

# IRONÍA EN LAS REDES

Un grupo de especialistas en lingüística y tecnología del lenguaje desarrolló un sistema para identificar el sarcasmo en la avalancha de mensajes de los diversos medios sociales.

Por José Luis Tapia

Ilustraciones: Edgar Gómez Álvarez



# E

n 2010, Toyota, el fabricante japonés de autos con mayor prestigio a nivel mundial, pasaba por la peor crisis en sus 70 años de historia. Las siguientes frases, escritas por dos clientes en una red social, son un ejemplo de la manera en que la gente se mofaba de la que tenía fama de ser una de las marcas más confiables del mercado: “El nuevo eslogan de Toyota: Avanzando (aunque tú no quieras)... ja ja ja” y “Toyota, avanzando... Sí porque tiene frenos defectuosos y aceleradores atorados”.

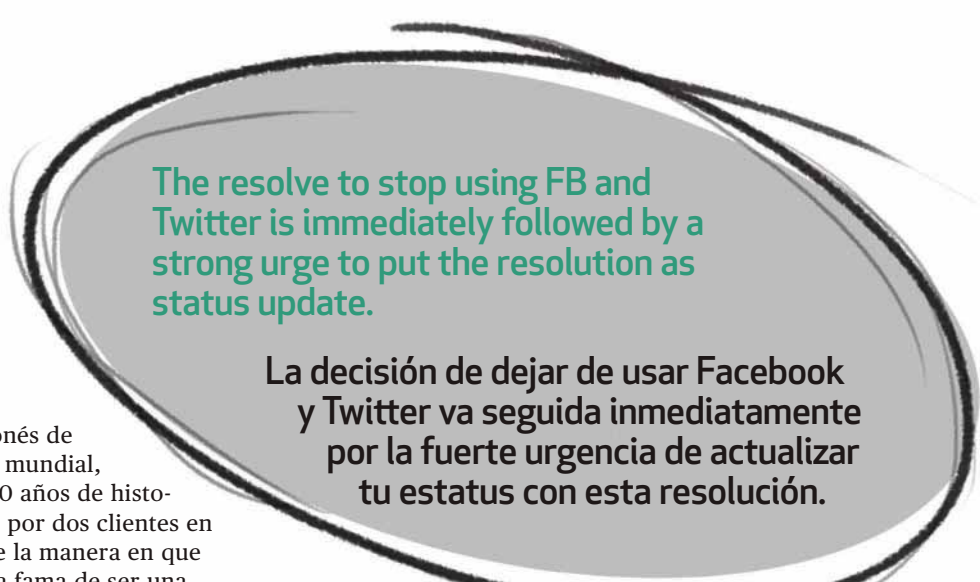
Ironías crueles –no hay ironía bondadosa, afirma el escritor y filósofo Mijail Malishev– que enmarcaron el momento en el que la armadora debió retirar de las calles del mundo alrededor de ocho millones de unidades para llevar a cabo las reparaciones correspondientes. En unos modelos el problema estaba en el pedal del acelerador, en otros, en el sistema de frenos. En efecto, los autos avanzaban... aún contra la voluntad del conductor.

A Antonio Reyes, lingüista egresado de la Facultad de Filosofía de la UNAM, y a Paolo Rosso, su asesor de doctorado en la Universidad Politécnica de Valencia (UPV) y experto en tecnología del lenguaje, se les ocurrió que en un mundo donde la gente puede expresarse con ayuda de cualquier medio digital –redes sociales, blogs, foros y páginas oficiales–, ya no era suficiente medir las opiniones positivas y negativas sobre un producto o servicio, sino ir más allá e intentar detectar la ironía, para lo que todavía no se había diseñado un sistema computarizado. Reyes, Rosso y su equipo de trabajo en el Grupo de Ingeniería del Lenguaje Natural y Reconocimiento de Formas de la UPV, se propusieron demostrar que era posible encontrar la ironía detrás de muchas frases o expresiones que solían agruparse como positivas o negativas, frases que sin pertenecer necesariamente a uno de estos dos rubros también brindaban información útil.

