

UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS
FACULTAD DE MATEMÁTICA, FÍSICA Y COMPUTACIÓN



Trabajo de Diploma para optar por el
Título de Licenciado en Ciencia de la Computación

Detección de ironía en textos cortos enfocada a la minería de opinión

Autor: Anakarla Sotolongo Peña

Tutores: Dra. Leticia Arco García

Lic. Adrián Rodríguez Dosina

2017

Dedicatoria

A mi familia por su apoyo en todo lo que he emprendido en mi vida.

Agradecimientos

En la vida no solo te marcan las metas que has alcanzado sino las personas que te acompañan en este camino. Tampoco se puede ver la universidad como un período aislado, todo sigue un curso y es bueno acordarse del pasado, pues cada momento es consecuencia del otro. Por eso quiero agradecer a todos aquellos que me acompañaron no solo en esta etapa universitaria sino también a lo largo de mi vida ya que nunca había tenido la oportunidad para hacerlo.

Agradezco a mi profesora de gimnasia porque aunque era muy pequeña lo recuerdo como uno de mis mejores momentos. A mis profesores de la escuela de música en especial a Aichel, María Hortensia y Lourdes Caro, quienes no solo me enseñaron a querer la música sino a disfrutar cada momento y a ponerle empeño a las cosas. A los de la primaria y la secundaria, en especial a Tere por su justicia y divertirse junto a nosotros. A los del preuniversitario en especial a Pipa por entendernos a nosotros los jóvenes (una tarea muy difícil!!!), a Luis Orlando quien me enseñó la historia como nadie lo había hecho, a Adolfo por darme esos viajes a Caballete de Casa y al Paso de las Damas (inolvidables), a Rolando por su locura que nos hizo pasar muy buenos momentos, a Chelo y a Clotilde por sus clases magistrales y a Margarita Otero no solo por enseñarnos matemática sino por hacer de sus clases un momento muy divertido. Por último, a todos mis profesores de la universidad, a Morell y Ramiro no solo por sus clases sino por sus charlas muy divertidas y llena de conocimiento.

A mi tutor Adrián por brindarme su apoyo en la tesis.

A Leticia por sus clases, por su dedicación en la tesis y los conocimientos que me transmitió. Además, por la bonita relación que establecimos durante todo este tiempo. Le deseo lo mejor.

A mis amigas de la escuela de música Leyanet, Lisbety y Roxana, y a mis otros compañeros con los que compartí momentos inolvidables.

A mis amigos del IPVC con los que viví momentos maravillosos, y por suerte permanecemos unidos: Lorraine, Mislede, Dayana, Javier (como dice él, mi amigo de 13 años), Lián, Johany, Monet y Alejandro.

A mis amistades de la UCLV por la diversión de estos 5 años: Beatriz, Katy, Loyda, Arletis, Claudia, Nelson, Adrián, Hansel, Angel, Yuniel, Carlos Guillermo, el Guajiro, Manuel, Alfredo, Elvis, Tony, Antony, Jorgito, Carlos Enrique, Lisandra, Yunet, Héctor, Santo.

A todos mis vecinos, mis amigos del barrio y amigos de la familia que se han convertido en parte de mi vida: Tere, Maypú, Mabel y sus niñas, Guillermo, María, Miguelito y Maylín.

A Robertico por su compañía y apoyo durante estos 5 años.

Y por último a mi familia: mi hermano, mi mamá, mi papá, mami, papi, Tania, mami Elida, Ñoño, caqui, tío Hochi, abuelo Israel, mis primos y primas, Mirian, Cerélida y Jenny.

A todos, Gracias!!!

Resumen

Las opiniones textuales imponen grandes retos a las aplicaciones de minería de opinión ya que varios problemas están presentes; entre ellos: la escritura de opiniones de manera irónica o sarcástica. Existen varias propuestas que permiten detectar ironía; sin embargo, la efectividad de la clasificación es baja, sobre todo debido a que aún no son suficientemente expresivos los rasgos identificados para discriminar entre oraciones irónicas y no irónicas. Es por ello que el objetivo general de esta investigación consiste en detectar de manera efectiva la ironía en textos cortos a partir de la identificación de nuevos rasgos lingüísticos y la aplicación de algoritmos de aprendizaje supervisado. Los principales resultados obtenidos son: (1) El estudio y consecuente identificación de las características de los rasgos utilizados en la clasificación de la ironía, así como los principales algoritmos de aprendizaje supervisado utilizados con tales propósitos. (2) La creación de nuevos rasgos que permiten caracterizar los textos irónicos y por tanto discriminar entre clases, principalmente basados en patrones específicos y generales, la ambigüedad, el desbalance contextual, las estadísticas de aparición de las palabras y el contraste en el texto. (3) Un módulo que permite detectar de manera efectiva la ironía a partir de los rasgos creados y seleccionados de la literatura, mediante el uso de varias técnicas de aprendizaje supervisado. La combinación de los rasgos creados, así como los algoritmos de clasificación empleados, permitieron obtener valores de calidad en la detección de la ironía superiores a los reportados previamente en la literatura.

Abstract

Textual opinions impose great challenges on opinion mining applications because several problems are present; such as: the opinion could be writing in an ironic or sarcastic way. There are several proposals to detect irony; however, the irony classification effectiveness is low, mainly because the features identified to discriminate between ironic and non-ironic sentences are not enough expressive. For that reason, this research is focus on the irony detect in short texts by identifying new linguistic features and applying supervised learning algorithms. The main results obtained are: (1) The study and consequent identification of the principal characteristics of features used in irony classification, as well as the main supervised learning algorithms used for such purposes. (2) The creation of new features that allow characterizing ironic texts and therefore discriminate between classes, mainly based on specific and general patterns, ambiguity, contextual imbalance, statistics on word appearances and contrast in the text. (3) A module which allows the effective irony detection from the created features and others selected from the literature, by applying supervised learning techniques. The combination of the created features, as well as the selected classification algorithms, allowed to obtain quality values in the irony detection better than those previously reported in the literature.

Índice

Introducción.....	1
Capítulo 1 Acerca de la detección de ironía en opiniones	6
1.1 Definición de ironía	6
1.2 Análisis de los rasgos frecuentemente utilizados para representar las opiniones irónicas.....	9
1.3 Consideraciones finales del capítulo	43
Capítulo 2 Representación de las opiniones para identificar la ironía	45
2.1 Definición de rasgos que permiten discriminar entre opiniones irónicas y no irónicas	45
2.2 Herramientas y recursos que contribuyen al pre-procesamiento textual y al cálculo de los rasgos diseñados	53
2.3 Módulo para la representación de las opiniones con el objetivo de detectar de manera efectiva la ironía.....	61
2.4 Conclusiones parciales	65
Capítulo 3 Evaluación de la representación propuesta para la detección de ironía.....	68
3.1 Principales propuestas para la detección de la ironía en documentos textuales	68
3.2 Principales medidas para validar la clasificación de las opiniones en irónicas o no	71
3.3 Descripción de las principales colecciones para el estudio de la ironía	74
3.4 Diseño y resultado experimental	75
3.5 Conclusiones parciales.....	78
Conclusiones	81
Recomendaciones	83
Referencias bibliográficas	84
Anexos 92	
Anexo 1. Glosario de términos	92
Anexo 2. Resultados de la clasificación de las opiniones de la colección proveniente de los estados California, Arizona, Texas, Oklahoma y Luisiana.....	93
Anexo 3. Resultados de la clasificación de las opiniones de la colección proveniente de los estados Georgia, Illinois, Minnesota, Nueva Jersey, Nueva York.....	94
Anexo 4. Resultados de la clasificación de opiniones de la colección 1 formada a partir de la usada en (Ling and Klinger, 2016)	95
Anexo 5. Resultados de la clasificación de opiniones de la colección 2 formada a partir de la usada en (Ling and Klinger, 2016)	96
Anexo 6. Resultados de la clasificación de opiniones de la colección 3 formada a partir de la usada en (Ling and Klinger, 2016)	97
Anexo 7. Resultados de la clasificación de opiniones de la colección 4 formada a partir de la usada en (Ling and Klinger, 2016)	98

<i>Anexo 8. Resultados de la clasificación de opiniones de la colección 5 formada a partir de la usada en (Ling and Klinger, 2016)</i>	<i>99</i>
<i>Anexo 9. Resultados de la clasificación de opiniones de la colección 6 formada a partir de la usada en (Ling and Klinger, 2016)</i>	<i>100</i>
<i>Anexo 10. Resultados de la clasificación de opiniones de la colección 7 formada a partir de la usada en (Ling and Klinger, 2016)</i>	<i>101</i>
<i>Anexo 11. Resultados de la clasificación de opiniones de la colección 8 formada a partir de la usada en (Ling and Klinger, 2016)</i>	<i>102</i>
<i>Anexo 12. Resultados de la clasificación de opiniones de la colección 9 formada a partir de la usada en (Ling and Klinger, 2016)</i>	<i>103</i>
<i>Anexo 13. Resultados de la clasificación de opiniones de la colección 10 formada a partir de la usada en (Ling and Klinger, 2016)</i>	<i>104</i>
<i>Anexo 14. Resultados de la clasificación de opiniones de la colección formada por 54 623 irónicos o sarcásticos y 5027 regulares (la colección completa usada en (Ling and Klinger, 2016)).....</i>	<i>105</i>

Introducción

Con el tiempo ha habido un desarrollo de la tecnología y la ciencia, y esto ha contribuido al surgimiento de redes sociales y sitios como Facebook¹, Instagram², Twitter³ y TripAdvisor⁴, entre otros. Estos sitios cada vez toman más auge y por tanto aumenta la cantidad de los usuarios en la red, quienes publican y comparten sus conocimientos, experiencias, emociones y opiniones.

Las personas para tomar una decisión, por ejemplo, en cuanto a qué producto quieren comprar o en cuál hotel van a hospedarse; además de basarse en sus experiencias desean tener en cuenta las opiniones de otros. De manera similar, en las empresas, es importante explorar las opiniones que se emiten acerca de los productos y servicios que ofrecen, para consecuentemente tomar decisiones y mejorar la calidad de ellos. Tanto empresarios, políticos, dirigentes y personas en general, necesitan considerar las opiniones para tomar decisiones, y por ello se dirigen a las redes sociales y sitios con el objetivo de valorar los criterios emitidos. Desafortunadamente, en algunos casos, el volumen de información disponible es tan grande que impide a los usuarios explorar manualmente las opiniones y llegar a buenas conclusiones.

En función de esta necesidad surgió la Minería de Opinión o Análisis de Sentimientos que es el campo de estudio que analiza las opiniones, sentimientos, evaluaciones, apreciaciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, problemas, sucesos, temas, y sus atributos. Existen también muchos nombres y tareas ligeramente diferentes relacionadas con esta área del saber; por ejemplo: extracción de opiniones, minería de sentimiento, análisis de subjetividad, análisis de afecto y análisis de emociones, entre otros. Sin embargo, todos ellos ahora están bajo la cubierta de análisis de sentimiento o minería de opinión. Mientras en la industria se emplea más comúnmente el término análisis de sentimiento, en la enseñanza se emplean frecuentemente ambos, análisis de sentimiento y minería de opinión. Ellos representan básicamente el mismo campo de estudio (Liu, 2012). De ahí

¹ <https://www.facebook.com/>

² <https://www.instagram.com/>

³ <https://twitter.com/>

⁴ <https://www.tripadvisor.es/>

que se han creado disímiles aplicaciones en el campo de la minería de opinión (Fernández et al., 2011; Casado and Cantador, 2013; Amores, Arco and Artiles, 2015; Henriquez, Guzmán and Salcedo, 2016); principalmente para determinar si las opiniones son positivas o negativas (Günther and Furrer, 2013; Hangya, Berend and Farkas, 2013; Levallois, 2013; Mohammad, Kiritchenko and Zhu, 2013; Amores, Arco and Artiles, 2015). Estas aplicaciones procesan opiniones que se emiten textualmente. Trabajar con textos impone grandes retos a cualquier aplicación de minería de textos, ya que son datos no estructurados y por tanto es necesario realizar una etapa de pre-procesamiento y representación textual. Las opiniones textuales imponen aún mayores retos a las aplicaciones de minería de opinión ya que varios problemas están presentes como son: el uso de lenguaje informal, la ironía y el sarcasmo, las abreviaturas, los errores ortográficos y tipográficos, la semántica composicional, el nivel cultural y el conocimiento del lenguaje (Amores, Arco and Barrera, 2016).

La ironía es definida por la Real Academia Española⁵ como “Figura retórica que consiste en dar a entender lo contrario de lo que se dice”. Existen dos tipos de ironía: verbal y situacional (Wilson and Sperber, 1992; Lucariello, 1994). La ironía verbal puede ser vista como una variedad de expresión que hace eco o que referencia la proposición de otro hablante, para poder expresar cierta actitud hacia esta proposición citada; involucra una actitud de desaprobación hacia esta proposición citada, usando desde el ridículo o burla leve, hasta un desprecio total. La ironía situacional es una condición de eventos opuestos respecto a lo que se espera naturalmente, lo que resulta en una contradicción; en general son eventos que se burlan del orden normal de las cosas. A pesar de que no son exactamente el mismo fenómeno, la ironía situacional se relaciona con la ironía verbal en que ambas son una oposición de términos y, además, ambas comparten un factor inesperado o sorpresa.

Los usuarios al emitir sus criterios en la web utilizan mucho la ironía verbal, la cual confunde a las aplicaciones que detectan la polaridad de las opiniones pues conocer su intención de forma automática es muy difícil al no seguir un esquema predeterminado y contener una idea implícita que no está presente en el significado textual de la expresión.

⁵ <http://www.rae.es/>

Los especialistas del laboratorio de Inteligencia Artificial del Centro de Investigaciones Informáticas (CII) de la Universidad Central “Marta Abreu” de Las Villas (UCLV), se dieron a la tarea de elaborar la aplicación PosNeg Opinion para la detección no supervisada de la polaridad de opiniones en Inglés y Español (Amores, 2013, 2016; Amores, Arco and Artiles, 2015; Amores, Arco and Borroto, 2016). Uno de los problemas que presenta esta herramienta en la clasificación de opiniones es que no tienen en cuenta la presencia de ironía. Como es política del CII-UCLV fomentar el desarrollo de investigaciones en el área de la minería de textos y minería de opinión, sobre todo con el objetivo de enriquecer el sistema PosNeg Opinion y su potencial aplicación en la solución de problemas reales, se justifica el planteamiento del **problema de investigación** siguiente:

Existen varias propuestas que permiten detectar ironía; sin embargo, la efectividad de la clasificación es baja, sobre todo debido a que aún no son suficientemente expresivos los rasgos identificados para discriminar entre oraciones irónicas y no irónicas.

El **objetivo general** de la investigación consiste en detectar de manera efectiva la ironía en textos cortos a partir de la identificación de nuevos rasgos lingüísticos y la aplicación de algoritmos de aprendizaje supervisado. Este se desglosa en los siguientes **objetivos específicos**:

1. Identificar, a partir del estudio de las principales propuestas para la clasificación de ironía, aquellos rasgos aplicables de manera efectiva cuando los textos a analizar son opiniones cortas, así como los principales algoritmos de aprendizaje supervisado utilizados con tales propósitos.
2. Crear nuevos rasgos que permitan caracterizar los textos irónicos y por tanto discriminar entre clases.
3. Desarrollar un módulo que permita detectar de manera efectiva la ironía a partir de los rasgos creados y seleccionados de la literatura, mediante el uso de varias técnicas de aprendizaje supervisado.
4. Valorar la repercusión del módulo desarrollado en la detección de ironía en colecciones disponibles en la Web.

Las **preguntas de investigación** planteadas son:

- ¿Cuáles rasgos permiten representar las opiniones de forma tal que los algoritmos de aprendizaje supervisado realicen una efectiva clasificación de la ironía?
- ¿Qué nuevos rasgos permiten caracterizar textos irónicos y por tanto discriminar entre clases?
- ¿Qué rasgos y qué técnicas de aprendizaje supervisado incluir en un módulo que permita la detección efectiva de la ironía?
- ¿Cómo evaluar el impacto del módulo desarrollado en la detección de la ironía en colecciones disponibles en la Web?

Después de haber realizado el marco teórico se formuló la siguiente **hipótesis de investigación** como presunta respuesta a las preguntas de investigación: La aplicación de técnicas de aprendizaje supervisado, a partir de la representación de las opiniones utilizando rasgos que logren distinguir entre textos irónicos y no irónicos, permite una efectiva detección de la ironía.

Para lograr los objetivos trazados y demostrar la hipótesis establecida se acometieron las **tareas de investigación** siguientes:

1. Análisis de los rasgos propuestos dentro del campo de investigación de la detección de la ironía en textos cortos.
2. Creación de nuevos rasgos que permitan discriminar entre las clases: texto irónico y no irónico.
3. Estudio de los principales algoritmos de aprendizaje supervisado utilizados para detectar la presencia de ironía.
4. Determinación de los rasgos que describirán los textos para la detección de la ironía.
5. Identificación y estudio de las principales herramientas para la minería de textos y selección de aquellas que contribuyan al pre-procesamiento textual y al cálculo de los rasgos diseñados.
6. Diseño e implementación de un módulo que permita representar opiniones con el objetivo de determinar si son irónicas o no.
7. Selección de las técnicas de aprendizaje supervisado a utilizar para detectar de manera efectiva la ironía y contribuir así al cálculo de la polaridad de opiniones cortas.

8. Búsqueda de colecciones de opiniones clasificadas en irónicas o no irónicas que permitan la validación externa de la propuesta que se presenta.
9. Estudio y selección de las medidas que permiten evaluar el esquema propuesto.
10. Estudio de la efectividad en la detección de ironía de los principales clasificadores existentes sobre la representación creada.

El **valor teórico** de la investigación radica en la concepción y definición de rasgos que permiten caracterizar las opiniones para la detección de ironía y contribuir así al análisis de sentimiento.

El **valor práctico** se relaciona con que los resultados de esta investigación pueden utilizarse como un módulo de una aplicación más general de minería de opinión, pues la ironía es uno de los problemas que se presenta en este campo, por ejemplo, al determinar la polaridad de opiniones. Los rasgos identificados pueden ser usados, además, para representar las opiniones y realizar otros análisis sobre ellas.

Los resultados que aquí se presentan tienen una aplicación potencial en sistemas de minería de opinión que requieran la detección de la ironía; por ejemplo, para procesar opiniones políticas, evidenciándose su **valor social**, y para procesar opiniones que emiten los turistas sobre nuestras instalaciones hoteleras, evidenciándose su **valor económico**.

La tesis está estructurada en tres capítulos. El Capítulo 1 aborda los principales conceptos asociados a la detección de ironía. Se presenta, además, un análisis de los rasgos frecuentemente utilizados para representar las opiniones irónicas. En el segundo capítulo se proponen nuevos rasgos que permiten discriminar entre opiniones irónicas y no irónicas, se describen las herramientas y recursos que contribuyen al pre-procesamiento textual y al cálculo de los rasgos diseñados y se presenta un módulo para la representación de las opiniones con el objetivo de detectar de manera efectiva la ironía. En el Capítulo 3 se describen las principales colecciones para el estudio de la ironía, los algoritmos supervisados que se utilizan para clasificar las opiniones en irónicas o no, así como las medidas que permiten su validación. Finalmente, se presenta el diseño experimental y se comentan los resultados obtenidos al detectar ironía a partir de opiniones representadas con los rasgos creados como resultado de esta investigación. Este documento culmina con las conclusiones, las recomendaciones, las referencias bibliográficas y los anexos.

Capítulo 1 Acerca de la detección de ironía en opiniones

La ironía es uno de los problemas que se presenta en la detección de la polaridad de una opinión. En este capítulo se mencionarán algunas de las definiciones que se han dado acerca de la ironía y se analizarán los rasgos frecuentemente utilizados para representar opiniones irónicas.

1.1 Definición de ironía

La presencia de ironía en el contenido generado por usuarios en la red representa un problema para las aplicaciones de Análisis de Sentimientos debido a su naturaleza subjetiva. Por ello, es necesario conocer las características de este fenómeno para identificarlo utilizando las técnicas actuales de clasificación de textos.

La Real Academia Española⁶ define la ironía como una figura retórica que consiste en dar a entender lo contrario de lo que se dice. Esto nos podría llevar a pensar que caracterizar este fenómeno es simple; sin embargo, esto no es tan sencillo ya que generalmente no es claro el significado opuesto e incluso a veces no es claro el significado literal del mensaje. A causa de esto, la ironía ha sido definida desde diversos puntos de vista (Gibbs and Colston, 2007).

Clark y un colectivo de autores en (Clark and Gerrig, 1984) analizaron dos teorías que buscan explicar los fundamentos de un mensaje irónico. Una corresponde a la teoría de la mención, en la que se plantea que la ironía puede ser explicada a través de la diferencia que existe en el uso y la mención de cierta expresión. Al ser irónico, un hablante no busca utilizar las palabras de una frase como si fueran propias, si no que busca solo mencionarlas haciendo referencia a otras frases (generalmente dichas por otra persona) que pueden formar parte de la sabiduría popular u opiniones aceptadas, con el fin de hacer notar desprecio hacia estas o de indicar lo opuesto a lo que se dice. A continuación se muestra una adaptación en Español del ejemplo que cita (Clark and Gerrig, 1984):

“Confía en la oficina meteorológica. Mira el maravilloso clima que hay: lluvia, lluvia, lluvia”.

La teoría de la mención asume que el hablante no utiliza la frase “mira el maravilloso clima que hay”, ya que él no busca que quienes le escuchan piensen que el clima es

⁶ <http://www.rae.es/>

maravilloso, sino que solo menciona la frase, aludiendo a lo que podría decir un meteorólogo, con el fin de hacer entender lo contrario.

La segunda teoría corresponde a la teoría de la pretensión en la que se plantea que al expresarse de forma irónica, el hablante pretende ser reconocido o percibido de cierta forma (sin dar a conocer esta pretensión para que no se pierda el efecto en el oyente). Esta teoría incluye algunas características fundamentales de la ironía como:

1. En general, la ironía corresponde a una pretensión positiva sobre algo negativo. O sea, el hablante dice algo con un sentido positivo para criticar algo negativo. Este caso es más frecuente que el contrario, dada la tendencia de las personas a ver el mundo basados en normas de éxito.
2. Suele ser dirigida a alguien, por lo que se dice que tiene víctimas. Estas son, primero, quien el hablante pretende ser (a quien replican o imitan sus palabras y/o actitudes), y segundo, los oyentes que reciben este mensaje en la ignorancia.
3. En su pretensión, el hablante no solo emite un mensaje sino que también le da un tono de voz adecuado, con el fin de exagerar o caricaturizar lo que pretende ser o representar.

Un ejemplo de ironía de pretensión puede ser la frase: “Qué bien me siento aquí!!”, imaginemos que ésta ha sido expresada en un tono irónico y, por tanto, lo que se pretende es que el oyente reciba la impresión de lo mal que se siente el hablante.

La teoría de la pretensión es aplicable a muchas expresiones de ironía en las que también concuerda la teoría de la mención, pero esta relación no se da en la otra dirección, por lo que los autores señalan que la teoría de la pretensión permite interpretaciones mucho más versátiles.

Además, existe un factor psicológico que incide en la expresión de ironía. Los mensajes son irónicos solo para algunos oyentes ya que la percepción de ironía es dependiente de ciertos juicios sutiles en base a cierta información que el hablante y los oyentes/lectores tienen en común, por lo que si a un oyente o lector no se le provee la información adecuada, no podrá emitir estos juicios correctamente. En otras palabras, si el hablante no es capaz de otorgar la información necesaria para que quienes reciben su mensaje puedan descubrir su pretensión, el mensaje no logra ser interpretado como irónico (Hernández and Velásquez, 2015).

La ironía también se ha clasificado en ironía verbal e ironía situacional (Wilson and Sperber, 1992; Lucariello, 1994). La ironía verbal es definida tradicionalmente como un tropo (una expresión en la que se utiliza el sentido figurado o metafórico) en el que la

expresión transmite un significado opuesto al significado literal de sus palabras. Esta definición de ironía no cubre todas las variantes que se pueden presentar en la ironía verbal (Wilson and Sperber, 1992); por ejemplo: el uso de la ironía como una subestimación exagerada, como una citación o como una interjección (las cuales no son una oposición) e incluso como una falsedad o algo contrario que podría cumplir con esta definición tradicional pero dado el contexto, no sería reconocida como una ironía por los oyentes o receptores del mensaje.

Por esto, Wilson y colectivo de autores (Wilson and Sperber, 1992) proponen una extensión de la teoría de la mención diciendo que la ironía verbal puede ser vista como una variedad de expresión que hace eco o que referencia la proposición de otro hablante, para poder expresar cierta actitud hacia esta proposición citada. Con esto, añade que la ironía verbal involucra una actitud de desaprobación hacia esta proposición citada, usando desde el ridículo o burla leve, hasta un desprecio total. Para finalizar, añade que en todos los casos de ironía se hace eco o se referencia otra opinión o frase y que al mismo tiempo el hablante se disocia de ésta (se desliga o la critica), generalmente porque cree que es falsa (y termina expresando lo opuesto) o porque es demasiado absurda. Algunos ejemplos de ironía verbal son los siguientes:

1. Mi autobús retrasado??? NOOOO me estás mintiendo, verdad? Pero si eso JAMÁS sucede!!!
2. Yo si te quiero... pero lejos, muy lejos de aquí.
3. *I HATE to admit it but, I LOVE admitting things.*
4. Antes tú la querías y ella no te quería, pero ahora ella te quiere y tú no la quieres.

El otro tipo de ironía considerado en esta clasificación es la ironía situacional. Lucariello en (Lucariello, 1994) la define como una condición de eventos opuestos respecto a lo que se espera naturalmente, lo que resulta en una contradicción. En general son eventos que se burlan del orden normal de las cosas. A pesar de que no son exactamente el mismo fenómeno, la ironía situacional se relaciona con la ironía verbal en que ambas son una oposición de términos y, además, ambas comparten un factor inesperado o sorpresa. A continuación, se muestran algunos ejemplos de ironía situacional:

1. Una pared en la que está escrito: "Prohibido escribir en la pared".
2. Una casa de arreglos que se jacta de poder arreglarlo todo, en donde no funciona el timbre.

3. En el colegio, lengua es la materia que más le costaba. Esta semana publicó su tercer libro.
4. Una persona que va en auto al gimnasio, para allí hacer bicicleta fija.
5. Una camioneta de la oficina de tránsito, parada sobre un lugar en el que está prohibido detenerse.
6. El barco más inmenso y más seguro del mundo acabó por hundirse.
7. El presidente dijo que no descansaría hasta alcanzar la paz. Desde que asumió, con ese objetivo declaró varias guerras.
8. El departamento de bomberos se ha incendiado.
9. El hijo de un policía está acusado por robar un departamento.
10. Quien era el más callado se convirtió en un reconocido locutor.

Estas clasificaciones y definiciones de tipos de ironía hacen referencia a la ironía de manera general, ya sea expresada por un hablante o expresada en un documento escrito. Es de interés de esta investigación el estudio de las formas que existen para detectar ironía en documentos escritos, lo que impone grandes retos, porque no es posible nutrirse de entonaciones, expresiones faciales y otros elementos que el hablante puede mostrar al expresarse verbalmente.

1.2 Análisis de los rasgos frecuentemente utilizados para representar las opiniones irónicas

La identificación de opiniones irónicas es sumamente importante en la detección de la polaridad de estas, porque pueden cambiar drásticamente el sentido de lo que se quiere expresar. Por eso es necesaria la definición de rasgos que permitan discriminar entre clases irónicas y no irónicas. En algunos estudios realizados muestran que es posible identificar opiniones irónicas en comentarios si los consideramos previamente positivos mediante el uso de patrones lingüísticos relativamente simples (Carvalho *et al.*, 2009). En otros artículos (Filatova, 2012) han demostrado que a veces es necesario analizar las frases juntas pues el contexto es muy importante al detectar ironía. A continuación, se presentan los rasgos utilizados con más frecuencia en los modelos creados hasta el momento para la detección de ironía.

En la ironía es usual encontrar oposición o contradicción en el texto, esto se puede identificar mediante la búsqueda de adverbios escogidos de forma puntual, que pueden dar señales de una negación, junto con sus sinónimos relacionados (también adverbios).

También se incluyen términos compuestos por más de una palabra como “sin embargo”, “por lo tanto” y “más o menos” (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015). No obstante, el hecho de que estas palabras o frases estén presentes en una opinión no necesariamente implica que ésta sea irónica pues es posible encontrarse dichos términos indistintamente en opiniones irónicas y no irónicas. Ejemplo de ello son las opiniones: “La conferencia fue un poco extensa, sin embargo, no fue aburrida” que es una opinión no irónica y “Que gente más educada, sin embargo, ni nos saludaron” que es una opinión irónica.

Otro elemento que se ha considerado en la literatura es la compresión temporal (*Temporal Compression*) que consiste en los cambios abruptos en la narrativa mediante la detección de términos que indiquen oposición en el tiempo (adverbios temporales junto con sus sinónimos). Esta lista incluye términos compuestos por más de una palabra, entre ellos: “en breve”, “desde entonces” y “dentro de poco”. Con este elemento ocurre algo similar a lo comentado anteriormente. El hecho de que estas palabras aparezcan no es una condición necesaria para que esté presente la ironía. Es posible que este razonamiento se base más en un análisis estadístico que lingüístico, quizás motivado por las estadísticas de la ocurrencia de dichos términos en una muestra determinada, lo cual no garantiza que esto sea una regularidad en los textos irónicos. Como se analiza más adelante existen otras formas de expresar la compresión temporal que sí resultan de mayor interés en la presente investigación.

