



Káñina

ISSN: 0378-0473

ISSN: 2215-2636

Universidad de Costa Rica

Casasola Murillo, Edgar; Leoni de León, Jorge Antonio  
Transferencia de la polaridad semántica de frases idiomáticas a comentarios de opinión  
Káñina, vol. 40, núm. s1, 2016, Julio-Diciembre, pp. 65-76  
Universidad de Costa Rica

DOI: <https://doi.org/10.15517/rk.v40i3.28813>

Disponible en: <https://www.redalyc.org/articulo.oa?id=44254856006>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH  redalyc.org

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto

## TRANSFERENCIA DE LA POLARIDAD SEMÁNTICA DE FRASES IDIOMÁTICAS A COMENTARIOS DE OPINIÓN

*Transfer of semantic polarity of idiomatic phrases to review comments*

*Edgar Casasola Murillo\**

*Jorge Antonio Leoni de León\*\**

### RESUMEN

Las reacciones en redes sociales a propósito de un comentario pueden ser clasificadas en positivas, negativas o neutras, según la opinión del usuario o emisor sobre el tema en juego. Las frases idiomáticas son portadoras de un valor emocional, que va de positivo a negativo (lo que conocemos como polaridad). En nuestra investigación, establecimos que la aparición de una frase idiomática negativa en una opinión en redes sociales está directamente vinculada con la polaridad de la totalidad del comentario. Es decir, que la polaridad de la expresión es transferida al comentario propiamente dicho.

**Palabras clave:** análisis de sentimientos, polaridad, redes sociales, opinión, expresiones idiomáticas.

### ABSTRACT

The reactions on social media about a specific comment can be classified in positive, negative o neutral, depending on the user's opinion about the issue at stake. Idiomatic phrases are bearers of emotional value, going from positive to negative (what we know as polarity). In our investigation, we established that in social media the appearance of a negative idiomatic phrase in an opinion is directly linked to the polarity of the entire comment. In other words, that the polarity is transferred to the comment itself.

**Key Words:** sentiment analysis, polarity, social networks, opinion, idioms.

---

\* Profesor de la Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica, Costa Rica. Correo electrónico: casasola@gmail.com

\*\* Profesor de la Escuela de Filología, Lingüística y Literatura e investigador del Instituto de Investigaciones Lingüísticas. Costa Rica. Correo electrónico: antonio.leoni@ucr.ac.cr

*Recepción: 2/3/16. Aceptación: 10/8/16.*

## 1. Introducción

Las expresiones idiomáticas son agrupaciones léxicas, frecuentes en la lengua, caracterizadas por una diversidad sintáctica que puede ir desde una fijación absoluta o fosilización, como en *saco de dormir o el cliente siempre tiene la razón*, hasta locuciones muy versátiles, como *meter la pata*, la cual incluso permite la cliticización de su argumento interno, “la pata” (Leoni de León, 2008), a condición de agregar un complemento circunstancial (“meterla hasta el fondo”). A menudo, estas expresiones están cargadas de connotaciones emocionales muy fuertes (lo que denominamos *sentimiento*), por lo que su realización nos permite identificar la tendencia positiva o negativa de una opinión. Esta tendencia es lo que llamamos *polaridad* y la *transferencia* la entendemos como la influencia ejercida por la connotación de la expresión idiomática en la opinión publicada. Es este aspecto del análisis de sentimientos, el que abordamos en la presente investigación y constituye uno de los eslabones de dos proyectos más amplios, uno llamado *Análisis de sentimientos: una aproximación híbrida*<sup>3</sup>, abocado al desarrollo de estrategias híbridas (es decir, lingüísticas y estadísticas) para el análisis de opiniones en redes sociales, y otro, llamado *Evaluación de algoritmos para análisis de sentimiento*<sup>4</sup>, enfocado en los aspectos meramente computacionales. En esta etapa, obtuvimos evidencias de una posible transferencia de la polaridad asociada a una frase idiomática hacia el texto que la contiene; en su mayoría se trata de frases idiomáticas negativas. Más adelante, ampliaremos la cobertura del método, tomando en consideración una gama mayor de polaridades.

