

Español: un idioma poco tecnológico

El español es la segunda lengua materna del mundo por número de habitantes. Sin embargo, se estima que apenas representa el 30% del mercado mundial de las tecnologías de procesamiento del lenguaje. La clave es concienciarse de que las tecnologías del lenguaje en español son un buen negocio, que debe ser impulsado desde España.

LUIS MENÉNDEZ

✉ luis.menendez@yahoo.es



Se estima que el español representa apenas el 30 por ciento del mercado mundial de las tecnologías de procesamiento del lenguaje

El anuario *El español en el mundo* 2019 del Instituto Cervantes revela que 483 millones de personas tienen el español como lengua materna, tres más que en 2018, aunque el número de usuarios potenciales de español supera los 580 millones.

El español es la segunda lengua materna del mundo por número de habitantes, solo por detrás del chino mandarín, y la tercera lengua en el cómputo global de hablantes. El 7,6% de la población mundial es hispanohablante y se prevé que siga subiendo. Además, continúa creciendo el número de personas que estudia español como lengua extranjera, hasta situarse en torno a 22 millones de personas.

En internet tampoco le va mal al español. Es la tercera lengua más utilizada en la red después del inglés y del chino, y la comunicación en internet producida en español representa el 8% del total. Es la segunda lengua más

En internet el español es la tercera lengua más utilizada, después del inglés y del chino.



Siete a uno

SE calcula que existen siete aplicaciones en inglés por cada una en español, por lo que la idea es liderar desde España el mercado el procesamiento de lenguaje natural antes de que vengan a hacerlo otros.

utilizada en Wikipedia, Facebook, Twitter y LinkedIn, y se estima que posee un alto potencial de crecimiento en internet, debido a que la penetración media en los países hispanohablantes, que es del 66%, queda aún muy lejos de la penetración en España, que supera el 92%.

Procesamiento del lenguaje. A pesar de lo dicho, se estima que el español apenas representa el 30% del mercado mundial de las tecnologías de procesamiento del lenguaje. Aunque quizá habría que empezar por entender qué son las tecnologías de procesamiento del lenguaje natural o PLN. Se trata de un campo de conocimiento de la inteligencia artificial (IA) que pretende que las máqui-



nas imiten el comportamiento de la inteligencia humana, intentar procesar el lenguaje humano.

El objetivo es que las máquinas comprendan el lenguaje mediante el que nos comunicamos y sean capaces de generarlo, y además hacerlo de la forma más parecida a como lo haría una persona. Es decir: que una persona que esté hablando con una máquina sea incapaz de distinguir si realmente está hablando con una máquina o con un ser humano.

La IA trabaja en distintos niveles: desde que la máquina sea capaz de interpretar simplemente los hechos que se están comunicando, hasta ejecutar una orden o una respuesta al ser humano. Por otra parte, se está trabajando para intentar detectar aquello que hay detrás del lenguaje, es decir, si aparte de la información que la persona aporta la máquina es capaz de saber si ésta lo hace en un tono enfadado o de broma. Básicamente se trata de hacer que la máquina sea capaz de comportarse a nivel comunicativo de la misma forma que un ser humano, tanto para recibir como para producir lenguaje.

Una vez que ya sabemos qué es el PLN, cabría preguntarse por qué el español sólo supone el 30% del mer-



Otros enfoques, mismo tema



► El Plan de Impulso de las Tecnologías del Lenguaje de la Secretaría de Estado para el Avance Digital, tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales.

► <https://cutt.ly/KrrAG7c> 



► En este artículo de *El País Retina*, José Ángel Plaza analiza por qué el español sigue siendo una lengua extranjera para la inteligencia artificial.

► <https://cutt.ly/jrrAH7l>



► Este comentario de Antonio Moreno en el *Huffington Post* aborda el concepto de procesamiento de lenguaje natural, así como sus modelos y componentes.

► <https://cutt.ly/lrrAXgw>



➔ cado de las tecnologías de procesamiento del lenguaje natural a pesar del tirón que esta lengua tiene en el mundo. La más evidente es que el inglés es la lengua en la que se publican más documentos de carácter científico. El inglés no solo es la lengua mayoritaria en procesamiento del lenguaje natural sino en inteligencia artificial en general.

En España, gran parte de las investigaciones comienzan a hacerse en inglés principalmente porque hay más recursos, y luego se van trasladando a otras lenguas como el español. Un aspecto relevante es que, cada vez más, la adaptación de las investigaciones que se están realizando actualmente para lenguas como el inglés u otras es muy inmediata, gracias al empleo de técnicas de IA, como las de aprendizaje automático o lo que se conoce como *deep learning*. La consecuencia positiva derivada del uso de estas técnicas es que cada vez necesitan menos recursos y por lo tanto cualquier desarrollo o investigación que se hace para lenguas como el inglés, es fácilmente trasladable a otros idiomas.

