

Análisis de sentimiento del español basado en corpus

(Spanish corpus-based sentiment analysis)

Antonio Moreno-Ortiz, Chantal Pérez-Hernández y Javier Fernández-Cruz

fernandezcruz@uma.es



¿Cómo citar?

Moreno-Ortiz, A., Pérez-Hernández, C., & Fernández-Cruz, J. (2022). Análisis de sentimiento del español basado en corpus. En G. Parodi, P. Cantos Gómez, & C. Howe (Eds.), *Lingüística de corpus en español: The Routledge handbook of Spanish corpus linguistics*. Routledge. pp. 511-529.

Doi: 10.4324/9780429329296-38

ISBN: 978-0-429-32929-6

Análisis de sentimiento del español basado en corpus

(Spanish corpus-based sentiment analysis)

1. Introducción

En este capítulo ofrecemos una visión panorámica del análisis de sentimiento, disciplina en la que los corpus textuales y su análisis son la pieza angular, y que se ha desarrollado en la última década como subárea del procesamiento del lenguaje natural (PLN), y núcleo principal de lo que se conoce como “minería de opinión”. Su auge, tanto en el ámbito de investigación como en la creación de aplicaciones comerciales, se debe a la necesidad de disponer de herramientas computacionales que ofrezcan métricas fiables sobre la opinión de los usuarios de productos y servicios, opinión que a diario es vertida en la red a través de múltiples medios sociales. El objetivo fundamental es obtener de forma automática una cualificación del texto de entrada en una escala (binaria o de mayor precisión), según su axiología, es decir, según la positividad o negatividad de la opinión que se refleja en el texto. Introduciremos los conceptos y metodologías fundamentales en el análisis de sentimiento, en los que el análisis textual y, por tanto, las técnicas y metodologías desarrolladas en la lingüística de corpus, juega un papel fundamental. Definimos los distintos enfoques según el objetivo que persiguen: clasificación a nivel de oración, de documento o de aspecto, y según los algoritmos empleados, tanto los basados en el léxico como los que emplean aprendizaje automático, ya sea mediante algoritmos supervisados o no supervisados; asimismo, incluimos una descripción de los rasgos que más comúnmente se emplean para entrenar los distintos clasificadores. Finalmente, exploramos en detalle el papel crucial que el corpus desempeña en todos estos enfoques y metodologías, tanto en los métodos supervisados, para la extracción de rasgos clasificatorios, como en los métodos no supervisados, como son los basados en semántica distribucional.

Palabras clave: análisis de sentimiento; minería de opinión; lenguaje evaluativo; corpus

In this chapter we provide an overview of sentiment analysis, a discipline that has evolved over the last decade as a branch of natural language processing, at the core of what is known as “opinion mining”. It has attracted great attention, both in academic research and commercial applications, due to the demand for computational tools that offer reliable metrics on users’ opinion

of products and services, which they express daily on a multitude of online sites and social media. Sentiment analysis seeks to automatically obtain an input text's qualification on a scale of a certain precision (binary or otherwise) according to its axiology, that is, the degree of positivity or negativity it expresses. We introduce the fundamental concepts in sentiment analysis where textual analysis, and particularly corpus linguistics methods and techniques, play a fundamental role. We define the various approaches according to the aim they pursue: sentence-level, document-level or aspect-based classification, and according to the types of algorithm employed, both by lexicon-based and machine learning approaches, which may use supervised and unsupervised classifiers. Finally, we explore in detail the crucial role that the corpus has in all of these approaches and methodologies, both in supervised ones, for the extraction of classifying features, and in unsupervised methods, such as those based on distributional semantics (*word embeddings*).

Keywords: sentiment analysis; opinion mining; evaluative language; corpus

2. Conceptos fundamentales

El análisis de sentimiento, desde sus principios, se ha centrado en extraer información sobre las actitudes emocionales a partir del análisis de grandes corpus textuales, que suelen estar conformados por textos en los que se vierten opiniones o valoraciones. De hecho, el título de este capítulo contiene en cierto modo una redundancia, pues todos los enfoques empleados en el análisis de sentimiento hacen uso en menor o mayor medida de los corpus textuales, y con frecuencia dependen de ellos por completo. Por ejemplo, la mayor parte de los sistemas propuestos hasta hace relativamente poco tiempo hacen uso de algoritmos de aprendizaje automático supervisados en los que un corpus, al que normalmente se le denomina *dataset* en este contexto, es anotado, ya sea de forma manual, automática o semi-automática, a partir del cual dichos algoritmos generan un *clasificador*, es decir, un modelo estadístico que es capaz de clasificar documentos no presentes en el texto en una de las categorías en las que el corpus de *entrenamiento* ha sido anotado. En cuanto a los enfoques basados en léxico, que funcionan mediante la identificación de palabras de opinión contenidas en un diccionario, también requieren de corpus para la obtención de esos diccionarios de sentimiento. Finalmente, los más recientes métodos para

el PLN en general y el análisis de sentimiento en particular, basados en el aprendizaje automático no supervisado, tales como los *word embeddings* y las arquitecturas *Transformer*, requieren de ingentes cantidades de texto a partir de las cuales se obtienen de forma automática los modelos de lenguas que luego son empleados para diversas tareas.

En comparación con otros ámbitos relacionados, como la *computación afectiva*, el análisis se enfoca de una manera más directa y efectiva, ya que no se proponen métodos de comprensión completa de los textos, sino que estos se tratan como conjuntos léxicos que incluyen características semánticas, p. ej., una polaridad positiva o negativa basada en la presencia o ausencia de determinados elementos léxicos y patrones gramaticales. Solo algunos sistemas de análisis de sentimiento incluyen también el análisis de emociones (p. ej., Toret Medina 2015), clasificando enunciados con un conjunto limitado de emociones (felicidad, ira, tristeza). El resultado de un análisis de sentimiento, por tanto, suele ser una clasificación: las opiniones son etiquetadas con una puntuación positiva, negativa o neutra (Waloszek y Waloszek 2017). La atención suscitada por el análisis de sentimiento justifica el rápido paso de disciplina meramente científica, hacia una enorme variedad de aplicaciones comerciales de todo tipo. La mayoría de las compañías con intereses en la inteligencia de mercado (por ejemplo, Microsoft, IBM o Google) ofrecen sus propias soluciones comerciales para la minería de opinión y muchas otras (Amazon, Facebook, Tripadvisor, etc.) cuentan con sus propios sistemas internos de análisis.

Prueba de la relevancia que el análisis de sentimiento ha tenido en los últimos años son las múltiples áreas de aplicación en las que se ha usado: el análisis de volumen de ventas, de destinos turísticos (Moreno-Ortiz *et al.* 2010) o la evaluación de las actitudes de los inversores (Bollen *et al.* 2011). Asimismo, los análisis de sentimiento se han aplicado con frecuencia al estudio del debate político (Wang *et al.* 2012) o como factor para predecir los resultados de elecciones a partir de los comentarios realizados por los usuarios en la red (Tumitan y Becker 2014). Otros campos de aplicación han sido estudios del discurso político (Gallardo Paúls 2017), el análisis de textos periodísticos (Godbole *et al.* 2007), jurídicos (Garofalo 2018), estudios en psicología (Tao *et al.* 2016) o la educación (Fernández-Cruz 2017).

