

# Tratamiento de la Negación en el Análisis de Opiniones en Español\*

## *Negation Scope Identification in Spanish Reviews*

Salud M. Jiménez Zafra, Eugenio Martínez Cámara, M. Teresa Martín Valdivia,  
M. Dolores Molina González

Departamento de Informática, Escuela Politécnica Superior de Jaén

Universidad de Jaén, E-23071 - Jaén

{sjzafra, emcamara, maite, mdmolina}@ujaen.es

**Resumen:** El análisis de opiniones es una tarea a la que le quedan muchos frentes abiertos aún para que se pueda considerar resuelta. Entre ellos destaca el tratamiento de la negación, dado que una opinión negativa puede ser expresada con términos positivos negados. La negación es una característica particular de cada idioma, por lo que su tratamiento debe ajustarse a las singularidades del idioma en cuestión. En este artículo se presenta una aproximación lingüística para la identificación del ámbito de la negación en español, que se ha aplicado en un sistema de clasificación de la polaridad de opiniones sobre películas de cine.

**Palabras clave:** Análisis de la opinión, clasificación de la polaridad, identificación del ámbito de la negación

**Abstract:** Sentiment Analysis is a task that still has several opened challenges. One of those challenges is the treatment of the negation, because a negative opinion can be built using negated positive words. Negation is a particular feature of each language, thus it must be considered differently per each language. In this article is shown a linguistic approach for the negation scope identification with the aim of integrating it in a polarity classification system in the domain of movie reviews.

**Keywords:** Sentiment analysis, polarity classification, negation scope identification

## 1 Introducción

La Minería de Opiniones (MO) o Análisis de Sentimientos (AS) es una disciplina que combina técnicas de Procesamiento del Lenguaje Natural (PLN) y de Inteligencia Artificial y que está recibiendo bastante atención por parte de la comunidad científica seguramente por el amplio abanico de aplicaciones en las que se está empleando. La MO estudia el tratamiento de textos que incluyen información cargada de subjetividad. Muchos son los trabajos que tratan las distintas tareas y aplicaciones de la MO, pero a pesar de ello al AS todavía le queda un largo camino que recorrer para resolver muchos de los problemas que tiene aún abiertos (Liu, 2012). El presente trabajo se centra en uno de estos retos, la

identificación del ámbito de la negación. La negación es un elemento clave en el análisis de opiniones, dado que son muy abundantes las opiniones negativas expresadas con términos positivos negados y viceversa. La oración “No me gusta la carcasa del teléfono” es un claro ejemplo de una opinión negativa con un término positivo (*gusta*) negado.

En el ámbito del AS, la mayor parte de la investigación se ha realizado para opiniones escritas en inglés. Sin embargo, la presencia cada vez mayor en Internet de otros idiomas pone de manifiesto la necesidad de desarrollar sistemas que traten lenguas diferentes al inglés. Así han aparecido algunos trabajos de investigación que tratan con textos escritos en chino (Zhang et al., 2009), en árabe (Rushdi-Saleh et al., 2011) o en español (Martín-Valdivia et al., 2013). En este artículo nos centraremos en un problema concreto del AS sobre textos en español. Nuestra aproximación consiste en la inclusión de un módulo de identificación del ámbito de la negación

\* Este trabajo ha sido parcialmente financiado por el Fondo Europeo de Desarrollo Regional (FEDER), el proyecto ATTOS (TIN2012-38536-C03-0) del Gobierno de España, el proyecto AORESCU (TIC - 07684) del Gobierno regional de la Junta de Andalucía y el proyecto CEATIC-2013-01 de la Universidad de Jaén.

en un sistema de clasificación de la polaridad en español, con el fin de mejorar la capacidad de predicción de la polaridad de textos de opinión.

Por otra parte, el tratamiento de la negación es un problema abierto dentro del PLN en general, y dentro de la MO en particular. Se trata de un fenómeno lingüístico que no ha sido suficientemente estudiado y que consideramos que requiere un análisis profundo. A pesar de las similitudes entre idiomas, la negación es un elemento lingüístico muy particular de cada lengua, por lo que para un tratamiento efectivo se debe realizar un estudio pormenorizado de los distintos elementos lingüísticos que intervienen en el proceso de negación de un enunciado. La investigación existente sobre la influencia de la negación es escasa, no estudia en profundidad estos aspectos lingüísticos y se centra casi en exclusividad en textos escritos en inglés. Este artículo tiene como una de sus principales aportaciones un estudio detallado del comportamiento de un conjunto partículas negativas, con el fin de determinar el conjunto de palabras que se ven afectadas para así decidir si se debe realizar una modificación de la orientación semántica de la palabra.