Entonces, ¿qué tiene que ocurrir para que el español adquiera un mayor peso a nivel global? “Lo primero es tomar conciencia de que es un negocio, es un buen negocio”, señala Patricio Martínez, vicepresidente de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN). Hasta hace poco más de una década la investigación que se hacía sobre procesamiento automático de lenguas quedaba en el plano académico y muy pocas empresas lo incorporaban a su producción de sistemas.



El inglés no solo es la lengua mayoritaria en procesamiento del lenguaje natural sino en inteligencia artificial

Pero estas técnicas han evolucionado hasta tal punto en los últimos años que la posibilidad de asistentes virtuales que utilicen cada uno su propia lengua empieza a ser del interés de muchas compañías, lo que significa que el desarrollo de este tipo de herramientas es un negocio. “Tanto es así, que el negocio del español, por desgracia, no lo están liderando empresas españolas, ni siquiera hispanohablantes, sino que se está desarrollando por empresas estadounidenses, canadienses y europeas que saben que hay mucho que vender en español”, añade el vicepresidente de la SEPLN.

Aprendizaje automático. En los inicios del procesamiento del lenguaje natural, se producían técnicas basadas en teorías lingüísticas y gramaticales, que eran muy específicas de cada lengua, por lo que la tecnología producida para el inglés no valía para el español. Pero actualmente

Iniciativas

LOS expertos afirman que las tecnologías de procesamiento del lenguaje natural van a transformar la manera en que las personas se comunican mediante cualquier sistema. El Gobierno de España, a través de la Secretaría de Estado de Avance Digital, presentó en 2015 el Plan de Impulso de las Tecnologías del Lenguaje para servir de catalizador y desarrollar una industria y ofrecer mejores servicios públicos en terrenos como la salud, la educación o la justicia. Con una inversión de 90 millones de euros en cinco años su objetivo es liderar desde España este mercado del español y asegurar la presencia digital del gallego, el catalán y el euskera.

se están trabajando técnicas de IA de aprendizaje automático, que consiste en darle ejemplos a la máquina, cuantos más mejor, para que la máquina sea capaz de aprender y a partir de ahí, construya su propio modelo de lenguaje.

Con estas técnicas ya no son tan decisivas las investigaciones que se hagan para una determinada lengua, sino que esencialmente lo mismo que se ha hecho para el inglés se puede aplicar con otra lengua y el resultado va a ser bueno. Simplemente, tenemos que conseguir esas colecciones de información para que la máquina aprenda. Cada vez es más fácil hacer que una lengua nueva para la cual nunca se hubieran desarrollado estas técnicas pueda obtener sistemas con resultados razonables utilizando metodologías empleadas para otras. La dificultad está en conseguir un volumen suficiente de información para que la máquina pueda entrenar.

Futuro. En opinión de Patricio Martínez, las empresas españolas que se dedican al PLN están viviendo sus momentos más felices, en el sentido de cada vez más se está contando con ellas para diferentes entornos, por ejemplo, el médico, donde las tecnologías del lenguaje humano son capaces de leer múltiples historias clínicas y determinar diagnósticos, que es básicamente lo mismo que hace un profesional de la medicina, pero contando con las altas capacidades de volumen y de tiempo que tienen las máquinas. “En cinco años vamos a empezar a ver a grandes empresas españolas trabajando con estas tecnologías, porque realmente van a ser fuentes de negocio por diferentes sectores”, vaticina.

Frente a las grandes compañías que vienen de fuera a *comernos la tostada*, el vicepresidente de la SEPLN anima a creerse que el español debería ser un negocio impulsado desde España y que realmente en el futuro próximo va a haber empresas que van a centrar su modelo de negocio en el español “y que si desde aquí no somos capaces de hacerlo pues vendrán de fuera y lo harán”. Para Martínez, lo importante es tomar la iniciativa, y no solamente en España. “Desde aquí se podría estar liderando todo el desarrollo del español en Hispanoamérica.” ●



CARMELO ENCINAS,
periodista

✉ carmeloencinas@hotmail.com

🐦 [@CarmeloEnc](https://twitter.com/CarmeloEnc)

Los acentos de la inteligencia artificial

LA nave enfilaba Júpiter como una bala en el espacio sideral. Una bala, sin atmósfera que le opusiera resistencia, en la que viajaban cinco tripulantes, tres de los cuales lo hacían en estado de hibernación, una suerte de animación suspendida. El viaje era largo y aburrido. Los dos astronautas que se mantenían activos afrontaban una monotonía atroz para la que habían sido debidamente entrenados. Comían, hacían algo de ejercicio, como los hámsteres en la noria de su jaula, y de cuando en cuando mantenían alguna charla con sus familiares y amigos de la tierra sobre asuntos intrascendentes, nunca nada que guardara relación con su misión.

