

**Día Europeo de las Lenguas,  
26 de septiembre**

## **La tecnología híbrida del proyecto QLeap mejora la calidad de traducción automática**



En los últimos días hemos visto llegar a millares de refugiados a Europa. Ahora, una de las tareas de integración más urgente es la de enseñarles las lenguas de los países de acogida. La envergadura de esta tarea nos debería hacer recordar que todavía el reto de la integración Europea no se ha abordado con seriedad: las barreras lingüísticas siguen siendo las fronteras más tenaces en Europa, las que entorpecen el comercio, la comunicación y la verdadera integración.

Aprender una lengua nueva no es una tarea sencilla. Los niños aprenden lenguas, ya sea la materna u otras, jugando o escuchando a adultos, pero éstos sólo suelen conseguir aprender reglas gramaticales y vocabulario con mucho esfuerzo.

La manera de aprender de los programas de ordenador, como son los conocidos sistemas de traducción automática accesibles en la red, se parece más a la de un niño. Utilizan métodos de aprendizaje automático aplicados a grandes cantidades de texto ("Big Data"), donde aprenden probabilidades estadísticas para traducir palabras y secuencias de palabras. Los sistemas ofrecen una traducción aproximada, cuya calidad varía considerablemente dependiendo del texto que se quiera traducir.

Sin embargo, las lenguas son más que meras secuencias de palabras. Tienen estructuras gramaticales, con elementos como sujeto, predicado y objetos. Existen sistemas de traducción automática que utilizan este tipo de información lingüística pero tienden a ser poco flexibles para traducir textos genéricos y se utilizan en contextos limitados. Por ejemplo, el periódico *La Vanguardia* se traduce al catalán todas las noches con un sistema de este tipo y un revisor lo repasa rápidamente antes de entrar en prensa.

En Internet, recursos como Wikipedia en formato legible para ordenadores forman la llamada Web semántica y ofrecen conocimiento sobre semántica (el significado de las palabras). Estos recursos codifican conocimiento estructurado sobre el mundo. Por ejemplo, "saben" que París es la capital de Francia, pero también que existe una ciudad llamada París en Texas.

En el marco del proyecto europeo QLeap (qtleap.eu), el grupo de investigación en procesamiento del lenguaje natural IXA, de la Universidad del País Vasco, ha desarrollado tecnologías híbridas para traducción automática que enriquece métodos estadísticos con conocimiento lingüístico y semántico de Internet para producir mejores traducciones.

Lehen prototipoak itxaropentsuak dira baina “oraindik ikerketa mota honen hasieran baino ez gaude” dio Eneko Agirrek, IXA-n proiektuko koordinatzaileak. “Pertsonak itzultzerantz jartzen garenean, jatorrizko testuaren esanahia ulertu, eta esanahi hori itzultzen dugu. Askotan itzulpen libreak ateratzen zaizkigu, eta itzulpen zehatza egitea asko kostatzen zaigu. Egungo teknologiek, aldiz, itzulpen literalak egiteko joera handia dute, esanahiei baino azaleko hitzei erreparatuz. Hitzen eta esaldien esanahia automatikoki atzitzea eta itzulpen automatikoaren prozesuan txertatzea erronka handia da gure alor honetan, eta IXA taldea gogotsu ari da horretan.”

Los primeros prototipos son prometedores pero “todavía estamos al comienzo de esta línea de investigación”, declara el Dr. Eneko Agirre, investigador principal del proyecto en IXA. “Cuando traducimos, tratamos de comprender el significado del texto original y traducir este significado. Para ello, muchas veces optamos por traducciones más libres y nos distanciamos de las palabras exactas del texto de origen. Sin embargo, las tecnologías de hoy en día tienden a ofrecer traducciones literales, centrándose más en la forma superficial de las palabras que en su significado. Extraer el significado de las palabras y oraciones e insertarlo en el proceso de traducción automática es uno de los grandes retos en nuestra línea de investigación, y en el grupo IXA, estamos trabajando en ello con entusiasmo.”

De la misma manera en que hoy en día nuestros hijos se preguntan cómo éramos capaces de hacer los deberes sin Wikipedia hace veinte años, puede que nuestros nietos se pregunten cómo, en 2015, todas aquellas personas que llegaron a Europa en busca de una vida mejor pudieron comunicarse sin el *omni-traductor*.

Para mayor información y contacto: [qtLeap.eu](http://qtLeap.eu).