La segunda aportación de este trabajo es el uso de la identificación del ámbito de la negación en un sistema de clasificación de la polaridad. Para ello se ha seguido un enfoque no supervisado, en el que el clasificador de polaridad necesita de un recurso semántico que indique la orientación semántica de los términos que aparecen en los textos. Como se verá en las siguientes secciones se han evaluado tres recursos, SentiWordNet, iSOL y eSOL. Los resultados que se han obtenido confirman nuestra hipótesis de que la identificación del ámbito de la negación es un elemento de relevancia para un clasificador de la polaridad.

El resto del artículo se organiza como sigue: la siguiente sección presenta una sucinta exposición de la investigación relacionada con el AS y el tratamiento de la negación. La sección tres se centra en el estudio lingüístico llevado a cabo del ámbito de las diferentes partículas negativas existentes en español. Posteriormente, la sección cuatro detalla el marco experimental realizado para comprobar la efectividad de nuestra hipótesis, y por último, se exponen las conclusiones y la línea de trabajo a seguir.

## 2 Trabajos relacionados

Hasta ahora, la mayoría de las investigaciones relacionadas con el tratamiento de la negación se han realizado sobre opiniones escritas en inglés. Las primeras aproximaciones comenzaron en el año 2001 y sugieren métodos relativamente sencillos. Das y Chen (2001) proponen añadir “NOT” a las palabras de la oración que se encuentren cerca de términos negativos, como por ejemplo “no” o “don’t”. Pang, Lee, y Vaithyanathan (2002) utilizan la técnica de Das y Chen asumiendo que las palabras afectadas por una palabra negativa (“not”, “isn’t”, “didn’t”, etc.) son todas aquellas que se encuentran entre dicha palabra y el primer signo de puntuación. Estos autores realizan experimentos empleando algoritmos de aprendizaje automático para la clasificación de las opiniones teniendo en cuenta el tratamiento de la negación (característica “NOT”) y sin tenerlo en cuenta, llegando a la conclusión de que con este método se produce una mejora insignificante. Polanyi y Zaenen (2004) dan un paso más allá, considerando además de la negación, intensificadores y atenuantes, introduciendo de esta manera el concepto de modificadores de valencia contextuales. Además, se trata del primer modelo computacional que asigna puntuaciones a expresiones polares, invirtiendo la polaridad de las expresiones negadas. Desafortunadamente este modelo no llegó a implementarse, por lo que sólo se puede especular sobre su efectividad. Kennedy e Inkpen utilizan un modelo de negación muy similar al de Polanyi y Zaenen (2004), definiendo como ámbito de una palabra negativa/intensificador/atenuante aquella palabra inmediatamente posterior. En el caso de las palabras afectadas por la negación, el enfoque que siguen es el de invertir la polaridad y en el caso de las palabras que se encuentran en el ámbito de los intensificadores/atenuantes, lo que hacen es incrementar/disminuir el grado de positividad/negatividad según sea el caso. Para clasificar las opiniones utilizan dos métodos, en el primero de ellos clasifican un comentario en base al número de términos positivos y negativos que contiene y en el segundo emplean el algoritmo de aprendizaje automático SVM, llegando a la conclusión de que el modelado de la negación es un hecho importante. Por otro lado, Wilson, Wiebe, y Hoffmann (2005) proponen utilizar una ventana fija de tamaño 4 para determinar el

ámbito de la negación, es decir, una palabra se ve afectada por la negación si existe una expresión de negación entre las 4 palabras anteriores a ella. En este trabajo también emplean un método de aprendizaje no supervisado para clasificar las opiniones en el que utilizan características para modelar las palabras afectadas por la negación, por intensificadores/atenuantes y por otras expresiones polares.

