

---

# Extracción de sentimientos y opiniones

---

PID\_00257771

Joaquim Moré

---

Tiempo mínimo de dedicación recomendado: 4 horas



Joaquim Moré

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por el profesor: Jordi Casas (2019)

Primera edición: septiembre 2019  
Autoría: Joaquim Moré  
Licencia CC BY-SA de esta edición, FUOC, 2019  
Av. Tibidabo, 39-43, 08035 Barcelona  
Realización editorial: FUOC



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-Compartir igual (BY-SA) v.3.0 España de Creative Commons. Se puede modificar la obra, reproducirla, distribuirla o comunicarla públicamente siempre que se cite el autor y la fuente (FUOC. Fundació per a la Universitat Oberta de Catalunya), y siempre que la obra derivada quede sujeta a la misma licencia que el material original. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-sa/3.0/es/legalcode.ca>

## Índice

<b>Introducción.....</b>	<b>5</b>
<b>1. Elementos de una opinión.....</b>	<b>7</b>
1.1. El <i>target</i> o <i>topic</i> .....	7
1.2. Los aspectos .....	7
1.3. La polaridad .....	7
1.4. El <i>opinion holder</i> y el momento .....	8
1.5. Ejemplo de los elementos de una opinión .....	8
<b>2. Detección automática de los elementos de una opinión.....</b>	<b>9</b>
2.1. Detección del <i>target</i> .....	9
2.1.1. Detección de las distintas denominaciones del <i>target</i> con una ontología .....	10
2.1.2. Detección de las distintas denominaciones del <i>target</i> con <i>word embeddings</i> .....	10
2.1.3. Detección de la referencia anafórica al <i>target</i> .....	11
2.2. Detección de los aspectos .....	12
2.2.1. Detección a partir de las <i>infoboxes</i> .....	13
2.2.2. Detección a partir de <i>opinion words</i> .....	14
2.3. Detección de la polaridad .....	17
2.3.1. Dificultades en la detección de la polaridad con criterios preestablecidos .....	17
2.3.2. Detección de la polaridad con <i>machine learning</i> .....	18
2.4. Detección del <i>opinion holder</i> y del momento .....	20
<b>3. Casos de uso del sentiment analysis.....</b>	<b>22</b>
3.1. Cosas que mejorar a través de las opiniones .....	22
3.1.1. La metodología .....	22
3.1.2. Recopilación del corpus de análisis .....	23
3.1.3. Creación de las colecciones .....	23
3.1.4. Identificación del <i>target</i> .....	24
3.1.5. Identificación de los aspectos del <i>target</i> .....	24
3.1.6. Aspectos vinculados a <i>opinion words</i> de polaridad negativa .....	27
3.1.7. Informe .....	28
3.2. Opiniones sobre un programa de televisión .....	29
3.2.1. Obtención de los tuits .....	30
3.2.2. Exploración .....	31
3.2.3. Preprocesado .....	34
3.2.4. Análisis de los datos .....	34
3.2.5. Informe .....	40

3.3. Detección de opiniones falsas .....	41
3.3.1. Tipificación del problema .....	42
3.3.2. ¿Cómo abordar el problema? .....	42
3.3.3. Análisis textual .....	44
3.3.4. Clasificador de opiniones falsas con <i>machine learning</i> ....	48
<b>Resumen.....</b>	<b>49</b>

## Introducción

En este módulo nos ocuparemos de cómo extraer los sentimientos y las opiniones a partir de los textos.

El análisis de los sentimientos y las opiniones se aplica a casi todos los negocios y actividades sociales. Las opiniones de los demás influyen en nuestro comportamiento y nuestra percepción del mundo. Quizá esta afirmación parece un poco exagerada pero, cuando reservamos un hotel, ¿no miramos antes sus evaluaciones? ¿No leemos las críticas de una película antes de verla? ¿No nos gusta escuchar tertulias políticas para sentirnos reafirmados en nuestras convicciones cuando los tertulianos las defienden o las atacan?

Las productoras cinematográficas, los partidos políticos, las empresas de servicios, etc., invierten mucho dinero en sacar partido de esta máxima: nuestras creencias y percepciones de la realidad y, sobre todo, nuestras decisiones están condicionadas por cómo los otros ven el mundo. Y ya sabemos el provecho económico que se puede sacar cuando uno sabe cómo influir en las decisiones. Pero no solo eso, si un Gobierno y las élites económicas son capaces de crear un estado de ánimo emocional positivo, los beneficios económicos y sociales son evidentes. Valga como ejemplo un estudio de la Universidad de Indiana en 2010 que demostraba que el estado de ánimo y los sentimientos de la población expresados en Twitter eran predictores de las subidas y bajadas de la bolsa en un día concreto.

El *sentiment analysis* no es algo nuevo. Siempre se ha hecho, pero antes de la era Web 2.0 se realizaba con encuestas y *focus groups*. Las cosas no eran como ahora, que todo el mundo tiene ganas de escribir para el público sus opiniones de manera desinteresada en los blogs, en las páginas web de hoteles o eventos, en Facebook, en Twitter, en las noticias de los diarios digitales, etc. Como las encuestas y los *focus groups* son caros de realizar, se ha recurrido al procesamiento del lenguaje natural para analizar, con un gasto muy bajo, un ingente número de opiniones publicadas cada día.

Ahora bien, el lenguaje natural tiene unas características que hacen que su análisis automático sea complejo. De algunas de estas características hemos hablado ya en el módulo anterior, la ambigüedad, por ejemplo. No es lo mismo decir *una brutal agresión* que decir *esta serie es brutal* o *este plato te ha quedado brutal*. En la primera opinión, se destaca algo negativo, en la segunda opinión, el adjetivo *brutal* puede interpretarse en sentido positivo y negativo y en la tercera opinión el adjetivo solamente se interpreta en sentido positivo.

### Bibliografía recomendada

Johan Bollen; Huina Mao; Xiao-Jung Zeng (2011). «Twitter mood predicts the stock market». *Journal of Computational Science* (vol. 2, núm. 1, págs. 1-8).

En este módulo explicaremos los elementos de una opinión y cómo estos se pueden detectar y analizar con procesamiento del lenguaje natural. Después aplicaremos el *sentiment analysis* para abordar una de las cuestiones más importantes a las que se debe enfrentar cualquier negocio, persona o colectivo que depende mucho de la reputación en línea. Nos referimos a la intervención de individuos cuya misión consiste en influir en la reputación de un colectivo o de un negocio.

Para explicar y ejemplificar los conceptos teóricos se utilizarán ejemplos de opiniones encontrados en la bibliografía sobre *sentiment analysis*. Una vez conocidos los principales elementos teóricos, trataremos los siguientes hipotéticos casos de uso.

- 1) Los propietarios de un hotel, cuyos huéspedes son mayoritariamente rusos y españoles, quieren saber qué servicios y características del hotel deben cuidarse especialmente según la nacionalidad del huésped.
- 2) Una cadena de televisión quiere analizar las opiniones sobre un programa dedicado al colectivo que se dedica a la prostitución voluntariamente.
- 3) Un buscador de hoteles, consciente del daño económico que puede suponer para su negocio la publicación de opiniones falsas que dañan la reputación de las cadenas hoteleras, quiere aplicar un detector de opiniones falsas antes de publicarlas.

Imaginemos que estos proyectos son encargos a la empresa Sense and Suitability (S&S), con los mismos protagonistas del módulo anterior. Anne y Henry trabajan en el proyecto del programa de televisión, y Joseph y Beth trabajan en los otros dos proyectos.

## 1. Elementos de una opinión

El objeto de estudio de este módulo son las opiniones que expresan sentimientos negativos o positivos. El primer reto es reconocer automáticamente una opinión. Para ello, hay que hacer una descripción estructurada del problema.

Las opiniones tienen dos componentes fundamentales:

- 1) la entidad sobre la que se expresa una opinión
- 2) el sentimiento expresado en esta opinión

La opinión también puede centrarse en expresar un sentimiento sobre uno o más aspectos de la entidad.

### 1.1. El **target** o **topic**

La entidad sobre la que se expresa una opinión se conoce con la denominación inglesa *target* o *topic*. Nosotros utilizaremos la denominación *target*.

Si la opinión se centra en un solo *target*, analizamos opiniones regulares, pero si la centramos en más de un *target*, diremos que analizamos opiniones comparativas. En este bloque nos centramos en las opiniones regulares, a las que llamaremos simplemente opiniones.

### 1.2. Los aspectos

Los aspectos son la parte o propiedades del *target* sobre la cuales se emite una opinión. Por ejemplo, hay opiniones que más que opinar sobre un modelo de teléfono móvil en su conjunto se centran en su pantalla, el teclado o sus funciones, como la fotografía, el asistente, etc.

### 1.3. La polaridad

Los sentimientos expresados en la opinión se conocen como orientaciones del sentimiento o de la opinión. También se conocen como polaridad, que es el término que utilizaremos aquí. Diremos que una opinión tiene polaridad positiva si el sentimiento que transmite es positivo, y polaridad negativa si el sentimiento transmitido es negativo.

### Bibliografía

- B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». En: C. Aggarwal; C. Zhai (eds.). *Mining Text Data*. Boston, MA: Springer.
- B. Liu (2010). «Sentiment Analysis and Subjectivity» (cap. 26). En: N. Indurkhy; F. J. Damerau (eds.). *Handbook of Natural Language Processing* (2.<sup>a</sup> ed.). CRC Press.

## 1.4. El *opinion holder* y el momento

El *target* y la polaridad son elementos fundamentales, pero hay otros componentes que también lo son. Nos referimos a quien expresa la opinión, conocido con la denominación inglesa *opinion holder*, y al momento en el que se expresa.

## 1.5. Ejemplo de los elementos de una opinión

Veamos un ejemplo de opinión donde se pueden encontrar los elementos comentados:

Figura 1. Ejemplo de opinión

I bought an **iPhone** a few days ago. **It** was such a **nice phone**. The **touch screen** was **really cool**. The **voice quality** was **clear** too. Although the **battery life** was not **long**, that is **ok** for **me**. However, **my mother** was mad with me as I did not tell her before I bought **it**. **She** also thought the **phone** was **too expensive**, and wanted me to return **it** to the shop...

Fuente: B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». En: C. Aggarwal; C. Zhai (eds). *Mining Text Data*. Boston, MA: Springer.

En rojo hemos marcado el *target*. Hay que tener en cuenta que su denominación no es siempre la misma. El *target* se puede expresar con su marca (*iPhone*), por el tipo de objeto que es (*phone*) o con un pronombre.

En violeta están marcados los aspectos de la entidad. La pantalla táctil, la batería y por extensión su duración son aspectos de la entidad, que es el *iPhone*. La calidad de sonido de la voz también es un aspecto del *smartphone* sobre el que se opina. Captar la relación que no se dice de manera explícita entre la entidad y sus aspectos, y detectar cuáles de estos aspectos se están evaluando, también es un objetivo muy complejo del procesamiento del lenguaje y de la inteligencia artificial.

En marrón hemos marcado los elementos que indican la polaridad. Lo difícil es seleccionar, entre los elementos que hacen referencia a polaridad, aquellos que se refieren al *target*. Es decir, elegir *nice*, *really cool*, *ok*, *long* o *too expensive*, pero no elegir *mad* porque se refiere a la madre, que no es el *target*.

Finalmente, en verde están marcados los dos *opinion holders*: el escritor de la opinión y su madre, que opina que el teléfono es demasiado caro.

### Bibliografía

Este ejemplo se ha extraído de B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». En: C. Aggarwal; C. Zhai (eds). *Mining Text Data*. Boston, MA: Springer.

## 2. Detección automática de los elementos de una opinión

En este apartado presentaremos los métodos que permiten detectar los elementos de las opiniones. Empezaremos por el *target* y sus aspectos. Continuaremos con la polaridad y finalmente nos ocuparemos del *opinion holder* y el tiempo.

### 2.1. Detección del *target*

Miremos el ejemplo de opinión que hemos presentado antes.

Figura 2. Ejemplo de opinión

I bought an **iPhone** a few days ago. **It** was such a nice **phone**. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought **it**. She also thought the **phone** was too expensive, and wanted me to return **it** to the shop...

Fuente: B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». En: C. Aggarwal; C. Zhai (eds). *Mining Text Data*. Boston, MA: Springer.

En rojo están marcadas las referencias al *target* de la opinión, es decir, la cosa sobre la que se opina. Aparentemente, las referencias son evidentes, pero desde el punto de vista del procesamiento del lenguaje natural, encontrar estas referencias y resolver que todas apuntan a la misma entidad no es tan sencillo ni mucho menos. Las referencias al mismo objeto con denominaciones diferentes es un campo de estudio muy activo en el procesamiento del lenguaje natural.

En principio, la opinión se centra en torno a un producto. Este producto tiene un nombre concreto, pero fijémonos en que también se puede hacer referencia al producto utilizando un hiperónimo –esto es, *phone*, que es la clase o tipo al que pertenece la entidad. Por otro lado, observemos que también se hace referencia al mismo objeto con dos palabras sinónimas (*phone* y *telephone*).

Si ya es complicado reconocer que todas estas denominaciones diferentes se refieren a la misma cosa, hay que añadir el reconocimiento de que el pronombre *it* en «*I bought it*» hace referencia al teléfono y no a la pantalla. Resolver correctamente a qué entidad hace referencia un pronombre se denomina resolución de la anáfora.

Por otro lado, *phone* es también la denominación del verbo *llamar*, así como una unidad de sonido en el campo de la fonética. Resolver que el sentido de *phone* es el mismo que el de *telephone* es tarea del desambiguador del sentido de una palabra, también conocido como WSD (*Word Sense Disambiguation*).

En aplicaciones de *sentiment analysis* el *target* es un nuevo modelo de producto, la marca de una empresa, un político, un partido político, un colectivo, etc. Es decir, en muchos proyectos el *target* ya está definido por el cliente. Por tanto, en estos casos, el reto consiste en reconocer las distintas denominaciones de este *target* en las opiniones. Veamos, por ejemplo, cómo encontramos las denominaciones del *target* en la opinión de ejemplo.

