos actuales.






Para ello, por un lado, utilizarán los bancos de árboles o treebank-s. "El objetivo es ayudar a los traductores automáticos en la sintaxis, gracias al empleo de bancos de árboles, es decir, una gran cantidad de frases previamente analizadas sintácticamente" cuenta Sarasola.

Por otro lado, la información necesaria para diferenciar una acepción de otra en una palabra concreta en una frase, no solo se encuentra en los diccionarios, sino que, en la actualidad, también se puede consultar en Internet. Por lo tanto, para diferenciar los significados de las palabras adecuadamente, emplearán ciertos recursos de Internet.

Por ejemplo, en grandes colecciones de textos, como loes la Wikipedia, la acepción de cada palabra esta propiamente definida. Es decir, "si debajo de un concepto hay un enlace, dicho enlace está orientado a uno u otro significado de la Wikipedia". Si se reúnen muchos de esos enlaces, el ordenador quizás aprenda de ellos a diferenciar entre un significado u otro. Wikipedia es, simplemente, una posibilidad.

"Hoy en día, existen cada vez más textos en Internet con ese tipo de enlaces; todas ellas se conocen como *Linked Open Data*" añade. Todo ello ayudaría, por ejemplo, a discernir si la palabra *baso* del euskera pertenece, en un caso determinado, a *bosque* o a *vaso* en castellano.

"Así mismo, trataremos de una manera muy especial los nombres propios, los nombres de las entidades y los sitios geográficos. A pesar de que es imposible tener todos los nombres propios etiquetados, tener controlados un gran grupo de ellos puede mejorar en gran medida la calidad de las traducciones" cuenta Sarasola. Con ello tendremos la opción, por ejemplo, de mantener el nombre de *Pilar del Castillo* como tal, sin obtener el resultado de *gatzeluko pilareak* en la traducción al euskera.

Dichos recursos se pueden obtener de dos maneras: off-line y on-line. Es decir, en el modo off-line,     



Iberoamérica | Opinión | Ciencia | Salud | Innova | Tierra | Tecnología | Sociedad

Por lo tanto, el Grupo IXA elaborará las tres líneas citadas anteriormente —los bancos de árboles, los recursos de Internet y los nombres propios— dentro del proyecto QTLeap a partir de noviembre, para dar una solución a los problemas de los traductores automáticos actuales, derivados de las grandes diferencias morfológicas y gramaticales, la falta de desambiguación de acepciones y la carencia de corpus extensos.

La larga trayectoria de la traducción automática

En la década de 1950 comenzaron a surgir las ideas para el desarrollo de la traducción automática. Desde entonces, se han planteado varias aproximaciones en la búsqueda de un sistema exitoso. El Grupo IXA de la UPV/EHU se empeña desde hace quince años en el desarrollo de dichas herramientas, en colaboración con Elhuyar.

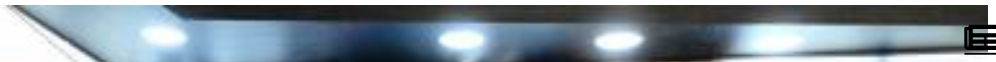
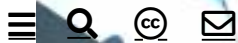
En los últimos años, ha aumentado el número de traductores en la red, así como las aplicaciones fáciles de descargar y de utilización gratuita. Una de ellas es Opentrad, una plataforma de código abierto que es válida para todas las lenguas oficiales del Estado: castellano, catalán, gallego y euskera. Detrás de Opentrad subyacen dos tecnologías: Matxin (para pares de lenguas divergentes como el castellano-euskera), desarrollado por el Grupo IXA de la UPV/EHU, y Apertium (para castellano-catalán-gallego, etc.).

Fuente: **Universidad del País Vasco**

Derechos: **Creative Commons**

TRADUCTOR QTLEAP TRADUCCIÓN AUTOMÁTICA EUSKERA

Artículos relacionados



Iberoamérica | Opinión | Ciencia | Salud | Innova | Tierra | Tecnología | Sociedad



La expresión 'inteligencia artificial' es la palabra del año según la FundéuRAE

SINC

Ha resultado ganadora por su presencia en los medios de comunicación y por el debate suscitado en torno a su uso en múltiples campos. Se ha puesto por delante de otras 11 candidatas, entre ellas, ucraniano, ecocidio, inflación y criptomoneda.



LO MÁS VISTO

- El epicentro del terremoto de Turquía
- Los avances científicos de 2022 según 'Science', y el Webb está el primero
- Los científicos descubren que algunas secuencias de ADN cambian según el tiempo
- El océano profundo acompañará a la Tierra durante las próximas décadas
- La anatomía de los yunnanozoos podría explicar el salto evolutivo que originó a los vertebrados
- La inteligencia artificial revela variaciones en la anatomía cerebral de personas con autismo
- Los científicos descubren las vías anatómicas que prepararon a nuestro cerebro para el lenguaje
- Los científicos descubren la estructura de la hecatomba atómica

Hallado en Navarra el texto más antiguo en lengua vascónica



Iberoamérica | Opinión | Ciencia | Salud | Innova | Tierra | Tecnología | Sociedad

Publica

Licencia Creative Commons



Síguenos



Creative Commons 4.0

Puedes copiar, difundir y transformar los contenidos de SINC. [Lee las condiciones de nuestra licencia](#)



SÍGUENOS EN



[Política de cookies](#)

[Accesibilidad](#)

[Aviso legal y Política de privacidad](#)

Desarrollado con eZ PublishTM