Además, es importante identificar lo inesperado con atributos como el desbalance contextual (*Contextual Imbalance*) y el desbalance temporal (*Temporal Imbalance*). El desbalance contextual es el grado de coherencia contextual que existe, aquí se estudian contradicciones que pudieran ocurrir en la oración midiendo la similitud semántica de las palabras del texto (Reyes, Rosso and Buscaldi, 2012). El desbalance temporal captura la presencia de oposición, pero acá se busca identificar oposición a través del tiempo, específicamente comparando el pasado con el presente. Para esto, se analizan las polaridades de sentimientos en los verbos conjugados en pasado y presente que componen cada texto (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015). Esta forma de analizar el desbalance temporal parece ser una aproximación más completa al problema de la compresión temporal pues no se restringe a la ocurrencia de ciertas palabras, sino que se enfoca más en la forma en que aparecen. En (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015) utilizan el tiempo gramatical para

detectar la compresión temporal, siguiendo de esta forma un enfoque más general y consecuente con las regularidades lingüísticas.

En (Buschmeier, Cimiano and Klinger, 2014) calculan el desbalance (*imbalance*) entre la polaridad global a priori de las palabras en la opinión y la clasificación de estrellas previamente establecida (Davidov, Tsur and Rappoport, 2010). Se asume que el desbalance se mantiene si la clasificación es positiva (es decir, 4 o 5 estrellas), pero la mayoría de las palabras tienen polaridad negativa y, viceversa, si la clasificación es negativa (es decir, 1 o 2 estrellas) pero ocurre que la mayoría de las palabras tienen polaridad positiva. Sin dudas la idea de capturar el desbalance de las polaridades que se propone en (Buschmeier, Cimiano and Klinger, 2014) aporta algunos elementos que pueden considerarse importantes y novedosos. Sin embargo, en este intento de capturar el desbalance no se identifican aquellas oraciones desbalanceadas con igual cantidad de palabras positivas y negativas. Este análisis refleja que la aproximación realizada puede refinarse enfocándose solamente en la dispersión de los valores de polaridad y no tanto en la polaridad global de la opinión y las polaridades individuales de las palabras que la conforman.

En (Reyes and Rosso, 2012) se usan como características los n -gramas de caracteres que corresponden a cadenas de n caracteres extraídos desde un texto, para encontrar términos frecuentes. Son consideradas específicamente cadenas entre tres y cinco caracteres en (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015). En (Ptáček, Habernal and Hong, 2014) se usa como atributo los n -gramas de caracteres desde 3-gramas hasta 6-gramas y se establece que para considerar un n -grama como un atributo es necesario que su frecuencia de aparición en el texto esté entre 5 y 50, de esta forma se reduce la dimensionalidad de la representación. En este artículo también se crean como atributos binarios los unigramas, bigramas y trigramas de palabras, para reducir el espacio de atributos se escogen los n -gramas que aparecen más de tres veces en el texto.

En (Forslid and Wikén, 2015) se le halló la raíz a todas las palabras del texto (*stemming*) y se crearon como atributos la cantidad de unigramas y bigramas de palabras que aparecían más de 10 veces en los datos de entrenamiento. Cada palabra que aparece en el texto analizado también es tomada como otro atributo en algunos artículos, siguiendo el modelo bolsa de palabras (*bag-of-words*) y por tanto solamente

considerando la frecuencia de aparición de los términos en los documentos (Buschmeier, Cimiano and Klinger, 2014). En (Ling and Klinger, 2016), además de considerar todas las palabras (unigramas), tienen en cuenta los 50 000 bigramas más frecuentes como atributos individuales para representar los textos.

En (Reyes and Rosso, 2011) fueron utilizados los n -gramas de palabras desde el orden dos hasta el siete para encontrar un conjunto de palabras recurrentes las cuales podrían expresar ironía. En (Bamman and Smith, 2015) se consideraron los unigramas y bigramas de palabras en minúsculas, resultando “*dare*”, “*shocked*”, “*clearly*”, “*#lol*” y “*gas*” los unigramas más indicativos y “*you*”, “*mean*”, “*how dare*”, “*i’m shocked*”, “*i’m sure*” y “*at all*” los bigramas más representativos. En este artículo también se calcularon unigramas y bigramas sobre los términos en un espacio reducido. Para disminuir la dimensionalidad se mapeó cada palabra en un vocabulario de 1000 grupos no solapados (Owoputi et al., 2013), el cual agrupa las palabras usadas en contextos similares en un mismo grupo. Además, fueron creados atributos booleanos para indicar la dependencia entre dos palabras y sus correspondientes grupos.

En un estudio realizado para la detección de ironía por n -gramas de palabras y de caracteres, considerando secuencias de una a tres palabras en el caso de n -gramas de palabras y de dos a cuatro caracteres en el caso de n -gramas de caracteres, se llegó a la conclusión de que los n -gramas de caracteres son más efectivos para textos cortos que los n -gramas de palabras (Jasso, 2015). Sin embargo, es necesario señalar que la definición de rasgos usando n -gramas, ya sean de palabras o de caracteres, tiene como inconveniente que al encontrar coincidencias en los textos basadas en el contenido que aparece explícitamente no ofrece una solución general al problema, pues no trata de descubrir regularidades o patrones que discriminen entre opiniones irónicas y no irónicas. Los n -gramas como rasgos son más efectivos en áreas como la recuperación de información, la detección de tópicos o la categorización de documentos donde la información explícita juega un papel fundamental.

Otro rasgo mencionado es la captura de *skipgrams* que corresponde a cadenas de las palabras que constituyen un texto, pero además incluyendo las cadenas producidas al omitir ciertas palabras (saltarlas). Se consideran solo bigramas, o sea, cadenas de dos palabras que consideren saltos de hasta tres palabras. Esto incluirá cadenas formadas por saltos de dos palabras, una palabra y ninguna palabra (n -gramas de palabras)

(Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015). En (Ptáček, Habernal and Hong, 2014) también se consideran los skip-bigramas con salto de dos a tres palabras y eliminando los skip-gramas con frecuencia menor o igual que 20.

Polarity Skipgrams es utilizado en los modelos propuestos en (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015). La idea aquí es utilizar el mismo concepto de *skipgram*, pero aplicarlo a una cadena formada en base a las etiquetas de polaridad de sentimientos de cada una de las palabras de un texto. Cada palabra es etiquetada como positiva o negativa, y luego se extraen los *skipgrams* de las cadenas resultantes. Se consideran saltos de hasta tres palabras. A diferencia de los *n*-gramas mencionados anteriormente, los *PolaritySkipgrams* son muy útiles para detectar ironía. Los *PolaritySkipgrams* pueden detectar los cambios bruscos de polaridad, ya que están enfocados en la polaridad de las palabras. Al detectar una palabra con un alto grado de positividad cerca de otra altamente negativa se puede inferir con bastante certeza que está presente el fenómeno de la ironía. Un ejemplo de ello es la opinión: “Qué bien, ahora si estamos perdidos”.

En (Reyes and Rosso, 2011) se diseñó un conjunto de rasgos para obtener secuencias recurrentes de patrones morfosintáctico. Se seleccionaron las *n* mejores secuencias de POS *n*-gramas desde el orden dos hasta el siete. Para la identificación de los POS *n*-gramas se tienen en cuenta las etiquetas POS asociadas a las palabras que lo componen. Para ello se utilizó un algoritmo de reducción (Lü, Zhang and Hu, 2004) que elimina las secuencias redundantes; por ejemplo, si las frases “he is going to look so hot in this shirt” y “he is going to look hot in this shirt” aparecían con una frecuencia similar en el corpus, entonces el algoritmo elimina la última porque es una subcadena de la primera. Luego se etiquetaron los documentos empleando el recurso FreeLing⁷ (Atserias *et al.*, 2006). En (Ptáček, Habernal and Hong, 2014) también fueron utilizados como rasgos aquellos POS *n*-gramas de 3-gramas a 6-gramas que aparecían como mínimo 5 veces.

Otra característica que ha sido utilizada en muchos modelos es el escenario emocional, donde se intenta capturar información emocional de las palabras a través de un diccionario, considerando los rasgos activación (*activation*), imaginación (*imagery*) y agrado (*pleasantness*). Activación es el grado de respuesta (activo o pasivo) que tiene un humano en un estado emocional respecto a la palabra en cuestión, imaginación

⁷ <https://github.com/TALP-UPC/freeling>

refleja que tan fácil o difícil es para un humano realizar una imagen mental frente a una palabra dada, y agrado cuantifica el grado de placer o agrado que representa una palabra para el humano (Reyes, Rosso and Buscaldi, 2012; Reyes, Rosso and Veale, 2013; Baca-Gómez *et al.*, 2014; Espinoza *et al.*, 2014; Hernández and Velásquez, 2015).

En (Reyes and Rosso, 2011) también se define una categoría para representar escenarios cognitivos ideales al expresar ironía. Para representar estos valores se empleó el diccionario de afectos en el lenguaje (*Dictionary of Affect in Language*)⁸ (Whissell, 1989). Este diccionario asigna valores de placer a aproximadamente 9 000 palabras en Inglés. Los valores se obtuvieron por clasificaciones humanas. El rango de los valores va desde el 1 (desagradable) hasta el 3 (placentero). Otra manera, también considerada en (Reyes and Rosso, 2011) para considerar los elementos afectivos en la detección de la ironía fue a partir de la definición de un conjunto de atributos para caracterizar los documentos en términos de palabras que simbolizan contenidos subjetivos como emoticonos, sentimientos, modos, entre otros. El recurso WordNet-Affect⁹ (Strapparava and Valitutti, 2004) fue empleado para obtener los términos afectivos. Este recurso contiene 11 clases para representar la afectividad.

En el sarcasmo se presenta el contraste de sentimiento y para detectarlo en (Rajadesingan, Zafarani and Liu, 2015) se crea un atributo que se basa en los valores de afecto. Para obtener los valores de afecto de las palabras se utiliza un conjunto de datos compilado por (Warriner, Kuperman and Brysbaert, 2013). El conjunto de datos contiene valores de afecto para 13 915 lemas en Inglés los cuales están en una escala de nueve puntos donde uno denota el menor placer. Este atributo se calcula según la expresión (1.1) donde A se define como se presenta en (1.2), t es el texto, w es una palabra en t y $affect(w)$ el afecto de la palabras w .

$$\Delta_{affect} = \max(A) - \min(A) \quad (1.1)$$

$$A = \{affect(w) | w \in t\} \quad (1.2)$$

En (Rajadesingan, Zafarani and Liu, 2015) también se incluye como atributo la posición de la primera palabra cargada de afecto, el cual tendrá valor cero si no está presente. Precisamente determinar el escenario emocional en ocasiones refleja claramente los

⁸ <https://www.god-helmet.com/whissel-dictionary-of-affect/index.htm>

⁹ <https://github.com/larsmans/wordnet-domains-sentiwords/blob/master/wn-domains/wn-affect-1.1>

elementos de humor en los textos, lo cual es muy recurrente en frases irónicas como: “jajaja que susto, pensé que me iban a ascender a jefe” o “que bien, me quedan solo dos días para la prueba”. Formalizar, o al menos establecer una hipótesis, sobre la relación entre las categorías mencionadas anteriormente con el fenómeno de la ironía es un proceso complejo. No obstante, vale la pena buscar expresiones que permitan capturar el escenario emocional, ya que esto pudiera abrir varias posibilidades, y por tanto, ser determinante en la detección de la ironía.

En el trabajo desarrollado en (Reyes and Rosso, 2012) se tuvo como objetivo encontrar secuencias de patrones morfosintácticos, ya que la ironía usa aspectos divertidos para crear su efecto. En (Reyes and Rosso, 2011) fueron utilizados los mejores atributos reportados en la literatura para representar aspectos divertidos: atributos estilísticos, centrados en el ser humano y llaves. Los atributos estilísticos son obtenidos para coleccionar todas las palabras etiquetadas con la etiqueta “sexuality” en WordNet Domains¹⁰ (Bentivogli *et al.*, 2004). El segundo atributo se enfoca en las relaciones sociales. Para extraer estas palabras, fueron recuperados los elementos registrados en el WordNet¹¹ (Miller, 1995), los cuales pertenecen a los *synsets* “*relation*”, “*relationship*” y “*relative*”. El último atributo se representa para obtener el valor clave de las palabras, este valor se calcula comparando las frecuencias de las palabras en los documentos irónicos contra sus frecuencias en un corpus de referencia. Google *n*-grams¹² (Brants and Franz, 2006) fue usado como corpus de referencia. Solo se preservaron las palabras cuyo valor clave era mayor o igual que 100. Este tipo de rasgo tiene la ventaja de considerar una supuesta relación entre el vocabulario utilizado y la presencia de ironía; sin embargo, al enfocarse en un conjunto reducido de etiquetas pudiera aportar falsos negativos en la clasificación.

Otro rasgo es el relacionado con la desviación típica de los valores de polaridad de los términos: la desviación típica total, la máxima, la mínima, la media, y la diferencia entre la máxima y la media. De igual forma se calcula para la ambigüedad (Rella, Saggion and Barbieri, 2015). En (Reyes, Rosso and Buscaldi, 2012) se plantea la ambigüedad relacionada con tres planos: estructural, morfosintáctico y semántico; donde la

¹⁰ <http://wndomains.fbk.eu/download.html>

¹¹ <https://wordnet.princeton.edu/>

¹² <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>

ambigüedad semántica está basada en la distancia de hiperónimos entre *synsets* de WordNet.

Para captar la ambigüedad en (Barbieri and Saggion, 2014; Barbieri, Ronzano and Saggion, 2014) hay tres atributos: *synset mean*, *max synset*, and *synset gap*. El primero es la media del número de *synsets* asociados a cada palabra del tweet. El segundo es el número de *synsets* que tienen una palabra simple, se considera que esta palabra es la que tiene la más alta posibilidad de ser usada irónicamente. El tercero se calcula como la diferencia entre el número de *synsets* de esta palabra (*max synset*) y el número promedio de *synsets* (*synset mean*). En el último artículo referenciado fueron calculados, además, estos atributos relacionados con la ambigüedad de la misma forma, pero solo teniendo en cuenta sustantivos, verbos, adjetivos y adverbios.

Los rasgos relacionados con el cálculo de la desviación típica de los valores de polaridad de los términos y aquellos que detectan la ambigüedad son consecuentes con una de las características principales de la ironía: el contraste. Los primeros analizan el contraste en el sentimiento expresado, mientras los segundos tratan de obtener el contraste manifestado en el vocabulario utilizado.

Elementos tipográficos como signos de puntuación (Reyes, Rosso and Veale, 2013; Barbieri and Saggion, 2014; Espinoza *et al.*, 2014; Ptáček, Habernal and Hong, 2014; Hernández and Velásquez, 2015; Rella, Saggion and Barbieri, 2015), palabras en mayúsculas (Reyes, Rosso and Veale, 2013; Ptáček, Habernal and Hong, 2014; Hernández and Velásquez, 2015), emoticonos positivos y negativos (Reyes and Rosso, 2012; Reyes, Rosso and Veale, 2013; Baca-Gómez *et al.*, 2014; Barbieri and Saggion, 2014; Ptáček, Habernal and Hong, 2014; Hernández and Velásquez, 2015; Rella, Saggion and Barbieri, 2015) y citas (Reyes, Rosso and Veale, 2013; Ptáček, Habernal and Hong, 2014) son utilizados en la ironía para resaltar la idea a expresar. La polaridad de los emoticonos se puede obtener a partir de varias listas¹³. Precisamente la implementación de este tipo de rasgos constituye una tarea indispensable en la presente investigación. Los elementos tipográficos tienen gran utilidad cuando se quiere sintetizar la información a expresar, lo cual es muy común en las opiniones publicadas en las redes sociales y otros sitios de debates en la Web.

¹³ www.netlingo.com, www.noslang.com, es.wikipedia.org/wiki/Anexo:Emoticonos

En (Carvalho *et al.*, 2009; Filatova, 2012) proponen un atributo relacionado con el uso de interjecciones que comúnmente expresan sentimientos positivos como son “bravo”, “muchísimas gracias” y “viva”. Se busca primero la existencia de una interjección positiva, seguida de un bloque formado por cero o muchas frases que presentan primero un pronombre demostrativo y luego un adjetivo positivo; a continuación de este bloque debe aparecer una entidad nombrada¹⁴ en una oración aparentemente positiva y luego símbolos de puntuación como ?, !,

El uso de frases formadas primero por un adjetivo o sustantivo positivo en una ventana de cuatro palabras excluyendo la presencia de cualquier elemento negativo, seguido por más de un signo de exclamación y/o de pregunta pueden denotar la presencia de ironía (Carvalho *et al.*, 2009; Filatova, 2012). En (Buschmeier, Cimiano and Klinger, 2014; Ling and Klinger, 2016), además de lo anterior, en las secuencias de cuatro palabras también se tienen en cuenta aquellas que contienen al menos un elemento negativo pero no palabras positivas. Además en (Buschmeier, Cimiano and Klinger, 2014; Ling and Klinger, 2016), análogamente a los atributos explicados anteriormente, se definen otros para encontrar frases positivas o negativas que termine con “...” y luego múltiples signos de exclamación o de pregunta o combinaciones de ellos. En (Ling and Klinger, 2016) definen además un atributo booleano para detectar si la polaridad cambia en una ventana de cuatro palabras.

En (Vanin *et al.*, 2013) también se buscan frases que contengan signos de puntuación y que tengan las siguientes estructuras: ...<expresión>!, !*, ?*, !*?*, ?* !*. En (Forslid and Wikén, 2015) se crearon atributos correspondientes a la proporción de la cantidad de signos de puntuación diferentes. Entre ellos, la razón de la cantidad de comas, puntos, dos puntos, puntos y comas, elipsis, comillas, guiones, signos de interrogación y de exclamación contra el número total de signos de puntuación que puede contener la opinión. Otro atributo formulado es el que se dirige a capturar exageraciones que se logra por el uso excesivo; por ejemplo, de los signos de exclamación. La cantidad de comas, puntos y comas, dos puntos, signos de exclamación e interrogación son tomadas como atributos en (Hernández *et al.*, 2015). En (Ptáček, Habernal and Hong, 2014) se

¹⁴ Una entidad nombrada (*Named Entity*; NE) se refiere a un valor que contiene tanto un valor como un tipo o categoría asociada al cual pertenece. En otras palabras, las entidades nombradas son pares <tipo, valor> que se extraen de un documento. También se les conoce como elementos o pares atributo-valor. También se define como una palabra o serie de palabras que contienen nombres de personas, organizaciones y lugares.

crean rasgos para la cantidad de palabras, signos de exclamación, de interrogación, comillas y palabras en mayúsculas normalizadas por la división de ellos por el valor máximo observado multiplicado por el valor promedio máximo de los otros grupos de atributos. Además, se utilizan atributos para capturar el número de ocurrencias de signos de puntuación y emoticonos, así como el número de palabras, signos de exclamación, de interrogación, comillas y palabras en mayúsculas normalizadas por el valor máximo observado. En (Rajadesingan, Zafarani and Liu, 2015) se crean siete atributos correspondientes a la distribución normalizada de los signos de puntuación comunes (.,!?"*"). Para comparar el uso actual de signos de puntuación del usuario al uso que él le dio a los signos en el pasado, se calcula el valor de la divergencia entre la distribución de signos de puntuación actual y en el pasado, y se incluye éste como atributo.

En (Ling and Klinger, 2016) se chequea si los tres puntos consecutivos “...” están directamente seguidos de un signo de puntuación. Además, se considera como otro atributo la aparición o no de símbolos de puntuación. Se utilizan representaciones considerando los atributos booleanos (i.e., solo considerando que se cumpla o no la característica definida y no teniendo en cuenta la cantidad de apariciones) y la otra variante es considerando atributos booleanos, definidos para intervalos de valores normalizados posibles del cumplimiento de las características (e.g., la aparición de signos de puntuación está entre 0.2 y 0.4).

Este tipo de rasgos analizado anteriormente también resulta de gran interés ya que, similar a otros como la desviación típica de las polaridades, se centra en la característica de contraste dentro de la ironía. Es novedosa la forma de análisis que aporta, ya que no sólo tiene en cuenta la polaridad de las palabras sino también una cierta estructura del texto, o sea, incluye algunos elementos concernientes al análisis sintáctico.

Algunos autores consideran la aparición de mayúsculas. Por ejemplo, en (Forslid and Wikén, 2015) se añade un contador para el número de palabras con letras mayúsculas. En (Rajadesingan, Zafarani and Liu, 2015; Ling and Klinger, 2016) cuentan el número de palabras totalmente en mayúscula en el texto. De este atributo se tienen cuatro versiones: la cantidad absoluta y la normalizada de palabras en mayúscula, la presencia o no de palabras en mayúscula, y una serie de rasgos booleanos considerando intervalos de valores posibles de la cantidad normalizada de palabras en mayúscula. En (Hernández *et al.*, 2015) se toma como atributo la cantidad de letras mayúsculas que

aparecen en el texto. En (Bamman and Smith, 2015) se tienen atributos para la cantidad de palabras con letra inicial mayúscula, la cantidad de palabras con todas las letras en mayúsculas y la cantidad de etiquetas POS con al menos la letra inicial en mayúscula. En (Rajadesingan, Zafarani and Liu, 2015) se obtuvo la etiqueta POS¹⁵ para cada palabra en mayúscula en el texto con TweetNLP¹⁶ (Owoputi *et al.*, 2013). Este etiquetador genera 25 etiquetas posibles, incluso identifica interjecciones y emoticonos. A partir del resultado del etiquetador, se calcula la distribución de probabilidad de las etiquetas POS para las palabras en mayúscula y se incluyen estos como atributos (en total 25 atributos, uno por cada etiqueta POS). En (Ptáček, Habernal and Hong, 2014) también se implementaron atributos relacionados con la búsqueda de letras en mayúsculas; por ejemplo, calcular el número de palabras en mayúsculas, el número de palabras con la primera letra en mayúscula normalizada por el número de palabras del texto y el número de caracteres en mayúsculas normalizados por el total de palabras del texto. Este rasgo tiene como ventaja que al basarse en un análisis estadístico pudiera descubrir alguna regularidad implícita dentro de la ironía que sea compleja de detectar mediante el método de observación, pero tiene como desventaja el hecho de basarse solamente en la frecuencia y no tener en cuenta otros elementos de análisis sintáctico y semántico.

Otro de los rasgos que utilizan para reconocer la ironía en (Carvalho *et al.*, 2009) es detectar la presencia del pronombre “tú” en lugar de “usted” para dirigirse a una entidad nombrada en una oración aparentemente positiva, pues el tipo de pronombre también puede ser importante para la detección de ironía al denotar el grado de familiaridad que existe entre la persona que opina y la entidad nombrada. El pronombre “tú” es usado en un contexto familiar. En este artículo se realizó el análisis solo para el verbo “ser” (*to be*). El pronombre “tú” puede estar explícito en el texto o puede estar embebido ya que el verbo “ser” se encuentra conjugado en la segunda persona del singular. En la expresión (1.3) se muestra la gramática utilizada por los autores para reconocer este rasgo.

$$\text{NE(tú)* ser}_{2s} \text{ 4-Gram}^+ \quad (1.3)$$

¹⁵ Etiqueta POS (*Part-of-Speech*) se refiere a la parte del discurso asociada a cada término; por ejemplo, sustantivo, adjetivo y verbo, entre otras.

¹⁶ <https://mvnrepository.com/artifact/edu.cmu.cs/ark-tweet-nlp/0.3.2>

Como se observa en la gramática al pronombre “tú” referido a la entidad nombrada (NE) le sigue el verbo “ser” en segunda persona del singular (que puede o no estar presente) y luego un *n*-grama de cuatro palabras no negativas que comience con un adjetivo o sustantivo positivo. Esta idea se presenta como una solución bastante convincente para un subconjunto específico de textos irónicos. Quizás para lograr mayor grado de generalización pudieran utilizarse otros pronombres dentro de esta gramática o también crear una nueva gramática que permita capturar la ironía no solo cuando se exprese en segunda persona.

La presencia de un adjetivo positivo o neutral modificando a un sustantivo a través de la preposición “de”, donde el sustantivo no es más que una entidad nombrada por humanos en una oración aparentemente positiva, es otro de los elementos que autores del artículo (Carvalho *et al.*, 2009) han determinado para identificar la presencia de ironía. Esto debe estar precedido por un artículo o un pronombre demostrativo.

En (Vanin *et al.*, 2013) también se definen atributos relacionados con las partes de la oración, ellos intentan encontrar frases que tengan las estructuras adverbio + adverbio o adjetivo + adjetivo. Además, se buscan frases que contengan tanto partes de la oración como expresiones constantes como el adverbio “tan” + adjetivo o “tan” + adverbio. Otros atributos definidos son los que intentan detectar en el texto estructuras como las siguientes: adjetivo + emoticono(s), pronombre demostrativo + entidades nombradas.

Sobre estos rasgos que se basan en un análisis sintáctico apoyado en ciertas gramáticas es necesario señalar que requieren de mayor generalización. Sin embargo, pueden utilizarse como un conjunto de reglas que contribuyan a un único rasgo que verdaderamente pueda discernir entre opiniones irónicas y no irónicas.

Las comillas son usadas frecuentemente en contenidos irónicos, especialmente si la frase a primera instancia parece de polaridad positiva, un ejemplo es cuando un adjetivo con polaridad positiva modifica a un objeto. En los artículos (Carvalho *et al.*, 2009; Filatova, 2012) se buscan las frases entre comillas compuestas por una o dos palabras donde al menos una de ellas sea un adjetivo o sustantivo con polaridad positiva. De manera similar en (Buschmeier, Cimiano and Klinger, 2014) se indica un atributo que tenga hasta dos adjetivos o sustantivos consecutivos entre comillas que tengan la misma polaridad. En (Ling and Klinger, 2016) también se buscan dos adjetivos, adverbios o sustantivos entre comillas con polaridad positiva o negativa. De esta manera se

aprovechan elementos tanto del análisis sintáctico como del semántico. No obstante, no siempre los textos irónicos contienen comillas por lo que éste no puede ser un rasgo para discriminar sino para tributar a otro más general.

En (Carvalho *et al.*, 2009; Filatova, 2012; Buschmeier, Cimiano and Klinger, 2014) se consideran para detectar la ironía la presencia de acrónimos como “lol” y las expresiones onomatopéyicas como “ah”, “eh” y “hi”, de todos estos se tienen en cuenta sus variaciones en mayúsculas. Además, se buscan emoticonos positivos como :) , ;-) y :P. En (Buschmeier, Cimiano and Klinger, 2014) consideran los emoticonos ya que estos capturan un conjunto de emociones como feliz, risa, guiñar el ojo o hacer señas, sorprendido, insatisfecho, triste, lloroso y burlón. En (Ptáček, Habernal and Hong, 2014; Hernández *et al.*, 2015) crean un atributo relativo a la cantidad de emoticonos positivos y otro a la cantidad de emoticonos negativos usados para expresar algunos tipos de emociones.

Este tipo de análisis es muy provechoso debido al lenguaje informal que es tan frecuente encontrar dentro de las opiniones publicadas en sitios web, por ejemplo, como Twitter donde se limita la cantidad de caracteres a utilizar. Con respecto a este tema los autores se han enfocado mayormente en recolectar listas de abreviaturas, emoticonos y de otras palabras que no se encuentran en los diccionarios existentes y que tienen un uso establecido dentro de la comunidad de usuarios de internet.

En (Barbieri and Saggion, 2014) siguen una idea similar, específicamente crearon un rasgo llamado risas para considerar la cantidad de expresiones como “hahah”, “lol”, “rofl” y “lmao”. Similarmente, en (Vanin *et al.*, 2013) fue usada una lista de expresiones de risas y de emoticonos.

En (Ling and Klinger, 2016) también se chequea si aparecen acrónimos como “lol”, “lawl”, “luls”, “rofl”, “roflmao”, “lmao” y “lmfao”; muecas (*grin*) como “*grin*”, “*gg*” y “*g*” y onomatopeyas como “[mu-|ba-]haha”, “hihi” y “hehe”. Tanto para los acrónimos, las muecas como para las onomatopeyas consideran dos tipos de rasgos: tipo de rasgo booleano donde solo se tiene en cuenta la presencia o no de estos símbolos y una serie de rasgos booleanos considerando intervalos de valores posibles de la cantidad normalizada de apariciones de estos símbolos.

En (Ling and Klinger, 2016) se presentan cuatro tipos rasgos asociados a los emoticonos: un rasgo para cuantificar la cantidad de emoticonos, un rasgo booleano para indicar la presencia o no de emoticonos en la opinión, otro para la cantidad normalizada que se calcula dividiendo el número de emoticonos presentes en el texto entre el total de *tokens* y una serie de rasgos booleanos considerando intervalos de valores posibles de la cantidad normalizada de apariciones de los emoticonos. Estos rasgos son calculados para cada uno de los tipos de emoticonos considerados: positivos, negativos, símbolos faciales y no faciales. Como se puede observar, estos autores consideran la ocurrencia de los emoticonos en las opiniones como señal de ironía. Sin embargo, esto no basta para poder discriminar entre opiniones irónicas y no irónicas, pues de esta manera se hace énfasis en el cálculo de la frecuencia de aparición de los emoticonos y no en la polaridad de los mismos, la cual analizada en conjunto a la polaridad general de la opinión pudiera aportar información más relevante para la clasificación.