## 2. ¿Qué es el análisis de sentimientos?

El análisis de sentimientos es un área de investigación que busca identificar y analizar automáticamente opiniones emitidas en foros de Internet o en redes sociales, a propósito de publicaciones específicas enunciadas por un emisor conocido. Siguiendo a Liu y Zhang

(2012), consideramos el análisis de sentimientos como una disciplina orientada al reconocimiento automático de la opinión, los sentimientos, las evaluaciones, las percepciones y las actitudes a propósito de productos, servicios, organizaciones, individuos, asuntos, eventos, temas o sus atributos. De forma general, toda manifestación verbal mediada por una emoción es abarcada dentro del término *sentimiento*; es importante recordar que un sentimiento está asociado, entonces, a un individuo que emite una opinión, es decir, que comparte públicamente su punto de vista sobre *algún aspecto* asociado a *alguna entidad*. Ahora bien, el término *entidad* está referido a algún producto, servicio, organización, individuo, asunto, evento, tema o sus atributos. La entidad es el sujeto sobre el que se emite la opinión. La característica específica sobre la cual la opinión es emitida es denominada *aspecto*. Si el aspecto al que se refiere la opinión es la entidad misma, entonces hablamos del *aspecto general*. De lo contrario, las partes de una entidad, o sus características concretas o abstractas pueden ser los aspectos sobre los que se emite la opinión. Por ejemplo, un producto, como un automóvil, tiene aspectos asociados a características físicas como el tipo de llantas y el color; otros aspectos, más abstractos, son la comodidad al conducir o un precio justo. Entonces, el *aspecto general* está asociado con una opinión sobre el automóvil propiamente dicho, tal como ocurre en el ejemplo (1):

(1) Me encanta el Audi Q7.

Ahora, si en una de las fuentes de información aparece una noticia como la que se muestra en la , el texto de la noticia expresa el punto de vista del medio de comunicación que la difunde, lo cual puede dar pie a reacciones de los lectores que son publicadas en línea en una serie de intercambios bajo la publicación original. Así, el acceso directo a las opiniones emitidas directamente por los lectores hace muy valioso el análisis de sentimientos, por cuanto permite tomar el pulso de la opinión pública sobre temas específicos; no obstante,

la gran cantidad de noticias y comentarios emitidos, cada día, en cada uno de los medios, imposibilita realizar manualmente esta labor, de

ahí que es necesario desarrollar métodos para llevarla a cabo automáticamente. El análisis



FIGURA 1.

Ejemplo de noticia publicada en algún medio de comunicación

de sentimientos pretende, entonces, identificar y obtener opiniones a partir de comentarios de usuarios de medios informativos a fin de estimar la tendencia del conjunto de opiniones. Siguiendo a Liu y Zhang (2012), en cada texto que contenga una opinión, esperamos obtener la siguiente información:

1. **La entidad** a la que se refiere el texto. Sea una persona o un objeto.
2. **El aspecto** de esa entidad sobre el que se está emitiendo la opinión. Puede ser general (si se refiere a la misma entidad) puede ser una característica propia de la entidad como su color, apariencia, olor o bien puede ser una parte de la entidad
3. **El autor** (denominado *opinion holder* en inglés) es la persona o usuario a la que se le atribuye la opinión.
4. **El sentimiento** (o **polaridad**) es el elemento esencial de una opinión, se refiere al hecho de si la opinión es una afirmación positiva o negativa sobre el aspecto en cuestión. Asigna uno de dos posibles valores (positivo o negativo) a la opinión. También se acostumbra manejar niveles de granularidad más específicos, por ejemplo:

como la pantalla de un televisor o su control.

CUADRO 1.

Indicadores de valor

Indicadores	Valor
+P	Muy positivo
-P	Algo positivo
N	Neutro
-N	Algo negativo
+N	Muy negativo

5. **La fecha y la hora** en que se emitió una opinión son importantes, puesto que un emisor o usuario puede cambiar de punto de vista; variación que se nota en el tiempo.

La noticia tiende a generar opiniones por parte de los lectores o personas que la leen. Esto genera una serie de comentarios como

los que aparecen en la Figura 2. De ellos se pueden obtener entonces opiniones como las expresadas en los Cuadros 2 y 3. El análisis de sentimientos tiene diferentes vertientes, por lo que su estudio ha sido dividido en tareas. El cálculo de la polaridad, como el efectuado en esta investigación, es una de sus tareas primordiales.



FIGURA 2.

Comentarios emitidos como producto del comentario inicial

CUADRO 2.