En relación con la lingüística, el análisis de sentimiento se relaciona con el estudio del lenguaje evaluativo, caracterizado como medio para expresar creencias personales y adoptar posiciones, para expresar acuerdo o desacuerdo con las personas o para reflejar un sistema de valores, disociar al hablante de determinados puntos de vista y comprometerse o no con lo expresado (Du Bois 2007; Hunston 2011). A lo largo de los años se han propuesto distintos modelos de observación para establecer un modelo de relación entre el contenido factual o referencial y el contenido afectivo o expresivo de la lengua. Sin embargo, la tarea de identificar el lenguaje evaluativo no es una tarea simple y la variedad de modelos lingüísticos propuestos para describirlo da cuenta de ello. Aunque la mayoría de los autores se han centrado en el análisis de distintos tipos de textos y de las características del lenguaje evaluativo, algunas teorías han ofrecido descripciones más exhaustivas de la naturaleza del lenguaje evaluativo, como *modalidad y modulación* (Halliday y Matthiessen 2014); *orientaciones evaluativas* (Lemke 1998); *posición o postura (stance)* (Biber *et al.* 1999; Hunston 2011, *inter alia*); *evaluación* (Hunston y Thompson 2000; Bednarek 2006) o *valoración (appraisal)* (Martin y White 2005).¹

En términos de comprensión del lenguaje natural, el análisis de sentimiento se puede considerar una subárea del análisis semántico, puesto que su interés se centra en observar el significado que ofrece la orientación semántica de los textos. Algo importante a tener en cuenta es que el análisis de sentimiento es, fundamentalmente, una tarea de *clasificación* de textos, oraciones, u otras segmentaciones: el objetivo fundamental es clasificar un texto o segmento en una de varias categorías, desde la clasificación binaria (positivo/negativo), lo que se le conoce como *Thumbs up/Thumbs down* (Pang *et al.* 2002) o con distintos niveles de intensidad, o *Seeing Stars*, donde se asigna una puntuación al sentimiento de un texto analizado dentro de una escala (Pang y Lee 2005), en la que “bueno” expresa una polaridad menor que “excelente” y “terrible” tiene mayor intensidad que “malo”. En cuanto al nivel de granularidad, lo más usual es la clasificación de documentos, entendido como un texto que incluye un mensaje completo, de longitud muy variable, desde un tuit hasta un artículo de opinión, pasando por una valoración de un producto o servicio; aunque también se intenta llegar a una comprensión más profunda y granular de las muchas opiniones que pueden ser reflejadas en un mismo documento, identificando

las entidades (personas, organizaciones, objetos) y la orientación semántica que se expresa sobre cada una de ellas, incluso sobre cada uno de los diferentes aspectos que tienen cada una de esas entidades. En buena medida, la elección del tipo de objetivo determina la metodología empleada, por lo que es necesario entender bien los distintos niveles de clasificación.

3. Estado de la cuestión

Los niveles de análisis que se suelen afrontar en las tareas de análisis de sentimiento son tres, con creciente complejidad: el nivel de documento, el nivel de oración y el nivel de entidad y sus aspectos.

3.1 Clasificación de sentimiento a nivel de documento

La clasificación de sentimiento a nivel documento se considera la tarea más simple de análisis de sentimiento y es, con mucho, la que más atención ha recibido, habiéndose empleado una amplia gama de algoritmos de clasificación que se aplican generalmente a un corpus de textos anotados que son utilizados como datos de entrenamiento. Este tipo de clasificación es la apropiada cuando se sabe de antemano qué se está valorando en el texto, por ejemplo, reseñas de productos. El objetivo no va más allá de la clasificación de la opinión general sobre esa entidad, sin dar cuenta de la opinión sobre los determinados aspectos que la conforman. Por ejemplo, en una reseña de un hotel el usuario puede hablar de distintos aspectos (la atención del personal, la comodidad de las camas, etc.), información que un clasificador a nivel de documentos es incapaz de recuperar.

3.2 Clasificación de sentimiento a nivel oración

En términos prácticos, es factible tomar el análisis de sentimiento a nivel oración como un análisis a nivel de documento a partir de documentos mono-oracionales. Sin embargo, Liu (2012) considera que esta tarea es más compleja, ya que la información contenida en una oración es mucho menor que la contenida en un documento típico, pues no suelen ser mensajes completos. Muchos

de los sistemas que realizan la clasificación a este nivel tienen dificultades para clasificar de manera precisa el sentimiento en tres valencias básicas (positivo, negativo y neutro) y con frecuencia obvian la clasificación de oraciones neutras, tendiendo a *positivizar* o *negativizar* la puntuación, lo que puede ser problemático, porque un documento puede contener oraciones que no expresan sentimiento, ya que una valencia neutra suele significar factualidad o ausencia de sentimiento.

Un término clave que se ha utilizado ampliamente en el análisis de sentimiento a nivel oracional es el de “subjetividad”, que no es equivalente a “sentimiento”, a pesar de que estén íntimamente interrelacionados. Una oración se considera subjetiva si se basa en la perspectiva mental de un individuo, en lugar de en una experiencia contrastable o del mundo físico. Por lo tanto, mientras que una oración objetiva ofrece la información de manera factual, una oración subjetiva expresa sentimientos personales, juicios o creencias. La tarea de determinar si una oración es objetiva o subjetiva se llama “clasificación de subjetividad” (Wiebe *et al.* 1999; Riloff y Wiebe 2003). Se debe tener en cuenta que las expresiones subjetivas relacionadas con el sentimiento expresan apreciación, evaluación, deseo, creencia, sospecha, especulación, etc. (Wiebe 2000). Sin embargo, es importante considerar que existen oraciones subjetivas que no contienen opinión ni sentimiento y que, por otra parte, las oraciones objetivas pueden implicar opiniones o sentimientos causados por hechos deseables o indeseables, como podemos ver en el siguiente ejemplo:

8.000 muertes sin contabilizar: así evoluciona el exceso de fallecidos en España y cada autonomía.²

Por ello, Liu (2015) considera obviar esta problemática y plantea diferenciar preferiblemente entre *opinionated* (si implica un sentimiento positivo o negativo), es decir “con carga de opinión” frente a “sin carga de opinión” (*non-opinionated*), si lo expresado no incluye un sentimiento positivo o negativo. Considerando estos factores, la clasificación a nivel oración se lleva a cabo mediante la resolución de dos tareas: la primera es la clasificación de la subjetividad, donde se trata de catalogar si una oración expresa una opinión o si por el contrario se trata de una información

objetiva o factual. La segunda es la clasificación de sentimiento, otorgando una valencia positiva o negativa a las oraciones. Entre las implementaciones prácticas, existen varios modelos basados en aprendizaje automático (véase sección 4.1) o en el uso de diccionarios, destacando los propuestos por Wiebe *et al.* (1999) o Riloff y Wiebe (2003) basados en la localización de patrones sintácticos de perspectivas subjetivas.

Es importante señalar que un sistema de análisis de sentimiento a nivel oración presume que una oración contiene una única expresión de sentimiento, algo que puede funcionar bien en oraciones simples (sujeto-verbo-objeto) con un único sentimiento. Sin embargo, su nivel de fiabilidad es menor cuando las oraciones son complejas e incluyen varias opiniones. En el caso de que una oración exprese varios sentimientos, puede asignársele una clase comodín denominada “mixta” (Liu 2015), o se puede detectar un tono general positivo o negativo, en el que obviamente se perderán detalles importantes, como en el siguiente ejemplo, donde el sentimiento negativo tiene mayor peso que el positivo:

Los buenos datos sobre la evolución de la pandemia no han logrado amortiguar las pérdidas del selectivo capitalino, después de que España registrase este lunes la cifra de muertos más baja desde el 18 de marzo.³

Actualmente, existen sistemas como Moreno-Ortiz (2017) que ofrecen soluciones para ponderar los elementos positivos y negativos en un mismo segmento.

