

## Blade Runner inspira un nuevo traductor de lenguas minoritarias

Un equipo de la Universidad Politécnica de Valencia ha desarrollado un sistema para traducir lenguas minoritarias basándose en su parecido con otros lenguajes. Los investigadores se han inspirado en la jerga de algunos personajes de la película *Blade Runner*.

UPV

25/5/2012 10:04 CEST



Los investigadores del ITI, Luis Leiva y Vicente Alabau. Imagen: UPV.

Investigadores del grupo *Pattern Recognition and Human Language Technologies (PRHLT)* en el Instituto Tecnológico de Informática (ITI) de la

Universidad Politécnica de Valencia (UPV) han desarrollado un nuevo traductor automático de lenguas minoritarias. El sistema permite superar la escasez de recursos lingüísticos de los traductores actuales y facilita la interpretación de los textos, sea cual sea la lengua en que estén escritos, aprovechando la similitud que tenga con otros lenguajes.

“Nuestro objetivo era ayudar a entender aquellas lenguas que actualmente no son capaces de abarcar los traductores, y contribuir así a la preservación de culturas y la eliminación de barreras lingüísticas”, destacan Luis Leiva y Vicent Alabau, responsables del nuevo método de traducción.

Se estima que actualmente más del 10% de los lenguajes del mundo no pueden ser asistidos por ningún sistema de traducción, ni siquiera empleando lenguajes intermedios para los que suele haber más recursos. Por ejemplo, para traducir de francés a zulú se suele pasar por el inglés (esto es, se traduce de francés a inglés y posteriormente de inglés a zulú). Incluso para los lenguajes mayoritarios en ocasiones es complicado abarcar todo el conocimiento lingüístico, dado su constante crecimiento y enorme evolución.

Según apuntan los investigadores, esto da lugar a un fenómeno bastante frecuente en los traductores automáticos: cuando el sistema no tiene información sobre algunas palabras, las deja sin traducir. “Por lo tanto, en el caso extremo de no tener datos acerca de ninguna de las palabras en el idioma origen, el sistema pierde toda utilidad y el mensaje no puede ser transmitido”.

Para resolver este déficit, el sistema incorpora en las traducciones información de lenguajes relacionados con el idioma destino, aprovechando la similitud gramatical y sintáctica que existe en las familias de lenguajes. “La idea consiste en escoger aquellas palabras de otros idiomas para los que se pueda generar una traducción automática, de forma que se parezcan lo más posible al idioma destino, que es el idioma del usuario que quiere entender el mensaje. Si bien las frases resultantes pueden parecer extrañas a primera vista, ya que gramaticalmente son una mezcla de varios idiomas, en realidad son lo suficientemente entendibles para un hablante de la lengua destino”, destaca Leiva.

### **Una frase de ejemplo**

Para explicar el funcionamiento y la utilidad del nuevo traductor, los investigadores exponen el siguiente caso: imaginemos que una persona que solamente hable español quisiera entender la frase '*another label with the same name already exists*'; y que el español fuera un idioma para el que no existieran traductores automáticos.

En este caso el sistema buscaría posibles similitudes en el italiano, portugués y francés –todas ellas lenguas romances– para las que sí puede generar traducciones automáticamente y construye en primer lugar traducciones de inglés a italiano, portugués y francés. A continuación, las tres traducciones se combinan estadísticamente para que el resultado sea parecido al español.

“Así, la frase anterior se traduce por '*un'altra étiquette con mesmo nome existe déjà*', que no es español pero permite a un español entender el mensaje original sin necesidad de dominar ninguno de los tres idiomas auxiliares; esto habría sido imposible de realizar mediante sistemas de traducción automática actuales”, apunta Alabau.

Entre sus ventajas, este nuevo traductor, en cuyo perfeccionamiento se sigue trabajando, facilitaría el acceso a contenidos digitales (como libros electrónicos o páginas web) que no pueden ser traducidos a ciertas lenguas, favorecería también la integración social de hablantes monolingües y ayudaría a los usuarios a adquirir soltura y familiarizarse con el vocabulario de otras lenguas.

### **Inspiración en 'Blade Runner'**

Para su desarrollo, los investigadores se inspiraron en la película *Blade Runner*, en la que algunos personajes utilizaban una jerga llamada *cityspeak* (*interlingua* en la versión española de la película), una mezcla de idiomas, que incluía, entre otros, japonés, español y alemán. Esta jerga surgía en Los Ángeles como consecuencia de un creciente entorno multicultural.

“A diario mucha gente incorpora palabras extranjeras en sus conversaciones, bien por familiaridad o frecuencia de uso, pero también porque a veces otro idioma tiene un término más preciso para definir un concepto”, explican los creadores.

“Eso nos llevó a pensar que sería buena idea usar lenguajes parecidos a los de un grupo de hablantes para suplir la falta de recursos en traducción automática; de manera que, al contrario que *cityspeak*, el lenguaje mezcla no tiene que ser aprendido porque está orientado al lenguaje de dichos hablantes”, concluyen los investigadores de la UPV.

Derechos: **Creative Commons**

TAGS

TRADUCTOR |

BLADE RUNNER |

LENGUA |

#### Creative Commons 4.0

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