Otra idea interesante se presenta en (Buschmeier, Cimiano and Klinger, 2014; Ling and Klinger, 2016) donde el atributo hipérbola (*Hyperbole*) es utilizado para la detección de ironía e indica la ocurrencia de una secuencia de tres palabras positivas o negativas en una fila. Quizás esta hipótesis no deba ser implementada como un rasgo independiente por su nivel de especificidad, pero sí se pudiera tener en cuenta como una regla gramatical que contribuya a algún rasgo existente relacionado con el análisis sintáctico.

En los artículos (Barbieri and Saggion, 2014; Barbieri, Ronzano and Saggion, 2014) se trabajó con las frecuencias de las palabras en el texto. Se calcularon tres atributos relacionados con ellas para todas las palabras de la oración: frecuencia de la palabra más rara que está en el texto, frecuencia media que es el promedio de todas las frecuencias de las palabras de la oración y la diferencia entre los dos atributos anteriormente mencionados. Estos atributos también fueron calculados pero considerando solo sustantivos, verbos, adjetivos y adverbios. Aunque los métodos propuestos por dichos autores obtienen resultados aceptables, este rasgo en particular no parece una buena opción, ya que no está respaldado con una hipótesis clara en cuanto al patrón lingüístico que se desea reflejar. Al respecto, una alternativa más lógica se propone en (Barbieri and Saggion, 2014) donde se buscan los sinónimos de cada palabra del texto a analizar, utilizando WordNet, y se calculan las frecuencias de aparición de los sinónimos en un

corpus de referencia, específicamente utilizan el corpus CoLFIS¹⁷. Las palabras son ordenadas decrecientemente según su frecuencia. Este ordenamiento es utilizado para definir cuatro rasgos, con el objetivo de buscar el uso de sinónimos comunes y raros.

El primer rasgo, llamado *syno_lower* es el número de sinónimos de la palabra w_i con frecuencia menor que la frecuencia de w_i . *Syno_lower* se calcula según la expresión (1.4).

$$sl_{w_i} = |\text{syn}_{i,k} : f(\text{syn}_{i,k}) < f(w_i)| \quad (1.4)$$

donde $\text{syn}_{i,k}$ es el sinónimo de w_i con ranking k , y $f(x)$ es la frecuencia de x en el corpus de referencia. También definen *syno_lower_mean* como la media de *syno_lower* para todas las palabras del tweet. Los otros dos rasgos propuestos por (Barbieri and Saggion, 2014) son *syno_lower_gap* y *syno_greater_gap*, y para su definición propusieron calcular previamente los parámetros *word_lowest_syno* y *word_greatest_syno*. *Word_lowest_syno* es el máximo *syno_lower* en un tweet, formalmente definido en la expresión (1.5); mientras que *word_greatest_syno* se calcula según la expresión (1.6).

$$wls_t = \max_{w_i} \{|\text{syn}_{i,k} : f(\text{syn}_{i,k}) < f(w_i)|\} \quad (1.5)$$

$$wgs_t = \max_{w_i} \{|\text{syn}_{i,k} : f(\text{syn}_{i,k}) > f(w_i)|\} \quad (1.6)$$

Posteriormente se calcula *syno_lower_gap* como la diferencia entre *word_lowest_syno* y *syno_lower_mean*, el cual detecta el desbalance que crea sinónimos comunes en un contexto de sinónimos raros. Análogamente es calculado *syno_greater_gap*, para detectar la presencia de varios sinónimos raros en un contexto de sinónimos comunes, como la diferencia entre *word_greatest_syno* y *syno_greater_mean*, donde este último se calcula según la expresión (1.7).

$$sgm_t = \frac{|\text{syn}_{i,k} : f(\text{syn}_{i,k}) > f(w_i)|}{\text{número de palabras del tweet}} \quad (1.7)$$

En (Barbieri, Ronzano and Saggion, 2014), teniendo en cuenta las expresiones calculadas en (Barbieri and Saggion, 2014), se calcula el mayor y el menor número de sinónimos con frecuencia mayor que uno presente en el texto a analizar, el número promedio de sinónimos con frecuencia mayor/menor que la frecuencia de la palabra relacionada presente en el texto. También se determina el mayor/menor número de sinónimos y el número medio de sinónimos de la palabra con frecuencia mayor/menor

¹⁷ http://linguistica.sns.it/CoLFIS/Home_eng.htm

que la que aparece en el texto. Además, se calculan estos atributos por separado para cada una de las categorías: sustantivo, verbo, adjetivo y adverbio. De esta manera, se intenta capturar la intención irónica del hablante a partir de la forma en que selecciona su vocabulario. Este enfoque puede considerarse bastante acertado, pues la única información con que se cuenta en este problema es el texto en sí, a diferencia de la detección de la ironía en las conversaciones presenciales donde se observan además otros elementos que pueden aportar mucho como los gestos y la entonación.

Otra forma de analizar el vocabulario, aunque quizás menos atractiva, se encuentra en (Ptáček, Habernal and Hong, 2014) donde se definen atributos que capturan patrones compuestos por palabras con alta frecuencia¹⁸ (*High Frequency Words*; HFW) y palabras de contenido¹⁹ (*Content Words*; CW). Los patrones deben contener al menos una palabra de alta frecuencia. Los patrones generalmente contienen de dos a seis HFW y de una a seis CW. En este artículo se fija en cinco el mínimo de ocurrencia de un patrón. Además, se crea un nuevo rasgo llamado *word-shape pattern* que se encarga de clasificar las palabras de contenido en una de las 24 clases consideradas en el análisis²⁰. Considerando las etiquetas POS así como las clases de palabras, se creó el tipo de atributo POS *word-shape* para contar la cantidad de unigramas de palabras que tienen asignada determinada etiqueta POS y una forma de palabra (*word-shape*) dada. Para incluir un atributo del tipo POS *word-shape* es necesario tener como mínimo cinco apariciones de unigramas que satisfagan esa clasificación.

En algunos artículos cuantifican la cantidad de hashtags positivos y negativos para detectar ironía en los tweets (Rella, Saggion and Barbieri, 2015). En este sentido es de vital importancia detectar algunos hashtags específicos como *#irony* o *#sarcasm*, sin embargo no todos los hashtags que expresen polaridad son indicio de ironía. Otros trabajos tienen en cuenta la longitud de la oración (Barbieri and Saggion, 2014; Espinoza *et al.*, 2014; Rella, Saggion and Barbieri, 2015; Ling and Klinger, 2016) y el conteo de las características gramaticales verbos, adverbios, sustantivos y adjetivos (Barbieri and Saggion, 2014; Espinoza *et al.*, 2014; Hernández *et al.*, 2015; Rella, Saggion and Barbieri, 2015) lo cual carece un tanto de lógica, pues no se refleja

¹⁸ Palabras cuya frecuencia en el corpus es mayor que 1000 palabras por millón más todos los caracteres de puntuación

¹⁹ Palabras cuya frecuencia en el corpus es menor que 1000 palabras por millón

²⁰ Esta clasificación se realiza a partir del resultado que ofrece la configuración WORDSHAPECHRIS1 en `edu.stanford.nlp.process.WordShapeClassifier`

claramente una hipótesis de cómo pudieran influir dichas cantidades en la presencia de ironía. No obstante, este tipo de rasgo basado en un análisis estadístico de las formas gramaticales que aparecen en la opinión, ha sido implementado de varias maneras en la literatura. A continuación se mencionan algunas variantes, aunque la mayoría de ellas presentan la misma deficiencia comentada anteriormente.

En (Barbieri, Ronzano and Saggion, 2014) son determinados ocho atributos para capturar el estilo del texto. Se cuenta el número de ocurrencias de las palabras que pertenecen a las siguientes categorías: verbo, sustantivo, adjetivo, adverbio, interjección, artículos, pronombre y adposición. En (Barbieri and Saggion, 2014; Rajadesingan, Zafarani and Liu, 2015) se calcula la porción de sustantivos, adjetivos, verbos y adverbios con respecto al número de palabras del texto. En (Forslid and Wikén, 2015), además de calcular la proporción de las cuatro categorías mencionadas anteriormente, se analizan las categorías artículos, preposiciones, interjecciones y verbos modales. En (Ptáček, Habernal and Hong, 2014) también fueron creados atributos para contabilizar el número de sustantivos, verbos y adjetivos; la cantidad de sustantivos respecto a los adjetivos y de los verbos respecto a los adverbios, así como la cantidad de verbos negativos. En (Bamman and Smith, 2015) se tiene un atributo booleano para saber si el texto contiene una palabra en una lista de 50 intensificadores²¹, por ejemplo, “*so*”, “*too*”, “*very*” y “*really*”. En (Rajadesingan, Zafarani and Liu, 2015) se crea un atributo para la cantidad de intensificadores usados. Además, se tienen atributos para la cantidad de pronombres de la primera persona del singular, primera persona del plural, segunda persona y tercera persona presentes en el tweet.

En (Rajadesingan, Zafarani and Liu, 2015) se generan las etiquetas POS para las palabras en el tweet usando el etiquetador TweetNLP. Se obtienen las etiquetas POS para cada palabra en el tweet y se construye una correspondencia con la distribución de probabilidad POS e incluyen estos como atributos (en total 25 atributos, uno por cada etiqueta POS). Una idea similar se sigue en (Bamman and Smith, 2015), donde a partir de las etiquetas generadas por TweetNLP, se diseñaron atributos basados en la cantidad absoluta y la porción de cada una de las 25 etiquetas, junto con la “densidad léxica” del texto, la cual modela la razón de sustantivos, adverbios, verbos y adjetivos respecto al total de palabras.

²¹ <http://en.wikipedia.org/wiki/Intensifier>

Además, en (Rajadesingan, Zafarani and Liu, 2015) se construye una distribución de probabilidad sobre las etiquetas POS presentes en el tweet actual y otra para las etiquetas POS presentes en los tweets anteriores y se incluye como atributo el valor de la divergencia entre las dos distribuciones.

Para capturar la posible idiosincrasia sintáctica presente en un tweet, en (Rajadesingan, Zafarani and Liu, 2015) se usan como atributos las etiquetas POS de las primeras tres palabras y las últimas tres palabras en el tweet (en total seis atributos).

En (Rajadesingan, Zafarani and Liu, 2015) se propone una idea muy interesante pues dichos autores suponen que una persona para escribir con sarcasmo tiene que dominar el lenguaje; por tanto, se proponen atributos para identificar las competencias de la persona al escribir un texto. Conocer la gramática de un usuario se hace difícil, pero se puede chequear el uso correcto de la gramática para palabras comúnmente usadas. En (Rajadesingan, Zafarani and Liu, 2015) se chequea el correcto uso gramatical para “your” y “its”, ambas usadas frecuentemente. En este estudio observaron que a veces los usuarios erróneamente usan palabras como “your” en lugar de “you’re” e “its” en lugar de “it’s”. Usando todos los tweets anteriores de un usuario, se obtiene el POS de la palabra usada inmediatamente después de “your” e “its”. Si la palabra ha sido usada gramaticalmente de forma correcta, el POS de la palabra que le sucede no debería ser un verbo, ni un adverbio, ni tampoco un artículo como “a” o “the”. Por ejemplo, “your doing great!” es incorrecto y “its freakin’ amazing” también es incorrecto. Por tanto, incluyen un atributo para contabilizar la cantidad de veces que se utiliza incorrectamente la palabra “your” y otro para la palabra “its”. Hay otros POS que pueden dar el uso incorrecto de “your” o “its”; sin embargo, por exactitud adoptaron una estrategia conservadora para chequear solo para verbos, adverbios y artículos. Este enfoque resulta provechoso específicamente cuando se quiere descartar la presencia de ironía en un texto con un mal uso de dichos elementos del lenguaje; sin embargo, el hecho de que se utilicen correctamente no asegura que esté presente la ironía.

En (Ling and Klinger, 2016) también se mide la cantidad de palabras gramaticales (*stopwords*), sustantivos, verbos, adverbios, adjetivos y pronombres (e.g., “I/we”, “my/our”, “me/us”, “mine/ours”). De todos estos atributos se tienen cuatro versiones: la cantidad absoluta y la normalizada de palabras que pertenecen a estas categorías, la presencia o no de este tipo de palabras, y una serie de rasgos booleanos considerando

intervalos de valores posibles de la cantidad normalizada de palabras que son de estas categorías.

En (Ling and Klinger, 2016) también crean un conjunto de atributos relacionados con la interjección donde tienen en cuenta la cantidad de ocurrencias de frases como “aha”, “brrr” y “oops”. De esto se tienen cuatro versiones: la cantidad absoluta y la normalizada, presencia o no de esta característica, y una serie de rasgos booleanos considerando intervalos de valores posibles de la cantidad normalizada.

La polaridad en lo concerniente a las palabras que denotan orientación semántica, ya sea positiva o negativa, es otro factor a tener en cuenta (Reyes, Rosso and Buscaldi, 2012). En (Rella, Saggion and Barbieri, 2015) son calculadas la polaridad total, la suma de polaridades positivas, la suma de polaridades negativas, la máxima, la mínima, la media, y la diferencia entre la máxima y la media, como rasgos a considerar en la clasificación de opiniones en irónicas o no. En (Buschmeier, Cimiano and Klinger, 2014) se determina la polaridad de las palabras con la ayuda de un diccionario el cual tiene 6 800 palabras con sus polaridades (Hu and Liu, 2004). En (Ling and Klinger, 2016) siguen una idea similar a la presentada en (Rella, Saggion and Barbieri, 2015), con la diferencia que construyen un diccionario basado en valores de sentimiento para clasificar un texto en positivo o negativo. Se calculan los rasgos: suma de todos los valores de polaridad de las palabras positivas, suma de los valores de polaridad de las palabras negativas y la diferencia absoluta entre los valores de polaridad de la palabra más positiva y la más negativa. En (Rajadesingan, Zafarani and Liu, 2015) se incluye un atributo para guardar la posición de la primera palabra cargada de sentimiento, el cual tendrá valor cero si no está presente.

Algunos recursos léxicos para buscar la polaridad de los términos son SentiWordNet²² (Baccianella, Esuli and Sebastiani, 2010) y PosNeg_Lexicon (Amores, 2016; Amores, Arco and Borroto, 2016). Existen recursos que devuelven la polaridad de toda la oración; por ejemplo, en (Forslid and Wikén, 2015) utilizaron la biblioteca TextBlob²³ para calcular la polaridad, ésta devuelve valores entre -1 y 1 para cada oración, donde -1 significa que es una oración muy negativa, 0 que es neutra y 1 que es muy positiva.

²² <http://sentiwordnet.isti.cnr.it/>

²³ <https://pypi.python.org/pypi/textblob>

En (Baca-Gómez *et al.*, 2014; Rella, Saggion and Barbieri, 2015) consideran como rasgos la presencia de palabras muy positivas, positivas, negativas y muy negativas. También consideran los modificadores, es decir, aquellas palabras que aumentan o disminuyen los valores de polaridad de otras.

Algunos autores han definido rasgos que consideran la estadística asociada a elementos léxicos, sintácticos y semánticos de los textos. En (Espinoza *et al.*, 2014) realizan el conteo de las características léxicas, tienen en cuenta la polaridad de los términos de la oración, el número de *slangs*, el número de contracciones, el número de prefijos, el conteo de las 100 palabras más utilizadas, la cantidad de palabras mal escritas, la cantidad de número, la longitud de la palabra más larga y la cantidad de palabras de longitud 1, 2, 10 y 15.

En (Carvalho *et al.*, 2009) proponen otro rasgo novedoso para la detección de ironía, en este caso relacionado con la presencia de diminutivos en las entidades nombradas, para ello, utilizan una ventana de cuatro palabras para recorrer el texto con el objetivo de identificar cuatro palabras, donde ninguna tenga polaridad negativa, que comiencen con un adjetivo o un sustantivo con polaridad positiva, seguida de una entidad nombrada por humanos en forma diminutiva en una oración aparentemente positiva, o en orden inverso. Este rasgo debe capturar con un buen grado de certeza el sentimiento de desprecio hacia personas o instituciones (generalmente representadas con entidades nombradas) que es común observar en una gran cantidad de frases irónicas.

Otro de los elementos que tienen en cuenta en (Carvalho *et al.*, 2009) para detectar la ironía es el uso de pronombres demostrativos como: este (*this*), ese, aquel (*that*), entre otros. Buscan que aparezca primero un pronombre demostrativo seguido de una entidad nombrada por humanos en una oración aparentemente positiva y a continuación una ventana de cuatro palabras, donde ninguna tenga polaridad negativa, que comience con un adjetivo o un sustantivo positivo. Aunque este rasgo también analiza el uso de entidades nombradas, no parece tan prometedor como el anterior, pues dicha estructura pudiera presentarse tanto en opiniones irónicas como no irónicas, indistintamente. Ejemplo de ello son las opiniones: “This World Baseball Classic was amazing because generally the games were closed” que no es irónica y “This World Baseball Classic ended hard-fought, 8-0 to USA” que es irónica y cumple con la misma estructura.

En (Barbieri and Saggion, 2014) se define un grupo de atributos relacionados con el sentimiento, para ello se asigna a cada *synset* del WordNet valores de sentimiento de positividad y negatividad. Se calcula la suma de todos los valores positivos/negativos en el texto, el promedio de los valores positivos/negativos en el texto, la diferencia entre la suma de los valores positivos en el texto y la suma de los negativos para detectar el desbalance positivo-negativo que hay en el texto, la diferencia entre la palabra más positiva/negativa y la media de todos los valores de sentimiento de todas las palabras del texto. Este tipo de rasgo es altamente discriminante ya que captura el contraste en el sentimiento que se expresa. Con el mismo objetivo de detectar el contraste se han propuesto otras variantes, algunas de ellas se abordan a continuación.

Un grupo de rasgos son creados en (Barbieri, Ronzano and Saggion, 2014) para detectar el sentimiento en un texto y lo inesperado, que se evidencia al aparecer una palabra negativa en un contexto positivo y viceversa. Para ello, se calcula la cantidad de palabras positivas y negativas, la suma de las intensidades de las palabras positivas y negativas, la media de las intensidades de las palabras positivas y negativas, el mayor valor positivo y negativo, la diferencia entre el mayor valor positivo/negativo y la media positiva/negativa. También estos rasgos son calculados por cada una de las categorías: sustantivo, verbo, adjetivo y adverbio.

En (Bamman and Smith, 2015) se crean varios atributos para detectar el sentimiento a nivel del texto. El primero es el valor numérico que representa el sentimiento de todo el texto según lo determinado por Stanford Sentiment Analyzer (Socher *et al.*, 2013). Como este analizador basado en frases también determina un valor que representa el sentimiento de cada nodo no terminal en su árbol de análisis sintáctico, también se define como otro rasgo la razón de no terminales con cada valor de sentimiento respecto al total, el cual permite capturar las diferencias en los sentimientos a través del árbol.

En (Bamman and Smith, 2015) también se incluyen atributos para detectar el sentimiento a nivel de palabra, donde se auxiliaron del diccionario de (Warriner, Kuperman and Brysbaert, 2013) y de los valores de emociones de (Thelwall *et al.*, 2010). Se crearon atributos para tener en cuenta el valor de la palabra con el mayor sentimiento, el valor de la palabra con el menor sentimiento y la diferencia entre los valores máximos y mínimos anteriores.

En (Rajadesingan, Zafarani and Liu, 2015) se crea un atributo basado en valores de sentimientos para detectar el contraste de sentimiento que es usual en el sarcasmo. El valor de sentimiento de una palabra es calculado con el uso de SentiStrength²⁴ que es una herramienta basada en el léxico optimizada para la detección de sentimiento de un texto basándose en el sentimiento que expresan las palabras individuales en el texto y devuelve una terna de sentimiento {positivo, negativo, neutral} correspondiente a todo el texto. Un valor de sentimiento negativo está entre -1 y -5 (desde no negativo a lo extremadamente negativo) y un valor de sentimiento positivo está entre 1 y 5 (desde no positivo a extremadamente positivo). Este atributo es calculado por la expresión (1.8) donde S se define como (1.9), t es el texto, w una palabra en el texto y $\text{sentiment}(w)$ el valor de sentimiento de la palabra w .

$$\Delta_{\text{sentiment}} = \max(S) - \min(S) \quad (1.8)$$

$$S = \{\text{sentiment}(w) | w \in t\} \quad (1.9)$$

En (Rajadesingan, Zafarani and Liu, 2015) crearon atributos para detectar el contraste entre tweets anteriores y el tweet actual. Para ello, se obtuvo el sentimiento expresado por el usuario (i.e., positivo, negativo, neutral) en los textos anteriores y en el actual utilizando SentiStrength. Luego, definen un rasgo que indica el tipo de transición de sentimiento que tiene lugar desde el tweet precedente al tweet actual (e.g., de positivo a negativo, de negativo a positivo); donde existen en total nueve transiciones diferentes al combinar los sentimientos positivo, negativo y neutral. Para tener una perspectiva histórica de la probabilidad del usuario en las transiciones de sentimiento se calcula la probabilidad de las nueve transiciones de los tweets anteriores del usuario. Las probabilidades de transición junto con la probabilidad de la transición actual son incluidas como atributos.

Tanto SentiStrength como el conjunto de datos proporcionado por (Warriner, Kuperman and Brysbaert, 2013) solo permiten obtener los valores de sentimiento y de afecto asociados a unigramas. Es por ello que en (Rajadesingan, Zafarani and Liu, 2015) decidieron construir un léxico de sentimientos positivos y negativos para bigramas y trigramas usado en Twitter de forma similar a (Kouloumpis, Wilson and Moore, 2011). Para ello siguieron los siguientes pasos:

²⁴ <http://sentistrength.wlv.ac.uk/>

1. Coleccionaron alrededor de 400 000 tweets con hashtags de sentimiento positivo como “*#love*”, “*#happy*”, “*#amazing*” y 400 000 tweets con hashtags de sentimientos negativos como “*#sad*”, “*#depressed*”, “*#hate*”.
2. En esos tweets se extrajeron los bigramas y trigramas junto a sus frecuencias y se seleccionaron aquellos con frecuencia menor que 10.
3. Para cada bigrama o trigrma b se halla el valor de sentimiento mediante la expresión (1.10) donde $POS(b)$ es el número de ocurrencias de b en el conjunto de tweets positivos y $NEG(b)$ es el número de ocurrencias de b en el conjunto de tweets negativos. Se escogieron aquellos bigramas y trigramas con valores de sentimiento entre -0.1 y 0.1.

$$\frac{POS(b) - NEG(b)}{POS(b) + NEG(b)} \quad (1.10)$$

Finalmente, usando esta herramienta generada se construyeron los atributos: cantidad de n-gramas con valores de sentimiento positivo, cantidad de n-gramas con valores de sentimiento negativo, la suma de los valores de los n-gramas positivos y la suma de los valores de los n-gramas negativos.

En (Rajadesingan, Zafarani and Liu, 2015), además, formularon varios atributos para representar el estado de emoción del usuario. Intuitivamente, el estado del usuario puede indicar cuán propenso está él a usar sarcasmo, si el usuario está mal humorado (estado negativo), entonces tal vez sea propenso a expresarse en la forma de un tweet sarcástico. Por tanto, para medir el estado del usuario se usa el sentimiento expresado en sus tweets anteriores; sin embargo, no se puede asumir que el estado del usuario está encapsulado en sus últimos n tweets. Por tanto, para capturar el estado usado en sus tweets anteriores se siguen los siguientes pasos:

1. Para cada tweet anterior t se calcula su valor de sentimiento positivo $pos(t)$ y el valor absoluto de su sentimiento negativo $neg(t)$ usando SentiStrength.
2. Se dividen los tweets anteriores del usuario en bloques solapados basándose en el número de tweets emitidos antes del tweet actual.
3. Cada bloque b_n consiste de los n tweets previamente emitidos por el usuario, donde n es seleccionado del conjunto $\{1,2,5,10,20,40,80\}$.
4. En cada b_n se captura el estado percibido del usuario usando dos tuplas. La primera tupla, como se presenta en la expresión (1.11), consiste de cuatro

atributos, donde Σ^+ y Σ^- son el total de sentimientos positivo y negativo en b_n : $\Sigma^+ = \sum_{t \in b_n} \text{pos}(t)$ y $\Sigma^- = \sum_{t \in b_n} \text{neg}(t)$. P es + o -, $P = +$ cuando $\Sigma^+ > \Sigma^-$ y $P = -$ en otro caso. La segunda tupla, representada en la expresión (1.12), consiste de seis atributos donde n_+ es el número de tweets positivos, n_- es el número de tweets negativos y n_0 es el número de tweets neutrales presentes en b_n , según valores de polaridad obtenidos de SentiStrength, n es el total de tweets presentes en b_n y Q indica la polaridad de la mayoría de los tweets donde $Q \in \{+, -, 0\}$. Por ejemplo, $Q = +$ cuando $n_+ = \max(n_+, n_-, n_0)$.

$$\langle \Sigma^+, \Sigma^-, P, \max(\Sigma^+, \Sigma^-) \rangle \quad (1.11)$$

$$\langle n_+, n_-, n_0, n, Q, \max(n_+, n_-, n_0) \rangle \quad (1.12)$$

Por cada b_n se incluyeron ambas tuplas como atributos; por tanto, considerando siete bloques solapados y cuatro atributos de una tupla y seis de la otra, resultaron 70 atributos.

Como el estado emocional de las personas permanece constante por una cantidad limitada de tiempo, esto sugiere medir el estado del usuario dentro una ventana de tiempo. Sin embargo, no se asume que el estado del usuario está encapsulado dentro de t minutos. Por tanto, se dividen los tweets anteriores del usuario en bloques b_t , los cuales consisten de todos los tweets emitidos por el usuario dentro de un rango de tiempo de t minutos en el tweet actual. Aquí $t \in \{1, 2, 5, 10, 20, 60, 720, 1440\}$ minutos, donde 1440 minutos es un día. Para cada bloque b_t se incluyen las tuplas descrita en las expresiones (1.11) y (1.12) como atributos, respectivamente.

Como el sarcasmo es una combinación de afecto y sentimiento, en (Rajadesingan, Zafarani and Liu, 2015) se examina cómo el afecto y el sentimiento son expresados en tweets sarcásticos para crear atributos que lo identifiquen. Con este fin, se construye una distribución de valores de sentimiento SD que consiste de 11 valores, cada valor es el número de palabras en el tweet con valor de sentimiento i , donde $i \in [-5, 5]$. También se construye una distribución de valores de afecto AD que contiene nueve valores, cada valor es el número de palabras en el tweet con valor de afecto j , donde $j \in [1, 9]$. Se normalizan los valores de SD y AD y se incluyen ambas distribuciones como atributos (en total 20 atributos). Por cada distribución se generan seis atributos correspondientes a la media, mediana, moda, desviación estándar, mínimo y máximo. Además, se incluye el número de palabras de afecto, el número de palabras de sentimiento y el sentimiento

en general del tweet que puede ser positivo, negativo o neutral como atributos. Para capturar diferencias en la expresión de sentimiento entre tweets sarcásticos y no sarcásticos, se compara la distribución de valores de sentimiento de tweets anteriores del usuario con la de su tweet actual. Se calcula el valor de la divergencia entre las distribuciones de valores de sentimiento anteriores y la actual e incluyen éste como un atributo más. Finalmente, para ganar una idea de cómo un usuario emplea Twitter para expresar emociones, se determina el rango de sentimientos expresados por el usuario en el pasado. Para realizar esto, para todos los valores de sentimiento $i \in [-5,5]$, se calcula el número de tweets con valor de sentimiento i en todos los tweets anteriores del usuario. Para normalizar estas cantidades se obtiene una distribución de probabilidad sobre los valores de sentimiento en los tweets anteriores, la cual se incluye como atributo (en total 11 atributos).

Cuando los individuos experimentan situaciones injustas, a veces se dirigen a los medios sociales para expresar sus quejas y frustraciones. Estas frustraciones son expresadas a veces en forma de sarcasmo. Es por eso que en (Rajadesingan, Zafarani and Liu, 2015) plantean una hipótesis que expresa que cuando usuarios se encuentran en situaciones desagradables en la vida real, ellos reaccionan espontáneamente mediante tweets para dar salida a sus frustraciones. Por tanto, ellos divergen de sus patrones de tweets regulares. Para capturar este patrón de conducta, se construye un supuesto que el usuario pudiera emitir siguiendo una distribución de probabilidad en el tiempo considerando los tweets anteriores del usuario. Para cada uno de los tweets anteriores del usuario, se extrae el tiempo de creación del tweet, y se construye una distribución normalizada considerando intervalos de una hora. Esta distribución aproxima la probabilidad de emisión de tweets del usuario cada hora. Teniendo en cuenta esta distribución se diseña un rasgo que consiste en calcular la probabilidad de que la emisión de un tweet en una hora se corresponda o no con la distribución aproximada de emisiones de tweets del usuario. Además, en este estudio observaron que los usuarios tienden a escribir tweets sucesivos en un corto período de tiempo para expresar sus frustraciones, por ello, incluyen la diferencia de tiempo entre el tweet que se está analizando y los tweets anteriores a él anunciados por el usuario como un atributo. Una forma de expresar frustración comúnmente es con el uso de palabras obscenas, por lo que mediante la lista de las palabras obscenas más comunes suministrada por (Wang *et*

al., 2014), se chequea la presencia de tales palabras en el tweet y se incluye esto como un atributo booleano.