Los enfoques presentados en el párrafo anterior son considerados los trabajos pioneros en el modelado de la negación en el AS en inglés. Sin embargo, estas soluciones no son suficientemente precisas. Por ello, se ha seguido trabajando en este tema para tratar de ofrecer mejores soluciones. Entre los últimos trabajos presentados cabe destacar los de Jia, Yu, y Meng (2009), Taboada et al. (2011) y Cruz Díaz (2014). Jia, Yu, y Meng (2009) proponen un sistema basado en reglas que utiliza información derivada de los árboles de dependencias de las oraciones de estudio, mejorando los enfoques existentes hasta el momento. En el completo y detallado trabajo de Taboada et al. (2011) se presenta una versión mejorada de su sistema SO-CAL (Taboada, Voll, y Brooke, 2008), para calcular la orientación semántica de una opinión, en el que marcan como negadas todas aquellas palabras que se encuentren después de una partícula negativa hasta llegar a un signo de puntuación, un conector o ciertas palabras pertenecientes a una determinada categoría gramatical (ej. “it” (pronombre)). Los autores introducen una nueva forma de tratar la negación que consiste en reducir el valor de polaridad de las palabras negadas en lugar de invertirlo. Cruz Díaz (2014) presenta un sistema para el tratamiento de la negación y de la especulación en textos médicos y opiniones, y el primer corpus de opiniones etiquetado a nivel de negación y especulación (Konstantinova et al., 2012). Por último, mencionar el trabajo de Wiegand et al. (2010) en el que se realiza una excelente revisión del estado del arte de la negación en inglés hasta ese momento.

Como se puede ver, la investigación sobre este tema en inglés es bastante amplia si la comparamos con la existente en español. Con respecto al tratamiento de la negación en documentos en español, el primer trabajo que conocemos es el de Brooke, Tofiloski, y Taboada (2009) en el que adaptan su primera

versión del sistema SO-CAL para análisis de opiniones en inglés (Taboada, Voll, y Brooke, 2008), a un nuevo idioma, el español. Con respecto al tratamiento de la negación siguen el mismo enfoque que el empleado en su versión en inglés, teniendo en cuenta que en español los adjetivos pueden aparecer antes y después de los sustantivos. Finalmente, Vilares Calvo, Alonso Pardo, y Gómez Rodríguez (2013) tienen en cuenta la estructura sintáctica del texto para el tratamiento de la negación, de la intensificación y de las oraciones subordinadas. Los resultados de los autores muestran una mejora en el rendimiento con respecto a los sistemas puramente léxicos.

Nuestra propuesta se basa en la estructura sintáctica del texto, al igual que el último trabajo mencionado pero difiere del mismo en las reglas definidas y en el hecho de que Vilares Calvo, Alonso Pardo, y Gómez Rodríguez (2013) limitan el tratamiento de la negación a los términos “no”, “nunca” y “sin”, mientras que en nuestro trabajo se incluyen además nuevos términos, como son: “tampoco”, “nadie”, “jamás”, “ninguno”, “ni” y “nada”.

### 3 Descripción de la arquitectura

En este trabajo se propone un sistema no supervisado para la clasificación de opiniones teniendo en cuenta la influencia de la negación. Así en primer lugar se describirá el enfoque utilizado para la identificación del ámbito de la negación, y posteriormente el sistema de clasificación de la polaridad.

#### 3.1 Identificación del ámbito de la negación

La negación es un fenómeno lingüístico que debería ser tenido en cuenta en un amplio abanico de tareas dentro del PLN. El AS es una de esas tareas, ya que una correcta identificación del ámbito de la negación permitirá clasificar correctamente opiniones negativas construidas con términos positivos negados y viceversa. Para ello hay que comenzar con un estudio pormenorizado de las distintas partículas negativas que existen en español. En este trabajo se ha comenzado con las partículas negativas más importantes señaladas por la RAE (Española, 2009): “no”, “tampoco”, “nadie”, “jamás”, “ni”, “sin”, “nada”, “nunca” y “ninguno”. Cada una de las partículas indicadas puede tener un comportamiento diferente y afectar a distintos elementos dentro de una oración, por lo que la

manera más adecuada de identificar el ámbito es definir reglas para cada una de las partículas negativas.

Para la construcción de las reglas se ha utilizado el analizador de dependencias de Free-ling<sup>1</sup>, que permite generar el árbol de dependencias de una oración en base a su estructura sintáctica. Así, para cada partícula se han generado los árboles de dependencias correspondientes a distintas oraciones extraídas tras una revisión de distintos sitios web, en las que se hace uso de alguna de las partículas negativas abordadas, y se ha realizado un estudio de los mismos llegando a la conclusión de que es posible generalizar su tratamiento. En la Tabla 1 se muestran las reglas obtenidas tras realizar el estudio.

| Partícula                          | Regla                                                                                        |
|------------------------------------|----------------------------------------------------------------------------------------------|
| no, tampoco, nadie, jamás, ninguno | Afecta al nodo padre y al árbol formado por el hermano de la derecha (incluido).             |
| ni, sin                            | Afecta a todos los hijos y a todos los árboles formados por ellos hasta llegar a nodos hoja. |
| nada, nunca                        | Afecta al nodo padre.                                                                        |

Tabla 1: Reglas ámbito de la negación.