3.3 Clasificación de sentimiento a nivel aspecto

El análisis de sentimiento a nivel aspecto⁴ parte de que la polaridad global de un documento no refleja necesariamente las diversas expresiones evaluativas sobre las entidades y sus aspectos que pueden aparecer en él. Este enfoque intenta hacer frente a esta complejidad y se centra particularmente en la identificación de sintagmas (en lugar de palabras), ya que estos son los fragmentos más significativos del texto. Para cada documento, se localizan los distintos asuntos o entidades (temas, personas, fenómenos o eventos) que son evaluados y que pueden portar opinión (Liu 2015) y, a continuación, los aspectos que se evalúan sobre estas entidades. En el diseño de un

clasificador de sentimiento primero han de categorizarse las entidades (*entity category*), que se representan por medio de un nombre exclusivo, por ejemplo, la entidad RESTAURANTE. A cada categoría le pueden corresponder distintas expresiones o menciones de entidad (*entity expression* o *entity mention*), instanciadas mediante palabras diferentes o frases sinónimas; por ejemplo, dentro de la categoría de la entidad RESTAURANTE entrarían las expresiones de entidad “hamburguesería”, “tapería” o “cervecería”. El proceso de agrupar las expresiones de entidad en categorías de entidad se llama *agrupación* o *resolución de entidades*. Para la extracción de aspectos de cada entidad, la tarea es básicamente la misma que para las entidades y utiliza los conceptos “categoría de aspecto” y “expresión de aspecto”. Por ejemplo, “cercano”, “céntrico” y “urbano” se refieren al mismo aspecto (ACCESIBILIDAD) para la entidad restaurantes (Moreno-Ortiz *et al.* 2019). Asimismo, al proceso de agrupar expresiones de aspecto en categorías de aspecto se le denomina “resolución de aspecto” (Liu 2015; Henríquez-Miranda y Guzmán-Luna 2016).

Tomando como base la descripción previa, las principales tareas de un sistema de análisis de sentimiento basado en aspecto pueden resumirse en seis (Liu 2012):

1. Extracción de entidades y su resolución: extraer todas las expresiones de entidad en un documento y reconocer y agrupar bajo una sola entidad a las expresiones de entidad sinónimas
2. Extracción de aspectos y su resolución: extraer todas las expresiones de aspecto de cada una de las entidades y agrupar las expresiones de aspecto sinónimas en grupos
3. Extracción del emisor de opinión y su resolución: resolver dentro la expresión quién es emisor de opinión de cada una de las opiniones del texto y agruparlos de manera análoga a las tareas 1 y 2
4. Extracción del tiempo: extraer el tiempo de publicación de cada publicación y homogeneizar los distintos formatos de tiempo de acuerdo con los estándares, p. ej.: *Sun, 15 May 2011 19:00:30 GMT*

5. Clasificación de sentimiento: Al clasificar determinamos si una opinión sobre un determinado aspecto a perteneciente a una entidad e es positiva, negativa o neutra en el caso de una clasificación “Thumbs up/Thumbs down”. En el caso de una clasificación tipo “Seeing Stars” (Pang y Lee 2005) se le asignará una valencia numérica en una escala (p. ej., de 0 a 5, siendo 5 la puntuación máxima)
6. Establecimiento de la relación entre los factores formales extraídos en las tareas anteriores, que constituyen las opiniones expresada en un texto d (entidad, atributo, sentimiento, emisor y tiempo), lo que Liu (2015) denomina la quintupla de opinión

Llevar a cabo esta serie de tareas es un proceso complejo que requiere no solo el uso de clasificadores de sentimiento que sean diseñados *ad hoc* y combinen distintas técnicas y herramientas; también implica, por ejemplo, el uso de técnicas de reconocimiento de entidades nombradas (NER, *Named Entity Recognition*) y de análisis de corpus para la extracción de n-gramas y palabras clave. En la Figura 34.1 mostramos un fragmento textual en el que se han identificado tanto entidades como aspectos y se han establecido sus relaciones, para el que se usó Brat (Stenetorp *et al.* 2012) como herramienta de anotación (véase Moreno-Ortiz *et al.* 2019).

[Insert 15031-4827-PIII-034-Figure-001 Here]

Figura 34.1. Ejemplo de anotación de entidades y aspectos.

Otra de las complicaciones que presenta este tipo de análisis es la clasificación de los tipos de opinión a la hora de generar un sistema de clasificación (Jindal y Liu 2006). El tipo más simple y frecuente de opinión expresa un sentimiento sobre una entidad determinada o un aspecto concreto de una entidad, por ejemplo, “El bar era genial”, pero una opinión comparativa es difícil de clasificar con una opinión final, pues se realiza una labor de contraste entre varias entidades que comparten aspectos. Por ejemplo, en el caso de “Los restaurantes son más caros en España que en Grecia”, el aspecto es PRECIO y compara dos entidades distintas que comparten aspecto: España y Grecia, por lo que la puntuación final de ese aspecto no sería concluyente para un clasificador automático.

Por otra parte, delimitar el objeto de opinión también es complejo, ya que no existen muchos modelos capaces de relacionar de manera efectiva entidades y/o aspectos con el objeto de opinión. Liu (2015) propone comenzar generando un conjunto de características dependientes jerárquicamente de la entidad de estudio, para después determinar el alcance de cada expresión de sentimiento y analizar si solo se refiere a una entidad determinada o a toda la oración. Por ejemplo, en “A Apple le está yendo muy bien en esta nefasta situación económica”, el alcance de aplicación de “nefasto” se refiere solo a “situación económica”, pero no a “Apple”. La propuesta de Boiy y Moens (2009) computa el peso de cada una de las características mediante el cálculo de la distancia entre la palabra que representa la entidad o el aspecto y el objeto de opinión relacionado. Otra solución para este problema es el uso de ontologías (Cadilhac *et al.* 2010; Lau *et al.* 2014), que se usan para definir las entidades y aspectos relevantes, darles estructura y definir restricciones de aplicación. También ofrece resultados satisfactorios la aplicación de la teoría de la estructura retórica (*rhetorical structure theory*, Taboada y Mann 2006) al análisis de sentimiento basado en aspecto. Usando esta teoría, Hoogervorst *et al.* (2016) proponen la generación de árboles discursivos a nivel de documento que representan formalmente las relaciones entre las diferentes oraciones de un texto. A continuación, clasifican y estructuran dichas relaciones siguiendo el listado de relaciones discursivas propuesto por Mann y Thompson (1988) para tener en cuenta, por ejemplo, el efecto acumulativo que la valoración de un determinado aspecto puede tener a lo largo del texto.

Se puede concluir que el análisis de sentimiento a nivel aspecto es el nivel más refinado, aunque la calidad de los resultados también depende en gran medida de la disponibilidad de corpus pre-annotados específicos de un dominio, ya que las entidades y aspectos relevantes a extraer son altamente dependientes de cada ámbito de especialidad (Moreno-Ortiz *et al.* 2019).

4. Consideraciones metodológicas

Las metodologías que se usan en el análisis de sentimiento se pueden agrupar en tres grandes categorías: el enfoque de “aprendizaje automático”, el “basado en léxico” y los “sistemas

híbridos”, que combinan características de ambos. El primero de ellos aplica algoritmos relacionados con la inteligencia artificial junto con información lingüística; el enfoque basado en léxico se fundamenta en el empleo de diccionarios de sentimiento, compilados usando diccionarios o, más frecuentemente, extrayendo información de corpus. Finalmente, existen enfoques híbridos que combinan técnicas de ambos enfoques, al ser muy común que los léxicos de sentimiento jueguen un papel clave en la mayoría de los métodos.

4.1 Aprendizaje automático

En la última década, la creciente disponibilidad de grandes cantidades de datos, sobre todo texto, junto con el aumento de la capacidad de computación disponible, ha hecho posible que los ordenadores actuales sean capaces de ejecutar de forma efectiva y práctica algoritmos de aprendizaje automático que, además, se encuentran convenientemente “empaquetados” en librerías de software. Esto aporta una accesibilidad a metodologías y datos hasta ahora desconocidas y ha resultado en una explosión de aplicaciones inteligentes que encontramos ya de forma rutinaria fuera del ámbito de la investigación y el desarrollo, pues estos algoritmos son la base de prácticamente todos los sistemas “inteligentes” actuales, desde los vehículos autónomos hasta la traducción automática, pasando por el reconocimiento facial. En el ámbito del análisis de sentimiento, el aprendizaje automático ha sido, desde el principio, una metodología ampliamente aplicada, especialmente en lo que se refiere a aprendizaje supervisado. En este apartado nos centramos en este tipo de aprendizaje, pues es el que se ha empleado “tradicionalmente”. En la última sección, dedicada a las direcciones futuras, describimos la tendencia actual, centrada en el aprendizaje no supervisado a partir de grandes corpus.