Un conjunto de rasgos relacionados también con la intensidad, esta vez solo para adverbios y adjetivos, fueron diseñados en (Barbieri and Saggion, 2014), para ello se utiliza la escala de valores de intensidad de (Potts, 2011) para adverbios y adjetivos. Se calcula la suma de los valores en la escala de adjetivo/adverbio para todos los adjetivos/adverbios del texto, la media de los valores en la escala de adjetivo/adverbio para todos los adjetivos/adverbios del texto, el máximo valor de todos los valores en la escala de adjetivo/adverbio para todos los adjetivos/adverbios del texto, y la diferencia entre el valor máximo y la media, tanto para los adjetivos como para los adverbios, para conocer en cuanto el adjetivo/adverbio más intenso está fuera de contexto.

En (Reyes and Rosso, 2011) un grupo de rasgos son creados para indicar la correlación entre los elementos positivos y negativos en los datos. Para etiquetar los datos se utiliza *Macquarie Semantic Orientation Lexicon*²⁵ (MSOL) (Saif, Dorr and Cody, 2009). En (Reyes and Rosso, 2012) también se calcula la relación entre elementos positivos y negativos en el texto y algunas otras emociones.

En (Ling and Klinger, 2016) se define un atributo booleano para detectar la ocurrencia de las negaciones “n’t” y “not”. Además, se crea un rasgo para medir la diferencia entre la sentencia más corta y la más larga y otro para conocer si las oraciones del texto tienen diferentes longitudes. También se tiene otro atributo para saber si una palabra aparece más de una vez en una ventana de cuatro palabras. En el caso específico de la detección de la ironía el tratamiento de las negaciones juega un papel fundamental. Sin embargo, no basta solo con identificarlas sino que se debe analizar cómo éstas modifican la polaridad del texto.

Otra forma de conocer si un texto puede ser sarcástico es si este contradice a hechos universales asumiendo que la persona está intentando negar el hecho. Para ello la oración es analizada y los tokens son escaneados por el analizador, el cual encuentra las palabras claves del tema y busca en todas las oraciones de la base de conocimiento las que contengan estas palabras. Si el texto contradice alguna de estas oraciones es considerada como sarcasmo. Por ejemplo, al analizar “Mixture of Green and Orange gave Blue” se obtienen las palabras claves <Mixture;Green;Orange;Blue> y al buscarlas

²⁵ <http://saifmohammad.com/WebPages/lexicons.html>

en la base de conocimiento la salida es “Mixture of Green and Orange gives Brown”, por lo que el texto es clasificado como sarcasmo. Un inconveniente es que el usuario puede desconocer el tema y se equivoca, en este caso el analizador lo tomará como un sarcasmo pues no conoce de esta situación (Ingle *et al.*, 2014). Otro inconveniente es la complejidad temporal que requiere la implementación de este tipo de rasgos.

Además, como los usuarios experimentados de Twitter a veces usan palabras cortas por el límite de 140 caracteres que establecen los mensajes de Twitter mediante la eliminación de vocales, el uso de números, entre otros; en (Rajadesingan, Zafarani and Liu, 2015) se crean atributos booleanos para detectar la presencia de palabras alfanuméricas y de palabras sin vocales, y otro para el porcentaje de palabras del diccionario presentes en el tweet. De esta manera se interpreta en cierto grado el lenguaje informal que es característico de los tweets; sin embargo, el hecho de contabilizar el grado de informalidad no implica la existencia de ironía, pues un tweet irónico puede tener el mismo grado de informalidad que uno no irónico.

Una idea similar se maneja en (Bamman and Smith, 2015) donde se modela el uso de un estilo de escritura específico para Twitter, para ello se incluyen como atributos la cantidad de palabras con solo caracteres alfabéticos pero no vocales (e.g., “btw”) y la cantidad de palabras con más de tres sílabas. El número de sílabas del texto, número de sílabas por palabra en el tweet derivado de *Flesch-Kincaid Grade Level Formula* (Flesch, 1948), el número de palabras con tres o más sílabas (polisílabas) y el número de polisílabas por palabra en el tweet derivado de SMOG (McLaughlin, 1969) son atributos que se tienen en cuenta en (Rajadesingan, Zafarani and Liu, 2015).

Otro de los atributos utilizados para la detección de ironía es el número de palabras (Barbieri and Saggion, 2014; Hernández *et al.*, 2015; Rajadesingan, Zafarani and Liu, 2015), la media de la longitud de las palabras del texto (Barbieri and Saggion, 2014) y la cantidad de caracteres que forman el texto (Hernández *et al.*, 2015; Rajadesingan, Zafarani and Liu, 2015), enfoque que posee algunas de las deficiencias comentadas anteriormente relacionadas con el análisis estadístico independiente de la semántica.

Los usuarios a veces repiten letras en las palabras para indicar un significado opuesto en lo que escriben. En (Rajadesingan, Zafarani and Liu, 2015) crean atributos booleanos para conocer la presencia de caracteres repetidos tres o más veces y la presencia de caracteres repetidos tres o más veces en palabras cargadas de sentimiento (e.g.,

loveeeeeee). Además, se incluye la proporción de la cantidad de caracteres distintos al total de caracteres del texto. La idea propuesta por estos autores aporta como elemento interesante la necesidad de identificar el énfasis como señal de ironía y sin dudas la variante implementada para lograr dicho objetivo parece un buen punto de partida.

En (Rajadesingan, Zafarani and Liu, 2015) se determina la amplitud del vocabulario del usuario en los tweets anteriores de él y se incluyen como atributos la cantidad total de palabras, la cantidad total de palabras distintas usadas y la proporción de palabras distintas al total de palabras usadas para medir la redundancia del usuario en el uso de la palabra.

Inspirado por el atributo media de la longitud de las palabras del texto mencionado anteriormente y usado en *Automated Readability Index* (Kincaid *et al.*, 1975), en (Rajadesingan, Zafarani and Liu, 2015) se formula un conjunto de atributos usando la distribución de la longitud de las palabras $L = \{l_i\}_{i=1}^{19}$ construida del texto t a través de los siguientes pasos:

1. Para cada palabra w en t , se calcula la cantidad de caracteres que la componen $|w|$. Se asume que serán ignoradas las palabras con longitud 20 o más. Luego, se construye la distribución de la longitud de las palabras $L = \{l_i\}_{i=1}^{19}$ para t , donde l_i denota el número de palabras en el texto con longitud de carácter i .
2. L puede ser representado usando la representación 6-tupla según la expresión (1.13), donde E es la media, med la mediana, $mode$ la moda y σ la desviación estándar.

$$\langle E[l_w], med[l_w], mode[l_w], \sigma[l_w], \min_{w \in t} l_w, \max_{w \in t} l_w \rangle \quad (1.13)$$

Cada componente de la 6-tupla es tomado como un atributo.

Además, en (Rajadesingan, Zafarani and Liu, 2015), dada la disponibilidad de los tweets anteriores del usuario, se examina si hay una diferencia notable en la distribución de la longitud de las palabras entre el tweet actual del usuario y sus tweets anteriores. En (Rajadesingan, Zafarani and Liu, 2015) hacen notar que mientras que los tweets sarcásticos también pueden estar presentes en los últimos tweets del usuario, debido a su relativa rareza, los tweets pasados cuando se toman en su totalidad, darían de salida una media posiblemente introducida por unos pocos tweets sarcásticos pasados. Por tanto, cualquier diferencia en la norma de la distribución de la longitud de las palabras

del tweet actual puede ser capturada. Para capturar las diferencias en la distribución de longitud de palabras, se realizan los siguientes pasos:

1. En el tweet actual del usuario, se construye una distribución de probabilidad D_1 basada en la longitud de las palabras del tweet.
2. En los tweets anteriores, se construye una distribución de probabilidad D_2 basada en la longitud de las palabras en todos los tweets anteriores.
3. Para calcular la diferencia entre la distribución de la longitud de la palabra del tweet actual y los tweets pasados, se calcula la divergencia Jensen-Shannon (JS) entre D_1 y D_2 por la expresión (1.14) donde $M = \frac{D_1 + D_2}{2}$ y KL es la divergencia KL expresada por (1.15)

$$JS(D_1 || D_2) = \frac{1}{2} KL(D_1 || M) + \frac{1}{2} KL(D_2 || M) \quad (1.14)$$

$$KL(T_1 || T_2) = \sum_i \ln \left(\frac{T_1(i)}{T_2(i)} \right) T_1(i) \quad (1.15)$$

Finalmente, el valor de la divergencia JS se toma como un atributo.

Este enfoque, a pesar de basarse en un análisis estadístico muy riguroso centrado en la distribución de las longitudes de las palabras en el historial de mensajes de los usuarios, no resulta del todo convincente en cuanto a su efectividad. Además, posee una alta complejidad computacional. Sin embargo, es necesario destacar que la idea de tener en cuenta el historial de los usuarios puede considerarse un aporte.

Otro rasgo ha sido definido para conocer si el texto analizado contiene o no un enlace a Internet (Barbieri and Saggion, 2014). En (Hernández *et al.*, 2015) se crea un rasgo para conocer la frecuencia o aparición de URL y en (Ling and Klinger, 2016) un rasgo es creado para saber si existe una URL. En (Ling and Klinger, 2016), además, se crean atributos para conocer si el texto contiene una cadena que comience con @ para referirse a un destinatario. No obstante, el análisis de las URL no debe limitarse solamente a su detección sino también a la exploración del contenido de las mismas pues existen algunos tweets que son irónicos por tener una referencia a una imagen irónica o a otro texto irónico localizado en dichas URL.

En (Barbieri and Saggion, 2014) se crean tres atributos para capturar el estilo escrito-hablado (*written-spoken*) utilizado: el primero y el segundo son la media de los valores

de frecuencia de todas las palabras del texto en la colección escrita y hablada, respectivamente, del corpus nacional americano (*American National Corpus*²⁶; ANC); el tercero, es el valor absoluto de la diferencia entre el primero y el segundo, el cual dice si el escrito irónico usa ambos estilos (en este caso el valor es pequeño), creando el desbalance, o solo uno de ellos. Sobre este particular, es pertinente señalar que un determinado estilo de redacción no implica que exista o no ironía. No es correcto afirmar, por ejemplo, que cada vez que se escribe un tweet utilizando el denominado estilo hablado, implica que es irónico. El uso del estilo hablado puede tener múltiples causas, no siempre para expresar una intención irónica, entre ellas: estilo inherente a las redes sociales, estilo y nivel cultural del usuario y características del tópico que se aborde. Por tanto, es importante concluir que el estilo se debe a un conjunto de factores entre los cuales se encuentran los mencionados anteriormente y donde la presencia de la ironía no juega un papel preponderante aunque pueda coocurrir en ciertas muestras de tweets con un determinado estilo.

Otros autores también han definido atributos para encontrar las palabras que pueden ser utilizadas para expresar ironía, por ejemplo “yeah right” (Filatova, 2012; Forslid and Wikén, 2015). Este tipo de rasgos que chequea si aparece una determinada frase considerada como irónica para clasificar la opinión. Lista de frases de referencia debe conformarse cuidadosamente.

En (Vanin *et al.*, 2013) definen un grupo de cadenas constantes a buscar, entre ellas se incluyeron varios hashtags, en este caso para el Portugués, como son: “só que”, “sim”, “seria”, “na boa”, “medo”, “Medo!”, “#medo”, “#ironia” o “#sarcasmo”, “#joking” o “#kidding”. En (Filatova, 2012) también se buscan algunos hashtags como “#sarcasm” y “#sarcastic”.

Algunos artículos plantean que el 90% de los tweets poseen hashtags, lo cual ayuda a detectar mensajes similares que pertenecen a un mismo grupo (Ingle *et al.*, 2014). En este estudio consideran la presencia de la cadena “#sarcasm” como un elemento que ayuda a clasificar un tweet en sarcástico. Pero este puede ser usado como parte de la oración y no como hashtag, en este caso la oración sería clasificada como no sarcástica.

²⁶ El corpus nacional americano *American National Corpus* (<http://www.anc.org/>) es una colección masiva de palabras en Inglés americano (15 millones)

En el artículo (Maynard and Greenwood, 2014) se hizo un estudio de los hashtags para identificar oraciones sarcásticas. En esta investigación coleccionaron manualmente una lista de hashtags sarcásticos en un corpus formado por tweets aleatorios, y luego lo extendieron automáticamente con el *plugin* GazetteerListCollector GATE²⁷ para coleccionar pares de hashtags donde uno de ellos es un hashtag sarcástico como “#sarcasm”; por ejemplo: “I love living with a 7 year old #NotReally #sarcasm”. El otro hashtag se adiciona a la lista de indicadores de sarcasmo; según la oración de ejemplo anterior, se adiciona a la lista el hashtag #NotReally. Estos indicadores se usan para significar cuándo el sarcasmo está presente en una oración. Si uno o más de esos indicadores de sarcasmo son encontrados en la misma oración de una opinión, o dentro del mismo tweet, la polaridad existente se invierte.

En (Maynard and Greenwood, 2014) presentan algunas reglas para obtener la polaridad de los tweets teniendo en cuenta los hashtags indicativos de sarcasmo, algunas de ellas se presentan a continuación. Si aparece un hashtag simple indicativo de sarcasmo, y el sentimiento original es positivo o neutral, se invierte la polaridad a negativa. Si se utiliza más de un hashtag, consideran el sentimiento contenido en todos los hashtags. Si dos hashtags son indicativos de sarcasmo, se tratan como un único hashtag simple e indicativo de sarcasmo. Si un hashtag positivo es seguido de un indicador de sarcasmo, y la polaridad del tweet es positiva o neutral, se cambia la polaridad del sentimiento del hashtag positivo a negativo, y se aplica ese sentimiento al texto. Para trabajar con aquellos hashtags que están formados por más de una palabra, propusieron un algoritmo para extraer los tokens individuales que lo formaban. Una vez que los hashtags están divididos en tokens se pueden usar para la detección del sentimiento.

En (Rajadesingan, Zafarani and Liu, 2015) para medir la familiaridad del usuario con el sarcasmo se tiene un atributo con la cantidad de veces que se usan los hashtags “#not” o “#sarcasm” en tweets pasados. Para capturar la familiaridad del usuario con el lenguaje de Twitter, se crean atributos para rescatar la cantidad de retweets, menciones y hashtags usados en los tweets anteriores. Además, el tweet se divide en tres partes basado en la cantidad de palabras presentes y se incluyen tres atributos para la cantidad de hashtags presentes en cada una de las tres partes del tweet.

²⁷ <http://gate.ac.uk/userguide/sec:gazetteers:listscollector>

En (Ling and Klinger, 2016) se mide la ocurrencia de hashtags. Se tienen cuatro versiones de este atributo: la cantidad de hashtags, la presencia o no de estos, la cantidad normalizada y una serie de rasgos booleanos considerando intervalos de valores posibles de la cantidad normalizada de hashtags.

En (Hernández *et al.*, 2015) se crean atributos booleanos para reconocer la presencia o ausencia de hashtags y menciones, y otros para la cantidad de estos en el texto. Además, se crea un rasgo para saber si el tweet es un retweet, y otro para revertir la polaridad de positiva a negativa, o viceversa, si el texto incluye los hashtags “#sarcasm” o “#not”.

Algunos estudios han demostrado que las personas en diferentes regiones perciben y usan el sarcasmo de forma diferente. En (Rajadesingan, Zafarani and Liu, 2015) trataron de inferir la localización del usuario; sin embargo, en Twitter ésta es suministrada por el usuario en un campo de texto libre y puede presentar equivocaciones, por lo que decidieron aproximar la localización del usuario con su zona horaria y lo incluyeron como un atributo. Este rasgo constituye un aporte para la detección de la ironía al contextualizarla como un fenómeno ligado a la cultura de cada región o país.

Otro aporte interesante se encuentra en (Rajadesingan, Zafarani and Liu, 2015) donde se plantea que un usuario expresa mejor el sarcasmo cuando está bien familiarizado con el ambiente, es por eso que decidieron medir la familiaridad del usuario en Twitter basándose en la familiaridad con el uso y la actividad social. La familiaridad con el uso fue capturada usando la cantidad de tweets registrados, el número de días en Twitter (i.e., la edad en Twitter) y el número promedio de tweets diarios. La frecuencia con que un usuario usa Twitter se obtuvo de los tweets anteriores del usuario, calculando la diferencia de tiempo entre todos los pares de tweets sucesivos. Luego se representa la distribución de esas diferencias de tiempo como una séxtupla similar a la expresión (1.13) y se incluyen sus componentes como atributos. Para medir la actividad social de un usuario, se incluyen como atributos la cantidad de amigos, la cantidad de seguidores, la división del número de amigos por los años del usuario en Twitter y la división del número de seguidores por la edad del usuario en Twitter. Este rasgo tiene como único inconveniente su alta complejidad computacional.

Algunos artículos no solo analizan los textos escritos por los usuarios, sino que además utilizan la información del autor del tweet, la audiencia y el contexto del tweet para la detección de la ironía (Bamman and Smith, 2015).

Los atributos que recogen la información del autor están relacionados con su historial de términos, temas, sentimientos y la información del perfil. Para cada autor se identifican los 100 términos en su historial de tweets que tengan el mayor valor de TF-IDF y se crean atributos binarios para cada uno de esos 100 términos. Además, se crean atributos basados en tópicos más amplios para inferir una proporción de temas del usuario, identificando con LDA (Blei, Ng and Jordan, 2003) los 100 temas principales que abordan los tweets, donde cada documento contiene hasta 1000 palabras de tweets del usuario. Los atributos creados con la información del perfil del usuario incluyen el género²⁸, el número de amigos, seguidores y estado, su duración en Twitter, el número promedio de mensajes por día, su zona horaria y si son o no verificados por Twitter (que designa un tipo de estado de celebridad). Para diseñar los atributos relacionados con el historial de sentimiento del autor se modela la distribución sobre los sentimientos en el historial de tweets del usuario usando los mismos diccionarios a nivel de palabra aplicado al sentimiento a nivel de tweet descrito anteriormente. Usuarios con un historial de sentimientos negativos tienen mayor probabilidad de sarcasmo. Además, en (Bamman and Smith, 2015) se crearon atributos binarios para todos los unigramas en el perfil del autor, donde los términos más representativos incluían “*sarcasm*”, “*chemistry*”, “*#atheist*” y “*humor*”.

Para tener en cuenta la participación de la audiencia se formularon los atributos relativos al historial de los tópicos que aborda el autor, los términos que emite, los unigramas en el perfil del autor y la información del perfil, para aquellos usuarios que emiten respuestas a un determinado autor.

Para capturar la similitud de los intereses entre el autor y el destinatario se crearon atributos que tienen en cuenta la distribución del historial de tópicos del destinatario. Estos rasgos obtienen altos valores si autores y destinatarios emiten tweets sobre los mismos tópicos.

De acuerdo a algunos autores es más probable que el sarcasmo ocurra entre dos personas que sean más familiares entre ellos (Bamman and Smith, 2015), por eso se diseñan atributos que modelan el grado de interacción entre dos usuarios incluyendo, por ejemplo, el número de mensajes previos enviados desde el autor al destinatario.

²⁸ El género puede inferirse a partir de las tendencias de los registros ciudadanos, por ejemplo, los registros de la Seguridad Social.

Como inconveniente cabe destacar que dicha afirmación se reduce solo a las conversaciones entre dos usuarios (mensajes directos), pues en ocasiones se brinda una opinión irónica sobre una noticia o evento y en tales casos no es ventajoso el análisis mencionado anteriormente.

Para capturar el contexto en el que se escribió el tweet, en (Bamman and Smith, 2015) se crearon atributos que recogen el contexto lingüístico original de éste y la relación que existe entre el mensaje original y la respuesta. Para modelar la relación entre el tweet original y la respuesta se incluyen indicadores binarios asociados la aparición de los términos en los dos tweets (original y respuesta). Además, fueron creados atributos binarios asociados a todos los unigramas del tweet original para conocer el contexto lingüístico original de éste. Los términos más representativos en el mensaje original incluyen marcadores claros que definen un ambiente sarcástico, por ejemplo, “*#sarcasm*”, “*sarcastic*” y “*sarcasm*” así como “*worry*”, “*defense*”, “*advice*”, “*vote*” y “*kidding*”.

Algunos autores, para clasificar una oración en sarcástica, realizan análisis de sentimiento y analizan si en la oración se contradicen o no los gustos y disgustos de la persona que emite el criterio. Para la colección de los gustos y disgustos se utiliza el aprendizaje supervisado basándose en el historial de tweets del autor. Cada tweet previo de una persona es analizado para extraer los tokens secuencialmente, concluir el tema del tweet y aplicando herramientas de análisis de sentimiento al tweet conocer si tiene polaridad positiva o negativa (Ingle *et al.*, 2014).

Otro de los elementos que se tienen en cuenta para la detección del sarcasmo son los temas, noticias, sucesos que ocurren en un período de tiempo por lo que se extrae conocimiento temporal. Para ello, el analizador obtiene el momento en el tiempo en que fue escrito el texto y las palabras claves del tema, los cuales son utilizados para verificar si es correcto lo que se menciona. Esto último se logra haciendo *crawling* para alimentarse de las noticias disponibles en Internet auxiliándose del tiempo y las palabras claves. Si los resultados revelan una negación de lo que se planteaba en el texto, éste es clasificado como sarcástico. Por ejemplo, en la oración “*India had an amazing win*” se habla de algún día en particular, por lo que es necesario buscar en las noticias de ese período para conocer si es verdadero el texto, en el caso de que no, será considerado

como sarcástico (Ingle *et al.*, 2014). Este enfoque parece bastante prometedor; sin embargo, conlleva un elevadísimo costo computacional.

1.3 Consideraciones finales del capítulo

Existe un conjunto de características que permiten identificar textos irónicos. Para ello, en la literatura se han definido varios rasgos que intentan expresar tales características. Se han propuesto varios tipos de rasgos, tales como aquellos que abordan el contraste en el sentimiento expresado, la estructura sintáctica de las opiniones, la selección del vocabulario utilizado, los estilos de redacción, la compresión temporal, el uso de elementos especiales, por ejemplo, emoticones, abreviaturas, signos de puntuación, onomatopeyas, acrónimos e interjecciones, el análisis del historial de los usuarios, la procedencia de las opiniones, el escenario emocional, el análisis estadístico de *n*-gramas de caracteres y palabras y el énfasis expresado en la opinión.

Aunque el análisis estadístico pudiera descubrir alguna regularidad implícita en la ironía que sea compleja de detectar mediante el método de observación, una deficiencia que presentan algunos de los rasgos analizados es que se apoyan en el análisis estadístico sin considerar elementos sintácticos y semánticos de la opinión.

En la definición de algunos rasgos solo se realiza un análisis parcial de la regularidad o patrón que se quiere representar. Por ejemplo, no basta con identificar las negaciones, modificadores o emoticones sin analizar cómo éstos modifican la polaridad del texto.

Algunos rasgos son muy específicos, ya que se basan en un análisis sintáctico apoyado en ciertas gramáticas y por tanto es necesario señalar que requieren de mayor generalización. Estos rasgos pudieran ser el punto de partida para definir un conjunto de reglas que los integren en un único rasgo y de esta forma captar elementos que logren discriminar entre oraciones irónicas y no irónicas.

Existen autores que consideran el estilo como una fuente directa para identificar la ironía. Sin embargo, es importante señalar que el estilo se debe a un conjunto de factores y donde la presencia de la ironía no juega un papel preponderante, aunque pueda coocurrir en ciertas opiniones con un determinado estilo.

Los rasgos relacionados con el cálculo de la desviación típica de los valores de polaridad de los términos y aquellos que detectan la ambigüedad son consecuentes con una de las características principales de la ironía: el contraste. Los primeros analizan el

contraste en el sentimiento expresado, mientras los segundos tratan de obtener el contraste manifestado en el vocabulario utilizado.

Si bien existen varios rasgos que contribuyen a identificar opiniones irónicas, aún no son suficientes para lograr verdaderamente discriminar entre opiniones irónicas y no irónicas. Por tanto, nuevos rasgos para caracterizar estas opiniones pudieran ser definidos de manera tal que combinen el análisis estadístico y los elementos sintácticos y semánticos de la opinión, partan de un análisis global de la opinión y descubran patrones generales.

Capítulo 2 Representación de las opiniones para identificar la ironía

La ironía es un fenómeno difícil de tratar, por ello numerosas investigaciones tributan a resolver problemas relacionados con este campo. En este capítulo se propone un conjunto de rasgos que permiten discriminar entre opiniones irónicas y no irónicas, algunos tomados de la literatura y otros definidos a partir del estudio de las características que identifican las opiniones irónicas. Además, se describen las herramientas y recursos que fueron necesarios para la implementación de este módulo. Finalmente, se presenta la estructura del módulo que permite pre-procesar las opiniones y obtener su representación estructurada a partir de los rasgos definidos con el objetivo de detectar de manera efectiva la ironía.

2.1 Definición de rasgos que permiten discriminar entre opiniones irónicas y no irónicas

La formulación de rasgos que distingan correctamente a las clases irónicas y no irónicas es imprescindible para detectar la ironía. En esta investigación se caracterizan los textos irónicos mediante rasgos. Se eligieron algunos rasgos definidos previamente en otras investigaciones y se crearon nuevos basándose en ideas planteadas por otros autores o mediante el estudio de ejemplos de ironía que permitió identificar las características distintivas de oraciones irónicas.

En varios artículos se crearon atributos para conocer el contraste entre los elementos positivos y negativos en un texto (Barbieri and Saggion, 2014; Barbieri, Ronzano and Saggion, 2014; Bamman and Smith, 2015; Rajadesingan, Zafarani and Liu, 2015; Rella, Saggion and Barbieri, 2015; Ling and Klinger, 2016). Aunque no se refleja explícitamente por dichos autores, es posible asumir que las opiniones irónicas deben tener un alto grado de contraste mientras que las no irónicas deben tener un valor casi nulo en este rasgo. Es necesario señalar que las variantes propuestas en las investigaciones mencionadas anteriormente tienen como desventaja que no consideran la cercanía de las palabras en el texto, o sea, se limita a un análisis semántico de cada palabra por separado sin tener en cuenta el contexto de las mismas. De esta manera se considera que existe contraste en las opiniones donde existan palabras con polaridades muy diferentes independientemente de donde se encuentren. Los ejemplos siguientes

permiten ilustrar esta afirmación: “Estamos muy mal y el profesor nos ha dicho que solo se aprueba con una buena preparación para el examen” y “Ayer nos fue bien, tanto que perdimos”. Se puede observar que en el primer caso existe contraste en cuanto a las polaridades; sin embargo, la opinión es no irónica pues las palabras con polaridades muy diferentes están relativamente lejanas entre ellas. Por otra parte, en el segundo ejemplo, las palabras con polaridades contrastadas están muy cercanas, lo que indica que la cercanía debe ser un elemento a considerar en este tipo de rasgo.