Con el fin de clarificar las reglas que se han definido, a continuación se muestran algunos ejemplos (Figuras 1, 2 y 3) utilizados para determinar el ámbito de cada partícula negativa. En cada figura se señala con una elipse la partícula analizada y con un recuadro su ámbito.

### 3.2 Sistema de clasificación de la polaridad

Una vez determinado el fragmento de la oración que se ve afectado por la negación, el siguiente paso es modificar su polaridad. Nuestra propuesta consiste en invertir la polaridad de las palabras pertenecientes a dicho fragmento que expresen opinión. Por ejemplo, en la oración “*La película no empieza mal con algunas escenas interesantes para introducirte en el drama vivido por el personaje principal.*” (Figura 1), la partícula negativa “no” afecta a las palabras “empieza” y “mal” pero sólo “mal” expresa opinión (-1, negativa)

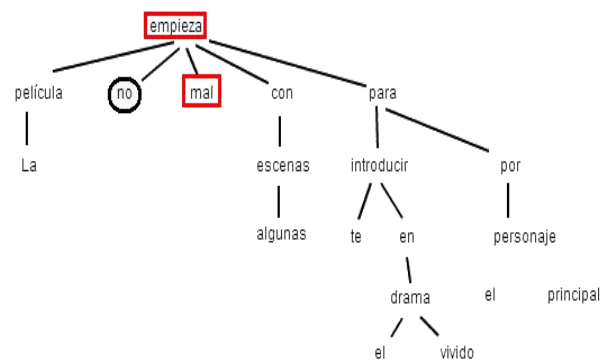


Figura 1: Árbol de dependencias en el que se analiza la partícula “no” en la oración: *La película no empieza mal con algunas escenas interesantes para introducirte en el drama vivido por el personaje principal.*

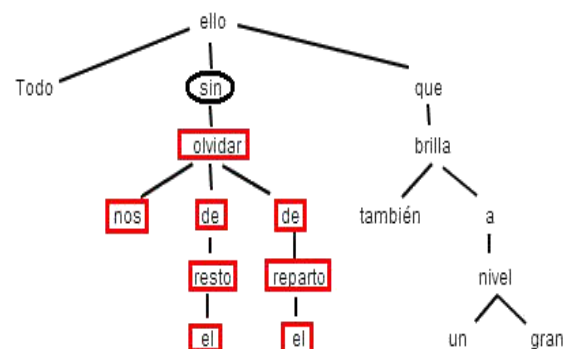


Figura 2: Árbol de dependencias en el que se analiza la partícula “sin” en la oración: *Todo ello sin olvidarnos del resto del reparto que brilla también a un gran nivel.*

por lo que su polaridad se verá invertida (1, positiva).

Para determinar la polaridad de una opinión se ha utilizado el enfoque basado en el cálculo de la orientación semántica de las palabras, que consiste en asignar a cada término que expresa opinión un valor que represente su polaridad, y en asignar a la opinión el valor correspondiente a la suma de estos valores. La Figura 4 muestra un esquema del sistema desarrollado.

El sistema propuesto permite identificar si una opinión es positiva o negativa sin llevar a cabo un entrenamiento previo, es decir, siguiendo un enfoque no supervisado. Para ello, cada opinión se divide en tokens y en oraciones para posteriormente realizar un análisis morfológico de cada una de las oraciones que forman la opinión. En este análisis se obtiene la categoría gramatical de cada una de las palabras de la oración. Tras realizar el análisis

<sup>1</sup><http://nlp.lsi.upc.edu/freeling/>



Figura 3: Árbol de dependencias en el que se analiza la partícula “nunca” en la oración: *Nunca he visto una película peor que esta*.