“Desde mi punto de vista y el de mucha gente, es un hecho que un sistema como el que desarrollamos es útil no solo para Toyota sino para cualquier empresa que se preocupe por conocer las opiniones de los consumidores-usuarios-público, a fin de tomar decisiones que puedan revertir los puntos débiles o malos que sus productos tienen”, señala el mexicano, cuya tesis de maestría se convirtió en libro y se vende en Amazon con el título de *Esquemas cognitivos para la representación de conocimiento: Semántica y pragmática en los sistemas de recuperación de información*.

## ¿Cómo enseñar a una máquina a entender la ironía?

Paolo Rosso es lingüista computacional, con especialidad en el procesamiento del lenguaje natural y recuperación de información. Antes había trabajado en proyectos para la detección de plagio y el análisis de textos cortos; su trabajo incluye tareas de procesamiento automático de textos, con conocimientos lingüísticos y estadísticos. “Sería poco práctico, y a veces imposible, hacerlo manualmente”, comenta. Su perfil coincidía con lo que Antonio estaba buscando para continuar sus estudios de doctorado en

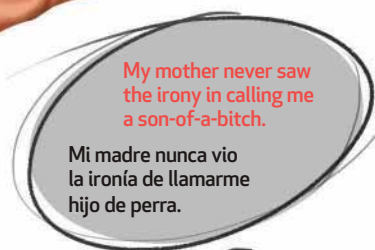


The resolve to stop using FB and Twitter is immediately followed by a strong urge to put the resolution as status update.

La decisión de dejar de usar Facebook y Twitter va seguida inmediatamente por la fuerte urgencia de actualizar tu estatus con esta resolución.



Estos son algunos ejemplos de frases irónicas con características que el sistema diseñado por Antonio Reyes y Paolo Rosso sí pudo identificar.



My mother never saw the irony in calling me a son-of-a-bitch.

Mi madre nunca vio la ironía de llamarme hijo de perra.



## Va por el maya

Antonio Reyes volvió a México y planea trabajar con una de las lenguas prehispánicas más importantes del país: el maya.





### ¿Qué significa "ironía"?

Es una burla fina y disimulada en que se da a entender lo contrario de lo que se dice.

If you find it hard to laugh at yourself, I would be happy to do it for you.

Si te resulta difícil reírte de ti mismo, yo estaría feliz de hacerlo por ti.



No es que la computadora entienda la ironía o el chiste. Lo que sí puede hacer es identificar patrones lo suficientemente discriminatorios y relevantes.”

la UPV, con apoyo de una beca del Conacyt. Cuando se integró, Rosso ya trabajaba en la desambiguación de las palabras, es decir, en determinar el sentido correcto de una palabra polisémica –con varios significados–, basándose en su contexto. En el proceso se dio cuenta del papel que la ambigüedad tenía en los textos humorísticos, en los cuales es un ingrediente principal, sobre todo en textos de doble sentido. “Siendo Antonio un lingüista nos pareció sumamente interesante colaborar para que pudiéramos investigar conjuntamente cómo se podría abordar la tarea del reconocimiento de humorismo, ya que además es importante para otras tareas como, por ejemplo, la traducción automática. Si un texto humorístico se traduce literalmente, la mayoría de veces pierde su gracia”, agrega.

En el proceso se dieron cuenta de la presencia de un mecanismo de lenguaje figurado, como la ironía, y de lo difícil que era determinar si una opinión era positiva o negativa.

Hubo que pasar por una fase de aprendizaje en la que se procesaron miles de ejemplos de opiniones con y sin ironía, y decidieron diseñar el sistema en inglés, idioma en el que se cuenta con mayor número de recursos léxicos en la red. “Se hicieron estudios para saber cuáles son

los patrones lingüísticos que se emplean mayormente en caso de opiniones irónicas, por ejemplo, ambigüedad a nivel de palabra, sintaxis, semántica, adverbios que implican un cambio, contradicciones en la primera parte de la opinión y en la segunda”, relata Rosso.

Incongruencias que podrían detectarse en una expresión como “Tengo un cuerpo de Dios, lástima que sea Buda”.