Por tanto, teniendo en cuenta estos atributos precedentes y analizando, además, ejemplos de oraciones irónicas, se propone un nuevo atributo para medir el contraste donde no solo se tiene en cuenta la polaridad, sino también la distancia entre los elementos de mayor contraste. Este nuevo atributo se nombra *ContrastePalabrasExtremas* y se basa en calcular la distancia entre la palabra más positiva y la palabra más negativa mediante la expresión (2.1), donde w_1 y w_2 son valores entre 0 y 1 que permiten ponderar la importancia de la posición de las palabras y la importancia de la diferencia de valores de polaridad, respectivamente, cumpliéndose que $w_1 + w_2 = 1$. Para asignarle igual importancia a la posición de las palabras y a la diferencia de polaridades extremas se sugiere asignar el valor 0.5 a ambos pesos. La distancia que hay entre las posiciones x e y que ocupan las palabras en el texto se denota por $distPosicion(x,y)$ y se calcula como el valor absoluto de la diferencia de x e y . Las variables $imax$ e $imin$ representan los índices de las palabras con polaridad máxima y mínima, respectivamente. Tanto en esta expresión como en las siguientes que se presentarán se considera que la polaridad de una palabra p , denotada por $polaridad(p)$, se calcula como la diferencia de la polaridad positiva y negativa de la palabra. La cantidad total de palabras del texto se denota por n .

$$\begin{aligned}
 &ContrastePalabrasExtremas & (2.1) \\
 &= w_1 * \left(1 - \frac{distPosicion(imax,imin)}{n} \right) + w_2 \\
 &\quad * \frac{polaridad(p_{imax}) - polaridad(p_{imin})}{2}
 \end{aligned}$$

En (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015) consideran para la detección de ironía la búsqueda de un conjunto de adverbios que indican contradicción. Al estudiar varios ejemplos de ironía se evidencia que las personas utilizan tanto adverbios como signos de puntuación para expresar oposición. Teniendo en cuenta estas

ideas, en este trabajo se crea el atributo *AntecDelimitadorConsec* basado en el contraste entre las dos partes de la oración que surgen al ser dividida ésta por un delimitador. Se consideran delimitadores los adverbios y los signos de puntuación, tales como: punto, dos puntos, coma, y punto y coma. El atributo *AntecDelimitadorConsec* se calcula mediante la expresión (2.2), donde w_1 y w_2 son valores entre 0 y 1 que permiten ponderar la importancia de la diferencia de la cantidad de palabras de ambas partes y la diferencia de la polaridad de ambas partes de la oración, respectivamente, cumpliéndose que $w_1 + w_2 = 1$. n_a es el total de palabras incluidas en el fragmento de texto que antecede al delimitador, $\overline{PolAntec}$ es el promedio de polaridad de las palabras del antecedente, n_c es el total de palabras incluidas en el fragmento de texto que sucede al delimitador, y $\overline{PolConsec}$ es el promedio de la polaridad de las palabras del consecuente.

$$\begin{aligned} &AntecDelimitadorConsec \\ &= w_1 * \frac{|n_a - n_c|}{n} + w_2 * \frac{|\overline{PolAntec} - \overline{PolConsec}|}{2} \end{aligned} \quad (2.2)$$

En investigaciones realizadas en el campo de la detección de la ironía se han creado atributos para tener en cuenta frases entre comillas que son usadas por muchos usuarios al expresar una ironía (Carvalho *et al.*, 2009; Filatova, 2012; Buschmeier, Cimiano and Klinger, 2014; Ling and Klinger, 2016), pero estos estudios tienen en cuenta frases que tengan una estructura fija. Por ejemplo, definen una sintaxis a identificar en la oración, pudiéndose quedar fuera frases irónicas que no respondan a tal sintaxis. Por tanto, en este trabajo se diseña un atributo para reconocer frases entre comillas sin tener en cuenta una sintaxis determinada basado en la polaridad. El atributo se llama *PromPolFrasesComillas* y consiste en el cálculo del promedio de la polaridad de las palabras que se encuentran en frases entre comillas según la expresión (2.3), donde n_f es la cantidad de frases entre comillas, $n_{f_j POS}$ es la cantidad de palabras que son sustantivos, adjetivos o adverbios en la frase f_j y $polaridad(p_i)$ es la polaridad de la palabra p_i .

$$PromPolFrasesComillas = \frac{\sum_{j=1}^{n_f} \sum_{i=1}^{n_{f_j POS}} polaridad(p_i)}{\sum_{j=1}^{n_f} n_{f_j POS}} \quad (2.3)$$

Otro atributo que se utiliza en este trabajo y ha sido empleado por otros autores (Rella, Saggion and Barbieri, 2015) es la desviación estándar de las polaridades de las palabras. *DesviaciónPolaridades* devolverá un valor alto cuando hay palabras en la opinión con polaridades extremas, y devolverá valores bajos en el caso contrario, es decir, cuando los valores de polaridad son homogéneos. *DesviaciónPolaridades* se calcula según la expresión (2.4), donde $\overline{polaridad}$ es la media de las polaridades de las palabras de la opinión.

$$DesviaciónPolaridades = \sqrt{\frac{1}{n} \sum_{i=1}^n (polaridad(p_i) - \overline{polaridad})^2} \quad (2.4)$$

Las personas al escribir irónicamente pueden usar caracteres repetidos para hacer énfasis en una determinada idea. Atributos para capturar si esta característica está presente o no fueron propuestos en (Rajadesingan, Zafarani and Liu, 2015). Estos atributos indican un indicio para explorar la presencia de ironía; no obstante, existen personas que utilizan caracteres repetidos para expresar emociones y no necesariamente están escribiendo irónicamente. No solo es importante tener en cuenta la presencia o no de caracteres repetidos, sino de alguna manera es necesario contabilizar en qué medida ellos están presentes en el texto. Por ello, se decide diseñar un atributo que no solo tenga en cuenta si hay caracteres repetidos, sino la proporción de estos con respecto al texto para dar la idea de en qué medida se está exagerando. El atributo se denomina *RazonRepeticion* y se calcula como la división de la cantidad de palabras que tienen al menos un carácter repetido tres o más veces de forma consecutiva en el texto, ya sean letras o símbolos, entre el total de palabras del texto.

El uso de sinónimos comunes y raros puede indicar la presencia de ironía. Para capturar esta característica, de igual forma que en los artículos (Barbieri and Saggion, 2014; Barbieri, Ronzano and Saggion, 2014), en este trabajo se calculan los rasgos *syno_lower_mean*, *syno_lower_gap*, *syno_greater_mean*, *syno_grater_gap*, basándose en un corpus de referencia donde se busca la frecuencia de cada palabra que aparece en el texto y sus sinónimos. De esta manera, se sigue el supuesto de que el uso de palabras que se utilizan con poca frecuencia, puede ser consecuencia de una intención de énfasis o de tratar de llamar la atención por parte del hablante, y por tanto, existe una probabilidad alta de que esté presente la ironía.

Muchos usuarios al expresar ironía utilizan emoticonos para hacer énfasis en la idea que se quiere transmitir. En varios artículos (Carvalho *et al.*, 2009; Filatova, 2012; Reyes and Rosso, 2012; Reyes, Rosso and Veale, 2013; Vanin *et al.*, 2013; Baca-Gómez *et al.*, 2014; Barbieri and Saggion, 2014; Buschmeier, Cimiano and Klinger, 2014; Ptáček, Habernal and Hong, 2014; Hernández and Velásquez, 2015; Rella, Saggion and Barbieri, 2015; Ling and Klinger, 2016) formulan rasgos que simplemente calculan la cantidad de emoticonos presentes en la oración o que identifican la presencia de emoticonos en determinado contexto, estableciendo sintaxis basadas en etiquetas POS. Estos rasgos no tienen en cuenta la polaridad que transmiten los emoticonos ni la polaridad del contexto donde se encuentran, es decir, no tienen en cuenta el contraste de la polaridad respecto al texto. Por tanto, se decide en este trabajo diseñar un atributo que se nombra *ContrasteEmotTexto* y se basa en el cálculo de la diferencia entre el promedio de la polaridad de los emoticonos y el promedio de la polaridad del texto según la expresión (2.5), donde $\overline{polaridadE}$ representa el promedio de la polaridad de los emoticonos en el texto y $\overline{polaridad}$ representa el promedio de la polaridad de las palabras en el texto. Una lista de emoticonos con su polaridad correspondiente se emplea en este caso.

$$ContrasteEmotTexto = |\overline{polaridadE} - \overline{polaridad}| \quad (2.5)$$

La oposición a través del tiempo es otro de los elementos que puede denotar ironía. De forma similar a las propuestas presentadas en (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015), en este trabajo se analizan las polaridades de sentimientos entre las dos partes que componen el texto delimitada por los verbos conjugados en pasado y presente. Para el cálculo del atributo es necesario localizar la parte de la oración que está en presente y en pasado a partir de los verbos que se encuentran en dichos tiempos, auxiliándose de una lista de verbos en pasado y presente. Luego, se calcula la diferencia entre ambas partes según la expresión (2.6), donde $\overline{PolPresente}$ es el promedio de la polaridad de las palabras que forman la parte de la oración en presente y $\overline{PolPasado}$ es el promedio de la polaridad de las palabras que forman la parte de la oración en pasado. En caso de que el texto no esté constituido por presente y pasado el valor del atributo es 0. Este atributo se denomina *ContrastePresentePasado*.

$$\text{ContrastePresentePasado} = |\overline{PolPresente} - \overline{PolPasado}| \quad (2.6)$$

Lo inesperado, es decir, el desbalance contextual es uno de los elementos que caracterizan la ironía y ha sido detectada por autores mediante la similitud semántica de las palabras (Reyes, Rosso and Buscaldi, 2012). Aunque el desbalance contextual ha sido trabajado para caracterizar la ironía, en los artículos publicados al respecto no queda explícitamente reflejado cómo es que miden el desbalance contextual. De ahí que en este trabajo se propone el atributo *RazonDominios* que calcula la razón de la cantidad de dominios presentes en el texto y la cantidad de palabras que pertenecen a estos dominios. Para buscar los dominios a los cuales pertenecen las palabras en el texto se sugiere el uso de WordNet Domains, recurso que devuelve el dominio de un término, al especificar el ID de su *synset* correspondiente. Este atributo nos da el grado de similitud semántica que hay en el texto analizado.

La ambigüedad puede ser utilizada en un texto para expresar ironía y varios autores han formulado atributos para capturarla (Reyes, Rosso and Buscaldi, 2012; Barbieri and Saggion, 2014; Barbieri, Ronzano and Saggion, 2014; Rella, Saggion and Barbieri, 2015). En las propuestas presentadas por dichos autores se utiliza generalmente algún estadígrafo sobre la cantidad de sentidos de cada palabra como la media y el máximo. En este trabajo se define un atributo nombrado *Ambigüedad* que cuantifica cuán ambiguo es un texto, asumiendo que mientras más significados diferentes tenga un texto, más ambiguo es. En este caso se seleccionó la diferencia del máximo y la media debido a que es más común encontrar que en opiniones irónicas existen pocas palabras con un alto grado de ambigüedad coocurriendo con muchas no ambiguas, que encontrar que la gran mayoría de las palabras son ambiguas. Se sugiere el uso del recurso WordNet para obtener la cantidad de sentidos o significados de una palabra dada su etiqueta POS. *Ambigüedad* se calcula según la expresión (2.7), donde n_{POS} es el total de ocurrencias de las palabras que son sustantivos, adjetivos, adverbios o verbos, $n_{sentidos(p_i)}$ es la cantidad de significados asociados a la palabra i y $max_{sentidos}$ es la mayor cantidad de significados que tiene una palabra en la opinión.

$$\text{Ambigüedad} = max_{sentidos} - \frac{\sum_{i=1}^{n_{POS}} n_{sentidos(p_i)}}{n_{POS}} \quad (2.7)$$

Los modelos propuestos en (Reyes, Rosso and Veale, 2013; Hernández and Velásquez, 2015) utilizan *Polarity Skipgrams* para detectar si existe contradicción en el texto. En este trabajo se crea un atributo denominado *RazonPolaritySkipgrams* y se calcula como la razón entra la cantidad de *Polarity Skipgrams* donde hay oposición y la cantidad de *Polarity Skipgrams* válidos. Los *Polarity Skipgrams* válidos son aquellos formados por dos palabras donde ambas son sustantivos, adjetivos, verbos o adverbios, con saltos de una o dos palabras. Es considerado que hay oposición en un *Polarity Skipgrams* si las dos palabras forman una de las siguientes combinaciones de valores de polaridad: positivo-negativo, positivo-neutro, negativo-neutro, y viceversa. Clasificar una palabra en positiva o negativa muchas veces se torna difícil, porque los límites de los valores de polaridad pueden ser borrosos y también el contexto influye para determinar si tienen polaridad positiva o negativa. Una variante para analizar si una opinión es positiva, neutra o negativa puede ser definiendo como neutras aquellas palabras que tienen polaridades negativas y positivas con valor cero, o aquellas palabras donde la diferencia del valor de la polaridad positiva respecto a la negativa es cero. Otro análisis pudiera ser más flexible, entonces considerando opiniones neutras aquellas que tengan bajos valores de polaridad positiva y negativa, o que la diferencia de las polaridades sea muy pequeña. Por ejemplo, una palabra pudiera ser considerada positiva si la diferencia de los valores de polaridad positiva y negativa es mayor que 0.1, neutra si su valor de polaridad está entre -0.1 y 0.1, y negativa si su polaridad es menor que -0.1. Sin embargo, siguiendo este análisis se corre el riesgo de considerar neutra una palabra que en un contexto sea muy positiva y en otro contexto sea muy negativa. En la presente investigación se utiliza el SentiWordNet para obtener la polaridad de las palabras. Debido a que esta herramienta proporciona dos valores de polaridad para cada palabra (un valor de positividad y otro de negatividad) y en los rasgos propuestos se utiliza un único valor de polaridad que pertenezca al rango [-1,1], fue necesario definir un criterio para determinar dicho valor a partir de los dos valores obtenidos del SentiWordNet. Este criterio establece que se debe seleccionar el máximo entre los dos valores, en el caso de que el mayor sea el de negatividad se halla su opuesto. No obstante, la definición de los rasgos pudiera adaptarse a otras formas para el cálculo de la polaridad, e incluso utilizando dos valores, uno para la polaridad positiva y otro para la negativa.

Las interjecciones son usadas por muchos usuarios en la ironía. En este trabajo se formula un atributo llamado *ExisteInterjFraseFalsaPositiva* relacionado con la

presencia de interjecciones de forma similar a como se define en (Carvalho *et al.*, 2009; Filatova, 2012), donde tienen en cuenta la estructura de la frase que lo acompaña. Este rasgo tiene valor uno si en el texto aparece una interjección acompañada de una frase aparentemente positiva formada por un sustantivo y un adjetivo positivo entre comillas o seguido de dos o más signos de exclamación y/o interrogación, en otro caso el atributo tiene valor cero.

El atributo creado anteriormente solo recoge un caso específico del uso de las interjecciones y los creados en los artículos (Barbieri, Ronzano and Saggion, 2014; Forslid and Wikén, 2015; Ling and Klinger, 2016) solo se detienen a encontrar la cantidad que hay y no la proporción de estas con respecto al texto. Por ello, se crea otro atributo relacionado con las interjecciones llamado *RazonIntejec*, el cual se calcula como la razón entre la cantidad de interjecciones y el total de palabras del texto.

Las palabras en mayúsculas generalmente son usadas en la ironía para decir lo opuesto a lo que se quiere expresar. En (Reyes, Rosso and Veale, 2013; Ptáček, Habernal and Hong, 2014; Bamman and Smith, 2015; Forslid and Wikén, 2015; Hernández and Velásquez, 2015; Rajadesingan, Zafarani and Liu, 2015; Ling and Klinger, 2016) solo crean atributos para la cantidad de palabras o letras en mayúsculas y no para tener en cuenta las polaridades asociadas a éstas y la polaridad que expresa el resto del texto. Por ello, en este trabajo se diseña un atributo denominado *ContrastePalabMayTexto* para capturar el contraste de las polaridades correspondientes a las palabras en mayúsculas y al resto del texto. Este atributo se calcula según la expresión (2.8), donde $\overline{polaridadM}$ denota el promedio de polaridad de las palabras en mayúsculas y $polaridad(p_{imin})$ es la polaridad mínima del texto.

$$ContrastePalabMayTexto = \overline{polaridadM} - polaridad(p_{imin}) \quad (2.8)$$

Una forma de expresar ironía es mediante frases positivas o negativas entre comillas o seguida de signos de exclamación y/o interrogación. Teniendo en cuenta esta idea, se han definido varios rasgos en (Carvalho *et al.*, 2009; Filatova, 2012; Buschmeier, Cimiano and Klinger, 2014; Ling and Klinger, 2016). En este trabajo se sigue una idea similar a las propuestas anteriores, pero marca la diferencia respecto a los presentados en la literatura en que no se centra en una estructura fija en cuanto a las formas gramaticales presentes entre comillas, ni tampoco en el orden en que aparecen, sino que

trata de percibir de manera general el sentimiento expresado a través de la polaridad. Para ello se define un atributo *PolFraseComillasPuntuac* para reflejar la polaridad dominante en la frase que está entre comillas o seguida de dos o más signos de exclamación y/o interrogación. Este rasgo se calcula según la expresión (2.9), donde $polaridad(p_{extremo})$ es la mayor polaridad (ya sea máxima o mínima) que expresa alguna palabra de la frase y $\overline{polaridadC}$ es el promedio de las polaridades que expresan las palabras que conforman la frase a analizar.

condición = "entre comillas o seguida de 2 o más ! o ?"

$$PolFraseComillasPuntuac = |polaridad(p_{extremo}) - \overline{polaridadC}| \quad (2.9)$$

2.2 Herramientas y recursos que contribuyen al pre-procesamiento textual y al cálculo de los rasgos diseñados

Al trabajar con opiniones, tal como ocurre al manejar datos textuales, es necesario realizar una etapa de pre-procesamiento y representación textual. Es por ello que fue necesario identificar qué etapas del pre-procesamiento textual son necesarias al manipular opiniones y aquellas herramientas que contribuyan a una eficaz representación de éstas. Algunas herramientas de código abierto en el lenguaje de programación Java que contribuyen al pre-procesamiento textual son Apache Lucene²⁹, Stanford CoreNLP³⁰, LingPipe³¹ y Apache OpenNLP³². En esta investigación se propone utilizar Apache Lucene para pre-procesar y representar las opiniones. Otros recursos como TreeTagger³³, WordNet³⁴, SentiWordNet³⁵ y WordNet Domains³⁶ fueron utilizados para un procesamiento posterior de las opiniones que permitieron extraer información útil para el cálculo de varios rasgos.

Como parte del pre-procesamiento es necesario extraer los tokens y eliminar algunas palabras gramaticales (*stopwords*), es decir, aquellas que no aportan ningún significado al texto. Al manipular opiniones, es preciso mantener algunas palabras gramaticales que sí aportan información al análisis; por ejemplo, aquellas que permiten delimitar partes de la oración o que indican negación. Se construye una lista con todos los tokens

²⁹ <http://lucene.apache.org/>

³⁰ <http://nlp.stanford.edu/software/corenlp.shtml>

³¹ <http://alias-i.com/lingpipe/demos/tutorial/sentences/read-me.html>

³² <http://opennlp.apache.org/>

³³ <https://github.com/reckart/tt4j>

³⁴ <http://wordnet.princeton.edu/>

³⁵ <http://sentiwordnet.isti.cnr.it/>

³⁶ <http://wndomains.fbk.eu/download.html>

identificados en las opiniones. Por tanto, los tokens se conservan en su forma original. Adicionalmente, los tokens son procesados; por ejemplo, se etiquetan, se identifica cuáles son modificadores y se calcula la polaridad de las palabras para el posterior cálculo de los rasgos. Varias herramientas, listas y recursos en general contribuyeron al cálculo de los rasgos diseñados para la clasificación de las opiniones en irónicas o no. Seguidamente se expondrán las ideas principales sobre estos recursos y cómo fueron utilizados.

Para la tokenización del texto se seleccionó a Apache Lucene porque brinda más flexibilidad que las otras bibliotecas para realizar el análisis léxico de los documentos. Posee diferentes analizadores que permiten eliminar palabras gramaticales, convertir las palabras a minúsculas, y presenta, además, un componente para corregir la ortografía³⁷. No obstante, en esta investigación se decidió no tratar los errores ortográficos debido a que se pudiera desvirtuar el texto original³⁸. El resto de las herramientas que permiten la identificación de los tokens no son igualmente configurables como Lucene; por ejemplo, Stanford CoreNLP lematiza, pero requiere simultáneamente etiquetar las partes del discurso, así como tokenizar y dividir el texto en oraciones, es decir, que las funcionalidades dependen unas de otras. Lucene, además, tiene la ventaja que brinda varios analizadores, de ellos se utilizaron en esta investigación: WhitespaceAnalyzer, que permite obtener los tokens en su forma original, en mayúscula, manteniendo los signos de puntuación, entre otras variantes que son significativas para el análisis a realizar; y SimpleAnalyzer que devuelve todos tokens en minúscula, sin números intermedios y sin signos de puntuación, información que es útil para los rasgos a calcular.

Después de identificar los tokens, fue necesario utilizar recursos que permitieran obtener información útil para calcular los rasgos definidos. Por ejemplo, fue necesario identificar las categorías gramaticales de cada palabra, la polaridad de las palabras, sus

³⁷ Lucene suggest API se basa en indexar palabras n-gram y el cálculo de la distancia Levenshtein entre palabras. Es posible utilizar un diccionario para verificar si las palabras de las opiniones se encuentran, si aparecen en el diccionario entonces son correctas, de lo contrario se sugiere el término utilizando un índice de Lucene construido sobre el diccionario. Un diccionario útil a tales efectos se encuentra disponible en /usr/share/dict/words.

³⁸ Las personas comenten muchos errores escribiendo los tweets, encontrar la palabra más cercana pudiera ayudar en el análisis, pero tal vez entorpecer la identificación de las palabras con repeticiones de caracteres. En futuras investigaciones se pudiera manejar esta situación considerando dos listas de tokens: originales y transformados.

sinónimos, los dominios a los que pueden pertenecer los términos, los posibles significados de las palabras y las emociones que transmiten.

La asignación de las categorías gramaticales a cada palabra del texto (*part-of-speech tagging*, *POS tagging* o *POST*) fue realizada con el TreeTagger. Esta es una herramienta para anotar textos con información de *part-of-speech* y lema, desarrollada en el Instituto para lingüística computacional de la Universidad de Stuttgart. Ha sido utilizada con éxito para etiquetar textos en Alemán, Inglés, Francés, Italiano, Español, Griego, y Francés antiguo, y es fácilmente adaptable a otros lenguajes si se dispone de un lexicón y corpus marcado manualmente. Algunas de las etiquetas que este recurso es capaz de reconocer son: sustantivo, adjetivo, verbo, adverbio, conjunción, signos de puntuación, símbolos, interjección y pronombre, entre otros.

El WordNet fue utilizado para obtener los sinónimos de las palabras y los significados o acepciones. El WordNet es una extensa base de datos léxica de la lengua inglesa, desarrollada por George A. Miller, a partir de la combinación de informaciones lexicográficas (relaciones léxicas y semánticas usadas para representar la organización del conocimiento léxico) y recursos computacionales (Fellbaum, 1998). En (Miller, 1995) se define el vocabulario de un lenguaje con un conjunto de pares palabra-sentido a partir de un conjunto de significados. En el WordNet un sentido es representado por un conjunto de uno o más sinónimos. La base posee más de 118000 palabras diferentes y más de 90000 sentidos (Miller, 1995). Esta herramienta está compuesta de sustantivos, verbos, adjetivos y adverbios agrupados en conjuntos de sinónimos (*synsets*), cada uno expresando un concepto distinto. Estos son organizados en un conjunto de lexicógrafos por categoría sintáctica y por otros criterios de organización. Los adverbios se registran en un archivo, mientras los sustantivos y verbos son agrupados de acuerdo con la semántica. Los adjetivos son divididos en descriptivos y relacionales (Miller *et al.*, 1990). Los *synsets* pueden estar conectados mediante relaciones semánticas (por ejemplo, hiperonimia, hiponimia, inclusión, antonimia, meronimia, etc.). En esta investigación se hizo uso del método `getAllSynsets(<palabra>, <etiqueta>)` que devuelve todos los sinónimos asociados a una palabra y `getSenseCount(<palabra>, <etiqueta>)` que retorna la cantidad de sentidos de una palabra, es decir, la cantidad de acepciones diferentes que tiene.

Otro recurso léxico que contribuyó al cálculo de los rasgos es el SentiWordNet. En esta investigación se utiliza la versión 4.1 (Amores, 2016) la cual presenta un total de 58969 términos anotados, de ellos 35520 etiquetados como sustantivos, 13190 adjetivos, 8298 verbos y 1961 adverbios. Este recurso léxico es el resultado de anotaciones automatizadas en todos los *synsets* del WordNet, con grado de positividad y negatividad. A cada palabra se le asocian dos valores numéricos que indican cuan positiva y negativa es. Cada uno de estos valores varían en el intervalo [0, 1]. Se toma como el valor de polaridad de una palabra el máximo entre el grado de positividad y el de negatividad, en caso de ser este último el máximo, se guarda el opuesto; de esta forma la polaridad de una palabra está en el intervalo [-1,1]. No obstante, se pudieran considerar en futuras investigaciones otras formas para calcular un valor único de polaridad a partir de los valores positivo y negativo. Por ejemplo, algunos experimentos fueron realizados calculando la polaridad como la resta de la polaridad positiva y negativa.

En esta investigación los dominios de las palabras del texto se emplearon para la formulación de rasgos. Para obtener esta información se utiliza el WordNet Domains, una extensión multilingüe (inglés e italiano) del WordNet, donde cada *synset* es apuntado en uno o más dominios; por ejemplo, deporte, política y medicina. Un dominio puede incluir *synsets* de diferentes categorías sintácticas (Strapparava, Valitutti and others, 2004).

Además, para el pre-procesamiento de las opiniones se utiliza una lista de modificadores. Las palabras modificadoras son aquellas que pueden incrementar, reducir o cambiar la polaridad de otro término dentro de la oración. Especial atención tienen las palabras intensificadoras las cuales aumentan la fuerza de otros términos y las atenuadoras que reducen la fuerza de otros términos.

En gramática, un intensificador es una palabra que hace hincapié en otra palabra o frase. También conocido como un refuerzo o un amplificador. Por ejemplo, algo no sólo es bueno, sino muy bueno, o incluso, terriblemente bueno. Los adjetivos intensificadores modifican a los sustantivos; los adverbios de intensificación modifican comúnmente verbos, adjetivos graduables, y otros adverbios.

Hasta cierto punto, un intensificador actúa como una señal: anuncia que la siguiente palabra que se lleva a cabo debe ser entendida como insuficiente. Algunos ejemplos de

palabras intensificadoras en el idioma Inglés son: “The woodwind has a slightly greater scope than the violin”, “A really good detective never gets married” y “Maggy knew that something surpassingly important had happened to him”.

Otras palabras modificadoras son las palabras atenuadoras; por ejemplo, el término “less” en Inglés o “poco” en Español. Desde el punto de vista semántico-pragmático, la atenuación se opone a la intensificación. En ambos casos los recursos pueden corresponder a ámbitos categoriales diferentes, como adverbios, adjetivos y frases.

En esta investigación se construyó una lista de palabras modificadoras, tomadas de diversas fuentes³⁹, que sigue la estructura que se muestra en la Figura 1. El fichero contiene en cada línea una palabra o frase, seguida por el carácter M o el I, para indicar si es un mitigador o intensificador, respectivamente.

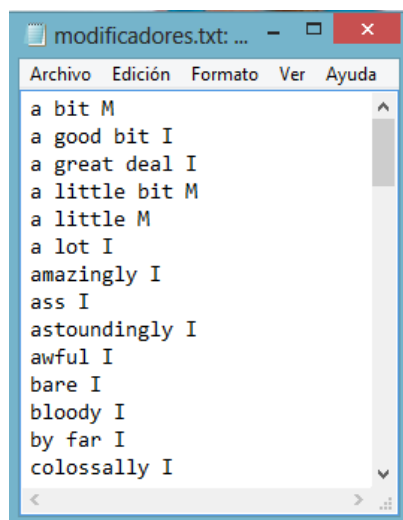


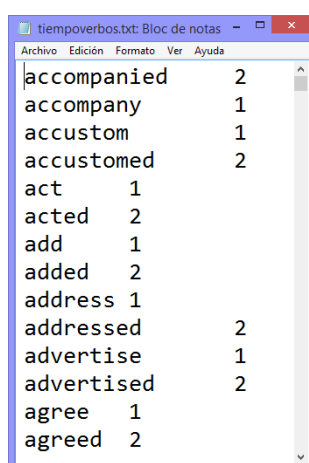
Figura 1. Fragmento de la lista de palabras modificadoras.

Esta lista se utiliza de la siguiente manera. Se buscan los modificadores que están presentes en el texto y a qué palabra están modificando, la cual puede ser un adjetivo, verbo, adverbio, y en último caso, sustantivo. Luego, se modifica la polaridad de esta palabra, sumando o restando el 50% de su polaridad en dependencia si está actuando sobre ella un intensificador o un mitigador, respectivamente. Finalmente, se verifica que la polaridad de la palabra continúe estando en el intervalo $[-1,1]$. En el caso de que la polaridad resultante sea mayor que 1, su polaridad se ajusta a 1, y si es menor que -1, se

³⁹ <http://www.shertonenglish.com/resources/es/adjectives/adjectives-intensifiers.php>
<https://en.wikipedia.org/wiki/Intensifier>
<http://es.stgeorges.com/blog/adjetivos-intensificadores-mitigadores>

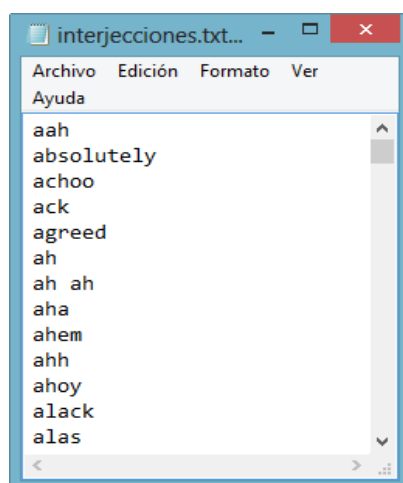
ajusta al -1. De esta forma se garantiza que los valores de polaridad estén en el intervalo [-1,1].

Otras listas fueron útiles para el cálculo de los atributos que permiten discriminar entre opiniones irónicas y no irónicas. Se creó una lista que contiene cada forma verbal y su tiempo, a partir de una lista de verbos publicada en internet⁴⁰. La lista está almacenada en un fichero texto con la siguiente estructura: cada línea corresponde a una forma verbal, acompañada de un código que indica si la forma verbal está en presente o en pasado. El presente está representado por un uno y el pasado por un dos, como se muestra en la Figura 2.



accompanied	2
accompany	1
accustom	1
accustomed	2
act	1
acted	2
add	1
added	2
address	1
addressed	2
advertise	1
advertised	2
agree	1
agreed	2

Figura 2. Fragmento de la lista de formas verbales.



aah
absolutely
achoo
ack
agreed
ah
ah ah
aha
ahem
ahh
ahoy
alack
alas

Figura 3. Fragmento de la lista de interjecciones.

⁴⁰ <https://sites.google.com/site/censenglish/verbos-regulares-e-irregulares-1>

Además, se crea una lista de interjecciones, a partir de la recolección de éstas⁴¹, que es utilizada para el cálculo de algunos rasgos. Este archivo contiene una interjección por línea como se muestra en la Figura 3.

También se tiene una lista de adverbios⁴² para ser utilizada en el cálculo de los rasgos. En la Figura 4 se muestra un fragmento del fichero construido, el cual contiene un adverbio por línea.

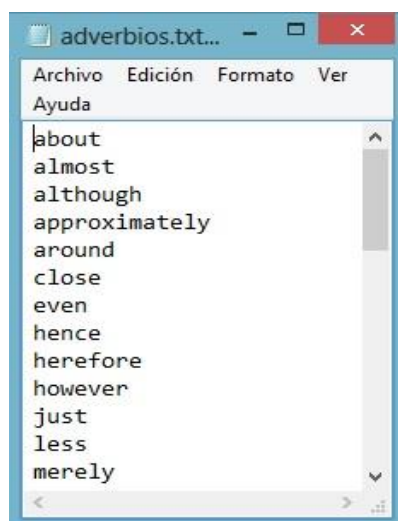


Figura 4. Fragmento de la lista de adverbios.

