

UN RIESGO PARA NUESTRAS LENGUAS Y UN RETO PARA LA TECNOLOGÍA LINGÜÍSTICA

Somos testigos de una revolución digital que está impactando drásticamente en la comunicación y en la sociedad. Esta revolución se ha comparado alguna vez con la invención de la imprenta por Gutenberg. ¿Qué nos puede decir esta analogía sobre el futuro de la sociedad de la información y de nuestras lenguas en particular?

Somos testigos de una revolución digital comparable a la invención de la imprenta por Gutenberg.

Después de la invención de Gutenberg, los avances reales en la comunicación y el intercambio de conocimientos se lograron mediante esfuerzos tales como la de la Biblia por Lutero a la lengua vernácula. En los siglos siguientes, se desarrollaron técnicas culturales para manejar mejor el procesamiento del lenguaje y el intercambio de conocimiento:

- la estandarización ortográfica y gramatical de las principales lenguas permitió la rápida difusión de nuevas ideas científicas e intelectuales;
- el desarrollo de las lenguas oficiales hizo posible que los ciudadanos pudieran comunicarse dentro de ciertos límites (a menudo políticos);
- la enseñanza de lenguas y las traducciones favorecieron los intercambios entre las distintas lenguas y culturas;

- la creación de editoriales y de directrices bibliográficas aseguraron la calidad y disponibilidad de material impreso;
- la creación de distintos medios de comunicación como periódicos, radio, televisión, libros y otros formatos satisfizo las diferentes necesidades de comunicación.

En los últimos veinte años, la tecnología de la información ha ayudado a automatizar y facilitar muchos de los procesos:

- el software de autoedición ha sustituido la mecanografía y la composición;
- Microsoft PowerPoint ha sustituido las transparencias para retroproyector;
- el correo electrónico envía y recibe documentos más rápido que una máquina de fax;
- Skype ofrece llamadas baratas de telefonía por Internet y organiza reuniones virtuales;
- los formatos de audio y de codificación de vídeo facilitan el intercambio de contenido multimedia;
- los motores de búsqueda basada en palabras clave proporcionan acceso a las páginas web;
- servicios en línea como Google Translate producen traducciones rudimentarias de forma rápida;
- las plataformas de medios sociales como Facebook, Twitter y Google+ facilitan la comunicación, la colaboración y el intercambio de información.

A pesar de que tales herramientas y aplicaciones son útiles, no son todavía capaces de dar soporte a una sociedad europea multilingüe y sostenible, donde la información y las mercancías puedan circular libremente.

2.1 LAS FRONTERAS LINGÜÍSTICAS SON UN OBSTÁCULO PARA LA SOCIEDAD DE LA INFORMACIÓN EUROPEA

No podemos predecir exactamente como será la sociedad de la información del futuro, pero muy probablemente la gran revolución en la tecnología de la comunicación será relacionar de maneras innovadoras a las personas que hablan diferentes idiomas. Más idiomas y más contenidos interactuarán con nuevos medios de comunicación en un mismo espacio económico y de información globalizado. La actual popularidad de las redes sociales (Wikipedia, Facebook, Twitter, YouTube, y, recientemente, Google +) es sólo la punta del iceberg.

Un espacio económico y de información globalizado supone una mayor cantidad de lenguas, hablantes y contenido.

Hoy en día, podemos transmitir gigabytes de texto alrededor del mundo en pocos segundos, incluso antes de reconocer que el mensaje es en un idioma que no entendemos. Según un reciente informe de la Comisión Europea, el 57% de los usuarios europeos de Internet adquieren bienes y servicios en un idioma distinto de su lengua materna. (El idioma más común es el inglés, seguido del francés, el alemán y el español.) El 55% de los usuarios leen contenidos en un idioma extranjero, mientras que sólo el 35% usa otra lengua para escribir comentarios, correos electrónicos o publicar contenido en la

Web [3]. Hace algunos años el inglés era la lengua franca de la Web -la gran mayoría de los contenidos estaba en inglés- pero la situación ha cambiado drásticamente. La cantidad de contenido en línea en otras lenguas europeas (así como asiáticas) se ha disparado.