Los diversos algoritmos de aprendizaje automático permiten que los sistemas “aprendan” a partir de datos existentes, con el fin de identificar patrones y clasificar información con mínima intervención humana. Los algoritmos de aprendizaje supervisado tratan de modelar relaciones y dependencias entre las características de entrada y las etiquetas de clasificación, de modo que se puedan predecir los valores de salida para los nuevos datos. Una vez que el modelo se entrena adecuadamente podrá etiquetar nuevos conjuntos de datos no catalogados.

Un ejemplo clásico es el trabajo de Pang *et al.* (2002), en el que los autores llevaron a cabo una clasificación automática de valoraciones sobre un corpus (que en el ámbito del aprendizaje automático se denomina normalmente *dataset*) del *Internet Movie Database* (IMDB), para lo que usaron los *ratings* aportados por los usuarios como etiquetas de clasificación y emplearon un conjunto reducido de palabras positivas y negativas como rasgos (“features”) para construir el clasificador (modelo estadístico), comparando el rendimiento de tres tipos de algoritmos distintos (Naive Bayes, máxima entropía y máquinas de soporte vectorial).

El corpus juega un papel fundamental en el análisis de sentimiento basado en aprendizaje automático supervisado y su presencia es esencial en los cinco pasos fundamentales del proceso de aprendizaje automático (Liu 2015):

1. Pre-procesado: se normalizan los datos a los formatos requeridos por los sistemas.
2. Extracción de rasgos de entrenamiento: se define a partir de qué, concretamente, se va a aprender. En análisis de sentimiento, la presencia de palabras positivas o negativas, por ejemplo, pero también determinados signos de puntuación, el uso de mayúsculas, emoticonos, etc.
3. Aprendizaje: se selecciona un algoritmo de aprendizaje y se genera un modelo (clasificador) a partir de los datos de entrenamiento.
4. Evaluación: mediante distintas técnicas se evalúan y comparan los resultados de los distintos algoritmos de aprendizaje para determinar cuál es el mejor método de análisis.
5. Predicción: una vez establecido cuál es el modelo con mejor rendimiento, se procede a la clasificación automática de nuevos textos.

Los conjuntos de datos necesarios para diseñar un modelo son al menos de dos tipos: un “conjunto de datos de entrenamiento” (*training set*) anotados previamente y un “conjunto de datos de evaluación” (*test set*) que se utiliza para proporcionar métricas de evaluación objetivas y el ajuste final del modelo en el conjunto de datos de entrenamiento. El proceso de evaluación del

rendimiento del sistema es crucial en toda tarea de PLN y para ello existen metodologías muy específicas sobre cómo medir el éxito de los resultados, en concreto las métricas de Precisión y Exhaustividad y su ponderación, el Valor F.

El problema del aprendizaje automático supervisado es que requiere un conjunto de datos (correctamente) etiquetados del que aprender y con el que evaluar. A veces estos datos etiquetados están disponibles, como en el caso del *dataset* IMDB, pero en la mayoría de los casos, para tareas concretas, los corpus deben ser anotados manualmente, un proceso que implica mucho tiempo y esfuerzo. Por contra, en el aprendizaje no supervisado el sistema se entrena con datos no etiquetados. Esta familia de algoritmos se basa en la detección de patrones y modelos descriptivos sin una referencia de variable de etiquetado conocida, aplicando distintas técnicas en los datos de entrada para buscar reglas, detectar patrones y agrupar los puntos de datos que ayudan a obtener información significativa y describir mejor los datos. En el caso del lenguaje, los datos de entrada no son otros que el texto, utilizándose corpus de muchos miles de millones de palabras para entrenar los modelos predictivos. En el apartado 5 describimos la extraordinaria evolución que estos métodos han experimentado en los últimos años.

4.2 Enfoque basado en léxico

Los sistemas basados en léxico dependen de la disponibilidad de diccionarios de sentimiento. Estas recopilaciones suelen tomar como punto de partida elementos extraídos de corpus que posteriormente son codificados a un formato determinado, ampliados y anotados mediante diversas técnicas. La fuente de conocimiento primordial son palabras cuyo significado contiene una polaridad positiva o negativa. Un lexicón prototípico incluirá mínimamente una serie de entradas (formas o lemas) junto con la polaridad que denotan (p. ej., “{malo: NEG}”). Además de utilizar fórmulas que agreguen el sentimiento asignado a cada una de las entidades de un texto, se debe tener en cuenta la distancia entre las distintas palabras que aparecen el texto y, en muchos sistemas, las unidades que se consideran de polaridad neutra. Además de las palabras, muchos lexicones incluyen también expresiones multi-palabra que denotan polaridad (Liu 2012; Moreno-Ortiz *et al.* 2013).

Los sistemas basados en léxico con clasificación a nivel documento obtienen las puntuaciones de sentimiento mediante algoritmos que bareman estadísticamente todas las expresiones con polaridad en un texto (Liu 2015). Por lo tanto, la clasificación de sentimiento será positiva o negativa dependiendo del resultado de dicha baremación, por lo que en muchos casos podemos encontrar sistemas que asignan un sentimiento neutro a un texto si los elementos positivos y negativos se cancelan mutuamente. En cuanto a los sistemas basados en léxico con clasificación a nivel de aspecto, se tiene en cuenta la distancia o la relación sintáctica entre las unidades léxicas con polaridad asignada. En este caso, el sistema debe hallar el ámbito de aplicación de cada expresión de sentimiento y determinar si este debe cubrir el aspecto o entidad objetivo en la oración. Así, al ser el ámbito de estudio supraoracional, la explotación del léxico de sentimiento debe llevarse a cabo a través de distintas estrategias mediante el análisis de las relaciones sintácticas entre las expresiones de sentimiento (sustantivo-adjetivo, verbo-adverbio, etc.) y los objetos de opinión (Ding *et al.* 2008; Liu 2015).

También es importante tener en cuenta que una serie de construcciones gramaticales modulan el sentimiento cuando se ponen en relación con otros recursos léxicos, por lo que el cálculo de la polaridad también depende de otros elementos como partículas de negación o intensificadores y atenuadores. Polanyi y Zaenen (2006) propusieron a nivel teórico lo que denominaron “modificadores de sentimiento” o “modificadores contextuales de valencia” (“sentiment shifters” o “contextual valence shifters”): una serie de reglas que revierten, aumentan o disminuyen la carga valorativa que se le asigna a una expresión cuando aparecen ciertas estructuras en su contexto. Por ejemplo, el sistema de Hu y Liu (2004) y Kim y Hovy (2004) asigna a cada expresión de sentimiento positiva el valor [+1] y a cada expresión de sentimiento negativa el valor [-1]. En este modelo, las palabras de negación como “no” y “nunca” invierten el valor de la orientación semántica. Otros de los pioneros en su implementación técnica fueron Kennedy y Inkpen (2006) y en español, el sistema de Moreno-Ortiz, Pérez-Hernández y Del-Olmo (2013), cuya descripción de *Context Rules* fue implementada primero en *Sentitext* (Moreno-Ortiz 2010) y posteriormente refinada y ampliada en el desarrollo de la herramienta *Lingmotif* (Moreno-Ortiz 2016, 2017).