### **2.1.1. Detección de las distintas denominaciones del *target* con una ontología**

Una estrategia posible es utilizar una ontología como la de Wordnet. La estrategia consiste primero en establecer el sentido o *synset* del *target*, y luego identificar las denominaciones que son o bien sinónimas o bien hiperónimas o hipónimas. Por ejemplo, teniendo como *target iPhone*, una referencia a este *target* podría ser su hiperónimo, como *phone* en la opinión de ejemplo.

Ahora bien, nombres de productos y marcas que se han popularizado de tal modo que más que una marca son un concepto, como *iPhone*, no están en Wordnet. Sin embargo, como ya explicamos en el módulo anterior, se puede consultar la ontología YAGO a través de la base de datos de la DBpedia y ver qué *synset* de Wordnet es su hiperónimo. En el *notebook PLA-2 1.1* se comprueba que *telephone* es hiperónimo de *iPhone*. De este modo, en la opinión anterior podemos establecer que los términos *iPhone* y *phone* se refieren al *target*, siempre y cuando se haya comprobado anteriormente que *phone* y *telephone* están en el mismo *synset*.

Hay que tener en cuenta que DBpedia se organiza como un grafo de conocimiento, conocido como *Knowledge Graph* o KG, y que el formalismo de representación de las relaciones existentes en este grafo es el RDF, empleado para la web semántica. Las relaciones consisten en un conjunto de triplets RDF. Cada tripla RDF tiene los siguientes elementos: un sujeto, un predicado y también una lista de objetos. En el *notebook PLA-2 1.1* se puede ver que, en nuestra consulta, el sujeto es la URL que apunta a la página de la DBpedia que describe *iPhone*, y el predicado es *type*, esto es, el tipo (hiperónimo) al que pertenece la entidad. En la respuesta a la consulta el objeto es una lista con las URL que apuntan a las relaciones con su hiperónimo según distintas ontologías. Elegimos las URL correspondientes a la ontología YAGO.

### **2.1.2. Detección de las distintas denominaciones del *target* con *word embeddings***

En principio, las distintas denominaciones a la misma entidad aparecen en contextos muy similares. Por lo tanto, con un cálculo de los términos más próximos al *target* con el método Word2Vec, se pueden detectar las referencias sinónimas al *target*. El *target* suele ser un sintagma nominal.

En el *notebook PLA-2*, 1.1 se muestra cómo se obtienen distintas denominaciones del *target Samsung*. Tras haber creado un modelo de opiniones sobre teléfonos móviles en Amazon con Word2Vec, obtenemos los términos más parecidos al *target* que son sintagmas nominales.

Es curioso, sin embargo, ver que con el método Word2Vec también se obtienen términos relacionados con el *target* por una relación de metonimia, esto es, son denominaciones de la parte por el todo. Como se ve a continuación, la marca *Samsung* se asocia a sus productos. Por lo tanto, la opinión de la marca se proyecta través de la opinión de sus productos (p. ej.: *samsung products*, *galaxy*, *touchwiz*, *touch wiz*). Ahora bien, debemos tener cuidado porque este método también revela una tendencia en las opiniones: referenciar al *target* contrastando sus productos con los de la competencia (*lg*, *htc*).

Figura 3. Términos similares a *Samsung* según un modelo de opiniones con Word2Vec

```
['samsung products', 'htc', 'lg', 'galaxy', 'touch wiz', 'touchwiz', 'samsungs']
```

### 2.1.3. Detección de la referencia anafórica al *target*

Una referencia anafórica es la relación de identidad que se establece entre un sintagma nominal con un nombre o un pronombre y una entidad que ha aparecido anteriormente en el discurso, que se llama antecedente. En el ejemplo que hemos presentado, la relación anafórica es la relación que se establece entre el pronombre *it* y *iPhone* en *I bought it* y en *return it*.

La identificación del antecedente de una referencia anafórica es tarea de los analizadores o *parsers* sintácticos y semánticos de la oración. Una vez que se construye el árbol sintáctico, se elige como antecedente de la expresión anafórica aquel sintagma nominal que ha aparecido anteriormente en el discurso que mejor se adecua a las restricciones semánticas a las que está sometida la expresión anafórica. Por ejemplo, para encontrar el antecedente de *it* en *I bought it*, el *parser* debe encontrar un antecedente que tenga rasgos semánticos de las cosas *comprables*. Por ejemplo, que sea un artefacto o, si es un ser animado, que no sea una persona, etc. Bien es cierto que hay *parsers* que resuelven el antecedente sin etiquetado semántico aplicando las restricciones puramente sintácticas entre la expresión anafórica y el antecedente establecidas por teorías gramaticales como el modelo de *Government and Binding* de Noam Chomsky.

### Bibliografía recomendada

Para más información sobre la obtención de denominaciones del *target Samsung*, consultad el *notebook PLA-2*, 1.1.

### Bibliografía

N. Chomsky (1981). *Lectures on Government and Binding: The Pisa Lectures*. Mouton de Gruyter.

También hay *parsers* que resuelven el antecedente una vez que aprenden esta tarea con métodos de *machine learning*. Es el caso de *neural-coref*, una extensión de spaCy que aprende la resolución de referencias anafóricas con una red neuronal. Una vez aprendida la tarea es capaz de, dada una frase *F* que contiene una expresión anafórica, y dado un contexto *C* que contiene el antecedente, generar una versión de *F* en la cual la expresión anafórica es sustituida por el antecedente. En la figura 4 y en el *notebook* PLA-2 1.1 presentamos el código en Python que genera una versión correcta del antecedente de *It* en *I bought an iPhone a few days ago. It was such a nice phone.* También genera una versión correcta del antecedente del pronombre *her* en *However, my mother was mad with me as I did not tell her before I bought it.* Sin embargo, en *bought it*, el pronombre queda por resolver.

La resolución anafórica todavía está lejos de ser un tema resuelto y las herramientas que la aplican aún no son muy fiables. Como ya hemos comentado anteriormente, la resolución anafórica automática es una de las tareas más complejas del procesamiento del lenguaje natural, por la complejidad cognitiva y las limitaciones que todavía existen en la inteligencia artificial para modelar las interpretaciones de lo que algunos llaman el «sentido común». Un sentido común que es el resultado de las experiencias individuales, el conocimiento del mundo, del entorno, etc.

Figura 4. Código de *neural-coref* para generar la versión de una opinión con las expresiones anafóricas resueltas.

```
import neuralcoref
import spacy
import pytorch

nlp = spacy.load('en_coref_sm')
# alternative: import en_coref_sm nlp = en_coref_sm.load()
# bigger corpus: en_coref_lg

doc = nlp(u'I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really
cool. The voice quality was clear too. However, my mother was mad with me as I did not tell her
before I bought it. Although the battery life was not long, that is ok for me')

doc._coref_resolved

'I bought an iPhone a few days ago. an iPhone was such a nice phone. The touch screen was really
cool. The voice quality was clear too. However, my mother was mad with me as I did not tell my
mother before I bought The touch screen . Although the battery life was not long, that is ok for me'
```

## 2.2. Detección de los aspectos

Nos ocupamos seguidamente de los aspectos del *target*. Esto es, las propiedades y características del objeto de la opinión.

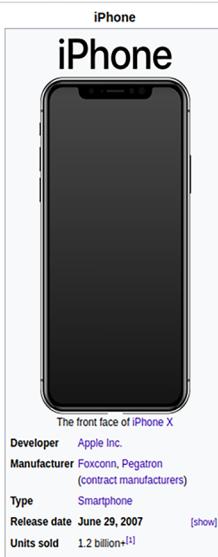
### Bibliografía recomendada

K. Clark; C. D. Manning (2016). «Deep Reinforcement Learning for Mention-Ranking Coreference Models», EMNLP.  
<https://github.com/huggingface/neuralcoref>  
<https://bit.ly/2xT4zDB>

## 2.2.1. Detección a partir de las infoboxes

Los aspectos de un producto o de una entidad se pueden encontrar en las *infoboxes* de Wikipedia. Por ejemplo, en la *infobox* de la entrada *iPhone* que vemos en la figura 5, aparece una relación de los aspectos de este producto (desarrollador, fabricante, cámara, teclado, batería, etc.).

Figura 5. Entrada de *iPhone* en la Wikipedia inglesa con su *infobox* en el margen derecho



The figure shows a screenshot of the iPhone page on Wikipedia. On the right side, there is a large image of the iPhone X. Below it is its infobox. The infobox contains the following information:

Developer	Apple Inc.
Manufacturer	Foxconn, Pegatron (contract manufacturers)
Type	Smartphone
Release date	June 29, 2007
Units sold	1.2 billion <sup>[1]</sup>

[show]

<sup>[1]</sup>

The original iPhone was described as "revolutionary" and a "game-changer" for the mobile phone industry. Newer iterations have also garnered praise, and the iPhone's success has been credited with helping to make Apple one of the world's most valuable publicly traded companies.

Fuente: Wikipedia

Por lo tanto, con la consulta a la *Knowledge Graph* de la DBpedia se pueden obtener las propiedades de la entidad. En el *notebook* PLA-2 1.2 se explica cómo. Las propiedades que hemos obtenido son las siguientes:

Figura 6. Propiedades de *iPhone* en su *infobox*

```
[unitssold', 'developer', 'reason', 'sound', 'date', 'dimensions', 'power', 'caption', 'display',
'related', 'storage', 'memory', 'soc', 'graphics', 'releasedate', 'connectivity', 'service']
```

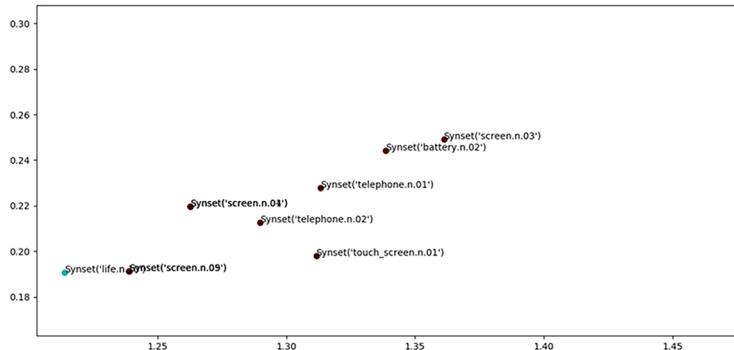
Si queremos identificar los aspectos del *target* de la opinión que nos sirve de ejemplo, esto es, *iPhone*, podemos agrupar los sentidos de los términos que son más cercanos a su hiperónimo, que es *telephone*.

En el *notebook* PLA-2 1.2 podemos ver cómo se hace esta agrupación. Se construye una matriz en cuyas filas están los posibles sentidos según Wordnet de los términos de la opinión. Por ejemplo, en las filas están los sentidos que puede tener la palabra polisémica *battery*. Uno de estos sentidos es *battery.n.02*, que se define como *a device that produces electricity; may have several primary or secondary cells arranged in parallel or series*, y también *battery.n.01*, que se define como *group of guns or missile launchers operated together at one place*.

Las columnas de la matriz son los sentidos que pueden tener los términos de las propiedades en el *infobox* de *iPhone*. Palabras que también son polisémicas, como *power* (p. ej.: *the rate of doing work; measured in watts* y *(of a government*

*(or government official) holding an office means being in power), entre otras. Cada celda contiene la distancia semántica entre el sentido de un término y el sentido de una propiedad. Con un método de agrupación o *clustering* (K-means) podemos ver que los sentidos más próximos a *telephone.n.01* (el sentido del hiperónimo del *target*) son *battery.n.02*, *telephone.n.02* (el sentido de *phone*) y *touch\_screen.n.01*, como se ve en la figura 7:*

Figura 7. Sentidos más próximos al sentido *telephone.n.01* en la opinión de ejemplo



El sentido *battery.n.02* (*a device that produces electricity; may have several primary or secondary cells arranged in parallel or series*) es compatible como aspecto del *target*.

Hay que tener en cuenta, sin embargo, que el término no es *battery* sino *battery life*. Sin embargo, ni en Wordnet ni en la DBpedia se ha encontrado un sentido de Wordnet para *battery life*.

### 2.2.2. Detección a partir de *opinion words*

La idoneidad del método anterior depende de la ontología que utilicemos. Ya hemos visto que *battery\_life* no se ha detectado como un aspecto porque no hemos encontrado su *synset* en WordNet. Por otro lado, con esta metodología, aspectos como la *voice quality* son difíciles de asociar a un teléfono porque la distancia ontológica entre un artefacto, una entidad física, como es un teléfono, y una entidad abstracta, como es la voz, y más su calidad, es considerable. Sin embargo, cuando leemos la opinión ambos conceptos nos parecen cercanos.

Entre los sistemas de *sentiment analysis* llamados *aspect-based opinion mining* (ABOM) *systems*, especializados en encontrar la valoración de los aspectos de la entidad, están los que detectan los aspectos a partir de elementos del discurso llamados *opinion words*. La *opinion word* es una palabra que tiene una carga negativa o positiva en la opinión. Cuando la *opinion word* califica positiva o negativamente un término que no está alejado en el discurso, este término es el aspecto.

En el caso de la opinión que estamos analizando, el adjetivo *clear* califica positivamente el aspecto *voice quality* y *long* califica positivamente *battery life*. Ahora bien, por la misma razón, *phone* debería tomarse como un aspecto porque el adjetivo *nice* lo califica positivamente, y *mother* también se debería tomar como un aspecto porque el adjetivo *mad* califica este término negativamente. Efectivamente, el adjetivo sirve para localizar un candidato a ser aspecto. Sin embargo, hay más probabilidades de que en las opiniones sobre productos el adjetivo *cool* se aplique a un aspecto del producto que el adjetivo *mad*. La idea es agrupar las *opinion words* con los candidatos que son más coherentes. *Mother* no debería tomarse como aspecto de *iPhone* porque una persona no puede ser el aspecto de un objeto y en cambio *touch screen* sí.