Ejemplo de opinión general

<b>Comentario:</b>	Me parece excelente
<b>Entidad:</b>	Mega terminal de transbordo de Moín
<b>Aspecto:</b>	General
<b>Autor:</b>	Usuario 2
<b>Sentimiento:</b>	+P
<b>Fecha:</b>	2 de noviembre 2014
<b>Hora:</b>	21:05

CUADRO 3.

Opinión sobre la corrupción

<b>Comentario:</b>	Y empezó el chorizo todos a concursar el que pague mas mordida se gana el contrato att : El Gobierno
<b>Entidad:</b>	Mega terminal de contenedores de Moín
<b>Aspecto:</b>	Corrupción
<b>Autor:</b>	Usuario 4
<b>Sentimiento:</b>	+N
<b>Fecha:</b>	18 de noviembre 2014
<b>Hora:</b>	13:44

### 3. Antecedentes

La literatura especializada sobre análisis de sentimientos es bastante amplia, en el ámbito de nuestra investigación cabe citar a Liu y Zhang (2012), a Cambria *et al.* (2013) y a Pang y Lee (2004). Ellos exponen el giro natural en la investigación sobre la lengua inglesa y justifican la necesidad de abarcar otras lenguas. Estas carencias y necesidades también han sido reconocidas por parte de empresas privadas, altamente dependientes de investigación y de desarrollo, como Google<sup>6</sup>, para la cual las *Tecnologías del Lenguaje*<sup>7</sup>

son una base importante para sus negocios, de ahí que la incursión en lenguas, otras que el inglés, sea una prioridad. No está de más mencionar que también hay propuestas en el sentido de esquemas para anotación de texto para el reconocimiento unificado de múltiples lenguas (McDonald *et al.*, 2013).

Es importante notar que sobre el español es bastante poco lo que ha sido producido. Generalmente, las propuestas van en el sentido de reproducir métodos desarrollados para el inglés, con resultados más pobres en español (Fernández Anta *et al.*, 2012) o métodos basados en el uso extensivo de traducción

automática, que fracasan por concatenación y multiplicación de errores (Brooke *et al.*, 2009). En la mayoría de los experimentos, los nuevos sistemas son comparados contra sistemas basados en clasificadores supervisados (Hutto y Gilbert, 2014). Investigaciones, como las de Fernández Anta *et al.* (2012), y Batista y Ribeiro (2013), convierten los textos en bolsas de palabras o grupos de pares de palabras (o incluso sus respectivos pesos) utilizando modelos matemáticos, principalmente estadísticos y vectoriales. Los resultados no son satisfactorios. Por lo tanto, es necesario desarrollar métodos propios, para el análisis de sentimientos, aprovechando las particularidades morfosintácticas y variantes de uso del español.

La extracción de las características para el análisis a partir del texto es uno de los aspectos más importantes de las investigaciones recientes, según lo resaltan Cambria *et al.* (2013). La calidad de los clasificadores automáticos y algoritmos de clasificación computacionales dependen de las características identificadas a partir del texto (Chenlo y Losada, 2014). En este caso, la idea fue explorar si la expresión idiomática es una característica importante que puede ser identificada y utilizada para la tarea de identificación de polaridad en análisis de sentimientos.

Nuestra investigación se centra en la identificación de frases idiomáticas a partir publicaciones en una red social, la que nos proveyó comentarios de opinión publicados

en línea. Además, procedimos a estimar la polaridad de dichas expresiones idiomáticas, la cual pusimos en relación con la polaridad de los textos (u opiniones) extraídas.

#### 4. Frases idiomáticas

Muchas son las definiciones posibles para el concepto de *expresión idiomática*, *locución o unidad fraseológica* (Leoni de León, 2008). Nosotros preferimos basarnos, operativamente, en una definición formal, de manera que para nuestra investigación:

**(2) Una Frase Idiomática** es una locución **L** con uno o más significados **S**, donde:

- (a)  $|L| \geq 2$
- (b) **S** es arbitrario<sup>8</sup>

Con la definición tenemos que el significado de una frase idiomática no depende del significado de los componentes que la conforman, y siempre estará compuesta por al menos dos términos.

Las frases idiomáticas con las que trabajamos fueron obtenidas en forma automatizada mediante herramientas computacionales desarrolladas *ad hoc*. En primer lugar recopilamos el corpus CR2013FB utilizando los recolectores de información de la herramienta Sentímetro, tal como se muestra en la Figura 3.