El modelo paradigmático en lengua inglesa es el *Sentiment Orientation Calculator* (SOCAL) de Taboada *et al.* (2011) que propone un enfoque más refinado, en el que la puntuación de sentimiento de cada elemento se mueve en un rango [-5, +5]. En este sistema, la intensificación o atenuación de las unidades con carga semántica funciona cuando un adjetivo o un sustantivo con una etiqueta de sentimiento positiva o negativa se coloca junto a un adverbio o un adjetivo modificador (p. ej., “muy bueno” o “totalmente”), y se aplica un sistema lineal de adición o resta, donde los intensificadores o atenuadores están anotados con determinado peso porcentual. Por ejemplo, se asigna a la partícula “ligeramente” un -50% con relación a la puntuación del sustantivo al que precede, o se suma un +25% si va precedido de “bastante” y un +90% de “extremadamente”, etc. ⁵

La inversión de la orientación de sentimiento también puede llevarse a cabo a través de recursos morfológicos (como los prefijos “i-/in-/im-”, “a-/an-”, “des-”) o por medio de elementos de polaridad negativa. Existen varios enfoques para tratar el efecto que los elementos de polaridad negativa tienen: mientras que Taboada *et al.* (2011) consideran que se les debe asignar una polaridad opuesta, otros autores, como Benamara *et al.* (2017) alegan que las acumulaciones de elementos de polaridad negativa refuerzan los sentimientos dentro de su ámbito (p. ej., “No hay falta de restaurantes en la Costa del Sol” sería una expresión positiva). Algunos estudios consideran que la negación en la oración es más notoria que la afirmación desde su propia esencia psicológica (Osgood y Richards 1973), a pesar de que las expresiones negativas suelen presentarse amortiguadas por recursos retóricos de carácter positivo, por lo que Taboada *et al.* (2011) propusieron que las expresiones de sentimiento negativas pesaran más a la hora de calcular el sentimiento de manera automática. Una solución propuesta por Moreno-Ortiz *et al.* (2010) es el caso de la unidad “excellent” con una puntuación de [+5]: su negación “not excellent”, está muy lejos del sentimiento de una palabra de sentimiento opuesto como puede ser “atrocious” [-5], sino que además su tonalidad será mucha más positiva que la negación “not atrocious”. Por esto, se propone que las partículas negativas tengan un desplazamiento fijo en el valor de sentimiento (p. ej., [-4]), en lugar de invertir el signo como proponen algunos modelos (p. ej., Wilson *et al.* 2005).

De esta manera, un adjetivo con valencia [+2] restaría 4 puntos y el clúster “not atrocious” pasaría a tener una puntuación final de [-2].

Como hemos visto, el análisis de sentimiento basado en léxico ofrece una respuesta robusta y versátil y el rendimiento de los resultados, además, es transparente y fácil de mejorar a partir de técnicas de recopilación de información lexicográficas o terminográficas. Los distintos desafíos que todavía quedan por resolver (ironía, anáfora, idiolectos, etc.) deberán afrontarse a partir de la profundización del estudio del lenguaje evaluativo en uso, en los que la lingüística de corpus ha tenido y tiene un papel central, por ejemplo, en la extracción de distintos patrones de las gramáticas locales de un determinado contexto, o la adaptación del léxico a una situación determinada.

4.3 Enfoques híbridos

Las distintas metodologías descritas pueden combinarse entre sí y también hacer uso de técnicas de clasificación no supervisadas que se pueden combinar entre sí. Por ejemplo, Martín-Valdivia *et al.* (2013) propusieron el uso de meta-clasificadores para desarrollar un sistema de clasificación de polaridad. Trabajaron en un corpus español de reseñas de películas junto con un corpus paralelo equivalente traducido al inglés. En primer lugar, generaron dos modelos individuales usando estos dos corpus como fuente de datos y aplicaron algoritmos de aprendizaje automático como SVM o NB. En segundo lugar, integraron la ontología de sentimiento *SentiWordNet* en el corpus en lengua inglesa, para generar un nuevo modelo no supervisado utilizando un enfoque de orientación semántica. Finalmente, combinaron los tres sistemas utilizando un meta clasificador. Sus resultados superaron los resultados del uso de corpus individuales y demostraron que su enfoque aportaba una buena estrategia para la clasificación de polaridad cuando hay corpus paralelos disponibles.

5. Direcciones futuras y conclusiones

Como hemos visto, los avances en la creación y análisis de diferentes tipos corpus textuales ha sido una pieza clave en la creciente actividad que el campo del PLN ha experimentado en los

últimos años, tanto a nivel de investigación y desarrollo como a nivel de aplicaciones industriales. De hecho, resulta cada vez más difícil estar al día de los últimos desarrollos, pues las técnicas, metodologías y arquitecturas computacionales evolucionan a un ritmo vertiginoso, movidas sin duda por la creciente demanda que existe para el tratamiento automatizado de las lenguas por parte de la industria (Bell y Rigau 2015). En esta sección tratamos de ofrecer, de forma muy sintética, una panorámica de las metodologías que han surgido en los últimos años y que están determinando una dirección muy clara hacia el uso de modelos lingüísticos matemáticos generados a partir de grandes corpus. Aunque estos modelos son de uso común en todas las tareas de procesamiento del lenguaje, nos centraremos en su aplicación concreta al análisis de sentimiento, ya que se ha convertido en uno de sus ámbitos de desarrollo fundamental. Como veremos, a pesar de la evolución imparable de técnicas y algoritmos, ésta viene acompañada de la necesidad de la creación y empleo sistemáticos de grandes corpus de los que poder generar los modelos lingüísticos.

El aprendizaje no supervisado a partir de grandes colecciones de textos lleva al extremo las grandes máximas de la lingüística de corpus, pues aspiran a construir modelos lingüísticos en los que la semántica de cada palabra se representa mediante un constructo matemático que se obtiene a partir de todos los contextos en los que esa palabra aparece en el corpus. El trabajo de Osgood *et al.* (1957) puede considerarse una primera aproximación a la representación del significado léxico mediante un espacio dimensional en el que las distancias entre palabras (*diferencial semántico*, usando el término acuñado por ellos) constituye una manera de medir de forma efectiva las relaciones semánticas entre ellas. Estos espacios semánticos multidimensionales van mucho más allá de las redes semánticas tradicionales, en las que una o más relaciones conceptuales (hiperonimia, etc.) determina una jerarquía o red conceptual y, sobre todo, no son solo constructos abstractos, sino que aportan la capacidad de medir de forma precisa distancias entre palabras, distancias que, por otra parte, han sido calculadas, exclusivamente, por los contextos en los aparecen.

Los primeros modelos computacionales que implementaron de forma efectiva estas ideas, en concreto el análisis semántico latente (*Latent Semantic Analysis*, LSA) de

(1997), tuvieron una gran repercusión y generaron aplicaciones concretas en todo tipo de ámbitos, incluyendo el análisis de sentimiento (por ej., Wang y Wan 2011). Hoy en día, sin embargo, las representaciones basadas en LSA se consideran ya “clásicas”, pues han surgido nuevas representaciones semánticas basadas en corpus que superan sus limitaciones, que se refieren a la enorme exigencia de memoria de almacenamiento y capacidad de proceso conforme aumenta el tamaño del corpus.