Además de las *opinion words* hay que añadir las *opinion phrases*, como *costar un ojo de la cara* o *fue una pesadilla*. Sin embargo, para facilitar la exposición, llamaremos *opinión words* tanto a las palabras como a las expresiones. Hay dos tipos de *opinión words*: los básicos y los comparativos. Unos adjetivos como *bueno* o *malo* serían *opinión word* básicos y los comparativos serían *mejor* y *peor*.

### Diccionarios de *opinion words*

Para algunas lenguas es fácil identificar las *opinion words* consultando diccionarios. Se distinguen tres tipos de diccionarios. El primer tipo es el diccionario de polaridad, el segundo tipo es el diccionario de intensidad y el tercer tipo es el diccionario de contexto.

Los diccionarios de polaridad son listas de *opinion words* clasificadas de manera binaria (positivo-negativo). El diccionario *General Inquirer* es el más antiguo de este tipo. También es conocido y utilizado el *Polarity Lexicon* de Hu y Liu de 2004, elaborado a partir de reseñas de productos. Este diccionario recoge expresiones de sentimiento en redes sociales, pero al igual que el *General Inquirer* no recoge emoticonos ni acrónimos. También hay que mencionar el lexicón del LIWC, validado por psicólogos, sociólogos y lingüistas.

En los diccionarios de intensidad las *opinion words* están asociadas a unos niveles de intensidad de la polaridad. Uno de los más utilizados es *SentiWordnet*. Se trata de un diccionario derivado de Wordnet. Cada *synset* de Wordnet tiene tres puntuaciones numéricas. Una indica el grado de positividad, la otra indica el grado de negatividad y la última el grado de neutralidad del *synset*. La suma de las tres puntuaciones da 1.

También están en esta categoría el diccionario *AFINN* y *SenticNet*. Términos como *ira*, *adoración* o *respeto* tienen un valor de polaridad en una escala continua entre el -1 y el +1, según su polaridad negativa o positiva. En *SenticNet* las *opinion words* están descritas como conceptos relacionados en forma de red con información de sentido común.

También contamos con el diccionario *ML-SentiCon*, que recoge *opinion words* del catalán, el español, el gallego y el euskera. Cada palabra tiene un valor que va del -1 al +1, según su polaridad, y un valor de ambigüedad que es el resultado de combinar los diferentes valores de polaridad de cada uno de los sentidos que puede tener una palabra.

## Aplicación del método

El método se aplica en el *notebook PLA-2 1.2*. Consiste, primero, en tener un modelo de opiniones. El modelo que utilizamos en el *notebook* es el de opiniones sobre *smartphones* realizado con Word2Vec.

Luego, dada una opinión O, cada frase de O que tiene alguna *opinion word*<sup>1</sup> se analiza sintácticamente. Los sintagmas nominales de O se consideran candidatos a ser aspectos. Con los candidatos y las *opinion words* se procede a modelar O. Los candidatos y las *opinion words* son el vocabulario, y el vectorizador crea, para cada término del vocabulario, un vector con las similitudes de este término respecto a los demás términos.

Recordemos que los términos del vocabulario son vectores en el modelo de opiniones de *smartphones* realizado con Word2Vec. Por lo tanto, la forma de calcular la similitud de términos del vocabulario será con una métrica de distancia entre vectores como la *cosine similarity*.

Una vez creados los vectores de similitud, se agrupan con un método de *clustering* en tres clases. La figura 8 muestra el resultado. Como podemos ver, la CLASS\_0 es la que agrupa más coherentemente las *opinion words* con los aspectos del *target*.

Figura 8. *Clustering* de aspectos y *opinion words* en una opinión

```
CLASS_0 ['day', 'phone', 'battery', 'life', 'phone', 'battery life']
CLASS_1 ['bought', 'mother', 'tell', 'bought', 'thought', 'return', 'shop', 'telephone', 'mad']
CLASS_2 ['iphone', 'touch', 'screen', 'voice', 'quality', 'ok', 'touchscreen', 'voice quality', 'nice', 'cool', 'clear']
```

### Ved también

Consultad el apartado 2.1.2.

<sup>(1)</sup>Las *opinion words* son las del diccionario AFINN.

### Bibliografía recomendada

Para profundizar en el tema de la *cosine similarity* y otras métricas basadas en espacios vectoriales, es recomendable leer el apartado 8.5.1 del libro de Manning y Schütze (2000) y el apartado 6.4 del libro de Jurafsky y Martin (2018).

Hay unos detalles que hemos de comentar. En primer lugar, hemos normalizado el término *touch screen* con la forma *touchscreen*, que es la más habitual en el modelo de opiniones de *smartphones*. Por otro lado, vemos que *voice* y *clear* se han agrupado coherentemente y que, en el modelo de opiniones sobre *smartphones*, *nice* y *cool* están más asociados a *touchscreen* que, en cambio, con *phone*, tal como está en la opinión. Finalmente, vemos que existen términos como *thought* y *tell* que se han desambiguado mal. Se ha tomado el sentido nominal de estas palabras. Aun así, podemos ver que el método nos permite identificar aspectos del *smartphone*, tales como su *touchscreen* y la calidad de voz.

Hay que reconocer que *battery* no se ha asociado con el adjetivo que la acompaña en la opinión, que es *long*. La razón es la siguiente: *long* no está en el diccionario de *opinion words*, ni tampoco *expensive*, como se habrá observado. Hay que ser conscientes de que las listas de *opinion words* no recogen palabras que, según el contexto en el que se usan, tienen o no tienen una polaridad. Cuando hablamos de la vida de una batería, *long* tiene una polaridad positiva, pero fuera de este contexto no tiene por qué tener una polaridad. Asimismo, *expensive*, que no aparece en ninguna CLASS, evoca un aspecto de la entidad que no se explicita en la opinión, que es el precio. Sin embargo, para un clasificador automático de opiniones negativas y positivas, que ha sido entrenado con un corpus de entrenamiento etiquetado según su polaridad, adjetivos como *long* o *expensive* pueden ser *features* importantes para clasificar la opinión.

### 2.3. Detección de la polaridad

La detección de la polaridad, esto es, el sentido positivo, negativo o neutral de una opinión, es el aspecto más complejo del *sentiment analysis*. Es compleja su detección en el nivel de palabra, como ya hemos apuntado en el apartado anterior, y lo es también en cuanto a la opinión en su conjunto.

#### 2.3.1. Dificultades en la detección de la polaridad con criterios preestablecidos

En el nivel de palabra, se puede considerar que los adjetivos y los adverbios cargan la polaridad de las opiniones con distinta intensidad, intensidad que está declarada en los diccionarios de *opinion words*. En estos diccionarios, la carga negativa o positiva de una palabra está preestablecida por su autor. Ahora bien, algunos adjetivos no son *opinion words per se*. Lo son en las opiniones de determinados dominios y contextos y el dominio y el contexto determinan también su polaridad (por ejemplo, el caso de *long*, ya comentado en el apartado anterior). Así pues, cuando un adjetivo o adverbio no está reconocido como una *opinion word* en un diccionario, se calcula su *tendencia* positiva o negativa según la probabilidad de coocurrencia con *opinion words* de una polaridad determinada. Por ejemplo, un adjetivo *A* tiene una tendencia negativa si suele coocurrir con *opinion words* de polaridad negativa. La tendencia se detecta con métricas como el *PMI*.

Los símbolos que no plantean ninguna duda sobre su polaridad son los emoticonos. El emotícono :) es universalmente interpretado como un símbolo de polaridad positiva y :( como un símbolo de polaridad negativa. Ahora bien, son símbolos que no se someten a las reglas gramaticales, como las palabras, y por ello tienen un uso menos previsible y su alcance –¿es un símbolo que marca la polaridad de toda la opinión o solo un aspecto del *target* de la opinión?– es difícil de determinar.

Otros elementos que pertenecen a una gramática de la expresión de opiniones, que es una gramática más líquida y mutante, son recursos tipográficos como las palabras en mayúscula, o la repetición de letras y signos de puntuación, que indican sorpresa, perplejidad, etc. (p. ej.: «esta película es HORRIBLE!!!!!!!, en serio??????»), y también el uso de expresiones oídas en algún medio de comunicación o anuncio publicitario de gran impacto (p. ej.: «COMM-MORRRRR????», «S'HA MATAO PACO»). No hay ninguna regla grammatical que establezca que hay que escribir «ES HORRIBLE», ni el número de veces que se tiene que repetir el signo de exclamación o la letra *R*. Tampoco hay un diccionario que recoja expresiones oídas en los medios de comunicación que momentáneamente se hacen populares.

Otra manifestación del uso «creativo» del lenguaje, sobre todo en las redes sociales, y que afecta a la normalización de los datos, es el uso de acrónimos como sinónimo de una palabra. Un ejemplo es DD para referirse a *derechos*.

En cuanto a la polaridad en el nivel de la opinión en su conjunto –esto es, saber por ejemplo si la opinión de una película es positiva o negativa–, su detección tampoco es sencilla. No se puede aplicar la siguiente regla: *en una opinión positiva, solo aparecen opinion words positivas y en una opinión negativa solo aparecen opinion words negativas*. La opinión que hemos analizado es un ejemplo de ello. La impresión global del comentario sobre el iPhone es positiva, pero hay *opinion words* positivas acompañadas por una negación (*the battery was not long*). Son los llamados *polarity flippers* porque cambian la polaridad positiva a negativa y viceversa. Además, a pesar de que la opinión global es positiva, aparecen también *opinion words* negativas, como *expensive*, si se aplica al precio, y *mad*.

Hay propuestas de cálculo global de la polaridad, que es el resultado de la suma de expresiones positivas y negativas de distintas intensidades, pero hay que tener en cuenta que cuando la opinión es irónica o sarcástica, un cálculo de este estilo puede dar un resultado opuesto completamente a la intención del opinador (p. ej.: *el iPhone me ha costado 900 euros. Baratísimo*). Por desgracia, el procesamiento del lenguaje natural está todavía muy lejos de conseguir detectar la ironía y el sarcasmo.

### 2.3.2. Detección de la polaridad con *machine learning*

Los métodos de *machine learning* han ido dominando la detección de la polaridad desde el momento en el que se vio que era mejor no especular sobre la complejidad lingüística de las opiniones, y dejar que un clasificador automático aprendiera a clasificar opiniones con comentarios reales etiquetados manualmente según su polaridad.

Como ya explicamos en el módulo anterior, en la clasificación con métodos de *machine learning* hay que realizar varios pasos. Para clasificar opiniones en foros, Twitter, etc., la ejecución de estos pasos puede tener una complejidad mayor que la que supuso el proyecto de la clasificación de titulares. Veamos por qué.

## Preprocesado

Recordemos que para entrenar un clasificador, hay que preprocesar los datos con el fin de que el sistema aprenda las *features* relevantes. Esto significa que hay que quitar de las opiniones los elementos que producen ruido e impiden que el sistema aprenda bien. En el módulo anterior, el clasificador de titulares aprendía de textos bien escritos. Sin embargo, en entornos como Twitter o foros de opinión, la labor de preprocesado es más compleja. Hay que eliminar *emojis* que no aportan nada excepto ruido, hay que aplicar correctores que corrijan errores de tecleo y faltas de ortografía. También hay que eliminar URL y expresiones recurrentes como RT, además de distinguir cuentas de usuario y eliminar datos personales, etc. En el *notebook* PLA-2, 1.3 vemos algunos ejemplos de preprocesado.

## Exploración y definición de las *features*

El resultado del preprocesado es un texto muy alterado. Se han quitado *emojis*, URL, etc., que estaban entre dos palabras y habrá que ir con cuidado para no procesar falsos n-gramas; esto es, que no se unan dos o tres palabras que estaban antes separadas y que tendrían que continuar estándolo. Además, los *hashtags* a veces se comportan como palabras en una oración bien formada, pero otras veces etiquetan el tuit entero. Hay que decidir, por tanto, si los *hashtag* se tratan aparte como etiquetas o se procesan como si fueran términos normales.

Por otro lado, en Twitter los *emojis* cada vez más reemplazan a las palabras. Es una tendencia general que está presente incluso en los tuits de instituciones, partidos políticos, etc., como en la figura 9, donde la palabra *España* en *una España mejor* está sustituida por el icono de la bandera española.

Figura 9. Ejemplo de *emojis* sustituyendo palabras en un tuit

The screenshot shows a tweet from the account @YoConSorayaPP. The profile picture is of a woman with glasses. The tweet text is:

Llegan momentos de regeneración política, ella @Sorayapp lo ha demostrado estos años al frente del Gobierno de @marienorajoy y es la persona idónea para liderar el @PPPopular en estos momentos de regeneración. #YoConSorayaPP ¿Y tú, te sumas? ¡Juntos por una 🇪🇸 mejor! #VamosSoraya

Fuente: Twitter

Los rasgos de polaridad que hemos comentado (mayúsculas, repetición de signos de interrogación, exclamación, etc.) son nuevas *features* que hay que tener en cuenta. Ser conscientes de que la gramática de la expresión de una opinión

es cambiante y muchas veces inesperada supone hacer siempre una exploración de los datos por si hay algún rasgo no previsto que pueda mejorar los resultados del clasificador. Trataremos esta cuestión con más detalle en el módulo siguiente.

## Tipos de clasificadores

Los clasificadores más habituales son los llamados clasificadores bayesianos. En el *notebook* PLA-2, 1.3 se muestra un clasificador que distingue tuits con opiniones positivas y negativas. También se aplican clasificadores SVM (*support vector machines*) y *logistic regression model*. Estos clasificadores asumen que el orden de los términos no influye en la clasificación (*bag of words*). Sin embargo, hay clasificadores de redes neuronales que sí tienen en cuenta el orden de los términos. Como en el método Word2Vec, los vectores que representan los términos de la opinión se construyen teniendo en cuenta el contexto de palabras ordenadas en las que aparece el término.