Después comenzaron a establecer patrones que no coincidían necesariamente con la definición literaria de la ironía: “La literatura dice que en la ironía lo esencial es que hay una negación de lo que estás diciendo, pero la gente la entiende mucho más en relación con un aspecto humorístico o una cuestión negativa hacia una tercera o segunda persona. No importa si los tratados de ironía establecen esas características, la gente va mezclando sus nociones de esta, de sarcasmo, de humor, hasta insulto, y eso es lo que a nosotros nos interesaba reconocer”.

Entonces se encontraron con el caso de Toyota, que sin solicitarlo, se volvió su materia de estudio. “No nos enfocamos en etiquetas respecto a si era o no irónico lo que se escribía de la marca, sino en tratar de evaluar nuestros modelos más allá de los datos que ya teníamos. La elección de qué era irónico y qué no la hicimos de forma manual y automática, y nuestro sistema tuvo un comportamiento muy parecido al análisis hecho por los evaluadores humanos: nunca al 100%, pero sí en un rango de 70%”.

Identificar siete de cada diez mensajes irónicos resulta relevante, sobre todo si se toma en cuenta que incluso como hablantes nativos de una lengua determinada es difícil percatarse de que el interlocutor está siendo irónico, destaca Antonio.

“No es que la computadora entienda la ironía o el chiste. Lo que sí puede hacer, y lo hace muy bien, es identificar patrones lo suficientemente discriminatorios y relevantes que, dado un modelo previo, serán representativos de la ironía”, explica. Lo ejemplifica con el siguiente esquema: Si establecemos que la ironía es lo contrario de lo que decimos literalmente, se puede crear un modelo en donde la negación (1), el orden (2), la presencia de marcas tipográficas (3) y la incongruencia (4), sean patrones de un texto irónico. Así, cuando el sistema encuentra un texto en el que aparece uno, algunos o todos estos valores, puede determinar, con cierta probabilidad, que el texto es irónico. No lo sería si el sistema señalara los valores 1, 4, 5, 8, 0 y 3, pero sí en el caso de 2, 5, 2, 4, 3 y 1.

No se trata de magia, explica Rosso, sobre todo si se toma en cuenta el componente subjetivo que los resultados de la búsqueda puedan tener. “La ironía cambia mucho en función del estilo del texto, del género. Una cosa es hablar de ironía en literatura y otra cosa es hablar de ella en medios sociales, donde además, en el caso de Twitter, hay limitaciones de espacio; se escribe con abreviaciones o con *slang*”.

Antonio reconoce que el sistema no es infalible. Al estar basado en lenguaje natural, es dinámico y siempre interactúa con las reglas gramaticales y cómo estas son utilizadas por los hablantes. “Hay ironías que escapan a nuestros patrones porque están mayormente sustentadas en información extralingüística que es muy difícil de representar en un sistema computacional. Además, hay que considerar que existen varios tipos de ironías: verbales, situacionales, discursivas o dramáticas, lo que hace que el espectro de identificación aumente y disminuya”.

Un par de ejemplos irónicos que el sistema muy probablemente no detectaría tienen que ver con los diferentes tipos de ironías que existen; por ejemplo, ironía dramática (cuando Romeo encuentra a Julieta que ha fingido su muerte al beberse una droga, Romeo asume que ella en verdad está muerta y se suicida; Julieta, al despertar, se da cuenta de la situación y termina por suicidarse), o ironía situacional (en *El mago de Oz*, cuando todos los personajes se dan cuenta de que todo lo que buscan lo tenían antes de empezar su búsqueda). Este tipo de ironías, explica Antonio, son muy complicadas de detectar porque implican que el sistema entienda y sea capaz de seguir todo un proceso comunicativo, el cual está dado por la capacidad de inferir, contextualizar, hacer implicaciones, etc., que muchas veces, nosotros mismos no somos capaces de seguir

## ¿Ingeniería lingüística?

En 1999 se creó el Grupo de Ingeniería Lingüística (GIL) dentro del Instituto de Ingeniería de la UNAM, para desarrollar sistemas informáticos, herramientas, métodos y productos que permitieran resolver problemas sociales mediante el estudio y procesamiento automatizado del lenguaje. Hoy se estudia además la terminología y terminológica, y la lingüística forense (para la detección de plagio y fraudes). Un ejemplo del trabajo que realizan los ingenieros lingüísticos son los traductores electrónicos. El trabajo más reciente del GIL fue la interfaz del corpus del Español Mexicano Contemporáneo ([www.iling.unam.mx/chem/](http://www.iling.unam.mx/chem/)), una colección de documentos diacrónicos de diferentes géneros textuales.