Curiosamente, esta brecha digital causada por las barreras lingüísticas no recibe mucha atención en los medios y, sin embargo, la cuestión que se plantea es crucial: ¿Qué idiomas europeos prosperarán en la red y en la sociedad digital, y cuáles están condenados a desaparecer?

2.2 NUESTRAS LENGUAS EN PELIGRO

Así como la imprenta ayudó a intensificar el intercambio de información en Europa, también llevó a la extinción de muchas lenguas europeas. Ciertas lenguas regionales y minoritarias, como el córnico y el dalmata, tuvieron una producción impresa escasa y su transmisión se limitó a las formas orales, lo que a su vez limitó su ámbito de uso. ¿Tendrá Internet el mismo impacto en nuestras lenguas?

Los aproximadamente 80 idiomas de Europa son uno de sus bienes culturales más ricos e importantes, y una parte vital de su singular modelo social [4]. Si bien algunos de sus idiomas, como el inglés y el español, probablemente sobrevivan en el mercado digital emergente, muchos otros podrían convertirse en irrelevantes en una sociedad en red. Esto debilitaría la posición global de Europa, e iría en contra del objetivo estratégico de garantizar la participación equitativa de todos los ciudadanos europeos, independientemente de su idioma.

De acuerdo con un informe de la UNESCO en materia de multilingüismo, las lenguas son un medio esencial para el disfrute de derechos fundamentales como la expresión política, la educación y la participación en la sociedad [5].

2.3 LA TECNOLOGÍA LINGÜÍSTICA ES CLAVE EN EL DESARROLLO DE MUCHAS APLICACIONES

En el pasado reciente, la mayoría de la inversión destinada a la preservación de las lenguas se centraban en la enseñanza de idiomas y en la traducción. Según una estimación, el mercado europeo de traducción, interpretación, localización de software y globalización de sitios web fue de 8.4 mil millones de euros en 2008 y se espera que crezca en un 10% anual [6]. Sin embargo, esta cifra sólo cubre una pequeña proporción de las necesidades actuales y futuras de la comunicación entre lenguas. La solución más efectiva para garantizar el multilingüismo en la Europa del futuro pasa por el uso de tecnologías apropiadas, de la misma forma que recurrimos a la tecnología para resolver nuestras necesidades de transporte, de energía y de atención a la discapacidad, entre otras.

La tecnología lingüística (orientada a todas las formas de texto escrito y discurso hablado) ayuda a las personas a colaborar entre ellas, a hacer negocios, a compartir conocimientos y a participar en los debates sociales y políticos, independientemente de las barreras idiomáticas y de las habilidades informáticas. A menudo, opera de forma invisible dentro de sistemas de software complejos para ayudarnos a:

- encontrar información con un motor de búsqueda por Internet;
- comprobar la ortografía y la gramática en un procesador de textos;
- ver recomendaciones de productos en una aplicación de comercio electrónico;
- escuchar las instrucciones verbales de un sistema de navegación para automóviles;
- traducir páginas web a través de un servicio en línea.

Las tecnologías de la lengua consisten en una serie de aplicaciones básicas que operan en un marco de aplicación más amplio. El propósito de la serie de White Papers de META-NET es describir el grado de preparación de cada una de las lenguas europeas en estas tecnologías básicas.

Europa necesita tecnología lingüística robusta y asequible para todos los idiomas europeos.

Para mantener nuestra posición en la vanguardia de la innovación global, Europa necesitará disponer de tecnologías lingüísticas adaptadas a todos los idiomas europeos, que sean sólidas, asequibles y estén perfectamente integradas en los entornos de software clave. En un futuro cercano, sin la tecnología lingüística, no seremos capaces de lograr una experiencia de usuario interactiva, multimedia y multilingüe realmente efectiva.

2.4 OPORTUNIDADES PARA LA TECNOLOGÍA LINGÜÍSTICA

En el mundo de la imprenta, el gran avance tecnológico fue la duplicación rápida de la imagen de un texto (una página), utilizando una prensa. Los seres humanos aún tenían que hacer el trabajo laborioso de buscar, leer, traducir y resumir el conocimiento. Tuvimos que esperar hasta Edison para grabar el lenguaje hablado, y en este caso también su tecnología simplemente hacía copias analógicas.