### 2.4. Detección del *opinion holder* y del momento

La detección del *opinion holder* es interesante para saber los aspectos de un producto más valorados y criticados según el tipo de cliente. Con un *parser* que recoja los sujetos de verbos de opinión como *think*, se puede saber el tipo de cliente que suele expresar una opinión favorable o desfavorable. En la opinión sobre el iPhone que hemos tomado de ejemplo tenemos el caso de una madre, representante de las generaciones no acostumbradas a comprar *smartphones*, que valora sobre todo el precio. En cambio, su hijo valora más el sonido, la cámara, la pantalla, etc.

El análisis del *opinion holder* también sirve para valorar los resultados de decisiones tales como abrir un producto a nuevos países o a nuevas tipologías de cliente. También para mejorar los servicios, como los de un hotel con clientela mayoritaria de un país, que sabe los aspectos que la gente de este país valora y critica más. Si tienes una clientela mayoritariamente inglesa, no basta con tener las habitaciones y el baño limpios, o un *buffet* variado y abundante; hay que asegurarse de que las instalaciones están adaptadas a clientes con movilidad reducida, si este es un aspecto muy valorado en las opiniones sobre hoteles por parte de opinadores ingleses. Este, precisamente, es el tema del primer caso de uso que veremos.

También es importante la detección del *opinion holder* cuando se analizan polémicas interesadas y campañas de desprestigio a una marca o a un colectivo. La información sobre el *background* del opinador, por ejemplo su partido político, es útil para saber la intencionalidad de estas campañas.

Cuando las polémicas interesadas y las campañas de desprestigio se producen en Twitter, la información del perfil de la cuenta del usuario es útil para obtener información sobre el *opinion holder*, asumiendo que no siempre en los

perfíles hay información explícita y veraz. La figura 10 muestra cómo se obtiene la información del perfil de la cuenta de Donald Trump con la librería Python Tweepy.

La información sobre los *opinion holders* debe tratarse con todas las precauciones porque los datos personales deben estar protegidos por la protección de datos. Los datos del perfil en las redes sociales son públicos, si así lo deciden los usuarios, y ellos deciden dar información real. Cuando buscamos perfíles de políticos o gente célebre y de prestigio, asumimos que esa información es veraz. Fuera de las redes sociales, pocas veces los usuarios exponen voluntariamente información sobre ellos mismos. Uno de los datos más fáciles de obtener es el país en el que han publicado una opinión en línea a través de la IP.

Figura 10. Script para obtener el perfil de la cuenta de Twitter de Donald Trump

```
import tweepy
from tweepy import API
from tweepy import OAuthHandler

consumer_key = XXXXX
consumer_secret = XXXXX
access_key = XXXXX
access_secret = XXXXX

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth)

description_user = api.get_user('@realDonaldTrump').description

print (description_user)
```

El cálculo de la influencia de ciertos opinadores en las redes sociales también tiene importancia para predecir el impacto de sus opiniones. Las métricas que miden la influencia de los usuarios de redes sociales quedan fuera del alcance de este texto.

En cuanto al tiempo, este dato está presente en todas las publicaciones en línea, desde Twitter a los foros de opiniones de viajes, de hoteles, etc. Con este dato se puede seguir el *timeline*, los picos de participación y el estado de opinión de los participantes en una controversia. También se pueden realizar análisis sobre la probabilidad de escribir un comentario negativo o positivo según el día de la semana o la hora del día. Análisis que pueden fundamentar incluso estudios psicológicos de los consumidores y de ciertos personajes influyentes del mundo de la política, la cultura, etc.

### 3. Casos de uso del *sentiment analysis*

Como ya hemos dicho en la introducción, expondremos tres casos de uso. El primero es el análisis de opiniones sobre hoteles para determinar sobre qué cosas los huéspedes son más exigentes según su nacionalidad. Aplicaremos a este caso de uso la teoría sobre los elementos de una opinión.

Los dos casos restantes ilustran el *sentiment analysis* aplicado a opiniones que tienen la intencionalidad de incidir en la reputación de un producto o del creador de este producto. Por un lado, analizaremos las opiniones sobre un programa de televisión y, por otro lado, trataremos la difusión de opiniones falsas.

Como en el módulo anterior, estos casos de uso se explicarán como si fueran encargos al grupo Sense and Suitability (S&S), que trabajó en el predictor de noticias del *New York Times* motivadoras de comentarios.

#### 3.1. Cosas que mejorar a través de las opiniones

El grupo S&S recibe un encargo de los directivos de una cadena hotelera: analizar opiniones de hoteles publicadas en TripAdvisor. Dado que un análisis manual sería demasiado costoso en tiempo y dinero, los propietarios del hotel quieren que S&S analice automáticamente las opiniones de huéspedes rusos y españoles con el fin de saber con qué servicios y características del hotel son más exigentes. Así sabrán qué aspectos deben cuidar especialmente y mejorar, teniendo en cuenta que la mayoría de sus clientes son de estas dos naciones.

Peter reúne al equipo. Están todos de acuerdo en que será un proyecto sin muchas complicaciones y que se podrá realizar en pocos días. Peter, el coordinador del grupo, asigna este proyecto a Joseph y a Beth.

##### 3.1.1. La metodología

En la segunda reunión con Peter los tres discuten la metodología que debe seguirse. Previamente han mirado los tipos de datos que ofrece TripAdvisor de cada opinión y discuten acerca de sobre qué datos, además de la nacionalidad del *opinion holder*, pueden hacer el análisis. Pronto llegan al acuerdo de tomar la valoración por puntos que pone el opinador como dato principal. El razonamiento es el siguiente: *analizando las opiniones que dan una baja valoración a un hotel sabremos sobre qué cosas los huéspedes rusos y españoles son más exigentes*.

Establecen, pues, que el corpus de análisis serán las opiniones negativas escritas por opinadores rusos y españoles. Con este criterio deciden realizar las siguientes tareas:

- 1) Recopilar el corpus de análisis.
- 2) Crear dos colecciones de documentos. La primera colección recoge las opiniones de *opinion holders* rusos. La segunda colección recoge las opiniones de *opinion holders* españoles.
- 3) Para cada colección de documentos:
  - identificar el *target* de la opinión,
  - identificar los aspectos de este *target*,
  - encontrar los aspectos más vinculados a *opinion words* de polaridad negativa.

A continuación explicaremos cómo Joseph y Beth van desarrollando cada paso. De este modo ilustraremos la aplicación de la detección de los elementos de una opinión, que ha sido el tema del apartado 2.

### 3.1.2. Recopilación del corpus de análisis

Por suerte, Joseph encuentra una base de datos pública con opiniones de hoteles de TripAdvisor. Esta base de datos tiene licencia para ser procesada.

La suerte no termina aquí. Cada opinión tiene asociada la nacionalidad del opinador y, además, ha sido etiquetada manualmente como una opinión favorable o desfavorable. Esto es importante porque Joseph ya se veía pasando las opiniones por un clasificador automático de polaridad para poder recoger, de todas las opiniones, las que eran desfavorables. Mucho mejor tener el criterio de un etiquetador humano que el de un etiquetador automático.

#### Ved también

Para más información sobre el clasificador automático de polaridad, vé el apartado 2.3.2.

### 3.1.3. Creación de las colecciones

Aprovechando el dato sobre la nacionalidad del *opinion holder*, Joseph crea las colecciones de opiniones para cada nacionalidad. La colección de opiniones de clientes rusos es de unas 3.900 opiniones y la de clientes españoles es de unas 4.800.

### 3.1.4. Identificación del target

Puesto que no se quieren analizar las opiniones de un hotel concreto, Joseph y Beth deciden tomar el término *hotel* como el término que hace referencia al *target* alrededor del cual pivotan los aspectos concretos que fundamentan la mala valoración de un hotel y que hay que detectar.

### 3.1.5. Identificación de los aspectos del target

Joseph y Beth discuten sobre qué metodología de las dos que hemos explicado en 2.2.1 y en 2.2.2 sería más adecuada para saber los aspectos más criticados por los clientes. Joseph, que tiene una tendencia ya conocida por solucionar las cosas vía *machine learning*, le dice a Beth que él siempre es partidario de *ensuciarse las manos* con los datos con los que trabajan antes que recurrir a recursos léxicos predefinidos e independientes del dominio que están estudiando. Beth, que también sabemos que tiene una predilección por las taxonomías, las gramáticas y los recursos que formalizan el lenguaje, al principio pensaba utilizar Wordnet. Sin embargo, cambia de opinión cuando Joseph le dice que tienen un corpus suficiente para hacer un modelo de opiniones de clientes rusos y españoles, lo cual les permitirá tener datos interesantes bastante pronto. Por lo tanto, deciden seguir el método basado en la presencia de *opinion words* con un modelo Word2Vec.

## Creación de los modelos

Joseph hace un modelo de las opiniones de *opinion holders* rusos que dan una baja valoración y hace otro modelo del mismo tipo de opiniones con *opinion holders* españoles. Los modelos los hace aplicando el método Word2Vec.

## Elección de los candidatos a aspectos a partir de un modelo de opiniones

Joseph y Beth establecen que los candidatos a aspectos del *target* serán los sintagmas nominales del modelo de opiniones que aparecen más de diez veces y tienen una proximidad con el término *hotel* (el *target*) superior a 0,8<sup>2</sup>. En la figura 11 presentamos una muestra de candidatos a aspecto tomando los términos cercanos a *hotel* en el modelo de opiniones ruso.

## Bibliografía recomendada

Para profundizar sobre este punto, véase *notebook PLA-1, 3.5.*

<sup>(2)</sup>Recordemos que el valor de proximidad según Word2Vec se encuentra entre el 0 y el 1.

Figura 11. Candidatos a aspectos de un hotel

```
'make', 'site', 'use', 'trip', 'hotel', 'place', 'window', 'ceiling', 'room', 'cost', 'way', 'day', 'time', 'view', 'noise', 'glass', 'entrance', 'stair', 'space', 'bathroom', 'cafe', 'staff', 'shower', 'problem', 'issue', 'floor', 'smoke', 'corridor', 'option', 'work', 'night', 'station', 'call', 'block', 'towel', 'wall', 'water', 'price', 'breakfast', 'paris', 'stay', 'morning', 'ground floor', 'payment', 'table', 'restaurant', 'dinner', 'pay', 'bar', 'suitcase', 'wifi', 'shampoo', 'personnel', 'reception', 'hour', 'arrival', 'coffee', 'city', 'people', 'bed', 'location', 'guy', 'cleaning', 'change', 'parking', 'walk', 'card', 'design', 'leave', 'hair', 'look', 'area', 'lift', 'food', 'check', 'system', 'furniture', 'bath', 'stuff', 'picture', 'balcony', 'request', 'luggage', 'lady', 'internet', 'gym', 'amount', 'rate', 'charge', 'money', 'family', 'street', 'smell', 'sauna', 'try', 'service', 'facility', 'renovation', 'elevator', 'euro', 'account', 'ventilation', 'carpet', 'face', 'start', 'building', 'train', 'minibar', 'taxi', 'order', 'slipper', 'tv', 'take', 'lobby', 'find', 'door', 'fix', 'help', 'air conditioner', 'toilet', 'offer', 'wash', 'air conditioning', 'speed', 'detail', 'closet', 'experience', 'sink', 'promise', 'double bed', 'connection', 'socket', 'park', 'pool', 'drink', 'menu', 'mention', 'paper', 'receptionist', 'sleep', 'phone', 'desk', 'london', 'reason', 'cause', 'person', 'guest', 'life', 'credit card', 'crowd', 'date', 'spa', 'deposit', 'repair', 'tea', 'pillow', 'client', 'neighbour', 'question', 'bathrobe', 'car', 'lock', 'hall', 'comfort', 'control', 'city center', 'conditioner', 'business', 'cup', 'neighbor', 'case', 'hot water', 'construction', 'return', 'heating', 'air', 'access', 'result', 'dish', 'visit', 'travel', 'laundry', 'mattress', 'tram', 'year'
```

Esta es una muestra de más de doscientos candidatos a ser aspectos de un hotel. Algunos términos, efectivamente, son características y servicios de un hotel (*room, shower, minibar, air conditioning, sauna, receptionist*, etc.), incluso podemos ver términos relacionados con estas características, como *noise* o *smell*. Sin embargo, Beth y Joseph consideran que es un número excesivo, y más teniendo en cuenta que hay muchos candidatos que no se pueden interpretar como aspectos de un hotel (*trip, lady, family, London, visit, tram, hour, hair*).

Joseph se sorprende de que se hayan «colado» tantos términos que no tienen que ver con hoteles. En la bibliografía que él ha leído, cuando se analizan opiniones sobre *smartphones*, y más cuando se publican en Twitter, no suele haber tantos aspectos falsos. Sin embargo, a Beth no le sorprende tanto. A ella le encanta viajar y suele publicar opiniones de los hoteles donde se ha alojado. Sabe que, cuando uno se queja, suele explicar las circunstancias que la han llevado a una situación frustrante. Por ejemplo, una vez publicó una queja explicando al principio que cayó un aguacero en la ciudad, que se mojó todo el pelo y la ropa, que no llevaba paraguas, y que un coche a toda velocidad pasó por un charco dejándola todavía más empapada. Cuando llegó al hotel tuvo la desgradable experiencia de no poder ducharse con agua caliente ni secarse el pelo porque el secador de pelo no funcionaba. Por tanto, en su queja, Beth hizo referencia a una ciudad, un aguacero, su pelo, su ropa, un paraguas, un coche y un charco antes de hacer mención a los aspectos del hotel causantes de la queja. Esto no suele pasar cuando uno opina sobre un *smartphone* y la restricción de caracteres de Twitter obliga a centrarse en los aspectos y no permite desarrollar todo un relato previo a la queja.