Plataforma Sentímetro. Extracción de comentarios.

FIGURA 3.

Los comentarios fueron recolectados en forma automática a partir de fuentes en línea.

Las características del corpus CR2013FB son las siguientes:

- *Conformación:* Publicaciones y comentarios de los foros costarricenses más populares en Facebook.
- *Localización:* Todos los comentarios refieren a Costa Rica.
- *Período:* Todas las publicaciones abarcan la totalidad del año 2013.
- *Tópicos:* El año 2013 precedió dos hechos relevantes en la vida política y deportiva del país: las elecciones presidenciales (febrero 2014) y la participación de la Selección Nacional en el Campeonato Mundial de Fútbol (Brasil, 2014). De ahí que los temas más frecuentes estén relacionados con la política y el deporte.
- *Tamaño:* El corpus consiste en 1 910 534 comentarios y 246 707 términos lematizados<sup>10</sup>.

El corpus consiste en comentarios de los foros de Facebook<sup>11</sup> más populares en el 2012 según el PROSIC (Arce Sanabria, 2012), específicamente, se trata de los siguientes perfiles:

- Al Día<sup>12</sup>
- Amelia Rueda (<https://facebook.com/ameliarueda>)
- Canal 9 (<https://facebook.com/canal9cr>)
- CR Hoy (<https://facebook.com/crhoy.comnoticias>)
- Diario Extra (<https://facebook.com/diarioextra>)
- El Financiero (<https://facebook.com/elfinanciero>)
- La Nación (<https://facebook.com/lanacion>)
- La República (<https://facebook.com/larepublica>)
- La Teja (<https://facebook.com/latejaoficial>)
- Repretel (<https://facebook.com/repretelcostarica>)
- Canal 13 (<https://facebook.com/sinartcanal13>)
- Teletica (<https://facebook.com/Teletica>)

La comparación entre los volúmenes de comentarios publicados en las versiones web de los medios de comunicación y sus correspondientes perfiles de Facebook, nos motivó a preferir la información en esta red social respecto de los sitios web oficiales. Así, comprobamos que la preferencia de los costarricenses, en el año 2012, decantó más por publicar en Facebook y menos por hacerlo en los foros de las páginas de cada una de las fuentes. Por esta razón, nuestro sistema de recolección (lo que llamamos, el recolector) estuvo focalizado en dichos perfiles de Facebook. Para mantener la privacidad de los usuarios se eliminaron los nombres de los usuarios que expresaron opiniones.

## 5. Metodología

El procedimiento seguido en este trabajo fue el siguiente:

- Selección manual de publicaciones, por parte de dos anotadores, a partir sitios de los candidatos presidenciales (corpus Candidatos2013).
- Identificación de expresiones idiomáticas (424 en total).
- Muestra aleatoria de expresiones idiomáticas (30 expresiones y sus variantes).
- Asignación manual, a las expresiones idiomáticas, de polaridades positivas o negativas.
- Búsqueda de comentarios usando las expresiones idiomáticas en el corpus CR2013FB.
- Inclusión de más variantes y nuevas búsquedas de comentario usando las variantes de las frases idiomáticas originales.
- Comprobación por muestreo de la polaridad en los comentarios.

Con el fin de no tomar la lista de frases idiomáticas del mismo corpus de donde se analizaría su polaridad, se utilizó un grupo de comentarios tomados de las páginas de



los diferentes candidatos presidenciales. A este grupo de comentarios se le denominó corpus Candidatos2013 y fue creado de forma similar al corpus CR2013FB. El corpus Candidatos2013, más pequeño, consiste en 227 866 comentarios. De la lista de comentarios fueron extraídas 30 expresiones idiomáticas, las cuales fueron identificadas según su polaridad (positiva o negativa). Únicamente extrajimos 30 expresiones idiomáticas por limitaciones de recursos humanos (anotadores de texto calificados) y de tiempo para llevar a cabo la anotación manual e identificación de la polaridad.