Las tecnologías lingüísticas son capaces de automatizar los procesos mismos de traducción, producción de contenidos y gestión del conocimiento para todos los idiomas europeos. También pueden servir para construir interfaces intuitivas, basadas en lenguaje humano (escrito o hablado) para electrodomésticos, maquinaria, vehículos, computadoras y robots. Las aplicaciones comerciales e industriales de estas tecnologías se encuentran todavía

en las primeras etapas de desarrollo, sin embargo, los éxitos de la I + D están creando una ventana real de oportunidades. Por ejemplo, la traducción automática ya es razonablemente precisa en áreas concretas, y existen aplicaciones experimentales que proporcionan información multilingüe y gestión del conocimiento, así como producción de contenidos, en muchas lenguas europeas.

Como ocurre con la mayoría de las tecnologías, las primeras aplicaciones lingüísticas, como las interfaces basadas en voz y los sistemas de diálogo, se desarrollaron para ámbitos altamente especializados, y con frecuencia su rendimiento es limitado. Sin embargo, existen grandes oportunidades de mercado en ámbitos como la educación o el ocio, con la integración de tecnología lingüística en juegos, en divulgación del patrimonio cultural, en paquetes de entretenimiento educativo, en bibliotecas, entornos de simulación y programas de capacitación. Los servicios de información móvil, el software de aprendizaje de idiomas, los entornos de e-learning, las herramientas de autoevaluación y el software de detección de plagio son sólo algunas de las áreas de aplicación en las que la tecnología lingüística puede desempeñar un papel importante. La popularidad de las redes sociales como Twitter y Facebook sugieren necesidades adicionales para esta tecnología, como por ejemplo, control de foros, resumen de discusiones, tendencias de opinión, detección de respuestas emocionales o identificación de infracciones de derechos de autor, entre otros.

La tecnología lingüística ayuda a superar la “discapacidad” de la diversidad lingüística.

La tecnología lingüística representa una gran oportunidad para la Unión Europea y puede ayudar a abordar el complejo tema del multilingüismo en Europa - el hecho de que distintas lenguas convivan de forma natural en empresas, organizaciones y escuelas europeas. Los ciudadanos tienen que comunicarse a través de estas fron-

teras lingüísticas que cruzan el Mercado Común Europeo, y la tecnología lingüística puede ayudar a superar esta barrera final, a la vez que apoya el uso libre y abierto de cada una de las lenguas individuales. Mirando aún más hacia adelante, una tecnología lingüística multilingüe e innovadora, proporcionará un punto de referencia a nuestros socios a nivel mundial, cuando empiecen a atender a sus propias comunidades multilingües. La tecnología lingüística puede verse como una forma de tecnología “asistencial” que ayuda a superar la “discapacidad” de la diversidad lingüística y hace que las comunidades lingüísticas sean más accesibles entre sí.

Por último, un campo activo de investigación es el innovador uso de la tecnología lingüística en operaciones de rescate en zonas de desastre, donde la eficiencia puede ser una cuestión de vida o muerte: los robots inteligentes del futuro, dotados de capacidades multilingües tendrán un potencial aumentado para salvar vidas.

2.5 RETOS DE LA TECNOLOGÍA LINGÜÍSTICA

Aunque la tecnología lingüística ha hecho progresos considerables en los últimos años, el ritmo actual de progreso tecnológico y la innovación en los productos es demasiado lento. Incluso tecnologías ampliamente utilizadas, como la corrección ortográfica y gramatical de los procesadores de texto, sólo están disponibles para unos cuantos idiomas.

El ritmo actual de progreso tecnológico y la innovación en los productos es demasiado lento.

Los servicios de traducción automática en línea, aunque son útiles para generar de forma rápida una aproximación razonable al contenido de un documento, todavía no resultan prácticos cuando la traducción debe ser muy exacta. Debido a la complejidad del lenguaje humano, el

desarrollo del software que lo modela y su puesta a prueba en el mundo real resulta un proceso largo y costoso que requiere de financiación continuada. Europa debe pues mantener su papel pionero a la hora de enfrentarse a los desafíos tecnológicos de una comunidad multi-lingüe inventando nuevos métodos para acelerar el desarrollo transversal de esta. Estos nuevos métodos incluyen tanto avances computacionales como técnicas emergentes, como por ejemplo, el crowdsourcing.