El trabajo de Mikolov *et al.* (2013) supuso sin duda un punto de inflexión y un auténtico revulsivo en el mundo del PLN. En él los investigadores de Google mostraban un método nuevo para crear modelos de lenguas mediante representaciones vectoriales de palabras basado en redes neuronales, que mejoraba ostensiblemente modelos anteriores de semántica distribucional (LSA, LDA⁶). Word2Vec se ha usado de forma generalizada para generar *embeddings* en todos los idiomas y todo tipo de aplicaciones y tareas de PLN. El éxito de Word2Vec se ha debido no solo a su novedad y eficacia para ciertas tareas, sino también a la disponibilidad de la implementación original, de acceso libre desde el principio, que permite crear *embeddings* a partir corpus, pero también otras implementaciones más versátiles, como es el caso de la librería de modelado semántico para Python Gensim.⁷ Además, se publicaron *embeddings* pre-entrenados con grandes corpus, como por ejemplo el conocido corpus de Google News de 100 mil millones de palabras, con una representación de 300 dimensiones y tres millones de palabras en su vocabulario. En español, el recurso más importante es el Spanish Billion Word Corpus and Embeddings (SBWCE)⁸ (Cardellino 2019) que ha sido empleado en numerosos trabajos de varias ediciones del Taller de Análisis de Sentimiento en Español (TASS) y otros foros relevantes (por ej., López-Solaz *et al.* 2016).

Sin embargo, los *word embeddings* generados mediante Word2Vec mostraron una importante limitación para su uso en análisis de sentimiento. La similitud coseno, es decir, la distancia entre dos palabras en un espacio vectorial, de dos palabras antónimas y de orientación semántica opuesta, como “bueno” y “malo” puede ser muy alta (cercana a 1), pues tienden a aparecer en contextos muy parecidos. Por ejemplo, empleando los *embeddings* del mencionado corpus SBWCE, la similitud coseno entre los adjetivos “estupendo” y “genial” es 0,783, mientras

que la distancia entre “bueno” y “malo” es solo algo menor (0,770). Esta limitación ha sido reconocida por otros autores (por ej., Tang *et al.* 2014; Ye *et al.* 2018). Incluso así, en general, la distancia de coseno es una buena medida de cercanía semántica y sintáctica. La Figura 34.2 muestra una visualización de un espacio vectorial de 100 palabras (50 positivas y 50 negativas) generada a partir de los *embeddings* del corpus SBWCE.⁹ Como se puede ver, palabras con semántica y comportamiento sintáctico similares tienden a agruparse.

Otra importante limitación de Word2Vec, que también queda patente en la Figura 34.2, es que su representación vectorial ignora la morfología de las palabras, asignando vectores diferentes y no relacionados entre sí a formas diferentes del mismo lema; por eso las agrupaciones en esta visualización tienden a estar compuestas por palabras correspondientes a la misma categoría gramatical. Esto es menos relevante en lenguas como el inglés, con una morfología flexiva pobre, pero multiplica el vocabulario en lenguas muy flexivas, complicando tanto su creación, como su uso y aplicación, pues los *embeddings* requieren recursos computacionales muy elevados. Para superar esta limitación, se han creado otros métodos de generar *embeddings*, tales como la librería de Facebook FastText (Bojanowski *et al.* 2017), en la que en lugar de usar palabras para generar el vocabulario, se toman n-gramas de caracteres. Otro generador de *embeddings* digno de mencionar es GloVe (Pennington *et al.* 2014).

Figura 34.2. Visualización de embeddings en dos dimensiones.

Por otra parte, el mayor problema que presentan estas representaciones vectoriales como modelo de lengua es que son incapaces de hacer frente a la polisemia. Recordemos que cada vector es una representación de todos los contextos en los que ha aparecido esa palabra, ya sea forma o lema, según el caso, pero en definitiva el vector de una palabra polisémica como “banco” estará en una posición vectorial a determinada distancia (dependiendo de cuántos casos existan en el corpus de cada acepción) entre “mueble” y “edificio”.

Para solucionar este problema se han propuesto recientemente modelos más complejos basados en múltiples capas de redes LSTM (*Long Short Term Memory*) que, frente al tipo de red

empleada en los algoritmos descritos (RNN, *Recurrent Neural Networks*), son capaces de capturar dependencias semánticas distantes y de este modo ofrecer posibilidades para gestionar la polisemia. El primer sistema en ofrecer este tipo de arquitectura fue ELMo (*Embeddings from Language Models*) (Peters *et al.* 2018), del Instituto AllenNLP (Gardner *et al.* 2018). Las representaciones de ELMo son, en palabras de sus creadores, “profundas y contextualizadas”, pero además los vectores se derivan de una LSTM bidireccional y multicapa, generando un modelo de lengua bidireccional (biLM). De esta forma, los vectores de ELMo para una palabra son una función de toda la oración que contiene esa palabra, por lo que la misma palabra puede tener múltiples vectores para contextos diferentes. A nivel práctico, ELMo consiguió mejorar los resultados y métricas en 6 tareas de PLN, incluyendo el análisis de sentimiento, cuya evaluación se realizó sobre el *benchmark* Stanford Sentiment Treebank (SST-5, Socher *et al.* 2013).

Poco después ULMFiT (Howard y Ruder 2018) mejoró considerablemente lo que ofrecía ELMo al integrar el denominado *aprendizaje por transferencia*. Este tipo de aprendizaje automático intenta “reciclar” conocimiento ya adquirido para una determinada tarea en otra para la que se carece de entrenamiento. En el ámbito del PLN, ULMFiT es el primer sistema con estas capacidades, permitiendo emplear un modelo de lengua preexistente generado sobre grandes corpus, el cual se “afina” mediante otro corpus más pequeño, específico del dominio sobre el que se desea trabajar. Las ventajas de este sistema son muchas. Para empezar, el tiempo y recursos computacionales necesarios para entrenar un modelo de lengua a partir de un corpus de muchos miles de millones de palabras son muy considerables y requieren la disponibilidad de procesamiento en paralelo, además de la obvia dificultad de compilar corpus de ese tamaño.

Este sistema de “pre-entrenamiento” y “ajuste” (*pre-training and fine-tuning*), sigue constituyendo la base de la tecnología actual. Sin embargo, las LSTM que empleaba ELMo han dado paso a otro tipo de red neuronal: el Transformer. En 2018 los ingenieros de Google presentaron BERT (*Bidirectional Encoder Representations from Transformers*) (Devlin *et al.* 2019), mejorando el estado del arte de once tareas de PLN, entre las que otra vez se incluye el análisis de sentimiento. El impacto de BERT en la investigación en PLN ha sido enorme y ha

conseguido que la mayoría de las publicaciones sobre comprensión del lenguaje en el último año hayan sacado partido de sus funcionalidades (Synced 2019).

Parece obvio que el aprendizaje no supervisado a partir de grandes corpus es el presente y el futuro, no solo del análisis de sentimiento, sino del PLN en general. La actual disponibilidad de modelos de lenguas pre-entrenados y listos para ser adaptados a tareas específicas es un sistema flexible que pone en nuestras manos unas herramientas muy potentes para la creación de aplicaciones. El éxito de esta arquitectura queda patente, además, por la multitud de implementaciones y modelos que se han presentado en el último año y medio: GPT y GPT-2 (OpenAI), XLNET (Google/CMU), Transformer-XL (Google/CMU), XLM (Facebook), DistilBERT (HuggingFace), CamemBERT (FIAR/Inria), ALBERT (Google), XLM-RoBERTa (Facebook AI), FlauBERT (CNRS), CTRL (Salesforce).

La vertiginosa evolución que hemos descrito en esta sección deja patente que estamos ante un campo en ebullición, en el que apenas hay tiempo de aprender y explotar las posibilidades de un nuevo sistema antes de que el siguiente aparezca y lo haga obsoleto. No obstante, siempre es mejor esta situación que la contraria, y desde luego es bienvenida en un campo de estudio que no ha cesado de evolucionar y desarrollarse desde que se crearan los primeros corpus lingüísticos informatizados a comienzos de los años 60 del siglo pasado.

6. Lecturas recomendadas

Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. Nueva York: Cambridge University Press.

Moreno-Ortiz, A. y C. Pérez-Hernández. 2014. “Form and Function in Evaluative Language. The Use of Corpora to Identify Contextual Valence Shifters in a Linguistically-motivated Sentiment Analysis System”. En *The Functional Perspective on Language and Discourse*, eds. M. A. Gómez-Castillo, F. J. Ruiz de Mendoza, F. González y A. Downing, 87–110. Ámsterdam y Filadelfia: John Benjamins Publishing Company.