Por tanto, cualquier término puede estar cerca de *hotel* en el modelo Word2Vec, pero la cercanía en un modelo de opiniones sobre hoteles no garantiza que este término sea uno de sus aspectos. Joseph tiene la esperanza de que las *opinion words* serán discriminantes de los aspectos. Los aspectos verdaderos tendrían que estar relacionados con adjetivos como *dirty, expensive, noisy*, etc. Beth le recuerda, sin embargo, que hay *opinion words* como *awful* que se pueden aplicar tanto a las características de un hotel como a una ciudad y también al viaje en su conjunto.

En definitiva, que el tema de detectar los aspectos no es tan fácil como parecía al principio.

### **Elección de los candidatos a aspectos con un modelo de lengua general**

Hemos visto que los términos más próximos a *hotel* en uno de los modelos de opiniones es una lista de términos heterogéneos. Joseph también prueba con *bathroom* y le salen términos como *ceiling* y *pillow*, que, a pesar de superar el umbral de cercanía según el modelo, no son próximos según el sentido común.

Beth ve en ello una consecuencia del hecho de que las opiniones sobre hoteles pertenecen a un dominio permeable en el cual entran términos de otros ámbitos de la vida cotidiana. Las almohadas no son una característica de los hoteles, están en otros sitios, aunque en las opiniones sobre hoteles los términos *bathroom* y *pillow* aparezcan en contextos parecidos. Así pues, la cercanía según el modelo no siempre lo es según nuestro sentido común.

Cuando Joseph y Beth pensaban que habían llegado a un callejón sin salida, Joseph se acuerda de que Google deja disponible un modelo Word2Vec de noticias en inglés. Este modelo se hizo a partir de un corpus de noticias de Google News, y contiene tres mil millones de términos (3,6 GB)<sup>3</sup>. Joseph cree que, al ser un corpus de uso de la lengua sobre muchos temas, este recurso les permitirá discriminar los aspectos más relacionados con el sentido común.

(3) Podéis descargarlos el modelo en este enlace: <https://github.com/mmihaltz/word2vec-Google-News-vectors>

En la tabla 1 se puede ver la comparación entre los términos más relacionados con *hotel* y *bathroom* según el modelo de opiniones y los términos más relacionados con *hotel* y *bathroom* según el modelo de Google News. Se puede comprobar que los términos más próximos según el modelo de Google News son los más reconocibles por el sentido común.

Viendo esto, deciden tomar como candidatos los sintagmas nominales de las opiniones que aparecen más de diez veces y que están próximos al término *hotel* en el modelo de Google News. Lo significativo es que, con este modelo, Joseph puede bajar el umbral de proximidad para tener resultados adecuados. En la figura 12 se muestran los términos más próximos a *hotel* con un umbral de 0,3. Como se puede ver, el número de candidatos se ha reducido considerablemente y la mayoría de ellos hacen referencia a aspectos de un hotel (*bar*, *bathroom*, *breakfast*, *restaurant*, *sauna*, *receptionist*, etc.).

Figura 12. Candidatos a aspectos de un hotel en las opiniones con el modelo de Google News

'room', 'bathroom', 'cafe', 'breakfast', 'restaurant', 'bar', 'bed', 'balcony', 'luggage', 'sauna', 'elevator',  
'minibar', 'taxi', 'lobby', 'receptionist', 'guest', 'spa', 'hall', 'travel', 'laundry'

Ahora bien, hay que tener en cuenta que el vocabulario del modelo de Google News no recoge términos multipalabra, con lo cual no se ha podido calcular la distancia de los términos *multiword* que ya habían sido recogidos en el vocabulario de opiniones. Beth y Joseph deciden añadirlos a la lista de candidatos. Los términos *multiword* que habían obtenido pero que no están en el modelo de Google News son los siguientes.

Figura 13. Candidatos *multiword* en el modelo de opiniones que no están en el modelo de Google News

'ground floor', 'air conditioner', 'air conditioning', 'double bed', 'credit card', 'city center', 'hot water'

La filtración de los candidatos a aspecto se hará por su relación con las *opinion words*. Puesto que esta relación debe calcularse para ver qué aspectos son valorados más negativamente, deciden abordar esta fase directamente con la lista de candidatos obtenidos, excepto *city center*, que descartan de entrada como aspecto.

Tabla 1. Términos parecidos a *hotel* y *bathroom* en los modelos de opiniones y de Google News

<b><i>Hotel</i></b>	
Modelo opiniones	[('order', 0,9998066425323486), ('hour', 0,9997938871383667), ('ready', 0,9997932314872742), ('time', 0,9997906684875488), ('minute', 0,9997877478599548), ('reception', 0,9997860789299011), ('ask', 0,9997844696044922), ('receptionist', 0,9997807145118713), ('anything', 0,9997801780700684), ('paid', 0,9997779726982117)]
Modelo Google News	[('hotels', 0,7709728479385376), ('Hotel', 0,7388719916343689), ('motel', 0,661712110042572), ('boutique_hotel', 0,660549521446228), ('Holiday_Inn', 0,6501166820526123), ('Radisson', 0,6451879739761353), ('luxury_hotels', 0,6348958015441895), ('Hilton_Hotel', 0,6312208771705627), ('hotel_rooms', 0,628795325756073), ('Comfort_Inn', 0,6243367791175842)]
<b><i>Bathroom</i></b>	
Modelo opiniones	[('ceiling', 0,9998364448547363), ('size', 0,9998249411582947), ('quite', 0,9998235106468201), ('carpet', 0,9998181462287903), ('pillow', 0,999807596206665), ('furniture', 0,9998031258583069), ('tiny', 0,9998018741607666), ('door', 0,9997978806495667), ('room', 0,9997972846031189), ('enough', 0,9997962117195129)]
Modelo Google News	[('restroom', 0,8025195598602295), ('bathrooms', 0,7190953493118286), ('lavatory', 0,6996769905090332), ('washroom', 0,6982483863830566), ('toilet', 0,6814740896224976), ('kitchen', 0,6556941270828247), ('downstairs', 0,6407461762428284), ('shower', 0,6207696199417114), ('bedroom', 0,617497444152832)]

### 3.1.6. Aspectos vinculados a *opinion words* de polaridad negativa

Para conocer los aspectos vinculados a *opinion words* de polaridad negativa, primero se toman como *opinion words* de polaridad negativa las palabras del diccionario AFINN que tienen un valor menor que cero. Luego, siguiendo el método explicado en 2.2.2, se crea un vocabulario con los aspectos y las *opinion*

*words* y luego un vectorizador crea, para cada término del vocabulario, un vector con las similitudes de este término respecto a los demás términos. Una vez creados los vectores de similitud de cada término del vocabulario, estos se agrupan para encontrar los aspectos más vinculados a *opinion words* negativos. La figura 14 muestra la agrupación de los aspectos vinculados a *opinion words* negativos en las opiniones de clientes rusos. Los aspectos están agrupados en cinco clases.

Figura 14. Agrupaciones de los aspectos más vinculados a *opinion word* negativos en opiniones de clientes rusos

CLASS0:	['cafe', 'breakfast', 'bar', 'bed', 'weird', 'pay', 'bad']
CLASS1:	['room', 'restaurant', 'luggage', 'lobby', 'guest', 'laundry', 'air conditioner', 'uncomfortable', 'problem', 'strange', 'limited', 'empty', 'noisy', 'leave', 'lobby', 'inconvenient', 'poor', 'broken', 'terrible']
CLASS2:	['minibar', 'travel', 'credit card', 'block', 'weak', 'wrong']
CLASS3:	['bathroom', 'balcony', 'sauna', 'elevator', 'taxi', 'receptionist', 'spa', 'hall', 'air conditioning', 'double bed', 'awful', 'dirty', 'horrible']
CLASS4:	['ground floor', 'hot water', 'hard']

Se puede ver que los términos más vinculados a *opinion words* negativos hacen referencia a aspectos relacionados con la restauración (*breakfast, restaurant, cafe, bar*), los servicios relacionados con el *spa* y la sauna, las camas y el aire acondicionado, y también apuntan a temas de movilidad (*luggage, elevator, taxi*).

### 3.1.7. Informe

A partir de la tabla comparativa de los *clusters* obtenidos con el análisis de las opiniones de clientes rusos y españoles, Joseph y Beth redactan un informe según el cual se pueden establecer los siguientes aspectos sensibles para los clientes de ambas nacionalidades:

- servicios de restauración (restaurante, bar, servicio de desayuno, etc.)
- elementos de la habitación (tipo de cama, aire acondicionado, baño, etc.)
- servicios de sauna y *spa* (sobre todo el tema del agua caliente es importante)
- gestión del equipaje
- movilidad dentro del hotel y transporte (por ejemplo, al aeropuerto)

De las *opinion words* se deduce que los clientes rusos y españoles se fijan en la limpieza, la comodidad, el ruido y si los aparatos funcionan.

Figura 15. Comparativa de *clusters* de las opiniones de clientes rusos y españoles (en rojo los términos comunes y en azul, los aspectos distintos)

Clientes rusos
CLASS0: ['cafe', 'breakfast', 'bar', 'bed', 'weird', 'pay', 'bad'] CLASS1: ['room', 'restaurant', 'luggage', 'lobby', 'guest', 'laundry', 'air conditioner', 'uncomfortable', 'problem', 'strange', 'limited', 'empty', 'noisy', 'leave', 'lobby', 'inconvenient', 'poor', 'broken', 'terrible'] CLASS2: ['minibar', 'travel', 'credit card', 'block', 'weak', 'wrong'] CLASS3: ['bathroom', 'balcony', 'sauna', 'elevator', 'taxi', 'receptionist', 'spa', 'hall', 'air conditioning', 'double bed', 'awful', 'dirty', 'horrible'] CLASS4: ['ground floor', 'hot water', 'hard']
Clientes españoles
['reservation', 'restaurant', 'bar', 'suite', 'bedroom', 'lobby', 'receptionist', 'luggage', 'taxi', 'bad', 'noisy', 'hard', 'uncomfortable', 'poor', 'empty', 'problem', 'limited', 'stop', 'pay', 'lobby', 'leave', 'awful', 'broken', 'terrible', 'wrong', 'horrible', 'strange'] ['worst'] ['difficult', 'negative'] ["concierge", "amenity", "airport", "waiter", "ground floor", "hot water", "neighbour", "disappointed", "cancel"] ['room', 'bed', 'breakfast', 'elevator', 'bathroom', 'guest', 'spa', 'sauna', 'lounge', 'terrace', 'travel', 'air conditioning', 'double bed', 'weird', 'weak', 'dirty', 'lack', 'badly']

Entre los aspectos presentados anteriormente, el minibar, el café y el balcón de la habitación merecen una especial atención entre los clientes rusos. Por su parte, el *lounge* y las terrazas merecen especial atención para los clientes españoles.

Los aspectos distintivos de los clientes rusos son:

- el tipo de pago (tarjeta de crédito)
- la lavandería

Por su parte, los aspectos distintivos de los clientes españoles son:

- la reserva
- los vecinos

Cabe destacar que, sobre todo entre los clientes españoles, es importante que no se sientan decepcionados ante las expectativas generadas tras ver las características anunciadas por el hotel.

El informe concluye que, a pesar de las diferencias, hay muchos aspectos en común cuya supervisión y mejora satisfarán tanto a los clientes rusos como a los españoles.

### 3.2. Opiniones sobre un programa de televisión

El lunes 4 de abril de 2016 se emitió por la noche, en el Canal 33 de la Televisió de Catalunya, un programa dedicado al tema de la prostitución voluntaria. Este fue uno de los programas de una serie llamada *La gent normal*, que se emitía en aquella cadena semanalmente. Durante este programa unas mujeres

que se dedicaban a la prostitución de manera voluntaria explicaban su día a día, las relaciones con los clientes, cómo habían decidido ellas mismas dedicarse a la prostitución, etc. La finalidad era mostrar que había mujeres que se dedicaban a la prostitución empoderadas, autónomas, a quienes les gustaba ejercerla y que reclamaban que se reconocieran sus derechos como a cualquier otra persona que se dedica a otra profesión.

#### Nota

Este caso de uso está inspirado en un estudio que se hizo sobre este programa, dentro del proyecto PROMEDIAR, liderado por Anna Maria Clua, de los Estudios de Ciencias de la Información y de la Comunicación de la UOC, y que se encuentra publicado en: Anna Clua Infante; Joaquim Moré (2016). «El debate público sobre prostitución. Estudio de caso de la repercusión de un programa televisivo en la esfera Twitter». En: Ramón Cota-relo García (coord.); Javier Gil (comp.). *Ciberpolítica: Gobierno abierto, redes, deliberación, democracia* (págs. 291-308). Madrid: INAP.

Televisió de Catalunya encarga a la empresa S&S un estudio sobre el impacto del programa. El estudio quiere comprobar si la temática provocó un debate en Twitter y qué cariz tuvo este, si favorable o, por el contrario, negativo hacia la cadena y hacia las personas que aparecieron en el programa defendiendo la prostitución voluntaria. Peter decide que este estudio lo realicen Anne y Henry.

Anne y Henry deciden realizar el estudio con la siguiente planificación:

- obtención de los tuits
- exploración de los tuits para recopilar datos sobre:
  - tema,
  - aspectos del tema,
  - *opinion holders*,
  - polaridad según los *opinion holders*,
  - tiempo en que se publicaron los tuits,
  - preprocessado.
- análisis de los datos
- informe

#### 3.2.1. Obtención de los tuits

Televisió de Catalunya creó la cuenta @lagentnormal y un *hashtag* con el nombre del programa (#lagentnormal). Anne y Henry deciden recoger los tuits que contienen este *hashtag* publicados desde el día en el que la cadena empezó la promoción del programa, una semana antes, hasta tres días después de su emisión.

De entre las distintas librerías que hacen peticiones a la API de Twitter para obtener tuits, Anne y Henry eligen la librería de Python llamada Tweepy. En la figura 16 se puede ver un ejemplo de *script* para realizar peticiones. Es importante señalar que en el *script* se han encriptado las credenciales que son necesarias para que un desarrollador pueda hacer peticiones a la API de Twitter<sup>4</sup>.