Los emoticonos (*emoticon*) son una secuencia de caracteres ASCII que, en un principio, representaba una cara humana y expresaba una emoción. Posteriormente, fueron creándose otros emoticonos con significados muy diversos.

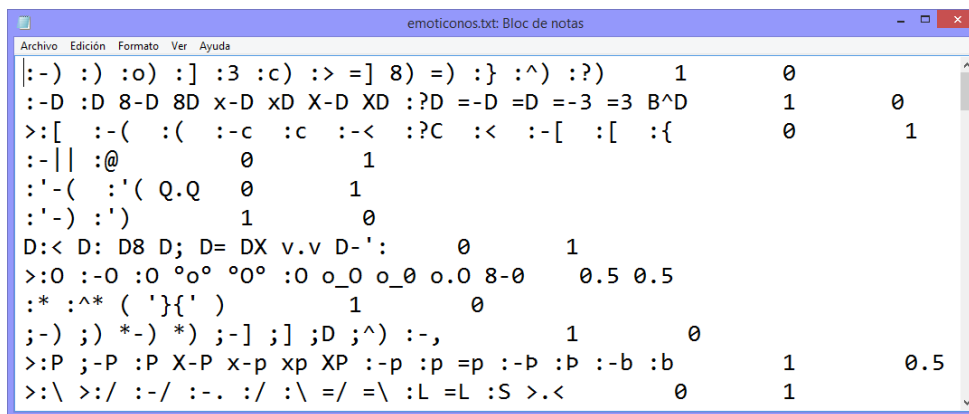
Existen varias listas que incluyen emoticonos; sin embargo, cada una adolece de términos que presentan otras, sobre todo debido a la velocidad y dinámica con que surgen nuevos emoticonos. De ahí que para tratarlos se utiliza una lista de emoticonos construidas en una investigación precedente (Amores, Arco and Barrera, 2016), a partir de varias listas previamente publicadas⁴³. Cada emoticón tiene asignados sus correspondientes valores de polaridad positiva y negativa.

⁴¹ <http://www.yourdictionary.com/index.php/pdf/articles/156.listof-interjections.pdf>
<http://www.english-grammar-revolution.com/list-of-interjections.html>
<http://www.enchantedlearning.com/wordlist/interjections.shtml>

⁴² <http://users.dsic.upv.es/grupos/nle/resources/Signatures.components>

⁴³ www.netlingo.com, www.noslang.com, es.wikipedia.org/wiki/Anexo:Emoticonos

En la Figura 5 se muestra un fragmento de la lista de emoticonos y los valores de polaridad positiva y negativa asignados, estos valores representan la emoción que expresa cada emoticón. Finalmente, se calcula para cada uno, un valor único de polaridad consistente en el máximo entre el valor de polaridad positivo y el negativo. En el caso de que los valores de polaridad positivo y negativo sean 0.5, la polaridad final del emoticón será cero. Si el valor de polaridad negativo es mayor que el positivo se guarda el opuesto del negativo; de esta forma la polaridad de una palabra está en el intervalo [-1,1].



Emoticono	Polaridad Positiva	Polaridad Negativa	Polaridad Final
:-) :o) :] :3 :c) :> =] 8) =) :} :^) :?)	1	0	1
: -D :D 8-D 8D x-D xD X-D XD :?D =-D =D =-3 =3 B^D	1	0	1
>:[:- (: (:-c :c :-< :?C :< :-[:[:{	0	1	0
:- :@	0	1	0
:'-(:'(Q.Q	0	1	0
:'-) :')	1	0	1
D:< D: D8 D; D= DX v.v D-':	0	1	0
>:O :-O :O °o° °O° :O o_o o_0 o.o 8-0	0.5	0.5	0
:* :^* ('){')	1	0	1
; -) ;) *-) *) ; -] ;] ; D ; ^) :- ,	1	0	1
>:P ; -P :P X-P x-p xp XP :-p :p =p :-p :p :-b :b	1	0.5	0.5
>:\ >:/ :-/ :- . :/ :\ =/ =\ :L =L :S >.<	0	1	0

Figura 5. Fragmento de la lista de emoticonos con sus valores de polaridad asociados.

Otro recurso que es importante tener en cuenta es un corpus de referencia para tener un estimado de la frecuencia de aparición de los términos en los documentos. Estas colecciones textuales de referencia generalmente ofrecen las palabras, sus etiquetas correspondientes y la frecuencia de aparición de éstas. El *Corpus of Contemporary American English* (COCA)⁴⁴ es una de las colecciones de referencia más utilizadas. Existen varias versiones de este corpus, fue escogida la versión de 5000 palabras⁴⁵ ya que es libre, a diferencia de otras versiones que no son gratuitas.

Además, se tiene un conjunto de reglas a aplicar si aparecen las palabras *not*, *no* o *never* (Na, Khoo and Wu, 2005). Estas reglas indican cuáles son las palabras que afectan por los negadores y, por tanto, se altera su polaridad durante el pre-procesamiento. Se tienen tres archivos, uno por cada negador y una regla por línea del documento.

⁴⁴ <http://www.wordfrequency.info>

⁴⁵ <http://www.wordfrequency.info/free.asp>

2.3 Módulo para la representación de las opiniones con el objetivo de detectar de manera efectiva la ironía

La creación de un módulo capaz de recibir un conjunto de opiniones en formato texto y devolver un archivo con el formato de entrada del Weka donde los valores de los atributos caractericen de la mejor forma a la ironía es uno de los objetivos de esta investigación. A continuación, se explicará la estructura de dicho módulo, las clases que lo forman y cada una de sus funcionalidades.

Para la construcción del módulo se utiliza el lenguaje Java. Este lenguaje orientado a objeto facilita el trabajo con ficheros al estar disponibles los paquetes `java.io` y `java.util` que contienen clases que permiten la lectura y escritura de archivos. Además, en el paquete `java.util` se encuentran las clases `HashMap` y `ArrayList` que fueron utilizadas para almacenar información y poder acceder rápidamente a ellas.

El módulo está formado por las clases `EntradaSalida`, `Rasgos`, `Recursos`, `Lucene`, `SentiWordNet`, `WordNetDomains` y `Diccionario`. En el diagrama de clases que se muestra en la Figura 6 se evidencian las relaciones entre ellas.

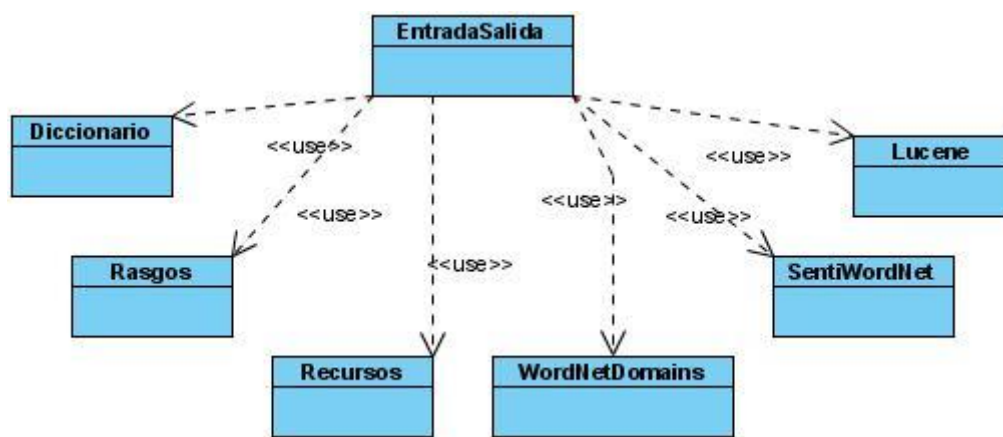


Figura 6. Diagrama de clases del módulo desarrollado.

La clase `EntradaSalida` es la encargada de leer desde un archivo de texto el conjunto de opiniones que serán procesadas. Aquellas opiniones que tienen hipervínculos no son procesadas. El valor resultante de calcular cada rasgo de una opinión se va escribiendo en un archivo de salida con extensión `arff` que sigue el formato de entrada del Weka. Para comenzar a procesar las opiniones se crea un objeto de esta clase, pasándole como parámetro tres cadenas que indican la dirección de los archivos donde se encuentran las

opiniones irónicas, las no irónicas y el fichero de salida. En la clase Recursos se cargan en memoria los archivos que contienen las listas de los verbos con su tiempo en pasado o presente, los emoticonos con su polaridad, las palabras y su frecuencia en el *Corpus of Contemporary American English*, la lista de los modificadores, las interjecciones, los adverbios y las reglas que indican a qué palabras hay que modificar en caso de que aparezca *not*, *no* o *never*.

Para procesar cada opinión se tiene la clase Rasgos. Los atributos de esta clase toman valor con la ejecución del método dedicado al pre-procesamiento. Como se muestra en la Figura 7, existen atributos para almacenar: el texto en su estado original como una cadena, el texto separado en tokens donde se mantienen las mayúsculas y signos de puntuación, los tokens que forman el texto sin signos de puntuación, todo en minúscula y con la corrección ortográfica realizada, una lista con las etiquetas devueltas por el TreeTagger para cada palabra, un arreglo con las polaridades suministradas por el SentiWordNet para cada palabra con su etiqueta POS, la frecuencia de las palabras del texto y sus sinónimos según el *Corpus of Contemporary American English* y un arreglo que nos indica si los tokens que forman el texto tienen al menos un carácter repetido tres o más veces. También se tiene el atributo *posInterj* donde se guardan las posiciones donde hay interjecciones, éste toma valor con el cálculo del rasgo *RazonIntejec* y es utilizado para hallar el valor del rasgo *ExisteInterjFraseFalsaPositiva*. Además, durante el pre-procesamiento se altera la polaridad de las palabras si están modificadas por algún intensificador o atenuador presente en la lista de modificadores de referencia. Si el modificador es un intensificador se suma a la polaridad de la palabra su 50%, si es un atenuador se resta dicho valor. Luego se verifica que su polaridad continúe en el intervalo [-1,1]; en caso contrario, su valor se lleva al extremo del intervalo más cercano. En esta etapa de pre-procesamiento se sustituyen los caracteres que se repiten tres o más veces por una sola ocurrencia del carácter, luego, se modifica la polaridad de estas palabras sumándole el 50% de su valor y se verifica que la polaridad continúe en el intervalo [-1,1]; en caso contrario, su valor se lleva al extremo del intervalo más cercano. Además, se sustituye toda aparición de “n’t” por “not” y se invierte la polaridad de aquellas palabras que están siendo modificadas por *not*, *no* o *never*. También en esta clase se encuentra el método *getValoresRasgos* que permite calcular el valor de los rasgos que discriminan entre opiniones irónicas y no irónicas utilizando la información almacenada en los atributos previamente calculados en la clase. Este

método llama a los métodos correspondientes a cada rasgo a calcular. En el caso de añadir nuevos rasgos para lograr una mayor discriminación entre opiniones irónicas y no irónicas, basta con crear un método para el cálculo de dicho rasgo, aumentar el tamaño del arreglo valoresRasgos y en el método getValoresRasgos asignar el valor del rasgo en la última posición del arreglo. En esta clase, además, se incluyeron otros métodos auxiliares que facilitan la organización y reutilización.

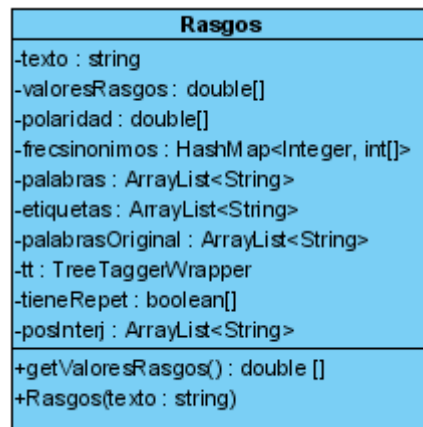


Figura 7. Atributos y métodos de la clase Rasgos.

En la Figura 8 se muestra el diagrama de actividad para el pre-procesamiento, haciendo énfasis en el orden utilizado para obtener la información necesaria para el cálculo de los rasgos.

En el módulo también se incluye la clase Lucene que posee dos métodos: uno donde el analizador SimpleAnalyzer devuelve los tokens que forman el texto en minúsculas sin signos de puntuación y números intermedios, y otro, que devuelve los tokens en su forma original con mayúsculas y signos de puntuación resultantes de aplicar el analizador WhitespaceAnalyzer.

También se decidió crear la clase SentiWordNet la cual tiene un atributo de tipo HashMap donde están las palabras con su etiqueta y el valor de la polaridad asociada. Este atributo toma valor en el constructor donde toda esa información es leída de un archivo y guardada en la tabla hash. Un método público es definido para obtener del atributo la información deseada, en este caso, la polaridad dada una palabra y su etiqueta.

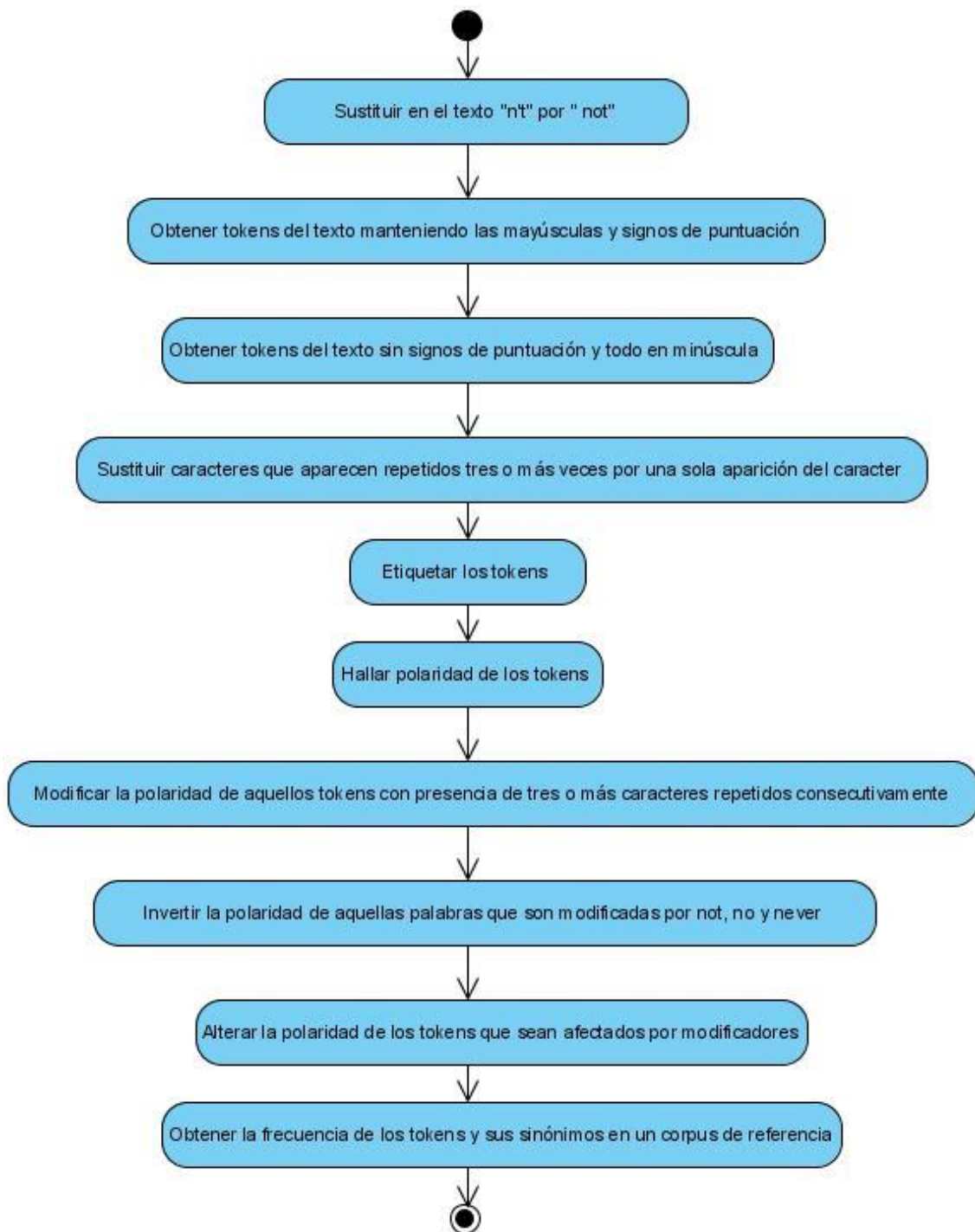


Figura 8. Diagrama de actividad para el pre-procesamiento.

Además, se crea la clase `Diccionario` donde se tiene un atributo del tipo `IDictionary` que toma valor en el constructor y es utilizado en el método público para buscar dada una palabra y su etiqueta, los identificadores de los *synset* que la representan.

Finalmente, la clase WordNetDomains se creó para acceder a este recurso. En ésta se declara un atributo del tipo HashMap para guardar el identificador del *synset* al que pertenecen las palabras con su etiqueta y el dominio asociado. Además, se incluyó un método público para obtener el dominio según el identificador del *synset* de la palabra y su etiqueta.

2.4 Conclusiones parciales

En este capítulo se propone un conjunto de rasgos que permiten discriminar entre opiniones irónicas y no irónicas. Los rasgos que permiten cuantificar cuán ambiguo es un texto, medir el uso de sinónimos comunes y raros como un indicador de la presencia de ironía, calcular la desviación estándar de las polaridades de las palabras y buscar una estructura fija basada en la presencia de interjecciones, fueron incorporados en esta investigación por los excelentes resultados que reportan. Ellos fueron calculados siguiendo las expresiones ya descritas previamente en la literatura. Además, se crearon nuevas expresiones para calcular rasgos enunciados en la literatura, entre ellos aquel que mide el contraste entre las dos partes de la oración que surgen al ser dividida ésta por un delimitador (adverbios o signos de puntuación), el que analiza las polaridades de sentimientos entre las dos partes que componen el texto delimitada por los verbos conjugados en pasado y presente y el rasgo que utiliza *Polarity Skipgrams* para detectar si existe contradicción en el texto.

Nuevos rasgos fueron definidos a partir del estudio de las características que identifican las opiniones irónicas, es por ello que se propuso un atributo para medir el contraste donde no solo se tiene en cuenta la polaridad, sino también la distancia entre los elementos de mayor contraste. Otro rasgo novedoso es aquel que mide el desbalance contextual mediante el cálculo de la razón de la cantidad de dominios presentes en el texto y la cantidad de palabras que pertenecen a estos dominios. Se definió un rasgo que permite reconocer las frases entre comillas sin tener en cuenta una sintaxis determinada basado en la polaridad. Se diseñó otro rasgo más general relacionado con las interjecciones que no busca una estructura específica, pues mide la proporción de estas con respecto al texto. Se creó un atributo que no solo tiene en cuenta la existencia de caracteres repetidos, sino la proporción de éstos con respecto al texto para dar la idea de en qué medida se está exagerando. El contraste también fue medido mediante el cálculo de la diferencia entre el promedio de la polaridad de los emoticones y el promedio de la polaridad del texto. Un nuevo rasgo fue definido para capturar el contraste de las

polaridades correspondientes a las palabras en mayúsculas y al resto del texto. Se creó un rasgo que analiza las frases positivas o negativas entre comillas o seguida de signos de exclamación y/o interrogación sin seguir una estructura fija en cuanto a las formas gramaticales presentes entre comillas, ni tampoco tener en cuenta el orden en que aparecen, sino que trata de percibir de manera general el sentimiento expresado a través de la polaridad.

Varios recursos fueron necesarios para pre-procesar las opiniones y extraer la información necesaria para calcular los rasgos seleccionados y creados. Lucene permitió la tokenización y eliminación de palabras gramaticales. TreeTagger ofreció la etiqueta POS de cada token obtenido. El WordNet fue utilizado para obtener los sinónimos de las palabras y los significados o acepciones, mientras que el SentiWordNet permitió obtener los valores de polaridad de los términos. WordNet Domains fue utilizado para obtener los dominios de las palabras del texto. Además, varias listas fueron útiles para el pre-procesamiento y el cálculo de los rasgos, entre ellas, la lista de modificadores, la lista de interjecciones, la lista que contiene cada forma verbal y su tiempo, la lista de adverbios y la lista de emoticonos. Otros recursos importantísimos en el análisis son: el corpus de referencia para tener un estimado de la frecuencia de aparición de los términos en los documentos y las reglas para buscar las palabras que son afectadas por la negación al estar presente los términos *no*, *not* o *never*.

El módulo diseñado e implementado para representar las opiniones mediante los rasgos que permiten discriminar entre textos irónicos y no irónicos está compuesto por las clases EntradaSalida, Rasgos, Recursos, Lucene, SentiWordNet, WordNetDomains y Diccionario. Las clases Lucene, SentiWordNet y WordNetDomains fueron creadas para manejar fácilmente los recursos del mismo nombre y conseguir la información necesaria de ellos. La clase Diccionario se utilizó para obtener los identificadores de los *synset* que ofrece el WordNet para una palabra, los cuales eran necesarios para la búsqueda de los dominios en WordNet Domains. La clase Recursos se creó para cargar en memoria y consultar rápidamente la lista de modificadores, la lista de adverbios, la lista de interjecciones, la lista que contiene cada forma verbal y su tiempo, la lista de emoticonos, las reglas para afectar la polaridad de las palabras que son modificadas por *no*, *not* o *never*, y el corpus de referencia para tener un estimado de la frecuencia de aparición de los términos en los documentos. Además, se creó una clase Rasgos para realizar el pre-procesamiento de cada opinión y el cálculo de los rasgos, para lo cual fue

necesario auxiliarse de todos los recursos mencionados anteriormente. En la clase EntradaSalida se leen las opiniones irónicas y no irónicas, se crean los objetos que manipulan los recursos necesarios para su procesamiento, se mandan a procesar las opiniones y se imprimen los valores de los rasgos en un archivo con formato de entrada al Weka.

Capítulo 3 Evaluación de la representación propuesta para la detección de ironía

Se han realizado numerosos trabajos orientados a la detección de ironía, pero la efectividad no es muy alta y conocer cuán expresivos son los rasgos propuestos en esta investigación es uno de los objetivos trazados. Por ello, primero se expondrán propuestas que se han hecho en el campo de la minería de opinión en cuanto a la detección de ironía, destacando el dominio, el idioma, el algoritmo con el cual se probó el modelo y los resultados obtenidos. Luego, se describirán las principales medidas y los corpus que permitirán evaluar la efectividad de los rasgos desarrollados para la discriminación entre opiniones irónicas y no irónicas. Finalmente, se aplicarán las pruebas diseñadas para esta etapa de validación que mostrarán cuán bueno son los rasgos diseñados e implementados y con qué algoritmos se obtuvieron los mejores resultados.

3.1 Principales propuestas para la detección de la ironía en documentos textuales

En los últimos años se han reportado varios trabajos orientados con la detección de ironía (Reyes, Rosso and Buscaldi, 2012; Reyes, Rosso and Veale, 2013; Barbieri and Saggion, 2014; Jasso, 2015; Rella, Saggion and Barbieri, 2015), a continuación se presentan algunos de ellos, haciendo énfasis en el dominio y el idioma en que fueron desarrollados, los algoritmos que fueron utilizados para la clasificación y los resultados obtenidos.

En la investigación realizada en (Baca-Gómez *et al.*, 2014) se propuso un modelo para clasificar en irónicos o no irónicos los comentarios en Español emitidos en Facebook. Este fue probado con el algoritmo Optimización Minimal de Secuencias (*Sequential Minimal Optimization*; SMO) donde se concluyó que la ironía tiene un impacto negativo del 10% sobre la calidad de la detección de polaridad.

En el trabajo (Espinoza *et al.*, 2014) se creó un motor de detección de ironía para mensajes de Twitter⁴⁶. Para probar su eficacia se utilizaron los clasificadores Naïve

⁴⁶ Twitter impone grandes retos al procesamiento textual, ya que los tweets son textos muy cortos donde como máximo se pueden utilizar 140 caracteres.

Bayes Multinomial (*Multinomial Naïve Bayes*) y Máquinas de Soporte Vectorial (*Support Vector Machine*; SVM). El mejor comportamiento se presentó con el clasificador SVM, con una precisión del 43.12%. Este resultado fue superado en (Mohammad, Kiritchenko and Zhu, 2013), alcanzando un valor de precisión superior al 60% clasificando con SVM los textos en irónicos o no. Se alcanzan distintos valores de precisión incluso utilizando los mismos clasificadores porque en este análisis es importante la selección de los rasgos que mejor caractericen a los textos irónicos. Los clasificadores Bayesianos así como SVM también fueron empleados para detectar ironía en opiniones de productos de un sitio de ventas en línea (Reyes and Rosso, 2012). Los resultados reportados son del 87% utilizando el algoritmo de Naïve Bayes y del 89.03% con SVM.

En (Jasso, 2015) se propone un modelo para detectar ironía y sarcasmo en textos cortos en Español de Twitter utilizando los clasificadores SVM y Random Forests. Los valores de calidad obtenidos en la detección de ironía solamente con n-gramas de palabras para los clasificadores Random Forests/SVM fueron de 0.6362/0.6520, 0.6262/0.6405, 0.6756/0.6932 y 0.65/0.6658 para las medidas Exactitud (*Accuracy*), Precisión (*Precision*), Exhaustividad (*Recall*) y la Medida-F (*F-measure*), respectivamente. Sin embargo, al utilizar n-gramas de caracteres con los mismos clasificadores los resultados fueron mejores: 0.8496/0.8490 de Exactitud, 0.8387/0.8387 de Precisión, 0.8658/0.8658 de Exhaustividad y 0.8520/0.8520 de la Medida-F.

En la investigación publicada en (Reyes, Rosso and Buscaldi, 2012) se alcanzaron valores de Exactitud, Precisión, Exhaustividad y Medida-F de hasta un 91%, 90%, 95% y 92%, respectivamente (mejor resultado alcanzado) al detectar ironía en textos en Inglés de Twitter utilizando clasificadores basados en la construcción de árboles de decisión. Reyes también desarrolló otro trabajo (Reyes, Rosso and Veale, 2013) para detectar la presencia de ironía en textos cortos de Twitter en Inglés. Para probar la calidad de los rasgos descritos utilizó los clasificadores *Naïve Bayes* y basados en árboles de decisión (*Decision Trees*). En los experimentos, la Exactitud varió desde el 44.88% hasta el 85.40%. Al clasificar teniendo en cuenta solo una pareja de atributos se obtuvo una Exactitud entre el 68% y el 74%. Es importante tener en cuenta que si existiera un clasificador que etiquetara todos los textos como no irónicos lograría una exactitud del 70%. Por tanto, este modelo difícilmente sobrepasa su línea base al considerar solo una pareja de atributos. La clasificación fue realizada empleando árboles

de decisión y evaluada usando validación cruzada con 10 particiones (*10-fold cross validation*). Al usar un corpus balanceado, la Exactitud es más baja pero la Precisión, la Exhaustividad y la Medida-F son claramente más altas, alcanzando los valores 72.30%, 73.6%, 69.5% y 71.5%, respectivamente. Por el contrario, al utilizar un corpus desbalanceado, la Exactitud es más alta pero la Precisión, la Exhaustividad y la Medida-F decaen, obteniéndose los valores 80.44%, 66.1%, 44.7% y 53.3%, respectivamente.

En el trabajo realizado por Barbieri y Saggion (Barbieri and Saggion, 2014) se propuso un modelo para la detección de ironía y humor en textos de Twitter en Inglés. Para ello se aplicaron los clasificadores Random Forest y aquellos basados en árboles de decisión. Barbieri and Saggion lograron valores de Precisión, Exhaustividad y Medida-F superiores a los alcanzados por Reyes en (Reyes, Rosso and Veale, 2013) al realizar la experimentación con el mismo corpus textual. Para colecciones textuales balanceadas los valores de estas medidas se comportaron alrededor del 80% al clasificar con árboles de decisión y del 83% al clasificar con Random Forest.

Los autores del artículo publicado en (Rella, Saggion and Barbieri, 2015) se basaron en el estudio realizado por Barbieri y Saggion (2014) para crear un modelo para la detección de ironía en mensajes de Twitter para el idioma Catalán. Dos algoritmos de clasificación fueron probados: Naïve Bayes y SVM. Los mejores resultados de esta propuesta se obtuvieron con SVM considerando todos los rasgos. El 75% del corpus de entrenamiento utilizado para la experimentación estaba conformado por instancias no irónicas, por lo que estaba desbalanceado y consecuentemente el sistema realizó predicciones incorrectas.

En (Hernández and Velásquez, 2015) se realiza una adaptación del modelo desarrollado por Reyes para el idioma Español y logran la clasificación del contenido generado por usuarios de Twitter. Para probar el modelo se utilizó el algoritmo Naïve Bayes, el cual es capaz de clasificar correctamente solamente el 60% de los casos positivos en el conjunto de prueba.

Algunos autores han usado el aprendizaje supervisado para construir una base de conocimiento, donde si una cantidad de tweets sustancial usan una frase determinada entonces esta frase es marcada como sarcástica y el analizador añade ésta a la base de conocimiento, es decir, se basa en la búsqueda de frases o palabras comunes. Luego se busca si un tweet dado contiene alguna de estas frases para ser declarado como

sarcástico. Un ejemplo de frase comúnmente usada en el sarcasmo es “Yeah right” (Ingle *et al.*, 2014).