Referencias citadas

- Bednarek, M. 2006. *Evaluation in Media Discourse: Analysis of a Newspaper Corpus. Research in Corpus and Discourse*. Londres: Continuum.
- Bell, N. y G. Rigau. 2015. “Informe sobre el estado de las tecnologías del lenguaje en España dentro de la Agenda Digital para España”. *Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información*. <https://plantl.mineco.gob.es/tecnologias-lenguaje/PTL/Bibliotecaimpulsotecnologiaslenguaje/Material%20complementario/Informe-Tecnologias-Lenguaje-Espana.pdf>.
- Benamara, F., M. Taboada y Y. Mathieu. 2017. “Evaluative Language Beyond Bags of Words: Linguistic Insights and Computational Applications”. *Computational Linguistics* 43 (1): 201–264.
- Biber, D., S. Johansson, G. Leech, S. Conrad y E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Londres: Longman.
- Blei, D. M., A. Y. Ng y M. I. Jordan. 2003. “Latent Dirichlet Allocation”. *Journal of Machine Learning Research* 3: 993–1022.
- Boiy, E. y M. Moens. 2009. “A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts”. *Information Retrieval* 12 (5): 526–558.
- Bojanowski, P., E. Grave, A. Joulin y T. Mikolov. 2017. “Enriching Word Vectors with Subword Information”. *Transactions of the Association for Computational Linguistics* 5: 135–146.
- Bollen, J., H. Mao y X. Zeng. 2011. “Twitter Mood Predicts the Stock Market”. *Journal of Computational Science* 2 (1): 1–8.
- Cadilhac, A., F. Benamara y N. Aussenac-Gilles. 2010. “Ontolexical Resources for Feature-based Opinion Mining: A Case-study”. En *Proceedings of the 6th Workshop on Ontologies and Lexical Resources*, 77–86. Beijing: Coling 2010 Organizing Committee.
- Cardellino, C. 2019. *Spanish Billion Words Corpus and Embeddings*. <https://crscardellino.github.io/SBWCE/>.

- Devlin, J., M. Chang, K. Lee y K. Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding”. En *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Volumen 1, 4171–4186. Minneapolis: ACL.
- Ding, X., B. Liu y P. Yu. 2008. “A Holistic Lexicon-based Approach to Opinion Mining”. En *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 231–240. Palo Alto, CA: ACM.
- Du Bois, J. W. 2007. “The Stance Triangle”. En *Stancetaking in Discourse. Subjectivity, Evaluation, Interaction*, ed. R. Englebretson, 139–182. Amsterdam: John Benjamins.
- Fernández-Cruz, J. 2017. “Can Sentiment in Our Words Be Quantified? An Introduction to Lingmotif, a Sentiment Analysis Software Tool and Its Educational Application”. *International Congress on the Didactics of the English Language 2* (1).
- Gallardo Paúls, B. 2017. “Pseudopolítica en la red: Indicadores discursivos de desideologización en Twitter”. *Pragmalingüística* 25: 189–210.
- Gardner, M., J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. H. S. Liu, M. E. Peters, M. Schmitz y L. Zettlemoyer. 2018. “AllenNLP: A Deep Semantic Natural Language Processing Platform”. En *Proceedings of Workshop for NLP Open Source Software*, 1–6. Melbourne: ACL.
- Garofalo, G. 2018. “Persiguiendo con imparcialidad “el total desprecio a la Constitución”: El léxico valorativo en la Querrela del Fiscal de Cataluña contra Carme Forcadell i Lluís”. *Quaderns de Filologia – Estudis Lingüístics* 22 (22): 79.
- Godbole, N., M. Srinivasaiah y S. Skiena. 2007. “Large-Scale Sentiment Analysis for News and Blogs”. *Icwsn* 7 (21): 219–222.
- Goźdz-Roszkowski, S. y S. Hunston. 2016. “Corpora and Beyond – Investigating Evaluation in Discourse: Introduction to the Special Issue on Corpus Approaches to Evaluation”. *Corpora* 11 (2): 131–141.
- Halliday, M. K. y C. Matthiessen. 2014. *An Introduction to Functional Grammar*. Abingdon: Routledge.

- Henríquez-Miranda, C. y J. Guzmán-Luna. 2016. “Las ontologías para la detección automática de aspectos en el análisis de sentimientos”. *Prospectiva* 14 (2): 90–98.
- Hoogervorst, R., E. Essink, W. Jansen, M. van den Helder, K. Schouten, F. Frasincar y M. Taboada. 2016. “Aspect-Based Sentiment Analysis on the Web Using Rhetorical Structure Theory”. En *Web Engineering*, eds. A. Bozzon, P. Cudre-Maroux y C. Pautasso, 9671: 317–334. Cham: Springer International Publishing.
- Howard, J. y S. Ruder. 2018. “Universal Language Model Fine-tuning for Text Classification”. En *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Volumen 1, 328–339. Melbourne: ACL.
- Hu, M. y B. Liu. 2004. “Mining and Summarizing Customer Reviews”. En *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. Nueva York: ACM.
- Hunston, S. 2011. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. Nueva York: Routledge.
- Hunston, S. y G. Thompson, eds. 2000. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Jindal, N. y B. Liu. 2006. “Identifying Comparative Sentences in Text Documents”. En *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 244–251. Nueva York: ACM.
- Jockers, M. L. 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Kennedy, A. y D. Inkpen. 2006. “Sentiment Classification of Movie Reviews Using Contextual Valence Shifters”. *Computational Intelligence* 22 (2): 110–125.
- Kim, S. y E. Hovy. 2004. “Determining the Sentiment of Opinions”. En *Proceedings of the 20th International Conference on Computational Linguistics*. Stroudsburg: ACL.
- Landauer, T. K. y S. T. Dumais. 1997. “A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge”. *Psychological Review* 104 (2): 211–240.

- Lau, R., C. Li y Stephen S. Y. Liao. 2014. “Social Analytics: Learning Fuzzy Product Ontologies for Aspect-Oriented Sentiment Analysis”. *Decision Support Systems* 65: 80–94.
- Lemke, J. 1998. “Resources for Attitudinal Meaning: Evaluative Orientations in Text Semantics”. *Functions of Language* 5 (1): 33–56.
- Liu, B. 2012. “Sentiment Analysis and Opinion Mining”. *Synthesis Lectures on Human Language Technologies* 5 (1): 1–167.
- Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. Nueva York: Cambridge University Press.
- López-Solaz, T., F. L. Cruz y F. Enríquez. 2016. “Ampliación de lexicones de opinión específicos de dominio usando representaciones continuas de palabras”. *Procesamiento del Lenguaje Natural* 57: 49–56.
- Mann, W. y S. A. Thompson. 1988. “Rhetorical Structure Theory: Toward a Functional Theory of Text Organization”. *Text* 8 (3): 243–281.
- Martin, J. R. y P. R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Hampshire: Palgrave Macmillan.
- Martín-Valdivia, M. T., E. Martínez-Cámara, J. Perea-Ortega, y L. A. Ureña-López. 2013. “Sentiment Polarity Detection in Spanish Reviews Combining Supervised and Unsupervised Approaches”. *Expert Systems with Applications* 40 (10): 3934–3942.
- Mikolov, T., K. Chen, G. Corrado y J. Dean. 2013. *Efficient Estimation of Word Representations in Vector Space*. ICLR Workshop Track 2013 Proceedings.
- Moreno-Ortiz, A. 2016. *Lingmotif* (versión 1.0). Málaga, España: Universidad de Málaga.
- Moreno-Ortiz, A. 2017. “Lingmotif: A User-focused Sentiment Analysis Tool”. *Procesamiento del Lenguaje Natural* 58: 133–140.
- Moreno-Ortiz, A., C. Pérez-Hernández y M. Del-Olmo. 2013. “Managing Multiword Expressions in a Lexicon-Based Sentiment Analysis System for Spanish”. En *Proceedings of the 9th Workshop on Multiword Expressions MWE 2013*, 1–10. Atlanta: ACL.