La búsqueda de comentarios que contienen las expresiones idiomáticas implicó la creación de un programa para la indexación de búsquedas por computadora y la realización de búsquedas automáticas de comentarios de opinión mediante consultas en formato de texto. Este motor fue desarrollado específicamente para este proyecto y es un producto importante de la investigación. El primer prototipo del motor permite extraer todos los comentarios que contienen las palabras pertenecientes a la frase idiomática buscada, sin tomar en cuenta ni el orden ni la distancia. Esto con el objetivo de identificar otras variantes posibles de la frase al momento de los analizar los resultados. Los resultados no son deterministas, por lo que no fueron producidas estadísticas a partir de ellos, y brindan, a su vez, una idea de nuevas variantes, por lo que, manualmente, el anotador procedió a identificar variantes para las 30 expresiones iniciales con el fin de expandir las búsquedas. Finalmente, realizamos una valoración de la polaridad asociada a los textos de los comentarios encontrados y determinamos la relación con la polaridad de las frases iniciales.

## 6. Resultados

De la masa de resultados, extrajimos una muestra de ejemplos típicos que ilustran la relación entre las expresiones idiomáticas y el sentimiento en las opiniones. Cada ejemplo consiste de la frase, la polaridad atribuida y

las variantes, si las hay. Luego, incluimos los comentarios donde aparece la expresión. Es muy importante aclarar que la ausencia o presencia de acentuación y los errores corresponden a la ortografía tal y como aparece en el corpus, los textos son copia original de lo digitado por los usuarios. Nuestro primer caso está relacionado con la frase *pura habladuría*:

(3) **Frase:** pura habladuría

**Polaridad atribuida:** negativa

**Consultas incluidas:** puro bla bla bla, pura paja, pura habladuría, pura hablada

En el comentario 1566375, aparece una variante de la frase idiomática, “puro bla bla bla”, ejemplo (4). El usuario claramente alude despectivamente a la falta de acción y verborrea inútil por parte de la entidad a la que se refiere. Este comentario es claramente negativo.

(4) Comentario 1566375: Pero como siempre todo se va a ir en pura hablada, no van a disparar ni un misil, esa vara da pereza no se que tanta hablada puro bla bla bla y nada de pla.

El segundo caso está relacionado con la corrupción, ahí aparece la frase “huele a chorizo”:

(5) **Frase:** huele a chorizo

**Polaridad atribuida:** negativa

**Consultas incluidas:** huele a chorizo

En el comentario 439153, se muestra como el uso de la frase manifiesta la sospecha de cierto grado de corrupción atribuible a la forma de proceder de la entidad a la que se refiere el comentario.

(6) Comentario 439153: Esto huele a chorizo podrido, esto no huele bien, quien sabe a quien quieren favorecer y volver a pisotear a nuestros agricultores, aqui hay gato encerrado...

En el comentario se critica a un jugador por regresar arrepentido de una decisión tomada en el pasado y solicitar una nueva oportunidad, lo que supone una cierta humillación personal atribuida por el usuario al jugador que en este caso es la entidad a la que se dirige la opinión.

(7) **Frase:** pidiendo cacao**Polaridad atribuida:** negativa**Consultas incluidas:** pedir cacao, pide cacao, piden cacao, pidiendo cacao**Variantes:** No se encontraron variantes.

- (8) Comentario 42795: es un tico que juega en rusia o alemania no se donde..pero en el 2010 dijo que nunca jugaria con Costa Rica y ahora viene a pedir cacao jajaja

La polaridad de (7) y (8) son negativas. No identificamos ninguna variante para la frase idiomática “pidiendo cacao”.

**6.1. Frases idiomáticas con alta transferencia de polaridad**

Un elemento que llamó la atención de las 30 frases idiomáticas identificadas originalmente fue su alta connotación negativa. Luego, también notamos que en todos los casos en que esas frases idiomáticas se realizaron, la polaridad fue transferida, de manera que los comentarios que las contenían también tenían una connotación negativa. Las 30 frases idiomáticas que se analizaron fueron las siguientes:

CUADRO 4.