<sup>(4)</sup>Para crear las credenciales, consultad <https://stackoverflow.com/questions/1808855/getting-new-twitter-api-consumer-and-secret-keys>

Anne y Henry deciden poner *lagentnormal* en la *query* de búsqueda, sin el símbolo de *hashtag*, porque de este modo podrán tener también los tuits que hacen referencia a la cuenta del programa @lagentnormal. Así obtienen un corpus de análisis con unos 2.200 tuits.

### 3.2.2. Exploración

Anne y Henry deciden que los datos con los que trabajarán serán los siguientes:

- momento de la publicación del tuit
- cuenta del usuario que publica el tuit o lo retuitea
- texto del tuit

El momento de la publicación del tuit es interesante para hacer un seguimiento temporal de las reacciones al contenido del programa. La cuenta del usuario es un dato que permite conocer el *background* del opinador y también las interrelaciones entre los usuarios que participaron en un debate que está todavía por demostrar si se produjo. El texto del tuit es la fuente de análisis de las opiniones que los participantes intercambiaron. En la figura 16 se ve cómo imprimir, precisamente, los textos de los tuits obtenidos.

Figura 16. Script para obtener mil tuits sobre *sentiment analysis*

```
import tweepy
CONSUMER_KEY = 'xxxxx'
CONSUMER_SECRET = 'xxxxxx'
ACCESS_KEY = 'xxxxxxxx'
ACCESS_SECRET = 'xxxxxxxx'

#Autentificación
auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
auth.set_access_token(ACCESS_KEY, ACCESS_SECRET)
api = tweepy.API(auth)

#Búsqueda de twits
query = 'sentiment analysis' # Término que se quiere buscar
max_tweets = 1000 #Número máximo de twits a obtener

#Obtener resultados
searched_tweets = [status for status in tweepy.Cursor(api.search, q=query).items(max_tweets)]
#Lista de resultados

#Imprimir los textos de los twits
for st in searched_tweets:
    print(st.text)
```

Anne y Henry se reparten la exploración de la siguiente manera: Anne se ocupa de la exploración de los textos y Henry se ocupa de las cuentas de los opinadores y del tiempo de publicación.

### **Exploración de los textos**

Los textos de los tuits retuiteados se repiten en los resultados de la búsqueda de tuits sobre el programa. Henry opina que hay que mantener los textos repetidos porque el conteo de términos en los mensajes repetidos será una buena métrica para valorar su relevancia. Cree que cuanto más se repite un término en los mensajes, incluyendo los retuiteados, más relevante es este término en el debate.

Anne propone procesar aparte los tuits originales (los que no tienen la expresión regular *RT*), con el fin de detectar a su autor y descubrir un patrón de los usuarios activos en el debate, frente a los pasivos, que son los que se limitan a retuitear. Henry está de acuerdo.

En los textos de los mensajes hay URL y *emojis* que Anne y Henry están de acuerdo en eliminar. Anne tiene dudas sobre si convertir el texto a minúsculas porque los cambios mayúscula-minúscula son un recurso para expresar sentimiento y emotividad. Henry propone unificar el texto y ponerlo en minúsculas, ya que el análisis está más centrado en los términos que aparecen que en la intensidad con la que se expresan. Además, se ha dado cuenta de que un mismo *hashtag* puede estar en mayúscula o minúscula (p. ej.: *#prostitución*, *#Prostitución*).

De todos modos, se plantean cuestiones nuevas, relacionadas con la lengua de los mensajes. Los mensajes están escritos en catalán pero hay también mensajes en español. Esto significa que hay que asegurarse de que los tuits están en una codificación unicode (UTF-8) y que no se han perdido las palabras con vocales acentuadas o con diéresis. Si no es así, hay que transformar el texto a UTF-8. Ahora bien, hay que ir con cuidado cuando se eliminan los *emojis*, pues también están en codificaciones distintas, y hay que verificar que tras la eliminación de los *emojis* las palabras conservan los acentos y las diéresis.

Otra cuestión relacionada con el uso de dos lenguas afecta a la normalización de los datos. Dos palabras distintas en los dos idiomas se contarán por separado aunque denoten el mismo concepto. La solución sería normalizar los textos traduciendo todos los mensajes a una sola lengua. Sin embargo, Henry, ante el volumen de tuits que habría que traducir, cree que la traducción manual no es abordable. Se debería recurrir a la traducción automática. Un detector automático de lengua detectaría la lengua del mensaje y la traduciría automáticamente a la lengua destino, si la lengua de origen y la de destino no coinciden.

Anne cree que si se adopta esta solución la probabilidad de añadir ruido a los datos es muy elevada. Teniendo en cuenta que muchos tuits tienen faltas de ortografía y errores de tecleo, el resultado de la traducción automática podría ser muy confuso. Por ejemplo, en catalán *espero que estiguis bé* se traduce al español como *espero que estés bien*, pero si el autor del mensaje olvida la tilde (*be*), algo que suele pasar muy a menudo, la traducción sería *espero que estés cordero*. Por ello, Henry y Anne confían en que la proximidad entre las dos lenguas permita identificar los datos relevantes para el análisis.

No poner acentos es una práctica general en la escritura de tuits y opiniones en los foros, pero esto afecta a la normalización de los datos. Es evidente que hay que normalizar el término *prostitución*, independientemente de que se ponga el acento o no. Por esta razón, Henry decide aplicar un corrector automático.

Los *hashtags* y nombres de cuenta también son un tema de debate. Como ya se ha comentado, hay *hashtags* que no son etiquetas al comentario, sino que son constituyentes sintácticos que no pueden tratarse aparte. Veamos el siguiente ejemplo:

#### Ejemplo de *hashtag*

Mujeres que hacen de la #prostitución su medio de vida porque así lo quieren.

Si se trata #prostitución aparte, se perdería un constituyente sintáctico necesario para entender la oración. Anne y Henry deciden tratar los *hashtags* como términos. De todos modos, Anne se da cuenta de que hay *hashtags* que aglutinan palabras que pueden ser procesadas por separado. Se trata de *hashtags* como, por ejemplo, #yonosoycómpline. Henry prefiere mantener el *hashtag* como está e incrementar en 1 la frecuencia de los términos aglutinados, esto es, se incrementaría en 1 la frecuencia de *cómpline*.

Por su parte, las cuentas se consideran datos personales cuya privacidad debe salvaguardarse. Por esta razón, se decide eliminarlas.

### Cuentas de los opinadores

Henry aplica el *script* que obtiene los perfiles de usuario de los opinadores. Por suerte, puede tener los perfiles públicos de la gran mayoría de los entrevistados y, también por suerte, la mayoría de los perfiles tienen datos sobre la profesión, las aficiones y los movimientos sociales a los que pertenecen.

En los perfiles se pueden encontrar casos de sustitución de palabras por *emojis*. Por ejemplo, *mujeres* y *hombres* en:

Figura 17

Feminismo defiende q las 🚨 deben tener los mismos DD reales y efectivos que los 🚨

#### Enlace de interés

Una librería Python que sirve para corrección automática es *pyenchant*. <https://github.com/rfk/pyenchant>.

#### Ved también

Para más información, vé el apartado 2.4.

También hay *hashtags* que son constituyentes de la oración y expresiones reducidas.

### Tiempo de publicación

El momento de la publicación está en el formato de Twitter, que es el siguiente:

Tabla 2

Nombre del día	Nombre del mes	Dígito del día	Hora: Minutos: Segundos:	+0000	Año

### Tiempo de publicación en Twitter

Fri Apr 08 01:27:13 +0000  
2016

Puesto que el objetivo del estudio no es tener un seguimiento de las opiniones al minuto, Henry decide tomar solamente el día y el mes.

### 3.2.3. Preprocesado

Los textos pasan por el corrector automático, se cambia el texto a minúsculas, y los acrónimos, los *emojis* y emoticonos se sustituyen por las palabras que representan. Se quita la expresión regular RT con el nombre de la cuenta del autor del tuit retuiteado, y también se quitan los nombres de las demás cuentas.

En los lugares donde se han eliminado *tokens* se pone una marca que lo indica. De este modo evitamos que el cálculo de n-gramas para encontrar colocaciones se haga con falsas palabras consecutivas. La marca se añade a la lista de *stopwords* del catalán y el español. También se le quita el símbolo de la almohadilla al *hashtag*.

Para el análisis de los perfiles se aplica un anonimizador que oculta el nombre y otros datos personales del usuario. El anonimizador sustituye los datos sensibles (DNI, cuentas bancarias) por unos códigos. Con un *script* de Python u otro lenguaje de programación, aprovechando su librería de expresiones regulares, se puede hacer la sustitución fácilmente. Lo complicado es cuando no se dispone de una lista de los datos sensibles ni cuando los datos son expresiones regulares. Sin embargo, hay herramientas que ayudan a detectar datos sensibles. Por ejemplo, las herramientas que realizan *named-entity recognition* (*NER recognition*), esto es, sistemas que detectan referencias a personas, entidades, calles, etc. *Parsers* como FreeLing tienen incorporado un *NER recognition*<sup>5</sup>. Es el sistema que utilizan Anne y Henry.

### Enlace de interés

En <https://github.com/Xangis/extrá-stopwords> se pueden obtener listas de *stopwords* extras a las que ya tiene NLTK, entre ellas el catalán.

<sup>(5)</sup>NER recognition <https://talp-upc.gitbooks.io/freeling-4-0-user-manual/content/modules/ner.html>.

### 3.2.4. Análisis de los datos

Los datos básicos son los términos. Dado que los textos que deben analizarse son multiidioma, Henry decide buscar los términos aplicando métodos independientes de la lengua. Esto supone no aplicar el método de detección de *patterns* sintácticos, ya que el *parser* tiene que configurarse de acuerdo con el

<sup>(6)</sup>Anne y Henry pueden filtrar los unigramas que son *stopwords* en español. No disponen de una lista de *stopwords* del catalán (p. ej.: són, què).



la difusión e impacto de este tuit. Por lo tanto, Henry decide aplicar análisis de redes para detectar a los usuarios más influyentes y concretar más el perfil apuntado por el *word cloud*.

El análisis de redes queda fuera del alcance de este curso, por lo que aquí no explicaremos con detalle cómo se detectan los participantes más influyentes. Pasamos directamente a presentar los perfiles de estos usuarios, que son los siguientes:

- 1) Los usuarios más influyentes y numerosos provienen, en su mayoría, del movimiento feminista.
- 2) Una usuaria muy influyente no explica su pertenencia a un movimiento o partido político concreto. Ahora bien, de sus tuits se desprende su activismo en la defensa de los derechos sociales y laborales de las mujeres que se dedican a la prostitución y de otros colectivos.
- 3) Solo una usuaria declara en su perfil que se dedica a la prostitución.
- 4) Una persona muy influyente ocupa la alcaldía de un municipio. Anne y Henry lo encuentran sorprendente, porque *a priori* no creen que el tema deba concernirle tanto como para tener un papel tan activo en la controversia

Henry y Anne deciden agrupar los tuits originales según tres grupos de usuarios, que son los siguientes: usuarios de perfil político, usuarios de perfil feminista y usuarios que se dedican a la prostitución. Como se puede ver en la gráfica de la figura 19, dominan los usuarios de perfil político y feminista, mientras que los usuarios que se dedican a la prostitución son minoría.

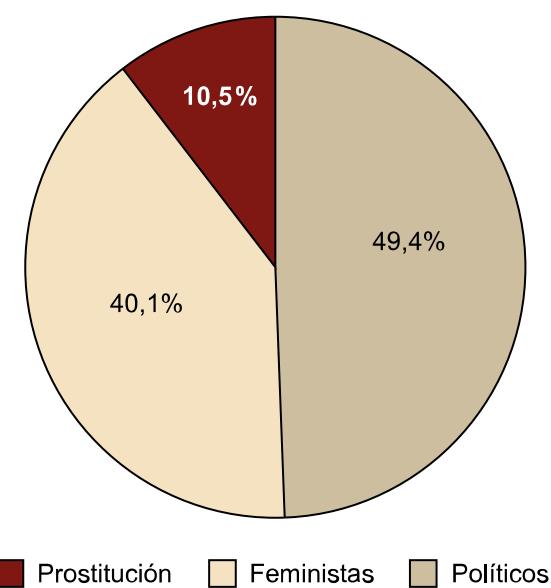
### Análisis de los temas y sus aspectos

La agrupación de los tuits originales según los usuarios sirve para responder a la siguiente pregunta: ¿los temas y los aspectos de los comentarios son distintos según el grupo al que pertenece el participante?

### Bibliografía recomendada

Para más información sobre el análisis de redes, consultad: C. Pérez; J. Casas (2016). «Análisis de datos de redes sociales». Barcelona: Editorial UOC.

Figura 19. Distribución de los tuits originales según los grupos de usuarios



Henry vectoriza los tuits originales de cada grupo con un *tf.idf vectorizer* para encontrar los términos más relevantes en cada grupo.

La figura 20 muestra los términos más relevantes del grupo de los políticos, ordenados por su valor de tf.idf.

Figura 20. Términos más relevantes en los tuits del grupo de los políticos

Término	TfIdf
tv3amagalatrata	0.401647
prostitució	0.355829
lagentnormal33	0.237219
què	0.213497
parlarà	0.200823
cara	0.160659
día	0.122185
pública	0.122185
veu	0.122185
normalitza	0.120494
trata	0.094888

Fijémonos en que el término más relevante en los tuits de las personas vinculadas a la política es *tv3amagalatrata*. Este es un *hashtag* cuya comprensión requiere la explicación del contexto en el que se emitió el programa.