- Moreno-Ortiz, A., F. Pineda Castillo y R. Hidalgo García. 2010. “Análisis de Valoraciones de Usuario de Hoteles con Sentitext: un sistema de análisis de sentimiento independiente del dominio”. *Procesamiento de Lenguaje Natural* 45: 31–39.
- Moreno-Ortiz, A., S. Salles-Bernal y A. Orrequia-Barea. 2019. “Design and Validation of Annotation Schemas for Aspect-Based Sentiment Analysis in the Tourism Sector”. *Information Technology & Tourism* 21 (4): 535–557.
- Osgood, C. E., G. J. Suci y P. H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Osgood, C. E. y M. M. Richards. 1973. "From yang and yin to and or but". *Language*, 380-412.
- Pang, B. y L. Lee. 2005. “Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales”. En *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 115–124. Ann Arbor: ACL.
- Pang, B. y L. Lee. 2008. “Opinion Mining and Sentiment Analysis”. *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.
- Pang, B., L. Lee y S. Vaithyanathan. 2002. “Thumbs Up?: Sentiment Classification Using Machine Learning Techniques”. En *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*. Volume 10, 79–86. Stroudsburg, PA: ACL.
- Pennington, J., R. Socher y C. Manning. 2014. “Glove: Global Vectors for Word Representation”. En *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Qatar: ACL.
- Peters, M., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee y L. Zettlemoyer. 2018. “Deep Contextualized Word Representations”. En *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Volume 1, 2227–2237. Nueva Orleans: ACL.
- Polanyi, L. y A. Zaenen. 2006. “Contextual Valence Shifters”. En *Computing Attitude and Affect in Text: Theory and Applications*, eds. J. G. Shanaham, G. Qu y J. Wiebe, 1–10. Dordrecht: Springer.

- Riloff, E. y J. Wiebe. 2003. "Learning Extraction Patterns for Subjective Expressions". En *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, 105–112. Stroudsburg, PA: ACL.
- Socher, R., A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng y C. Potts. 2013. "Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank". En *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1631–1642. Seattle: ACL.
- Stenetorp, P., S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou y J. Tsujii. 2012. "BRAT: A Web-Based Tool for NLP-Assisted Text Annotation". En *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 102–107. Avignon, Francia: ACL.
- Synced. 2019. "Google's ALBERT Is a Leaner BERT; Achieves SOTA on 3 NLP Benchmarks". *Medium*. <https://medium.com/syncedreview/googles-albert-is-a-leaner-bert-achieves-sota-on-3-nlp-benchmarks-f64466dd583>.
- Taboada, M., J. Brooke, M. Tofiloski, K. Voll y M. Stede. 2011. "Lexicon-Based Methods for Sentiment Analysis". *Computational Linguistics* 37 (2): 267–307.
- Taboada, M. y W. Mann. 2006. "Rhetorical Structure Theory: Looking Back and Moving Ahead". *Discourse Studies* 8 (3): 423–459.
- Tang, D., F. Wei, N. Yang, M. Zhou, T. Liu y B. Qin. 2014. "Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification". En *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Volume 1, 1555–1565. Baltimore, MD: ACL.
- Tao, X., X. Zhou, J. Zhang y J. Yong. 2016. "Sentiment Analysis for Depression Detection on Social Networks". *Lecture Notes in Computer Science* 7867: 807–810.
- Toret Medina, J. 2015. *Tecnopolítica y 15M: La Potencia de Las Multitudes Conectadas. Un Estudio Sobre La Gestación y Explosión Del 15 M*. Barcelona: Editorial UOC.
- Tumitan, D. y K. Becker. 2014. "Sentiment-Based Features for Predicting Election Polls: A Case Study on the Brazilian Scenario". En *Proceedings of the 2014 IEEE/WIC/ACM*

- International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. Volume 2, 126–133. Varsovia: IEEE Computer Society.
- Waloszek, A. y W. Waloszek. 2017. “SACAM – A Model for Describing and Classifying Sentiment Analysis Methods”. En *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*, 196–206. Oporto: SCITEPRESS – Science and Technology Publications.
- Wang, H., D. Can, A. Kazemzadeh, F. Bar y S. Narayanan. 2012. “A System for Real-Time Twitter Sentiment Analysis of 2012 US Presidential Election Cycle”. En *Proceedings of the ACL 2012 System Demonstrations*, 115–120. Jeju Island, Korea: ACL.
- Wang, L. y Y. Wan. 2011. “Sentiment Classification of Documents Based on Latent Semantic Analysis”. En *Advanced Research on Computer Education, Simulation and Modeling*, eds. S. Lilly y X. Huang, 356–361. Berlin: Springer.
- Wiebe, J. 2000. “Learning Subjective Adjectives from Corpora”. En *Proceedings of the AAAI/IAAI 00, 17th Conference of the American Association for Artificial Intelligence*, 735–740. Cambridge, MA: AAAI Press/The MIT Press.
- Wiebe, J., R. Bruce y T. O’Hara. 1999. “Development and use of a gold-standard data set for subjectivity classifications”. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* 246–253.
- Wilson, T., J. Wiebe y P. Hoffmann. 2005. “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis”. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)* 347–354.
- Ye, Zhe, F. Li y T. Baldwin. 2018. "Encoding Sentiment Information into Word Vectors for Sentiment Analysis". *Proceedings of the 27th International Conference on Computational Linguistics*, 997-1007. Association for Computational Linguistics.

1 ([Hunston \(2011\)](#)) revisa de forma detallada las diferencias y similitudes entre diferentes aproximaciones lingüísticas al estudio del lenguaje evaluativo y delimita los puntos de consenso que se pueden encontrar entre esos diferentes enfoques, ampliados después en [Goźdź-Roszkowski y Hunston \(2016\)](#).

2 Titular del periódico digital El País. 11 de mayo 2020.
https://elpais.com/sociedad/2020/04/25/actualidad/1587831599_926231.html?rel=friso-portada

3 Diario El País. 11 de mayo 2020. <https://elpais.com/economia/2020-05-11/las-bolsas-abren-con-modestas-ganancias-animadas-por-la-progresiva-desescalada.html>

4 En distintos dominios de aplicación el análisis de sentimiento basado en aspecto se denomina también *análisis de sentimiento basado en temas (topic based sentiment analysis)* o *análisis de sentimiento basado en entidades (entity-based sentiment analysis)*.

5 En el caso de que estos elementos estén encadenados, estos se aplican de manera recursiva tomando como punto de partida la unidad léxica más próxima a la palabra de sentimiento (p. ej., ‘realmente excelente’). De esta manera, el sentimiento de ‘realmente muy bueno’ se modelaría de una forma más eficiente a partir de multiplicaciones, en lugar de a través de adiciones, de la siguiente forma $[3 * [100\% + 25\%]]$ ([Benamara et al. 2017](#)).

6 Latent Dirichlet Allocation ([Blei et al. 2003](#)) es una técnica de aprendizaje automático especialmente adecuada para tareas de topic modeling. Ha sido llamada “la madre de todas las herramientas de colocaciones” ([Jockers 2013](#), 123).

7 https://radimrehurek.com/gensim/auto_examples/tutorials/run_word2vec.html

8 La primera versión fue publicada en 2016. La última versión está disponible en

<https://crscardellino.github.io/SBWCE/>

9 Para la visualización, las 300 dimensiones de los *embeddings* originales han sido reducidas a 2 mediante un algoritmo t-SNE.