Frases idiomáticas analizadas

1. a lajazo limpio	15. pura habladuría
2. cuatro gatos que estan en contra	16. qué payasada
3. digan lo que digan	17. que payasos
4. embarrialar la cancha	18. que pegas
5. hacerse el loco	19. saco charita saco
6. hasta aqui	20. salgo por dentro
7. hay gente que vive en la luna	21. se la pela
8. hilar muy fino	22. se pela el culo
9. huele a chorizo	23. siga soñando de ese lado
10. (les) vale culo	24. sin vergüenzas
11. más perdido que el chiquito de la llorona	25. son de la misma calaña
12. meterán las patas	26. tiran piedras
13. pidiendo cacao	27. todos los políticos son iguales
14. pocos gatos	28. tormenta vaso agua

Al incluir las variantes de las frases se logró aumentar la cobertura, pero se hace necesario establecer un mecanismo que permita de forma automática el aumento de la cobertura, ya que estas frases tienen a ser un excelente indicador de polaridad (en este caso negativa). Entre las propuestas para extender la cobertura está la transformación de todo el

texto, ya que existen variantes lexicográficas, como abreviaturas, emoticonos que no son elementos gramaticales en sí, mayúsculas, errores de acentuación, errores de digitación, entre otros, que deben homogeneizarse. Véase por ejemplo algunos de los casos asociados al perfeccionamiento de la identificación de variantes (Cuadro 5).

## CUADRO 5.

## Perfeccionamiento de la identificación de variantes

<b>Código</b>	meterse en -> meter c en
<b>Variantes</b>	a) más perdido que {el chiquito de la Llorona, el hijo de Lindbergh} b) meter {la pata, las de andar}
<b>Ortografía</b>	ahí, ahí, ai
<b>Metátesis y sustituciones</b>	sustitución -> sustitución

En el ejemplo (9) resaltamos en negrita algunos de los fenómenos mencionados en el Cuadro 5, donde se muestra que no solo los usuarios de la lengua utilizan algún tipo de código en sus comentarios, sino que, además, es algo frecuente:

- (9) [...] Con **respto** a **tdos** **ustdes**, soy pequeño agricultor, y para nosotros si es **d** suma **imxtancia**, yo no opino **d** dom pèrignom, caviar, filet mignón, **xq** no conosco estar en ese nivel, pero para **nstros**, **q** **n** compran el **kintal d** maíz en 9mil colones y aparte **d** eso **tgmos q** comprar un **kintal d** semilla en 100mil, es [...]

No podemos dejar de mencionar la necesidad de adecuarse a diversos contextos socioculturales, por ejemplo, de un país a otro, de un tema a otro, etc. Esto nos mueve a identificar, en otra etapa de la investigación, más *punteros* que denoten emoción en el texto. Por ejemplo, valdría la pena analizar en detalle el uso de emoticonos, como los citados en (10).

- (10) (a) Caritas felices: :-), XD  
(b) Guiño: ;-)

Para determinar claramente la entidad y el aspecto al que se refieren los comentarios también es necesario contemplar el comentario individual en su contexto. En este caso, el contexto parece ser la noticia original y los comentarios que la preceden, debido a que se trata un elemento inmediatamente consultable

por un lector/usuario, al momento de emitir su opinión. Otro elemento que debemos considerar más adelante y que está mencionado en la literatura, es la posible alteración del sentido original de la expresión y la importancia del manejo de la negación como en (11).

- (11) no hizo una tormenta en un vaso de agua

En nuestro estudio no registramos ningún caso en donde el uso de la frase idiomática fuera utilizado de una forma tal que su intensidad cambiara una polaridad negativa a una positiva; sin embargo, no podemos hacer una aseveración tan fuerte sin estudios posteriores más profundos. Debido a que las frases en su mayoría fueron negativas, nos encontramos en un punto donde es necesario encontrar marcadores *positivos*.

## 7. Conclusiones y trabajo futuro

Nuestra investigación desarrolló un método basado en el análisis de sentimientos asistido por herramientas computacionales y lo aplicó a un corpus representativo de comentarios de opinión en línea en Costa Rica.

Encontramos evidencia que favorece la hipótesis de que la polaridad presente en las frases idiomáticas se transfiere a los textos de los comentarios donde están presentes, sobre todo si se trata de expresiones idiomáticas negativas. Queda pendiente identificar expresiones idiomáticas positivas, en las que el mismo fenómeno se produzca. Probablemente, esto se

debe a fuerte carga connotativa de las propias unidades fraseológicas.

Tenemos muchas tareas pendientes, aparte de la identificación de expresiones ideológicas positivas, como, por ejemplo, reconocer patrones léxico-sintácticos y nuevos caracteres gráficos (como emoticones) asociados con valores específicos positivos o negativos. Esto es natural, tomando en cuenta la complejidad del tema. Así, además, aún debemos aumentar la cobertura del método y utilizarlo como un elemento para determinar toda la gama de la polaridad de comentarios en línea en español, producidos en perfiles de redes sociales frecuentados por costarricenses. Idealmente, esta labor debe ser efectuada en tiempo real.