Los modelos creados hasta el momento para la detección de la ironía de forma automática no caracterizan completamente este fenómeno, pues los mejores resultados alcanzados hasta el momento están entre un 70% y 90% de efectividad.

Los algoritmos utilizados para la clasificación de las opiniones fueron Naïve Bayes, máquinas de soporte vectorial, Random Forest y árboles de decisión. Estos clasificadores generalmente son evaluados realizando validación cruzada y analizando las medidas de calidad precisión, exactitud y medida-F.

3.2 Principales medidas para validar la clasificación de las opiniones en irónicas o no

La evaluación de la efectividad de los rasgos propuestos para la distinción entre opiniones irónicas y no irónicas se basa en el análisis de la matriz de confusión y los valores de las medidas de calidad: precisión, exhaustividad, exactitud, medida-F y la medida kappa. Seguidamente se describen estos términos que indican cuán expresivos son los rasgos diseñados e implementados en esta investigación.

Para analizar los resultados del clasificador en la detección de la ironía primero es necesario conocer las siguientes definiciones (Han and Kamber, 2012):

- Verdaderos positivos (VP): Son aquellos datos del conjunto positivo (irónico), que son etiquetados correctamente como positivos (irónico) por el clasificador.
- Verdaderos negativos (VN): Son aquellos datos del conjunto negativo (no irónico), que son etiquetados correctamente como negativos (no irónico) por el clasificador.
- Falsos positivos (FP): Son aquellos datos del conjunto negativo (no irónico), que son etiquetados de forma errónea por el clasificador. Es decir, estos datos son en realidad del conjunto negativo, pero fueron etiquetados como positivos (irónico).
- Falsos negativos (FN): Son aquellos datos del conjunto positivo (irónico), que son etiquetados de forma errónea por el clasificador. Es decir, estos datos son en realidad del conjunto positivo, pero fueron etiquetados como negativos (no irónico).

La cantidad de VP, FN, FP y VN son los valores que se tienen en cuenta para construir la matriz de confusión, la cual permite ver a grandes rasgos qué tan bien se está comportando el clasificador al reconocer opiniones irónicas y no irónicas. En la Tabla 1 se muestra la estructura de la matriz de confusión donde se permite contrastar lo real con lo inferido por el clasificador, a partir de la distribución de los valores VP, FN, FP y VN.

Tabla 1. Estructura de la matriz de confusión.

Matriz de confusión		Clases Inferidas	
		Irónica	No Irónica
Clases Reales	Irónica	VP	FN
	No Irónica	FP	VN

A partir de estas definiciones básicas es posible caracterizar las medidas para evaluar el desempeño de los clasificadores. A continuación, se describen las más significativas (Han and Kamber, 2012):

- **Exactitud:** Esta medida representa el porcentaje de datos que son etiquetados correctamente por el clasificador, como se aprecia en la ecuación (3.1). Refleja qué tan bien el clasificador reconoce los datos pertenecientes a las distintas clases. Esta medida por sí sola no es suficiente ya que algunos problemas de clasificación son susceptibles a un desequilibrio de clases en la que la cantidad de elementos de una clase es menor que en la otra, por lo que analizar solo esta medida podría dar una falsa percepción de que el clasificador tiene un buen desempeño, cuando en realidad lo que sucede es que una clase representa una mayor cantidad de los datos con los que se está probando el clasificador.

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.1)$$

- **Precisión:** Corresponde a una medida de exactitud, que refleja el porcentaje de datos positivos (irónicos) que fueron etiquetados correctamente. Se calcula según la expresión (3.2) donde se comparan los verdaderos positivos con todos los datos etiquetados como positivos por el clasificador.

$$Precisión = \frac{VP}{VP + FP} \quad (3.2)$$

- Exhaustividad: Es una medida que representa qué tan completa fue la clasificación. Refleja el porcentaje de datos positivos que fueron clasificados como tal. Como se ve en la ecuación (3.3), se obtiene la cantidad que representan los verdaderos positivos respecto al total de datos positivos reales.

$$Exhaustividad = \frac{VP}{VP + FN} = \frac{VP}{P} \quad (3.3)$$

- Medida-F: Esta medida combina la precisión y la exhaustividad en una sola. Se define según la expresión (3.4), donde β es un número real no negativo. Corresponde a la media armónica entre la medida de precisión y exhaustividad. Generalmente se asigna el mismo peso para ambas medidas, definiendo $\beta = 1$.

$$Medida - F = \frac{(1 + \beta^2) * Precisión * Exhaustividad}{\beta^2 * Precisión + Exhaustividad} \quad (3.4)$$

- Medida kappa de Cohen (Fleiss, 1971): se utiliza para evaluar la concordancia entre dos anotadores de un corpus, mediante el cálculo del grado de acuerdo en la clasificación, comparándose con el azar, siendo 1 el máximo valor posible. Se define k de Cohen según la ecuación (3.5), donde p_0 es el acuerdo observado relativo entre los evaluadores, o sea, la suma de acuerdos sobre el total de anotaciones, y p_e es la probabilidad hipotética de acuerdo por azar, utilizando los datos observados para calcular las probabilidades de que cada observador clasifique aleatoriamente cada categoría. Si los evaluadores están completamente de acuerdo, entonces $k = 1$. Si no hay acuerdo entre los evaluadores distinto al que cabría esperar por azar (según lo definido por p_0), $k = 0$. Se considera que valores cercanos o menores a cero implican un acuerdo pobre, mientras que los cercanos a 1 son considerados buenos.

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (3.5)$$

3.3 Descripción de las principales colecciones para el estudio de la ironía

Escoger colecciones textuales con tamaños distintos ayuda a evaluar la efectividad de los rasgos implementados en este trabajo para la detección de la ironía. A continuación, se describirán las colecciones textuales que se utilizarán en la etapa de validación especificando de dónde provienen y la cantidad de opiniones que pertenecen a cada clase, lo cual indicará si los corpus utilizados son balanceados o no, un elemento de interés en esta investigación.

Un corpus es un conjunto estructurado de textos almacenados en forma electrónica y agrupados con el fin de estudiar fenómenos particulares del lenguaje. En el contexto de minería de opinión, minería de texto o de cualquier procedimiento que implique identificar patrones en textos, toma una gran relevancia ya que representa la herramienta a utilizar para identificar los patrones que hacen posible el uso de algoritmos de aprendizaje supervisado (Hernández and Velásquez, 2015).

En esta ocasión, se presentan colecciones en el idioma Inglés provenientes de Twitter, que permiten identificar patrones propios de la ironía utilizando herramientas de minería de textos, procesamiento del lenguaje natural y aprendizaje automatizado.

Para la creación de colecciones de los mensajes de Twitter se consideraron las facilidades que otorga la existencia de hashtags que categorizan los tweets en función de distintos contenidos o tópicos. Además, se crearon colecciones a partir de la descarga de aquellos tweets correspondientes a un conjunto de identificadores de tweets publicados y utilizados en otras investigaciones, lo que facilita la comparación con resultados previamente publicados en la literatura. Para descargar los tweets y consecuentemente construir estas colecciones, fue necesario crear una aplicación en lenguaje Java que permitiera interactuar con la API de Twitter y obtener los tweets a partir de los identificadores o hashtags. La API de Twitter permite al usuario escribir y leer los datos disponibles en esta red social. Para tener acceso desde Java a ésta se utiliza la biblioteca `twitter4j` que permite gestionarla.

La colección de textos ⁴⁷, usada en (Ling and Klinger, 2016), es uno de los corpus que se utiliza en la etapa de validación de esta investigación. Este corpus está formado por

⁴⁷ <http://www.romanklinger.de/ironysarcasm/>

132 000 mensajes de Twitter, incluyendo 33 000 tweets por cada categoría: ironía, sarcasmo y regular, y otros 33 000 que son sarcasmo o ironía. Los tweets de las categorías ironía y sarcasmo fueron seleccionados en (Ling and Klinger, 2016) con los hashtags #irony/#ironic y #sarcasm/#sarcastic. Otros hashtags frecuentes son #drugs, #gopdebate, #late, #news y #peace los cuales fueron usados como términos de búsqueda para reunir el conjunto de tweets regulares junto con los hastags #education, #humor y #politics. Por política de la red social no se permiten publicar los tweets, por lo que suministraron los identificadores de estos y a partir de la aplicación se extrajeron las opiniones. Debido a que algunos tweets por antigüedad no estaban disponibles en la red o los usuarios que escribieron los tweets estaban suspendidos, además, considerando que no se incluyeron aquellos textos que poseían hipervínculos, la colección se redujo a un total de 59 650 tweets, de ellos 17 767 irónicos, 18 727 sarcásticos, 18 129 irónicos o sarcásticos y 5027 regulares. Para la experimentación se tiene esta colección desbalanceada, ya que está formada por 54 623 tweets irónicos o sarcásticos y 5027 regulares. A partir de los 54 623 tweets irónicos o sarcásticos, así como los tweets regulares, se formaron 10 colecciones balanceadas donde se incluyeron 5000 textos irónicos y los 5 000 regulares.

Además, fueron creadas dos colecciones formadas por textos irónicos y no irónicos pertenecientes a estados de Estados Unidos de América. La colección que incluye opiniones de los estados Minnesota, Georgia, Nueva Jersey, Nueva York, y Illinois está formada por 201 textos irónicos y 201 no irónicos. La segunda colección de este tipo también es balanceada y tiene un total de 504 opiniones, estas pertenecen a los estados Arizona, California, Texas, Oklahoma y Luisiana. Se consideraron irónicos los textos que contenían el hashtag #ironic o el hashtag #sarcasm, y no irónicos, los que no tenían estos hashtags. Aunque en esta tesis no se tuvo en cuenta, es importante señalar que algunos autores consideran también como tweets irónicos aquellos que contienen #not, aunque necesariamente no incluyan #ironic y #sarcasm, ya que #not pudiera estar indicando ironía al ser lo contrario de lo que se está expresando. Por tanto, al seleccionar los tweets no irónicos, se quitaron aquellos que contienen #not.

3.4 Diseño y resultado experimental

La etapa de validación demuestra cuán expresivos son los rasgos propuestos para la discriminación entre opiniones irónicas y no irónicas. Para probar la efectividad de dichos rasgos para discriminar entre opiniones irónicas y no irónicas, se utilizarán

clasificadores y consecuentemente se evaluará el desempeño de éstos mediante el análisis de los valores de la matriz de confusión y de las medidas de calidad precisión, exhaustividad, medida-F y la medida kappa. Se utilizarán las colecciones descritas en el epígrafe 3.3.

Tal como fue descrito en el epígrafe 3.1, los clasificadores más utilizados para detectar oraciones irónicas han sido los clasificadores Bayesianos, las máquinas de soporte vectorial, Random Forest y los basados en árboles de decisión; de ahí que serán esos los clasificadores que se utilizarán en esta experimentación.

Para la aplicación de estos clasificadores se utilizarán las implementaciones que aparecen disponibles en Weka⁴⁸. De estos tipos de clasificadores, se utilizarán las implementaciones siguientes: Naïve Bayes, SMO, J48 y Random Forest. Para el clasificador SMO se utilizará un kernel polinomial y una tolerancia de 0.001. Para el algoritmo J48 se usará un factor de confianza de 0.25 para la poda pues valores menores incurren en más podas y se establecerán dos instancias como mínimo por hoja. Para Random Forest no se limita la profundidad de los árboles.

Estos clasificadores fueron aplicados directamente a los conjuntos creados a partir de los rasgos calculados, y también fueron aplicados a los conjuntos de datos después de aplicar el filtro Discretize, pues en ocasiones al aplicar este filtro se obtienen mejores resultados. Luego, se determina con qué clasificador y en qué condiciones se obtuvieron los mejores resultados y con esta misma configuración y el filtro AttributeSelection se realiza otra prueba.

Estos clasificadores generalmente son evaluados realizando validación cruzada y analizando las medidas de calidad precisión, exactitud y medida-F. Precisamente estas fueron las medidas de validación aplicadas, además de utilizar la medida kappa en caso de colecciones desbalanceadas, a partir de una validación cruzada considerando 10 particiones para indicar el conjunto de entrenamiento y el de prueba.

A continuación, se presentan los resultados de la etapa de validación de las dos colecciones de textos irónicos y no irónicos escritos en los diferentes estados de Estados Unidos de América.

⁴⁸ <https://sourceforge.net/projects/weka/>

Al clasificar las opiniones recolectadas del estado de California, Arizona, Texas, Oklahoma y Luisiana el mejor resultado se obtuvo con J48 al aplicar el filtro Discretize pues la exactitud, es decir, el porciento de instancias clasificadas correctamente fue mayor, al igual que la medida-F, lo cual evidencia lo completa que fue la clasificación para ambas clases, la irónica y la no irónica. Los valores observados en la matriz de confusión, de la exhaustividad y precisión fueron buenos. Con el algoritmo Naïve Bayes y el filtro Discretize se obtuvo una mayor cantidad de verdaderos positivos, lo cual se expresa en la exhaustividad. Al aplicar el filtro AttributeSelection y los algoritmos con los que se habían obtenido los mejores resultados, en algunos casos fueron mejores y en otros peores, pero los valores de las medidas de calidad no estuvieron tan distantes.

El mejor resultado en la clasificación de las opiniones recolectadas de los estados de Georgia, Illinois, Minnesota, Nueva Jersey, Nueva York se obtuvo con Random Forest y el filtro Discretize pues la exactitud, es decir, el porciento de instancias clasificadas correctamente y la medida-F fueron mayor en este caso. Además, los valores de las medidas exhaustividad y precisión fueron buenos, al igual que la cantidad de verdaderos positivos. Al aplicar el algoritmo Random Forest con el filtro Discretize y AttributeSelection la clasificación no fue tan acertada.

A continuación se presentan los resultados para la colección desbalanceada formada por 54 623 textos irónicos o sarcásticos y 5027 textos regulares obtenida de (Ling and Klinger, 2016) y las 10 colecciones balanceadas de 5000 textos irónicos y 5 000 regulares adquirida de dividir la colección anterior.

Al realizar la clasificación de las 10 colecciones de 10 000 textos obtenidas de fragmentar la usada en (Ling and Klinger, 2016), el mejor resultado se obtuvo con SMO y el filtro Discretize para todas, pues la exactitud, es decir, el porciento de instancias clasificadas correctamente, y la medida-F, fueron mayor en estos casos. Los valores de las medidas exhaustividad, precisión y de la matriz de confusión fueron buenos. Además, con el clasificador SMO sin utilizar ningún filtro fue con el que se clasificó la mayor cantidad de verdaderos positivos para las 10 colecciones, esto se ve reflejado en el valor de la exhaustividad. Al aplicar el algoritmo SMO con el filtro Discretize y AttributeSelection para todas ellas la clasificación no fue tan buena, al obtenerse valores menores de las medidas de calidad, pero no están tan alejados de los obtenidos al aplicar

SMO solamente con el filtro Discretize, y por tanto, no hay tanta diferencia con el mejor resultado.

Al aplicar varios algoritmos para la colección desbalanceada obtenida a partir de las opiniones usadas en (Ling and Klinger, 2016) se obtuvieron buenos resultados y bastantes cercanos. El mejor se alcanzó con Random Forest sin aplicar filtros pues en este caso se obtuvo una menor cantidad de fallos al clasificar las opiniones de la clase irónica, y por tanto, una mayor exhaustividad. Además, la exactitud y medida-F fueron mayores. Se obtuvo un buen valor de precisión y la medida kappa tiene un valor cercano a uno. Al aplicar el filtro AttributeSelection los valores de estas medidas de calidad fueron menores, pero no están tan alejados de los mejores resultados obtenidos.

Al aplicar el clasificador J48 y el filtro AttributeSelection para todas las colecciones los rasgos de mayor ganancia fueron *ContrastePalabMayTexto*, *AntecDelimitadorConsec*, *syno_greater_gap*, *ContrastePresentePasado*, *RazonPolaritySkipgrams*, *PromPolFrasesComillas*, *ContrasteEmotTexto*, *RazonInterjec*. Al utilizar J48 sin filtro AttributeSelection se observó que los rasgos de mayor ganancia fueron *ContrastePalabMayTexto*, *ContrastePalabrasExtremas*, *AntecDelimitadorConsec*, *DesviaciónPolaridades*, *syno_lower_mean*, *ContrasteEmotTexto*, *syno_lower_gap*, *syno_greater_mean*, *ContrastePresentePasado*, *Ambigüedad*, *RazonPolaritySkipgrams*, *RazonDominios*. De ellos seis fueron tomados de la literatura y el resto creados en esta investigación.

3.5 Conclusiones parciales

Los modelos creados hasta el momento para la detección de la ironía de forma automática no caracterizan completamente este fenómeno, pues los mejores resultados alcanzados hasta el momento están entre un 70% y 90% de efectividad. Estos trabajos dirigidos a la clasificación de opiniones en irónica o no, se han realizado para el idioma Inglés, Español, Catalán, Portugués, Italiano, entre otros; y para textos principalmente de Twitter, aunque también se han elaborado para las opiniones emitidas sobre los productos en Amazon. Los algoritmos más exitosos utilizados para la clasificación de las opiniones son los probabilísticos, las máquinas de soporte vectorial, Random Forest y los basados en árboles de decisión. Estos clasificadores generalmente fueron evaluados realizando validación cruzada y analizando las medidas de calidad precisión, exactitud y medida-F.

La evaluación de la efectividad de los rasgos propuestos para la distinción entre opiniones irónicas y no irónicas en esta investigación se basó en el análisis de los valores de la matriz de confusión y de las medidas de calidad: precisión, exhaustividad y medida-F. En el caso de las colecciones desbalanceadas, se incorporó la medida de validación basada en el estadígrafo Kappa. Estos valores dan la medida de cuán completa fue la clasificación de las opiniones, y principalmente, de aquellas que pertenecen a la clase irónica, la de mayor interés. Los algoritmos usados para la clasificación de los textos en irónicos o no fueron Naïve Bayes, Random Forest, J48 y SMO, sin aplicar filtros y utilizando el filtro Discretize y AttributeSelection, considerando una validación cruzada con 10 particiones.

Las colecciones que se utilizaron en la etapa de validación están conformadas por opiniones en idioma Inglés, tienen diversos tamaños, algunas son balanceadas y otras no, y todas provienen de Twitter. Tener en cuenta esta diversidad en la conformación de los conjuntos de entrenamiento y prueba, permitió demostrar la efectividad de los rasgos en la detección de ironía para textos cortos.

Al clasificar las dos colecciones correspondientes a opiniones en Twitter de 10 estados de Estados Unidos de América, los mejores resultados se alcanzaron aplicando J48 para la colección balanceada con un total de 504 opiniones y Random Forest para la formada por 201 textos irónicos y 201 no irónicos. El mejor valor de exactitud, precisión, exhaustividad y medida-F para la primera fue de 88.4921%, 0.962, 0.802 y 0.874, respectivamente. Los valores alcanzados para la segunda colección fueron 86.8159%, 0.878, 0.856 y 0.866, respectivamente. Estas dos colecciones son muy pequeñas, por tanto, se cuenta con pocos ejemplos de entrenamiento, lo que repercute negativamente en la calidad de la clasificación.

Al aplicar los algoritmos de clasificación supervisados para las 10 colecciones de opiniones de Twitter obtenidas de las usadas en (Ling and Klinger, 2016) formadas por 5 000 textos irónicos y 5 000 textos no irónicos, el que brindó siempre el mejor resultado fue SMO y los mejores valores de exactitud, precisión, exhaustividad y medida-F oscilan entre 92.26% y 92.87%, 0.896 y 0.927, 0.924 y 0.958, 0.923 y 0.93, respectivamente.

Al clasificar las opiniones de la colección desbalanceada formada por 54 623 textos irónicos o sarcásticos y 5027 textos regulares de Twitter, obtenidos a partir de la

colección usada en (Ling and Klinger, 2016), Random Forest fue el ganador con valores de exactitud, precisión, exhaustividad, medida-F y medida kappa de 97.84%, 0.980, 0.996 y 0.988 y 0.848, respectivamente.

Al aplicar el filtro AttributeSelection los valores de las medidas de calidad disminuyeron, pero no estuvieron tan alejados de los mejores resultados obtenidos para cada colección.

Los rasgos de mayor ganancia fueron: *ContrastePalabMayTexto*, *AntecDelimitadorConsec*, *syno_greater_gap*, *ContrastePresentePasado*, *RazonPolaritySkipgrams*, *PromPolFrasesComillas*, *ContrasteEmotTexto*, *RazonInterjec*, *ContrastePalabrasExtremas*, *DesviaciónPolaridades*, *syno_lower_mean*, *syno_lower_gap*, *syno_greater_mean*, *Ambigüedad* y *RazonDominios*. De ellos, seis fueron tomados de la literatura y el resto creados en esta investigación, lo que evidencia el poder discriminante entre clases que tienen los rasgos creados y, por tanto, su capacidad para caracterizar la ironía. El rasgo *ContrastePalabMayTexto* fue el de mayor ganancia para todas las colecciones.

Conclusiones

Como resultado de esta investigación se diseñó un conjunto de rasgos para la detección supervisada de la ironía, obteniéndose valores promedios de exactitud, precisión, exhaustividad y medida-F de 92.19%, 0.920, 0.926 y 0.922, respectivamente, siendo estos superiores a los referidos en la literatura, cumpliéndose de esta forma el objetivo general propuesto, ya que:

- Existen rasgos reportados en la literatura que no caracterizan adecuadamente los textos irónicos, entre ellos los que realizan un análisis parcial del patrón que se quiere representar o aquellos que son muy específicos; lo cual limita el poder discriminante entre opiniones irónicas y no irónicas, reflejado en los resultados alcanzados por los principales clasificadores utilizados al detectar ironía: clasificadores bayesianos, Random Forest, máquinas de soporte vectorial y aquellos basados en árboles de decisión. No obstante, existen rasgos reportados en la literatura que resultan muy interesantes para caracterizar opiniones irónicas ya que consideran el contraste en el sentimiento expresado, la estructura sintáctica de las opiniones y la selección del vocabulario utilizado.
- Los rasgos creados a partir de la búsqueda de patrones específicos y generales, el desbalance, el contraste, la razón de determinados elementos con respecto al texto y el cálculo de la polaridad de frases entre comillas; así como aquellos seleccionados de la literatura que cuantifican cuán ambiguo es un texto, miden el uso de sinónimos comunes y raros, calculan la desviación estándar de las polaridades de las palabras y buscan una estructura fija basada en la presencia de interjecciones, permitieron discriminar entre opiniones irónicas y no irónicas, así lo evidenciaron los resultados de los clasificadores y los atributos ganadores al construir los árboles de decisión.
- Las clases EntradaSalida, Rasgos, Recursos, Lucene, SentiWordNet, WordNetDomains y Diccionario conforman el módulo diseñado e implementado para representar las opiniones, partiendo de opiniones irónicas y no irónicas hasta obtener los valores de los rasgos que logran discriminar entre este tipo de opiniones en un archivo con formato de entrada al Weka.

- Los algoritmos de aprendizaje supervisado J48, Random Forest y SMO reportaron un buen rendimiento al clasificar opiniones en irónicas o no considerando los rasgos seleccionados y creados; destacándose el clasificador Random Forest sin utilizar filtros que permitió alcanzar valores de exactitud, precisión, exhaustividad, medida-F y medida kappa de 97.84%, 0.980, 0.996, 0.988 y 0.848, respectivamente, al ejecutarlo con la colección reportada en (Ling and Klinger, 2016).

Recomendaciones

Derivadas del estudio realizado, así como de las conclusiones generales emanadas del mismo, se recomienda:

- Añadir otras tareas de pre-procesamiento como la corrección ortográfica, la desambiguación del sentido de las palabras y el tratamiento de las jergas que pudieran tributar a una mayor calidad de los resultados obtenidos.
- Crear nuevos rasgos que consideren otros elementos que no se incluyeron en la presente propuesta, como por ejemplo el escenario emocional, la detección de acrónimos y otras variantes para el cálculo de la polaridad.
- Integrar el módulo realizado en esta investigación a la herramienta PosNeg Opinion de forma tal que se calcule la polaridad de las opiniones tratando el problema de la ironía.

Referencias bibliográficas

Amores, M. A. (2013) *Detección no supervisada de la polaridad de las opiniones*. Universidad Central ‘Marta Abreu’ de Las Villas.

Amores, M. A. (2016) *Detección de la polaridad de las opiniones basada en nuevos recursos léxicos*. Universidad Central ‘Marta Abreu’ de Las Villas.

Amores, M., Arco, L. and Artiles, M. (2015) ‘PosNeg opinion : Una herramienta para gestionar comentarios de la Web’, *Revista Cubana de Ciencias Informáticas*, 9(1), pp. 20–31. Available at: <http://rcci.uci.cu>.

Amores, M., Arco, L. and Barrera, A. (2016) ‘Efectos de la Negación, Modificadores, Jergas, Abreviaturas y Emoticonos en el Análisis de Sentimiento’, in *Proceedings of the 2nd International Workshop on Semantic Web (IWSW)*. La Habana: CEUR. Available at: <http://ceur-ws.org/Vol-1797>.

Amores, M., Arco, L. and Borroto, C. (2016) ‘Unsupervised Opinion Polarity Detection based on New Lexical Resources’, *Computación y Sistemas*, 20(2), pp. 263–277. doi: 10.13053/CyS-20-2-2318.

Atserias, J., Casas, B., Comelles, E., González, M., Padró, L. and Padró, M. (2006) ‘FreeLing 1.3: Syntactic and semantic services in an open-source NLP library’, in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy: ELRA. Available at: <http://www.lsi.upc.edu/~nlp/freeling>.

Baca-Gómez, Y. R., Castro-Sánchez, N. A., Martínez, A., Hernández, D. I. and Rosso, P. (2014) ‘Impacto de la ironía en la minería de opiniones basada en un Léxico Afectivo’, pp. 41–44. Available at: <http://ceur-ws.org/Vol-1199/>.

Baccianella, S., Esuli, A. and Sebastiani, F. (2010) ‘SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining’, in *LREC*, pp. 2200–2204.

Bamman, D. and Smith, N. A. (2015) ‘Contextualized Sarcasm Detection on Twitter’, in Cha, M., Mascolo, C., and Sandvig, C. (eds) *ICWSM*. AAAI Press, pp. 574–577.

Available at: <http://dblp.uni-trier.de/db/conf/icwsm/icwsm2015.html#BammanS15>.

Barbieri, F., Ronzano, F. and Saggion, H. (2014) ‘Italian Irony Detection in Twitter: a First Approach’, in *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014*, p. 28.

Barbieri, F. and Saggion, H. (2014) ‘Automatic Detection of Irony and Humor in Twitter’, in *Proceedings of the Fifth International Conference on Computational Creativity*. Ljubljana, Slovenia: Josef Stefan Institute, Ljubljana, Slovenia. Available at: http://computationalcreativity.net/iccc2014/wp-content/uploads/2014/06/9.2_Barbieri.pdf.

Bentivogli, L., Forner, P., Magnini, B. and Pianta, E. (2004) ‘Revising the Wordnet Domains Hierarchy: Semantics, Coverage and Balancing’, in *Proceedings of the Workshop on Multilingual Linguistic Resources*. Association for Computational Linguistics, pp. 101–108.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) ‘Latent Dirichlet Allocation’, *J. Mach. Learn. Res.*, 3, pp. 993–1022.

Brants, T. and Franz, A. (2006) ‘Web 1T 5-gram Version 1’, *Linguistic Data Consortium*.

Buschmeier, K., Cimiano, P. and Klinger, R. (2014) ‘An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews’, in *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Baltimore, Maryland: Association for Computational Linguistics, pp. 42–49. Available at: <http://www.aclweb.org/anthology/W14-2608>.

Carvalho, P., Silva, M. J., Sarmento, L. and de Oliveira, E. (2009) ‘Clues for Detecting Irony in User-Generated Contents : Oh ...!! It " s “ so easy " ; -)”’, in *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pp. 53–56. doi: 10.1145/1651461.1651471.

Casado, Á. J. and Cantador, I. (2013) *Sistema de extracción de entidades y análisis de opiniones en contenidos Web generados por usuarios*. Universidad Autónoma de Madrid.

Clark, H. and Gerrig, R. (1984) 'On the pretense theory of irony', *Journal of Experimental Psychology: General*, 113(1), pp. 121–126. doi: 10.1037/0096-3445.113.1.121.

Davidov, D., Tsur, O. and Rappoport, A. (2010) 'Semi-supervised recognition of sarcastic sentences in Twitter and Amazon', in *CoNLL '10 Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pp. 107–116.

Espinoza, P., Vilariño, D., Tovar, M. and Beatriz, B. (2014) 'Detección del lenguaje figurativo e ironía en textos cortos', *Research in Computing Science*, 85, pp. 85–92.

Fellbaum, C. (1998) *WordNet*. Wiley Online Library.

Fernández, J., Gómez, J. M., Boldrini, E. and Martínez-Barco, P. (2011) 'Análisis de Sentimientos y Minería de Opiniones: el corpus EmotiBlog', *Procesamiento del Lenguaje Natural*, 47, pp. 179–187.

Filatova, E. (2012) 'Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing', in Calzolari, N., Choukri, K., Declerck, T., Uğur Doğan, M., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Stelios (eds) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA), pp. 392–398.

Fleiss, J. L. and others (1971) 'Measuring nominal scale agreement among many raters', *Psychological Bulletin*, 76(5), pp. 378–382.