La televisión pública catalana emitió el programa el Día Internacional contra la Explotación Infantil. Esta coincidencia motivó que se relacionara el tema del programa –la prostitución– con la explotación sexual. Esto explica que aparezca el término *trata* y que *trata* esté en el *hashtag* más relevante. Un *hashtag* de denuncia a la televisión porque, según los opinadores que lo incluyeron en

sus tuits, el contenido del programa se podía interpretar como una forma de normalizar y esconder el drama que es la prostitución en todas sus manifestaciones. Un mensaje que no toleraban, sobre todo vieniendo de una televisión pública.

La figura 21 muestra los términos más relevantes del grupo feminista, ordenados por su valor de tf.idf.

Figura 21. Términos más relevantes en los tuits del grupo feminista

Término	TfIdf
lagentnormal33	0.545041
què	0.292461
prostitució	0.225993
tv3complicetratamenors	0.225082
tv3complicettractamenors	0.222535
prostitución	0.172818
dones	0.154062
protesta	0.135049
mujeres	0.132937
són	0.119826
infantil	0.106350
trata	0.106350
día	0.102708

Vemos que también se manifiesta la protesta y la indignación contra la televisión catalana. Esta vez se la acusa de ser cómplices de la trata de menores. También, y en coherencia con la naturaleza del grupo de opinadores, se relaciona el tema infantil con el de las mujeres.

Finalmente, la figura 22 muestra los términos más relevantes del grupo de personas relacionadas con la prostitución, ordenados por su valor de tf.idf.

Figura 22. Términos más relevantes en los tuits del grupo de personas relacionadas con la prostitución

Término	TfIdf
lagentnormal33	0.480986
putafeminista	0.296138
putafemista	0.259121
què	0.174904
putas	0.112610
abolicionistas	0.111052
censurar	0.111052
derechos	0.111052
estigma	0.111052

El foco no está en la denuncia al programa sino en las personas que se denominan feministas abolicionistas, que ven la prostitución como una forma de explotación en todas sus manifestaciones y que son favorables a no exponer un planteamiento contrario en la televisión pública.

También vemos un término específico de este grupo –*putafeminista*– que se refiere a las mujeres que se dedican a la prostitución y que se sienten empoderadas y libres de practicarla, que era precisamente el tema del programa. Más que a la explotación, hacen referencia al estigma social que para ellas representa la visión de la prostitución de las abolicionistas.

### Análisis de la polaridad

En los tres grupos la polaridad es negativa. El grupo de los políticos y de las feministas opinan negativamente sobre el programa y sobre la cadena de televisión que lo emite. Incluso difunden *hashtags* que tienen apariencia de consigna de campaña de des prestigio (*tv3amagalatrata*). La polaridad de los tuits del grupo de las personas relacionadas con la prostitución también es negativa. Opinan negativamente sobre las abolicionistas porque las censuran y fomentan el estigma social de las personas que se dedican a la prostitución.

Anne ha distinguido de los tuits de cada grupo las *opinion words* positivas y negativas. Ha utilizado como recurso la traducción al catalán y al castellano del diccionario de Hiu y Liu. Según este diccionario, la palabra *estigma* tiene una polaridad negativa, que es consecuente con la queja del grupo de las mujeres que se dedican a la prostitución.

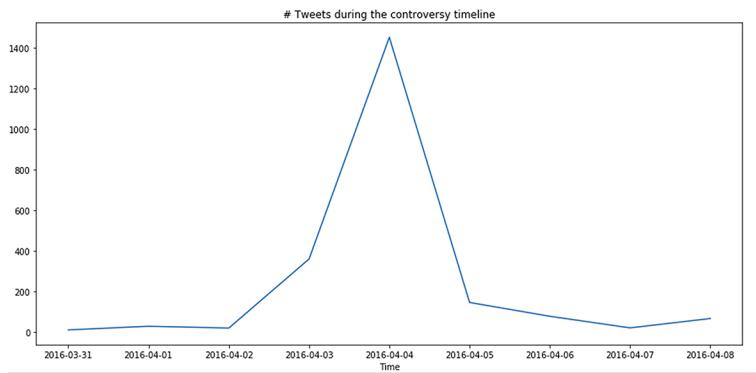
Ahora bien, algo que le llama la atención a Anne es que el diccionario no recoge palabras de las opiniones que ella considera *opinion words*. Por ejemplo, *humillaciones*, *pedófilas*, *discriminatoria* y *mercantilización*. Por otra parte, también se da el caso contrario: *infantil* se toma como una palabra negativa porque el diccionario recoge el sentido de esta palabra equivalente a *irresponsable*, *inconsciente*, etc. Un sentido que no tiene nada que ver con el sentido que tiene *prostitución infantil*.

Anne también ve algo muy curioso. La palabra *puta* tiene una polaridad negativa según el diccionario, pero en las opiniones de las personas relacionadas con la prostitución esta palabra tiene una carga positiva, usada con orgullo por pertenecer a un colectivo. La palabra se relaciona con *empoderadas* (con un tf.idf de 0,07 en los tuits de las personas relacionadas con la prostitución, y sin valor en otros grupos). En resumen, el análisis de la polaridad con un diccionario de uso general no siempre se ajusta del todo al contexto de las opiniones.

## Análisis del tiempo

La siguiente gráfica muestra el número de tuits durante los días que duró la controversia. Evidentemente, el pico de tuits corresponde al día en el que se emitió el programa. Pero lo que llama más la atención es el hecho de que la controversia ya se iniciara un día antes de su emisión.

Figura 23. Tuits publicados durante los días que duró la controversia (ved notebook PLA-2, 1.5)



Henry echa un vistazo a los tuits y comprueba que el día antes ya se calentaban motores en contra del programa y de la cadena, y también en contra de los que defienden las posturas abolicionistas. Vemos aquí algunos ejemplos:

### Ejemplos de tuits iniciadores de una campaña de desprecio

Proxenetisme de "la gent normal"...

4 abril Dia contra la prostitució infantil @lagentnormal33 @tv3cat normalitza la prostitució com a professió

@tv3cat @lagentnormal33 la normalització ens involuciona com a societat!  
#tv3amagalatrata @TradxAbolicion

Instamos a @tv3cat a no ser cómplice de la trata de niñas

#Prostitución por voluntad propia. Que el abolicionismo no censure las voces de las trabajadoras sexuales.

qué os parece si mañana le damos caña al HT #PutaFeminista junto al HT del programa para comentar?

### 3.2.5. Informe

Henry y Anne redactan un informe en el que se explican los datos que hemos presentado anteriormente. Llegan a conclusiones que son de interés para los estudiosos en ciencias políticas, pero no las expondremos aquí. No obstante, sí que es relevante el dato que muestran sobre cómo se parecen o se alejan los tuits entre los grupos. Para obtener este dato, crean una colección de tres documentos, cada uno con los tuits de un grupo, los vectorizan con un *tf.idf vectorizer* y calculan la *cosine similarity* entre los vectores. El resultado es el siguiente.

Tabla 3

<b>Distancia políticos-feministas</b>	<b>Distancia políticos-personas relacionadas con la prostitución</b>	<b>Distancia feministas-personas relacionadas con la prostitución</b>
0,341	0,170	0,367

Como se ve, los mensajes de los políticos se acercan a los mensajes del grupo feminista, mientras que mantienen una posición más alejada respecto al grupo de las personas relacionadas con la prostitución. Sin embargo, la distancia entre el grupo de personas relacionadas con la prostitución y el grupo feminista es similar a la que tiene este último con los políticos. Esto apunta a que el debate se centró más en el discurso feminista, lo cual se confirmó al ver que en el total de tuits y retuits los términos relevantes de este grupo eran los más frecuentes. Por otro lado, demuestra que el hecho de que se hable más de las ideas de un grupo ideológico no significa que se simpatice con él. Muy a menudo, los grupos ideológicos recurren a la metarreferencia de los conceptos del grupo ideológicamente contrario para reafirmar sus propias posiciones.

### 3.3. Detección de opiniones falsas

La práctica muy extendida de consultar las opiniones antes de tomar una decisión nos ha hecho muy permeables al criterio de los demás. Conscientes del efecto que se puede producir cuando una crítica negativa se extiende por las redes sociales, entendiendo por redes sociales no solo Twitter y Facebook, sino también foros como TripAdvisor, hay opinadores que se dedican a publicar de forma deliberada opiniones falsas con el objetivo de dañar la reputación de una empresa, institución, etc. También existe el opinador que publica opiniones injustificadamente positivas con el fin de promocionar y mejorar la reputación en línea.

Es evidente el daño que pueden hacer las opiniones falsas. Una campaña orquestada en contra de una empresa, un político, etc., puede arruinar su reputación en muy poco tiempo. Sobre todo cuando la reputación está centrada actualmente en la reputación en línea. En cambio, si la campaña es positiva, aumentar el grado de expectativas que pronto se verán frustradas tampoco es beneficioso.

Detectar las opiniones falsas es, por tanto, una necesidad imperiosa y actualmente es una de las aplicaciones del *sentiment analysis* más demandadas.

En este subapartado nos ocuparemos del siguiente caso de uso:

Un buscador de hoteles, consciente del daño económico que puede suponer para su negocio la publicación de opiniones falsas interesadas en dañar la reputación de cadenas hoteleras, quiere aplicar un detector de opiniones falsas antes de publicarlas.

La empresa S&S es la que recibe este encargo. Peter, el coordinador del grupo de lingüística computacional, asigna a Beth y a Joseph este proyecto.

### 3.3.1. Tipificación del problema

Beth se ocupa de consultar la bibliografía sobre el tema y observa que, generalmente, se distinguen tres tipos de opiniones falsas (Jindal y Liu, 2008).

- 1) Opiniones que dan información no fiable sobre el producto con el fin de promover o, por el contrario, perjudicar y dañar su reputación.
- 2) Opiniones que se focalizan en la empresa, la marca, los fabricantes o los vendedores del producto, pero que no cuentan una experiencia con un producto concreto. Por ejemplo, *Odio Microsoft. Nunca comprará ningún producto que lleve esta marca.*
- 3) Opiniones que indirectamente hacen mención al producto en forma de anuncios no explícitos y textos que no contienen opiniones.

### 3.3.2. ¿Cómo abordar el problema?

Principalmente hay dos métodos. El primero es detectar opiniones falsas encontrando rasgos (*features*) de los textos de las opiniones. El segundo consiste en encontrar comportamientos de los opinadores que salen del patrón normal.

El primer método da soporte sobre todo a la detección de opiniones falsas del primer tipo, esto es, opiniones que responden a la intención de favorecer o perjudicar una reputación. También da soporte a las opiniones falsas del segundo tipo. Con este método, los dos tipos se tratarían detectando el *target* y sus aspectos y su cercanía con *opinion words*. Además, se analizarían otros rasgos, como la frecuencia de determinadas palabras y de categorías gramaticales que ya se han estudiado, como característicos de los mensajes *spam*. Sin embargo, las opiniones del tercer tipo, que no son explícitas, serían más difíciles de detectar.

Para Beth, los pros y contras del primer método son los siguientes:

Tabla 4

Pros	Contras
Se pueden aplicar los métodos de análisis de textos que ya se han aplicado en proyectos anteriores (titulares de <i>The New York Times</i> , opiniones sobre hoteles, controversia en Twitter).	No se tiene una idea muy clara de cuáles son las <i>features</i> distintivas de una opinión falsa.
Los elementos de la opinión no deberían ser difíciles de detectar.	Las opiniones falsas de tipo 3 son más difíciles de detectar.
En las opiniones de tipo 2, el contraste entre las menciones a la empresa en general pero no al producto es abordable.	En las opiniones de tipo 1 es difícil de detectar la fiabilidad.

### Bibliografía recomendada

N. Jindal; B. Liu. «Opinion Spam and Analysis». *Proceedings of First ACM International Conference on Web Search and Data Mining (WSDM-2008)*, 11-12 de febrero de 2008, Stanford University, Stanford, California, USA.

J. Piskorski; M. Sydow; D. Weiss (2008). «Exploring Linguistic Features for Web Spam Detection: A Preliminary Study». *AIRWeb '08 Proceedings of the 4th international workshop on Adversarial information retrieval on the web*.

El segundo método, centrado en detectar comportamientos en el usuario que se salen de lo normal, podría detectar opiniones falsas de los tres tipos, si la mención del *target*, sus aspectos y los rasgos lingüísticos y estilísticos de sus opiniones conformaran un modelo temporalmente dispar respecto al modelo del usuario a lo largo de un tiempo prolongado. La disparidad sería una consecuencia del comportamiento inusual del opinador. Los pros y contras del segundo método son los siguientes:

Tabla 5

Pros	Contras
Es una oportunidad para adentrarse en el mundo de la lingüística forense.	No tienen experiencia en lingüística forense ni en modelar comportamientos.
El modelo de comportamiento de un opinador que difunde opiniones falsas puede basarse en datos cuantitativos (número de opiniones en un espacio de tiempo, longitud de las opiniones, tiempo de conexión de la misma IP), que no requieren la complejidad del análisis de texto.	Hay un trabajo extra verificando que el modelo de comportamiento de los opinadores que difunden opiniones falsas es aplicable en cualquier escenario.
El análisis del comportamiento puede ir acompañado de acciones concretas que afectan solamente al que difunde opiniones falsas (bloquearlo, advertirlo, etc.). De todos modos, es difícil saber si son opinadores individuales o actúan en grupo.	Definir qué es lo que sale de lo normal significa también modelizar el comportamiento de los opinadores honestos. Y hay que preguntarse si los opinadores que siguen un patrón aceptable no son en realidad gente que también opina obedeciendo intereses de terceros.

La decisión sobre con qué método trabajarán depende sobre todo de los datos de los que disponen para trabajar. Tienen un conjunto de opiniones de polaridad positiva y negativa, ya clasificadas previamente como reales y falsas, pero sin información sobre el opinador. Por lo tanto, desarrollarán el trabajo analizando los datos del texto.