## 8. Agradecimientos

Les agradecemos el apoyo al Instituto de Investigaciones Lingüísticas (INIL) y al Centro de Investigaciones en Tecnologías de la Información y la Comunicación (CITIC), ambos de la Universidad de Costa Rica, así como al Ministerio de Ciencia, Tecnología y Comunicaciones (MICITT), quien a través del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT) ha brindado un soporte fundamental para estas investigaciones.

## Notas

1. Profesor de la Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica.
2. Profesor de la Escuela de Filología, Lingüística y Literatura e investigador del Instituto de Investigaciones Lingüísticas.
3. Inscrito con el código 745-B4048 en el Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica.
4. Inscrito con el código 326-B4-149 en el Centro de Investigaciones en Tecnologías de la Información y la Comunicación.
5. Fuente: <https://es-la.facebook.com/ameliarueda/posts/10153261400017923>.

6. URL: <http://www.google.com>
7. Programas de computación diseñados para el procesamiento de lenguaje humano.
8. Es decir, el significado no es composicional.
9. URL: <http://citic.ucr.ac.cr/proyecto/mineria-opinion>
10. La versión en formato de texto del corpus se encuentra disponible en la dirección URL: <http://citic.ucr.ac.cr/proyecto/analissentimiento>
11. URL: <http://www.facebook.com>
12. Fuera de circulación. URL original: <https://facebook.com/aldiacr>

## Referencias

- Arce Sanabria, José Luis. (2012). Medios de comunicación de masas en Costa Rica: Entre la digitalización, la convergencia y el auge de los "new media". *Hacia la Sociedad de la Información y el Conocimiento, capítulo Medios de Comunicación de Masas en Costa Rica*: 283–308. Programa Sociedad de la Información y el Conocimiento, Universidad de Costa Rica.
- Batista, Fernando y Ribeiro, Ricardo. (2013). Sentiment analysis and topic classification based on binary maximum entropy classifiers. *Procesamiento del Lenguaje Natural* 50: 77-84. Recuperado de <http://hdl.handle.net/10045/27867>.
- Brooke, Julian *et al.* (2009). Cross-linguistic sentiment analysis: From English to Spanish. *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*: 50–54. Borovets, Bulgaria. Recuperado de <http://anthology.aclweb.org/R/R09/R09-1.pdf#page=74>.

- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems* XXVIII (2): 15-21.
- Chenlo, José M. y Losada, David E. 2014. An empirical study of sentence features for subjectivity and polarity classification. *Information Sciences* 280: 275-288. Recuperado de <http://dx.doi.org/10.1016/j.ins.2014.05.009>.
- Fernández Anta, Antonio *et al.* 2012. Sentiment analysis and topic detection of Spanish tweets: A comparative study of NLP techniques. *Procesamiento del Lenguaje Natural*. 50: 45-52.
- Hutto, Clayton y Gilbert, Eric. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*.
- Leoni de León, Jorge Antonio. 2008. *Modèle d'analyse lexico-syntaxique des locutions espagnoles*. Tesis de doctorado, Université de Genève, Ginebra, Suiza, 24 de mayo 2008. Recuperado de <http://www.unige.ch/cyberdocuments/theses2008/LeonideLeonJA/meta.html>
- Liu, Bing y Zhang, Lei 2012. A survey of opinion mining and sentiment analysis. C.C. Aggarwal and C.X. Zhai (editores.), *Mining Text Data*, Springer Science+Business Media LLC: 415-463. Recuperado de <http://link.springer.com/book/10.1007/978-1-4614-3223-4>.
- McDonald, Ryan *et al.* 2013. Universal dependency annotation for multilingual parsing. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*: 92-97. Sofía, Bulgaria, 4-9 de agosto 2013, Association for Computational Linguistics. Recuperado de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.386.8061&rep=rep1&type=pdf#page=140>.
- Pang, Bo and Lee, Lillian. 2004, July. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*: 271. Association for Computational Linguistics. Recuperado de <http://dl.acm.org/citation.cfm?doid=1218955.1218990>.