2.6 ADQUISICIÓN DEL LENGUAJE POR LOS SERES HUMANOS Y POR LAS MÁQUINAS

Para ilustrar cómo manejan el lenguaje los ordenadores y por qué es difícil programarlos para usarlo, vamos a examinar brevemente la manera en que los seres humanos adquieren sus primeras y segundas lenguas, y luego cómo funcionan las tecnologías lingüísticas.

Los seres humanos adquieren capacidades lingüísticas de dos maneras diferentes: aprendiendo de los ejemplos y aprendiendo las reglas gramaticales subyacentes.

Los seres humanos adquieren capacidades lingüísticas de dos maneras diferentes. Los bebés aprenden un idioma al escuchar las interacciones reales entre sus padres, hermanos y personas del ámbito familiar. Desde los dos años, los niños producen sus primeras palabras y frases cortas. Esto sólo es posible porque los seres humanos tienen una predisposición genética a imitar y racionalizar lo que escuchan.

Aprender un segundo idioma a una edad avanzada requiere más esfuerzo, sobre todo si la persona no se encuentra inmersa en una comunidad lingüística de hablantes nativos. En la escuela, los idiomas extranjeros se

adquieren a través del aprendizaje explícito del vocabulario y la gramática, haciendo ejercicios, aplicando reglas abstractas y viendo ejemplos. El aprendizaje de una lengua extranjera se hace más difícil con la edad.

Existen un paralelismo claro entre estas dos maneras de aprender lenguas en los seres humanos y los dos tipos principales de sistemas de tecnología lingüística: los sistemas estadísticos o basados en los datos, y los sistemas basados en reglas.

Los sistemas estadísticos obtienen conocimiento lingüístico a partir de grandes colecciones de textos concretos. El algoritmo de aprendizaje automático “aprende” cómo se traducen las palabras, las frases cortas y, por último, las oraciones completas. Los textos no son siempre fáciles de obtener. Si bien para entrenar ciertas aplicaciones, por ejemplo, un corrector ortográfico, es suficiente usar texto en un solo idioma, para entrenar un sistema de traducción automática se precisan textos paralelos, en dos o más lenguas. El enfoque estadístico puede requerir millones de frases y su rendimiento aumenta con la cantidad de texto analizado. Esta es una de las razones por la que los proveedores de motores de búsqueda están dispuestos a recoger tanto material escrito como sea posible. Servicios como Google Search y Google Translate se basan en métodos estadísticos. La gran ventaja de la estadística es que la máquina aprende rápidamente en una serie sucesiva de ciclos de entrenamiento, a pesar de que la calidad del resultado puede variar de forma aparentemente arbitraria.

La estrategia alternativa es la construcción de sistemas basados en reglas. Expertos lingüistas e informáticos primero tienen que codificar el conocimiento gramatical o las reglas de traducción y compilar diccionarios y listas de vocabulario. Esto requiere una gran inversión en tiempo y esfuerzo. Algunos de los principales sistemas de traducción automática basados en reglas han sido objeto de constante desarrollo durante más de veinte años. La gran ventaja de los sistemas basados en reglas es que

los expertos tienen un control directo sobre el procesamiento del lenguaje. Esto hace que sea posible corregir errores sistemáticos y dar información detallada al usuario, especialmente cuando los sistemas de reglas se utilizan para el aprendizaje de idiomas. Sin embargo, debido a su alto coste, este tipo de sistemas sólo han sido desarrollados para los idiomas mayoritarios.

Como las ventajas e inconvenientes de los sistemas estadísticos y de los basados en reglas tienden a ser complementarios, la investigación actual se centra en los enfoques híbridos, que combinan las dos metodologías. Sin embargo, estos métodos de momento han tenido más

éxito en los centros de investigación que en las aplicaciones industriales.

Como hemos visto en este capítulo, muchas aplicaciones de las más comunes dependen en gran medida la tecnología lingüística. Por su condición multilingüe, esto es particularmente crítico en el espacio europeo de la información. Aunque la tecnología lingüística ha hecho considerables progresos en los últimos años, todavía hay un enorme potencial para mejorar su calidad. En las siguientes secciones, vamos a describir el papel del español en la sociedad europea de la información y evaluar el estado actual de la tecnología lingüística para el español.

The Spanish Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 79 p. 24 illus. in color., Softcover

ISBN: 978-3-642-30840-6