Let's pray that the human race never escapes from Earth to spread its iniquity elsewhere.

Recemos para que la raza humana nunca escape de la Tierra para propagar su iniquidad en otro lugar.

I spilled precious coffee on myself because I hadn't had enough coffee to function properly.

Me eché el café encima porque no tenía suficiente café para funcionar adecuadamente.



WEB quo.mx

Manifiéstate en nuestras redes

@RevistaQuo y Facebook/RevistaQuo.mx

## ¿Qué sigue?

Dice Mijail Malishev, autor de *Amigos, la vida es irónica*, que la ironía es un desprecio fino que oculta sus garras poniéndose guantes de seda. Para el profesor de la Universidad Autónoma del Estado de México, ser irónico es parte inagotable del ser humano. En su libro incluye un aforismo: “El deseo de expresarnos es más fuerte que el deseo de aprender, porque no exige mucha energía, y la ‘epidemia’ de los teléfonos celulares lo confirma”. Siempre estamos opinando. Nunca estamos contentos. El irónico, agrega, es un burlón obsesionado por la manía de lo inesperado.

Esta quizá sea una justificación más para la existencia del detector de ironía, un producto en el que Rosso ve gran potencial, sobre todo desde un punto de vista empresarial. “Puede ser interesante también aplicarlo en otros entornos, como en el análisis de opiniones con respecto a un político y su reputación entre la gente”, dice.

A partir de que su invento se dio a conocer en junio de 2012, después de cuatro años de investigación, varias empresas se han interesado, sobre todo las dedicadas a crear *software* a partir del procesamiento del lenguaje natural. Pero según admiten, la fase de verlo como negocio es algo a lo que aún no dedican los esfuerzos necesarios. “A nivel de trabajo de investigación nos cuentan más publicaciones en revistas con factor de impacto, que sacar o no un producto al mercado, sobre todo en

este momento de crisis que vivimos en España y Europa”, indican. Reyes es optimista y cree que es cuestión de tiempo y del trabajo que hagan los expertos en negocios y administración. Menciona un caso de éxito que podría replicarse con su invención: “Google: hace algunos años su traductor era muy malo y la mayoría de las traducciones automáticas eran hechas en otros buscadores (Altavista o Yahoo). Hoy su traductor es de los más usados y su efectividad va en aumento dada la cantidad de datos lingüísticos con que cuenta y que le sirven para crear sus modelos. Otro par de ejemplos son los sistemas de reconocimiento y síntesis de voz (usados hoy día por casi todos los teléfonos inteligentes, tabletas y computadoras) y los de reputación *online*. Basta recordar cómo, en las pasadas elecciones, el tema de la popularidad era un factor que todos los días se medía para estimar la repercusión de los candidatos”, dice.

El *software* se puede usar en cualquier computadora, y como con otros programas, ofrecerá resultados más rápidos si se tiene un buen procesador y una buena cantidad de memoria. En 2013 esperan poder centrarse en el desarrollo de la interfaz para los usuarios. El siguiente paso en la investigación será hacer más poderosa la herramienta. Otra meta será habilitar el detector de ironía en otros idiomas. En el español de México implica un reto complicado, reconoce Reyes: “Depende, primero, de la existencia y disponibilidad de recursos léxicos que nos permitan desarrollar el modelo”. ❏

## La ironía del detector

El Dr. Lauro Zavala, autor de *Ironías de la ficción y la meta-ficción en cine y literatura*, duda de la eficiencia de un detector de ironía, un recurso que depende del contexto. Agrega además que la ironía posmoderna es una maquinaria de producción de frases cuyo sentido no es posible decidir de manera definitiva, ni siquiera por quien las escribió. “Una tecnología que pretende reconocer la intención irónica en un contexto tecnológico definido por la ambigüedad semántica, sintáctica y pragmática... es algo muy irónico, es decir, un oxímoron absurdo”.