Flesch, R. (1948) 'A new readability yardstick', *Journal of Applied Psychology*, 32(3), pp. 221–233. Available at: <http://libezproxy.open.ac.uk/login?url=http://search.ebscohost.com.libezproxy.open.ac.uk/login.aspx?direct=true&db=pdh&AN=apl-32-3-221&site=ehost-live&scope=site>.

Forslid, E. and Wikén, N. (2015) *Automatic irony- and sarcasm detection in Social media*. Uppsala Universitet.

Gibbs, R. and Colston, H. (2007) *Irony in Language and Thought: A Cognitive Science Reader*. Psychology Press.

Günther, T. and Furrer, L. (2013) 'GU-MLT-LT: Sentiment Analysis of Short Messages using Linguistic Features and Stochastic Gradient Descent', in *Proceedings of the*

seventh international workshop on Semantic Evaluation Exercises (SemEval-2013). Atlanta, Georgia, USA.

Han, J. and Kamber, M. (2012) *Data Mining: Concepts and Techniques*. 3ra edn. Morgan Kaufmann Publishers.

Hangya, V., Berend, G. and Farkas, R. (2013) ‘SZTE-NLP: Sentiment Detection on Twitter Messages’, in Diab, M. T., Baldwin, T., and Baroni, M. (eds) *Proceedings of the 7th international workshop on Semantic Evaluation (SemEval-2013)*. Atlanta, Georgia, USA: Association for Computational Linguistics, pp. 549–553.

Henriquez, C., Guzmán, J. and Salcedo, D. (2016) ‘Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles’, *Procesamiento de Lenguaje Natural*, 56, pp. 25–32.

Hernández, D. I., Sulis, E., Patti, V., Ruffo, G. and Bosco, C. (2015) ‘ValenTo : Sentiment Analysis of Figurative Language Tweets with Irony and Sarcasm’, in *Proceedings of the 9th International Workshop on Semantic Evaluation*. Denver, Colorado, USA, pp. 694–698. Available at: <http://aclweb.org/anthology/S/S15/S15-2117.pdf>.

Hernández, V. A. and Velásquez, J. D. (2015) *Identificación de la presencia de ironía en el texto generado por usuarios de Twitter utilizando técnicas de opinion mining y machine learning*. Universidad de Chile.

Hu, M. and Liu, B. (2004) ‘Mining and summarizing customer reviews’, in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. USA: ACM, pp. 168–177. doi: 10.1145/1014052.1014073.

Ingle, A., Maheshwari, N., Sutrave, N., Akumarthi, S., Bhitre, T. and Keskar, R. B. (2014) *Sentiment analysis: sarcasm detection of tweets*. Visvesvaraya National Institute of Technology.

Jasso, G. (2015) ‘Hacia la detección de ironía en textos cortos’.

Kincaid, J. P., Fishburne, R. P. J., Rogers, R. L. and Chissom, B. S. (1975) ‘Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel’, in *Technical report, DTIC Document*.

Kouloumpis, E., Wilson, T. and Moore, J. D. (2011) ‘Twitter Sentiment Analysis: The Good the Bad and the OMG!’, in Adamic, L. A., Baeza-Yates, R. A., and Counts, S. (eds) *ICWSM*. The AAAI Press. Available at: <http://dblp.uni-trier.de/db/conf/icwsml/icwsml2011.html#KouloumpisWM11>.

Levallois, C. (2013) ‘Umigon: sentiment analysis for tweets based on terms lists and heuristics’, in *Second Joint Conference on Lexical and Computational Semantics (*SEM)*. Atlanta, Georgia, USA: Association for Computational Linguistics. Available at: <http://www.aclweb.org/anthology/S13-2068>.

Ling, J. and Klinger, R. (2016) ‘An Empirical , Quantitative Analysis of the Differences between Sarcasm and Irony’, in Sack, H., Rizzo, G., Steinmetz, N., Mladenicić, D., Auer, S., and Lange, C. (eds) *The Semantic Web*. Springer International Publishing, pp. 203–216. doi: 10.1007/978-3-319-47602-5_39.

Liu, B. (2012) *Sentiment Analysis and Opinion Mining*.

Lü, X., Zhang, L. and Hu, J. (2004) ‘Statistical Substring Reduction in Linear Time’, in Su, K.-Y., Tsujii, J., Lee, J.-H., and Kwong, O. Y. (eds) *Natural Language Processing-IJCNLP 2004*. Springer Berlin Heidelberg, pp. 320–327. doi: 10.1007/978-3-540-30211-7_34.

Lucariello, J. (1994) ‘Situational irony: A concept of events gone awry’, *Journal of Experimental Psychology: General*, 123(2), pp. 129–145.

Maynard, D. and Greenwood, M. A. (2014) ‘Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis’, in *Proceedings of the Ninth International Conference on Language Resources and Evaluation, {LREC} 2014*. Reykjavik, Iceland, pp. 4238–4243. Available at: <http://www.lrec-conf.org/proceedings/lrec2014/summaries/67.html>.

McLaughlin, H. G. (1969) ‘SMOG grading - a new readability formula’, *Journal of Reading*, 12(8), pp. 639–646.

Miller, G. A. (1995) ‘WordNet: A Lexical Database for English’, *Communications of the ACM*, 38(11), pp. 39–41. doi: 10.1145/219717.219748.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. J. (1990)

‘Introduction to wordnet: An on-line lexical database*’, *International journal of lexicography*. Oxford Univ Press, 3(4), pp. 235–244.

Mohammad, S. M., Kiritchenko, S. and Zhu, X. (2013) ‘NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets’, in *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*. Atlanta, USA.

Na, J.-C., Khoo, C. and Wu, P. H. J. (2005) ‘Use of negation phrases in automatic sentiment classification of product reviews’, *Library Collections, Acquisitions, & Technical Services*, 29(2), pp. 180–191.

Owoputi, O., Dyer, C., Gimpel, K., Schneider, N. and Smith, N. A. (2013) ‘Improved part-of-speech tagging for online conversational text with word clusters’, in *Proceedings of NAACL*.

Potts, C. (2011) ‘Developing adjective scales from user-supplied textual metadata’, *NSF Workshop on Restructuring Adjectives in WordNet*.

Ptáček, T., Habernal, I. and Hong, J. (2014) ‘Sarcasm Detection on Czech and English Twitter’, in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, Ireland, pp. 213–223.

Rajadesingan, A., Zafarani, R. and Liu, H. (2015) ‘Sarcasm Detection on Twitter : A Behavioral Modeling Approach’, in *8th ACM International Conference on Web Search and Data Mining, WSDM 2015*. Shanghai, China: Association for Computing Machinery, Inc, pp. 97–106. doi: 10.1145/2684822.2685316.

Rella, J., Saggion, H. and Barbieri, F. (2015) *TwIrony: Identificación de la ironía en Tweets en Catalán*. Universitat Pompeu Fabra. Available at: <https://repositori.upf.edu>.

Reyes, A. and Rosso, P. (2011) ‘Mining Subjective Knowledge from Customer Reviews : A Specific Case of Irony Detection’, in *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 118–124.

Reyes, A. and Rosso, P. (2012) ‘Making objective decisions from subjective data: Detecting irony in customer reviews’, *Decision Support Systems*, 53(4), pp. 754–760. doi: 10.1016/j.dss.2012.05.027.

Reyes, A., Rosso, P. and Buscaldi, D. (2012) 'From humor recognition to irony detection: The figurative language of social media', *Data & Knowledge Engineering*. Elsevier B.V., 74(0), pp. 1–12. doi: 10.1016/j.datak.2012.02.005.

Reyes, A., Rosso, P. and Veale, T. (2013) 'A multidimensional approach for detecting irony in Twitter', *Language Resources and Evaluation*, 47(1), pp. 239–268. doi: 10.1007/s10579-012-9196-x.

Saif, M., Dorr, B. and Cody, D. (2009) 'Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus', in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*. Singapore.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, Christopher D. Ng, A. and Potts, C. (2013) 'Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank', in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 1631--1642. Available at: <http://www.aclweb.org/anthology/D13-1170>.

Strapparava, C. and Valitutti, A. (2004) 'WordNet-Affect: An affective extension of WordNet', in *Proceedings of the 4th International Conference on Language Resources and Evaluation*. ELRA, pp. 1083–1086.

Strapparava, C., Valitutti, A. and others (2004) 'WordNet Affect: an Affective Extension of WordNet.', in *LREC*, pp. 1083–1086.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. and Kappas, A. (2010) 'Sentiment in Short Strength Detection Informal Text', *J. Am. Soc. Inf. Sci. Technol.*, 61(12), pp. 2544--2558. doi: 10.1002/asi.v61:12.

Vanin, A. A., Freitas, L. A., Vieira, R. and Bochernitsan, M. N. (2013) 'Some Clues on Irony Detection in Tweets', in *Proceedings of the 22nd International Conference on World Wide Web*, pp. 635–636. doi: 10.1145/2487788.2488012.

Wang, W., Chen, L., Thirunarayan, K. and Sheth, A. P. (2014) 'Cursing in English on twitter', in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work*. ACM, pp. 415–425. doi: 10.1145/2531602.2531734.

Warriner, A. B., Kuperman, V. and Brysbaert, M. (2013) 'Norms of valence, arousal, and dominance for 13,915 English lemmas', *Behavior Research Methods*, 45(4), pp. 1191–1207. doi: 10.3758/s13428-012-0314-x.

Whissell, C. (1989) 'The Dictionary of Affect in Language', in Plutchik, R. and Kellerman, H. (eds) *Emotion: theory, research and experience*. London: Academic Press.

Wilson, D. and Sperber, D. (1992) 'On Verbal Irony', *Lingua*, 87(1), pp. 53–76. doi: 10.1016/0024-3841(92)90025-E.

Anexos

Anexo 1. Glosario de términos

- p_i palabra i
- $polaridad(p_i)$ polaridad de la palabra i
- $polaridad(p_{imax})$ máxima polaridad del texto que se encuentra en la posición $imax$
- $polaridad(p_{imin})$ mínima polaridad del texto que se encuentra en la posición $imin$
- $polaridad(p_{extremo})$ es la polaridad global que se encuentra en alguna palabra de la frase en la posición $extremo$ (puede ser el mínimo o el máximo)
- $n_{sentidos(p_i)}$ cantidad de sentidos o significados de la palabra i
- n total de palabras del texto
- n_{POS} total de palabras del texto que son sustantivos, adjetivos, verbos y adverbios
- $n_{f_j POS}$ es la cantidad de palabras que son sustantivos, adjetivos o adverbios en la frase f_j
- $max_{sentidos}$ es la mayor cantidad de significados que tiene una palabra en la opinión
- n_f total de frases que están entre comillas
- n_j cantidad de palabras que forman la frase f_j que está entre comillas
- $\overline{polaridadC}$ promedio de las polaridades de las palabras que conforman la frase con una condición C.
- $\overline{polaridad}$ media de las polaridades de las palabras del texto
- $\overline{polaridadE}$ promedio de la polaridad de los emoticonos en el texto
- $\overline{polaridadI}$ promedio de la polaridad de las interjecciones del texto
- $\overline{polaridadM}$ promedio de la polaridad de las palabras en mayúsculas del texto

Anexo 2. Resultados de la clasificación de las opiniones de la colección proveniente de los estados California, Arizona, Texas, Oklahoma y Luisiana

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	76.98	a b <-- classified as 206 46 a = ironico 70 182 b = no_ironico	0.746	0.817	0.780
	Random Forest	85.32	a b <-- classified as 216 36 a = ironico 38 214 b = no_ironico	0.850	0.857	0.854
	J48	85.52	a b <-- classified as 208 44 a = ironico 29 223 b = no_ironico	0.878	0.825	0.851
	SMO	78.37	a b <-- classified as 220 32 a = ironico 77 175 b = no_ironico	0.741	0.873	0.801
Se aplicó filtro Discretize	Naïve Bayes	84.52	a b <-- classified as 227 25 a = ironico 53 199 b = no_ironico	0.811	0.901	0.853
	Random Forest	85.52	a b <-- classified as 203 49 a = ironico 24 228 b = no_ironico	0.894	0.806	0.848
	J48	88.49	a b <-- classified as 202 50 a = ironico 8 244 b = no_ironico	0.962	0.802	0.874
	SMO	85.91	a b <-- classified as 182 70 a = ironico 1 251 b = no_ironico	0.995	0.722	0.837
Se aplicó filtro Discretize y AttributeSelection	J48	87.50	a b <-- classified as 209 43 a = ironico 20 232 b = no_ironico	0.913	0.829	0.869
	Naïve Bayes	87.50	a b <-- classified as 209 43 a = ironico 20 232 b = no_ironico	0.913	0.829	0.869
Se aplicó filtro AttributeSelection	Random Forest	84.92	a b <-- classified as 209 43 a = ironico 33 219 b = no_ironico	0.864	0.829	0.846

Anexo 3. Resultados de la clasificación de las opiniones de la colección proveniente de los estados Georgia, Illinois, Minnesota, Nueva Jersey, Nueva York

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	71.89	a b <-- classified as 157 44 a = ironico 69 132 b = no_ironico	0.695	0.781	0.735
	Random Forest	83.09	a b <-- classified as 166 35 a = ironico 33 168 b = no_ironico	0.834	0.826	0.830
	J48	81.84	a b <-- classified as 151 50 a = ironico 23 178 b = no_ironico	0.868	0.751	0.805
	SMO	75.37	a b <-- classified as 175 26 a = ironico 73 128 b = no_ironico	0.706	0.871	0.780
Se aplicó filtro Discretize	Naïve Bayes	84.33	a b <-- classified as 172 29 a = ironico 34 167 b = no_ironico	0.835	0.856	0.845
	Random Forest	86.82	a b <-- classified as 172 29 a = ironico 24 177 b = no_ironico	0.878	0.856	0.866
	J48	85.57	a b <-- classified as 168 33 a = ironico 25 176 b = no_ironico	0.87	0.836	0.853
	SMO	84.58	a b <-- classified as 141 60 a = ironico 2 199 b = no_ironico	0.986	0.701	0.82
Se aplicó filtro Discretize y AttributeSelection	Random Forest	83.33	a b <-- classified as 135 66 a = ironico 1 200 b = no_ironico	0.993	0.672	0.801

Anexo 4. Resultados de la clasificación de opiniones de la colección 1 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	76.37	a b <-- classified as 3894 1106 a = ironico 1257 3743 b= no_ironico	0.756	0.779	0.767
	Random Forest	91.73	a b <-- classified as 4717 283 a = ironico 544 4456 b = no_ironico	0.897	0.943	0.919
	J48	91.01	a b <-- classified as 4571 429 a = ironico 470 4530 b = no_ironico	0.907	0.914	0.91
	SMO	81.75	a b <-- classified as 4920 80 a = ironico 1745 3255 b = no_ironico	0.738	0.984	0.844
Se aplicó filtro Discretize	Naïve Bayes	90.81	a b <-- classified as 4753 247 a = ironico 672 4328 b = no_ironico	0.876	0.951	0.912
	Random Forest	92.15	a b <-- classified as 4622 378 a = ironico 407 4593 b = no_ironico	0.919	0.924	0.922
	J48	91.99	a b <-- classified as 4585 415 a = ironico 386 4614 b = no_ironico	0.922	0.917	0.92
	SMO	92.40	a b <-- classified as 4620 380 a = ironico 380 4620 b = no_ironico	0.924	0.924	0.924
Se aplicó filtro Discretize y AttributeSelection	SMO	90.03	a b <-- classified as 4310 690 a = ironico 307 4693 b = no_ironico	0.934	0.862	0.896

Anexo 5. Resultados de la clasificación de opiniones de la colección 2 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	77.99	a b <-- classified as 4095 905 a = ironico 1296 3704 b = no_ironico	0.76	0.819	0.788
	Random Forest	91.89	a b <-- classified as 4710 290 a = ironico 521 4479 b = no_ironico	0.9	0.942	0.921
	J48	91.35	a b <-- classified as 4628 372 a = ironico 493 4507 b = no_ironico	0.904	0.926	0.915
	SMO	81.98	a b <-- classified as 4944 56 a = ironico 1746 3254 b = no_ironico	0.739	0.989	0.846
Se aplicó filtro Discretize	Naïve Bayes	91.15	a b <-- classified as 4761 239 a = ironico 646 4354 b = no_ironico	0.881	0.952	0.915
	Random Forest	92.03	a b <-- classified as 4618 382 a = ironico 415 4585 b = no_ironico	0.918	0.924	0.921
	J48	92.17	a b <-- classified as 4636 364 a = ironico 419 4581 b = no_ironico	0.917	0.927	0.922
	SMO	92.86	a b <-- classified as 4713 287 a = ironico 427 4573 b = no_ironico	0.917	0.943	0.93
Se aplicó filtro Discretize y AttributeSelection	SMO	89.88	a b <-- classified as 4359 641 a = ironico 371 4629 b = no_ironico	0.922	0.872	0.896

Anexo 6. Resultados de la clasificación de opiniones de la colección 3 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	76.70	a b <-- classified as 3897 1103 a = ironico 1227 3773 b = no_ironico	0.761	0.779	0.770
	Random Forest	92.12	a b <-- classified as 4722 278 a = ironico 510 4490 b = no_ironico	0.903	0.944	0.923
	J48	91.28	a b <-- classified as 4613 387 a = ironico 485 4515 b = no_ironico	0.905	0.923	0.914
	SMO	82.09	a b <-- classified as 4954 46 a = ironico 1745 3255 b = no_ironico	0.740	0.991	0.847
Se aplicó filtro Discretize	Naïve Bayes	91.34	a b <-- classified as 4752 248 a = ironico 618 4382 b = no_ironico	0.885	0.950	0.916
	Random Forest	92.10	a b <-- classified as 4600 400 a = ironico 390 4610 b = no_ironico	0.922	0.92	0.921
	J48	91.46	a b <-- classified as 4577 423 a = ironico 431 4569 b = no_ironico	0.914	0.915	0.915
	SMO	92.68	a b <-- classified as 4631 369 a = ironico 363 4637 b = no_ironico	0.927	0.926	0.927
Se aplicó filtro Discretize y AttributeSelection	SMO	90.50	a b <-- classified as 4347 653 a = ironico 297 4703 b = no_ironico	0.936	0.869	0.901

Anexo 7. Resultados de la clasificación de opiniones de la colección 4 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	77.02	a b <-- classified as 4003 997 a = ironico 1301 3699 b = no_ironico	0.755	0.801	0.777
	Random Forest	91.34	a b <-- classified as 4715 285 a = ironico 581 4419 b = no_ironico	0.890	0.943	0.916
	J48	90.97	a b <-- classified as 4640 360 a = ironico 543 4457 b = no_ironico	0.895	0.928	0.911
	SMO	81.92	a b <-- classified as 4937 63 a = ironico 1745 3255 b = no_ironico	0.739	0.987	0.845
Se aplicó filtro Discretize	Naïve Bayes	91.21	a b <-- classified as 4776 224 a = ironico 655 4345 b = no_ironico	0.879	0.955	0.916
	Random Forest	91.94	a b <-- classified as 4622 378 a = ironico 428 4572 b = no_ironico	0.915	0.924	0.920
	J48	91.88	a b <-- classified as 4678 322 a = ironico 490 4510 b = no_ironico	0.905	0.936	0.920
	SMO	92.55	a b <-- classified as 4673 327 a = ironico 418 4582 b = no_ironico	0.918	0.935	0.926
Se aplicó filtro Discretize y AttributeSelection	SMO	89.53	a b <-- classified as 4312 688 a = ironico 359 4641 b = no_ironico	0.923	0.862	0.892

Anexo 8. Resultados de la clasificación de opiniones de la colección 5 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	78.71	a b <-- classified as 4292 708 a = ironico 1421 3579 b = no_ironico	0.751	0.858	0.801
	Random Forest	91.77	a b <-- classified as 4760 240 a = ironico 583 4417 b = no_ironico	0.891	0.952	0.92
	J48	91.04	a b <-- classified as 4646 354 a = ironico 542 4458 b = no_ironico	0.896	0.929	0.912
	SMO	82.02	a b <-- classified as 4948 52 a = ironico 1746 3254 b = no_ironico	0.739	0.990	0.846
Se aplicó filtro Discretize	Naïve Bayes	90.71	a b <-- classified as 4814 186 a = ironico 743 4257 b = no_ironico	0.866	0.963	0.912
	Random Forest	91.53	a b <-- classified as 4609 391 a = ironico 456 4544 b = no_ironico	0.91	0.922	0.916
	J48	91.42	a b <-- classified as 4645 355 a = ironico 503 4497 b = no_ironico	0.902	0.929	0.915
	SMO	92.45	a b <-- classified as 4683 317 a = ironico 438 4562 b = no_ironico	0.914	0.937	0.925
Se aplicó filtro Discretize y AttributeSelection	SMO	89.19	a b <-- classified as 4565 435 a = ironico 646 4354 b = no_ironico	0.876	0.913	0.894

Anexo 9. Resultados de la clasificación de opiniones de la colección 6 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	78.08	a b <-- classified as 4172 828 a = ironico 1364 3636 b = no_ironico	0.754	0.834	0.792
	Random Forest	91.78	a b <-- classified as 4745 255 a = ironico 567 4433 b = no_ironico	0.893	0.949	0.920
	J48	91.11	a b <-- classified as 4614 386 a = ironico 503 4497 b = no_ironico	0.902	0.923	0.912
	SMO	81.88	a b <-- classified as 4934 66 a = ironico 1746 3254 b = no_ironico	0.739	0.987	0.845
Se aplicó filtro Discretize	Naïve Bayes	90.65	a b <-- classified as 4801 199 a = ironico 736 4264 b = no_ironico	0.867	0.960	0.911
	Random Forest	91.85	a b <-- classified as 4604 396 a = ironico 419 4581 b = no_ironico	0.917	0.921	0.919
	J48	91.35	a b <-- classified as 4658 342 a = ironico 523 4477 b = no_ironico	0.899	0.932	0.915
	SMO	92.26	a b <-- classified as 4646 354 a = ironico 420 4580 b = no_ironico	0.917	0.929	0.923
Se aplicó filtro Discretize y AttributeSelection	SMO	88.49	a b <-- classified as 4713 287 a = ironico 864 4136 b = no_ironico	0.845	0.943	0.891

Anexo 10. Resultados de la clasificación de opiniones de la colección 7 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	77.7	a b <-- classified as 4115 885 a = ironico 1345 3655 b = no_ironico	0.754	0.823	0.787
	Random Forest	91.43	a b <-- classified as 4738 262 a = ironico 595 4405 b = no_ironico	0.888	0.948	0.917
	J48	90.97	a b <-- classified as 4630 370 a = ironico 533 4467 b = no_ironico	0.897	0.926	0.911
	SMO	81.80	a b <-- classified as 4926 74 a = ironico 1746 3254 b = no_ironico	0.738	0.985	0.844
Se aplicó filtro Discretize	Naïve Bayes	90.60	a b <-- classified as 4791 209 a = ironico 731 4269 b = no_ironico	0.868	0.958	0.911
	Random Forest	91.29	a b <-- classified as 4594 406 a = ironico 465 4535 b = no_ironico	0.908	0.919	0.913
	J48	91.28	a b <-- classified as 4661 339 a = ironico 533 4467 b = no_ironico	0.897	0.932	0.914
	SMO	92.42	a b <-- classified as 4717 283 a = ironico 475 4525 b = no_ironico	0.909	0.943	0.926
Se aplicó filtro Discretize y AttributeSelection	SMO	88.90	a b <-- classified as 4254 746 a = ironico 364 4636 b = no_ironico	0.921	0.851	0.885

Anexo 11. Resultados de la clasificación de opiniones de la colección 8 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	78.36	a b <-- classified as 4176 824 a = ironico 1340 3660 b = no_ironico	0.757	0.835	0.794
	Random Forest	91.51	a b <-- classified as 4744 256 a = ironico 593 4407 b = no_ironico	0.889	0.949	0.918
	J48	91.38	a b <-- classified as 4662 338 a = ironico 524 4476 b = no_ironico	0.899	0.932	0.915
	SMO	81.99	a b <-- classified as 4944 56 a = ironico 1745 3255 b = no_ironico	0.739	0.989	0.846
Se aplicó filtro Discretize	Naïve Bayes	90.87	a b <-- classified as 4812 188 a = ironico 725 4275 b = no_ironico	0.869	0.962	0.913
	Random Forest	91.45	a b <-- classified as 4603 397 a = ironico 458 4542 b = no_ironico	0.910	0.921	0.915
	J48	92.03	a b <-- classified as 4704 296 a = ironico 501 4499 b = no_ironico	0.904	0.941	0.922
	SMO	92.39	a b <-- classified as 4704 296 a = ironico 465 4535 b = no_ironico	0.910	0.941	0.925
Se aplicó filtro Discretize y AttributeSelection	SMO	89.59	a b <-- classified as 4605 395 a = ironico 646 4354 b = no_ironico	0.877	0.921	0.898

Anexo 12. Resultados de la clasificación de opiniones de la colección 9 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	79.86	a b <-- classified as 4288 712 a = ironico 1302 3698 b = no_ironico	0.767	0.858	0.81
	Random Forest	92.01	a b <-- classified as 4761 239 a = ironico 560 4440 b = no_ironico	0.895	0.952	0.923
	J48	92.11	a b <-- classified as 4726 274 a = ironico 515 4485 b = no_ironico	0.902	0.945	0.923
	SMO	81.97	a b <-- classified as 4931 69 a = ironico 1734 3266 b = no_ironico	0.740	0.986	0.845
Se aplicó filtro Discretize	Naïve Bayes	91.36	a b <-- classified as 4794 206 a = ironico 658 4342 b = no_ironico	0.879	0.959	0.917
	Random Forest	91.47	a b <-- classified as 4610 390 a = ironico 463 4537 b = no_ironico	0.909	0.922	0.915
	J48	92.57	a b <-- classified as 4727 273 a = ironico 470 4530 b = no_ironico	0.910	0.945	0.927
	SMO	92.87	a b <-- classified as 4756 244 a = ironico 469 4531 b = no_ironico	0.910	0.951	0.93
Se aplicó filtro Discretize y AttributeSelection	SMO	91.32	a b <-- classified as 4757 243 a = ironico 625 4375 b = no_ironico	0.884	0.951	0.916

Anexo 13. Resultados de la clasificación de opiniones de la colección 10 formada a partir de la usada en (Ling and Klinger, 2016)

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	79.58	a b <-- classified as 4291 709 a = ironico 1333 3667 b = no_ironico	0.763	0.858	0.808
	Random Forest	91.73	a b <-- classified as 4729 271 a = ironico 556 4444 b = no_ironico	0.895	0.946	0.920
	J48	91.92	a b <-- classified as 4687 313 a = ironico 495 4505 b = no_ironico	0.904	0.937	0.921
	SMO	81.78	a b <-- classified as 4921 79 a = ironico 1743 3257 b = no_ironico	0.738	0.984	0.844
Se aplicó filtro Discretize	Naïve Bayes	90.95	a b <-- classified as 4809 191 a = ironico 714 4286 b = no_ironico	0.871	0.962	0.914
	Random Forest	91.69	a b <-- classified as 4632 368 a = ironico 463 4537 b = no_ironico	0.909	0.926	0.918
	J48	92.47	a b <-- classified as 4770 230 a = ironico 523 4477 b = no_ironico	0.901	0.954	0.927
	SMO	92.38	a b <-- classified as 4792 208 a = ironico 554 4446 b = no_ironico	0.896	0.958	0.926
Se aplicó filtro Discretize y AttributeSelection	SMO	90.44	a b <-- classified as 4714 286 a = ironico 670 4330 b = no_ironico	0.876	0.943	0.908
Se aplicó filtro Discretize y AttributeSelection	J48	90.97	a b <-- classified as 4723 277 a = ironico 626 4374 b = no_ironico	0.883	0.945	0.913

Anexo 14. Resultados de la clasificación de opiniones de la colección formada por 54 623 irónicos o sarcásticos y 5027 regulares (la colección completa usada en (Ling and Klinger, 2016))

Pre-procesamiento	Clasificador	Exactitud (%)	Matriz de confusión	Kappa	Precisión	Exhaustividad	Medida-F
No se aplicó filtro	Naïve Bayes	94.63	a b <-- classified as 53140 1483 a = ironico 1720 3307 b = no_ironico	0.645	0.969	0.973	0.971
	Random Forest	97.84	a b <-- classified as 54419 204 a = ironico 1084 3943 b = no_ironico	0.848	0.980	0.996	0.988
	J48	97.81	a b <-- classified as 54259 364 a = ironico 944 4083 b = no_ironico	0.850	0.983	0.993	0.988
	SMO	95.91	a b <-- classified as 53952 671 a = ironico 1767 3260 b = no_ironico	0.706	0.968	0.988	0.978
Se aplicó filtro Discretize	Naïve Bayes	97.65	a b <-- classified as 54347 276 a = ironico 1123 3904 b = no_ironico	0.836	0.980	0.995	0.987
	Random Forest	97.62	a b <-- classified as 54289 334 a = ironico 1083 3944 b = no_ironico	0.835	0.980	0.994	0.987
	J48	97.64	a b <-- classified as 54327 296 a = ironico 1112 3915 b = no_ironico	0.835	0.980	0.995	0.987
	SMO	97.64	a b <-- classified as 54331 292 a = ironico 1116 3911 b = no_ironico	0.835	0.980	0.995	0.987
Se aplicó filtro AttributeSelection	J48	97.05	a b <-- classified as 54091 532 a = ironico 1226 3801 b = no_ironico	0.796	0.978	0.990	0.984