Ahora bien, hay una discrepancia entre Beth y Joseph. Beth quiere desarrollar procesos de análisis de texto, mientras que Joseph considera que, ya que las opiniones están clasificadas en verdaderas y falsas, se podría aplicar *machine learning* para entrenar directamente un clasificador de opiniones falsas. Beth dice que, en todo caso, el análisis de texto previo es imprescindible para saber con qué *features* se entrenará al clasificador. Joseph piensa que con unas *features* básicas, como la frecuencia de aparición o el tf.idf de los n-gramas, basta para clasificar las opiniones. Al menos probarlo. Al final deciden que Beth hará el análisis textual, Joseph aplicará las técnicas de *machine learning* y luego contrastarán resultados.

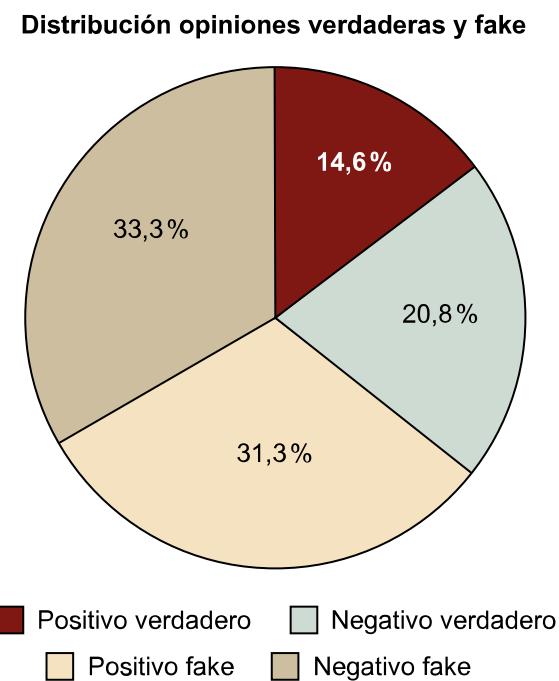
### 3.3.3. Análisis textual

#### Mención a la marca

Beth empieza por verificar si es cierto que la mención de la marca, el fabricante o el vendedor, sin mencionar un producto concreto, es un rasgo de las opiniones falsas. En el dominio en el que están trabajando, el de la hostelería, el equivalente a la marca es el nombre de la cadena hotelera. Beth considera como producto concreto los servicios (habitación, bar, wifi, etc.).

Beth toma las opiniones verdaderas y falsas (*o fake*) de veinte cadenas hoteleras, y las organiza de la siguiente manera: opiniones verdaderas de polaridad positiva, opiniones verdaderas de polaridad negativa, opiniones falsas de polaridad positiva y opiniones falsas de polaridad negativa. La proporción de estos tipos de opiniones que contienen el nombre de la cadena hotelera se muestra en la figura 24:

Figura 24. Proporción de opiniones verdaderas y falsas (*fake*) que mencionan la cadena hotelera



El resultado es lo suficientemente significativo como para que Beth no se ocupe de distinguir las referencias al nombre de la cadena hotelera de las referencias a los servicios. Las opiniones con más referencias al nombre de la cadena están mayoritariamente en las opiniones falsas. Es interesante notar que la mención a la cadena es más característica de las opiniones negativas, sean estas falsas o no.

## El *opinion holder*

Beth continúa analizando las opiniones con datos textuales. Ahora tomando datos sobre el *opinion holder*. Dado que las opiniones negativas falsas son las más nocivas, Beth se concentra en ver las diferencias entre las opiniones negativas falsas y las negativas verdaderas.

En estas opiniones los pronombres de primera persona indican el punto de vista del opinador. Beth sitúa estas opiniones en un espacio vectorial con un *tf.idf vectorizer*. A continuación presentamos la comparativa de distribución de los valores de tf.idf de los pronombres de primera persona en los dos tipos de opiniones.

Tabla 6. Valores de relevancia (tf.idf) de los pronombres de primera persona en opiniones negativas verdaderas y falsas

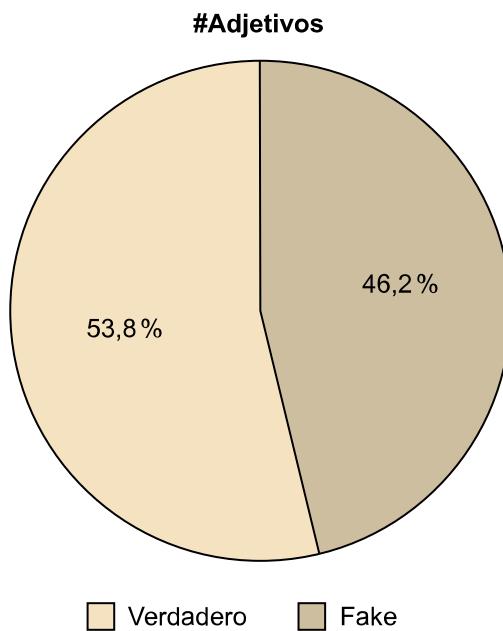
Pronombres	Tf.idf opiniones verdaderas	Tf.idf opiniones falsas
<i>I</i>	0,570	0,678
<i>me</i>	0,096	0,088
<i>my</i>	0,181	0,240
<i>myself</i>	0,002	0,004

Según esta tabla, vemos que los pronombres *I* y el posesivo *my* son más relevantes en las opiniones falsas negativas. Beth cree que esto se debe a que, en las opiniones falsas, la voluntad expresa de influir se manifiesta en una mayor presencia del «yo» en las opiniones.

## Polaridad

Los adjetivos son las palabras que tienen más probabilidad de expresar la polaridad de la opinión. Beth obtiene el número de adjetivos de los dos tipos de opiniones y se da cuenta de que el número de adjetivos es menor en las opiniones falsas, como se ve en la figura 25:

Figura 25. Proporción de adjetivos en las opiniones verdaderas y falsas (*fake*)



## Un poco de estilometría

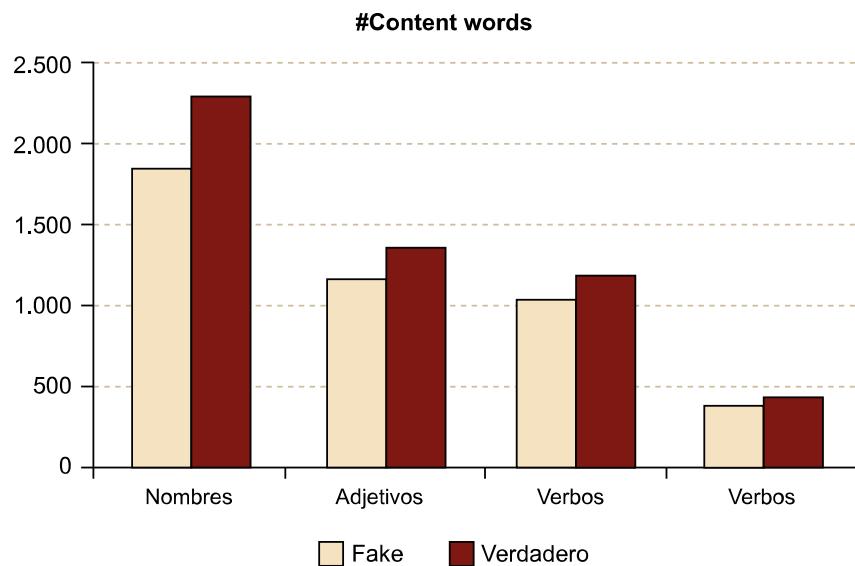
La estilometría es una disciplina que analiza textos con el fin de dar datos numéricos que son característicos de un estilo o de un tipo de texto. Los datos típicos de un análisis estilométrico son el número de adjetivos o adverbios, por ejemplo, o bien la frecuencia de las palabras, como datos explicativos de un tipo de texto. La estilometría se aplica para determinar la autoría de un texto, y en este caso Beth cree que se puede aplicar para caracterizar un cierto estilo en las opiniones falsas negativas.

Beth, tras analizar las referencias a la marca y la polaridad, tiene la intuición de que en las opiniones falsas no hay tantas *content words* (palabras con contenido semántico) como en las opiniones reales, esto es, hay más palabras que cohesionan el discurso pero no aportan contenido. Esta intuición la lleva a realizar un análisis de estilometría comparando, en los dos tipos de opiniones, el número de sustantivos, verbos, adjetivos y adverbios, tomados como *content words*. El resultado se muestra a continuación:

### Enlace de interés

Para una aplicación de la estilometría con una librería Python véase: <https://github.com/jpotts18/stylo-metry>.

Figura 26. *Content words* en las opiniones verdaderas y falsas (*fake*)



Como se puede ver, el número de *content words* en las opiniones verdaderas supera el número de *content words* en las opiniones falsas. El nombre es la categoría gramatical que supera en mayor medida a las *fake opinions*.

### Conclusiones del análisis textual

Del análisis textual de las opiniones *fake* negativas, realizado por Beth, se pueden sacar las siguientes conclusiones

- 1) En las opiniones falsas hay más menciones a la marca (el nombre de la cadena hotelera).
- 2) En las opiniones falsas hay más pronombres y determinantes posesivos de primera persona.
- 3) En las opiniones falsas hay sobre todo menos nombres.
- 4) En las opiniones falsas hay también menos adjetivos y adverbios, lo cual tiene incidencia en la expresión de la polaridad.
- 5) Se puede inferir que en las *fake opinions* hay más palabras que no son *content words*, esto es, preposiciones, artículos, etc.

Beth comenta estos resultados con Joseph y este cree que podrían ser *features* que tener en cuenta para el clasificador de opiniones falsas que ha estado elaborando. De todos modos, él ha trabajado con otras *features*, tal como explicaremos en el siguiente apartado.

### 3.3.4. Clasificador de opiniones falsas con *machine learning*

Joseph aprovecha las opiniones ya etiquetadas como opiniones falsas y opiniones verdaderas para entrenar un clasificador automático de opiniones falsas. Como Beth, se centra en la clasificación de opiniones negativas falsas por ser las más sensibles.

#### Establecimiento de las *features*

Joseph ha ido tomando nota de los comentarios que le ha hecho Beth para establecer las *features*. El hecho de que aparezcan más nombres y adjetivos en las opiniones falsas le hace considerar el tf (frecuencia de un término en un documento) como una *feature* que considerar.

Por otro lado, Joseph se da cuenta de que cuando trata unigramas, en el ránquin de términos con mayor tf.idf aparecen prácticamente los mismos términos (*room*, *hotel*, *stay*, *service*, *desk*, etc.) en los dos tipos de opiniones, salvo algunas excepciones en las opiniones verdaderas, como *bathroom*. Por esta razón, quiere añadir como *feature* la longitud de los n-gramas. Decide considerar n-gramas de hasta cuatro *tokens* para ver si el orden de los n-gramas es una *feature* importante a la hora de clasificar las opiniones.

La frecuencia de no *content words* en las opiniones falsas será una *feature* que considerar en el caso de que las *features* basadas en n-gramas no fueran suficientemente significativas para entrenar el clasificador.

#### Método de clasificación

Joseph decide aplicar el método bayesiano por ser un método que la literatura reconoce que funciona bien para tareas de *sentiment analysis*. Quiere probar si también es así para clasificar opiniones falsas y verdaderas. El resultado y su evaluación serán tratados en el siguiente módulo, dedicado a la evaluación del método de clasificación de tipos de opiniones.

## Resumen

Este módulo se ha dedicado a los métodos para extraer los sentimientos y las opiniones a partir de textos. En primer lugar, hemos presentado los elementos de una opinión y hemos ofrecido métodos de detección de estos elementos de forma automática.

Hemos explicado métodos de detección del tema, de los aspectos y de la polaridad aplicando los procedimientos tratados en el módulo anterior. Hemos aplicado métodos basados en ontologías y también en diccionarios. Ante la evidencia de que los aspectos y la polaridad dependen en gran medida del dominio temático de la opinión y también de su contexto, además de otros factores todavía no resueltos, como la intención del opinador por ser irónico, hemos profundizado también en la detección de los elementos de la opinión a partir de un modelo del dominio y con métodos de *machine learning*, clasificando la polaridad de las opiniones directamente de los datos, sin asunciones predeterminadas.

También hemos tratado el análisis de la opinión desde el punto de vista del *opinion holder*, aprovechando herramientas de análisis de las redes sociales y tomando en consideración el grado de influencia de los usuarios en la red social.

A continuación hemos presentado tres casos de uso. El primer caso ha sido el análisis de las opiniones de los clientes según su nacionalidad, con el fin de mejorar los servicios de un hotel. El segundo caso se ha centrado en una controversia en Twitter sobre el tema de un programa emitido por una cadena de televisión pública. Este caso ha servido para ejemplificar los métodos de detección de los elementos de una opinión y también como caso de uso ilustrativo de cómo se aplican los métodos explicados en el módulo anterior. Se ha visto que los elementos de una opinión son relevantes porque permiten descubrir cómo los participantes dominan el discurso y difunden el mensaje que quieren en una red social. El caso de uso también ejemplifica la importancia y el esfuerzo necesario para desarrollar el preprocesamiento de las opiniones. Una tarea costosa pero absolutamente necesaria para obtener datos fiables.

Finalmente, el tercer caso de uso ha sido la detección de opiniones falsas. Además de ser un tema que preocupa y mucho a todos los que dependen de una reputación en línea, este caso nos ha servido para introducir aspectos lexicométricos. También ha resultado útil para ver la necesidad de flexibilizar nuestro criterio para encontrar las *features* con las que entrenar un clasificador. Si estábamos acostumbrados a que las *features* estuvieran en las *content words*, en casos como la detección de opiniones falsas las palabras que no son *content*

*words* también son importantes, además de considerar como posibles *features* datos que no son tan lingüísticos como de muy bajo nivel (presencia de mayúsculas, o la longitud del n-gram, por ejemplo).

De todos modos, la flexibilización del criterio debe acompañarse de un método de clasificación que también sea flexible. Trataremos esta cuestión en el próximo módulo, cuando exemplifiquemos la mejora o el empeoramiento de los resultados de un clasificador con la adición de nuevas *features*.