En realidad, con quien más conversaban era con HAL 900 quien, por cierto, destacaba por su habilidad jugando al ajedrez. La voz de HAL, en un inglés académico, era grave y metálica; no pretendía ser autoritaria pero su seguridad y aparente eficiencia lograba transmitir a los viajeros ese rigor. HAL no era humano, se trataba de un super computador de última generación con ojos y oídos en toda la nave que gestionaba los aspectos técnicos de la misión.

Este relato tampoco sale enteramente de mi imaginación, lo que aquí les cuento surge del recuerdo de *2001: Una Odisea del Espacio*, la película que revolucionó en su día el género de ciencia ficción y que, con unos efectos especiales igualmente revolucionarios, trató de proyectar una idea de futuro sobre la evolución del hombre y la carrera espacial. Su director, Stanley Kubrick, autor también del guion basado en un cuento de Arthur Clarke, la estrenó en abril de 1968, cuando aún faltaba más de un año para que el norteamericano Neil Armstrong pusiera pie en la luna.

Han pasado casi dos décadas de la fecha en que Kubrick situaba la acción de sus viajeros del espacio y es obvio que no se ha avanzado tanto como auguraba su relato, lo que no le resta un solo gramo de mérito a la cinta. Ni hay una base subterránea en la Luna, ni las naves de ahora son tan sofisticadas como él imaginó para el arranque del tercer milenio, ni la tecnología de hibernación nos permite ir mucho más allá de congelar un filete en la nevera para comérselo con plenas garantías de salubridad. En lo que sí acertó la visión de aquel guion cinematográfico en tiempo y casi en forma fue en el papel de HAL 900, la computadora parlante que facilitaba la vida a los astronautas.

A cualquier usuario corriente de los modernos dispositivos electrónicos, la voz de HAL le recordará a SIRI, Alexa, Cortana o cualquier otro asistente virtual de voz

ideados por las plataformas digitales y que, poco a poco, van sustituyendo a los teclados a la hora de reclamar información a internet u ordenar el cierre de puertas o el encendido de luces en una casa. La inteligencia artificial que lo hacía posible en la ficción de *2001: Una Odisea en el Espacio* es ya realidad en una vivienda domótica con bombillas, cerraduras y elementos electrónicos capaces de automatizar la casa y gobernarla con un simple impulso de voz. Los comandos parlantes son pues un hecho medio siglo después de la visión de Kubrick pero aún queda mucho por hacer en lo relativo a la capacidad de esta tecnología a la hora de interpretar los contextos, los tonos y los acentos de quienes emiten una orden o hacen preguntas.

Tengo un amigo gaditano con un acento tan cerrado que cuando le llevaron a urgencias por una reacción alérgica el médico, tan hispano parlante como él, pidió un intérprete para lograr entender lo que decía. No quiero imaginar cómo pueden ser sus conversaciones con los asistentes de voz del móvil o cualquiera de las operadoras virtuales que emplean las empresas de servicios.

En nuestro país, la variedad de dialectos, acentos y modismos del lenguaje es tan variada que se impone el uso de comandos de voz muy precisos y, desde luego, no pretender que los contestadores finjan una naturalidad en la conversación que termina resultando desesperante para el interlocutor de carne y hueso.

Otro tanto puede decirse de los países de habla hispana cada cual con un acento propio y toda suerte de modismos, vocablos y tonos de expresión autóctonos que complican las entendederas de la inteligencia artificial. Es obvio que el hecho de que los grandes avances en la tecnología digital hablen en inglés dificulta aún más el entendimiento con las máquinas para los hispano hablantes. El que la inmensa mayoría de los ensayos e investigaciones que tienen que ver con la inteligencia artificial sean en el idioma de Shakespeare y no en el de Cervantes determina el que los asistentes virtuales tengan en el inglés su lengua natural y que el castellano sea una aprendida.

Lo deseable sería convertir nuestra lengua en un impulsor de la tecnología en este ámbito de la ciencia introduciendo datos y creando algoritmos en español. Un idioma que hablan casi 600 millones de habitantes y cuya implantación e interés por su aprendizaje crece de forma exponencial.

Situarse a la vanguardia de la inteligencia artificial es de una importancia estratégica, aunque confío que los robots nunca lleguen a gestionar nuestra existencia hasta el punto de gobernarla. Según escribía estas líneas, recordé que los astronautas de la película de Kubrick, poco antes de concluir su misión, tuvieron que enfrentarse a HAL 900 porque éste tomaba sus propias decisiones. Tanto fue así que el super computador de marras terminó por cargarse a cuatro de los cinco tripulantes de la nave, y no acabó con el quinto porque al final, felizmente, el hombre resultó ser más astuto que la máquina. Menos mal.

«Lo deseable sería convertir nuestra lengua en un impulsor de la tecnología introduciendo datos y creando algoritmos en español»

‘El móvil en la nevera’