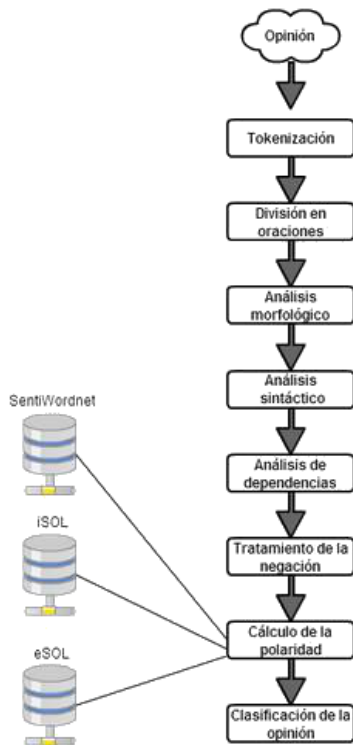


Figura 4: Arquitectura del sistema.

morfológico se procede a determinar las funciones de las palabras dentro de la oración mediante un analizador sintáctico. Una vez obtenida toda esta información sobre las palabras se realiza un análisis de dependencias que será de gran utilidad para determinar el ámbito de la negación.

El siguiente paso, fundamental en nuestro sistema, es el tratamiento de la negación. Aquí es donde se emplean las reglas obtenidas en el estudio realizado (Tabla 1). Cada palabra afectada por la negación se marca, de manera que a la hora de calcular la orientación semántica de la opinión se tenga en cuenta que su polaridad se tiene que invertir. Tras obtener el sentido de cada una de las palabras de la opinión y tras determinar cuáles

forman parte del ámbito de la negación, el siguiente paso es calcular la polaridad de la opinión para poder clasificarla. Para ello se han utilizado tres recursos diferentes (SentiWordNet, iSOL, eSOL) que serán explicados posteriormente.

#### 4 Marco experimental

A continuación se presentan los recursos que han hecho posible la consecución de este trabajo, así como los experimentos realizados y los resultados obtenidos.

Para la realización de las tareas de tokenización, análisis morfológico, análisis sintáctico, desambiguación y análisis de dependencias se ha utilizado Freeling (Padró, 2012). Se trata de una librería de código abierto que proporciona una amplia gama de herramientas para PLN en varios idiomas, entre ellos el español, razón por la cual hemos decidido emplearla en este trabajo.

Por otra parte, para probar la funcionalidad del sistema desarrollado se ha utilizado un corpus en español formado por 3878 críticas de cine recogidas de la web MuchoCine<sup>2</sup> (Cruz et al., 2008). Las críticas que componen el corpus están escritas por usuarios de la web. Esto aumenta la dificultad de la tarea, ya que los textos pueden contener errores gramaticales e incluso expresiones informales. Es preciso señalar que las críticas están puntuadas en un rango de 1 a 5, donde 1 indica que la película es muy mala y 5 que es muy buena. Para la realización de los experimentos se han considerado dos clases, “positiva” y “negativa”. Las críticas con una valoración inferior a 3 se han clasificado como “negativas”, mientras que las puntuadas con un valor superior a 3 se han etiquetado como “positivas”. Las películas valoradas con un 3 (ni buenas ni malas) no se han tenido en cuenta en este estudio. Por tanto, la experimentación se ha llevado a cabo sobre un total de 2625 opiniones, de las cuales 1351 se corresponden con críticas “positivas” y 1274 con “negativas”.

Debido al hecho de que el objetivo de nuestra propuesta es comprobar si la predicción del grado de subjetividad de un conjunto de opiniones mejora con la inclusión de un módulo para el tratamiento de la negación, resulta de gran utilidad analizar el número de opiniones de este corpus que utilizan alguna partícula de negación. De las 2625 opi-

<sup>2</sup><http://www.lsi.us.es/~fermin/corpusCine.zip>

niones utilizadas en la experimentación, 2616 utilizan algún marcador negativo de los analizados (Tabla 2). En la Tabla 3 se puede ver la frecuencia de aparición de las diferentes partículas abordadas en este estudio reafirmando la importancia de su tratamiento.

| Opiniones | Con partículas negativas | % opiniones con negación |
|-----------|--------------------------|--------------------------|
| Positivas | 1345                     | 99,56                    |
| Negativas | 1271                     | 99,76                    |

Tabla 2: Opiniones del corpus MuchoCine con partículas negativas.

| Partícula | Frecuencia |
|-----------|------------|
| no        | 15517      |
| sin       | 3816       |
| ni        | 2107       |
| nada      | 1583       |
| nunca     | 671        |
| tampoco   | 515        |
| nadie     | 458        |
| jamás     | 123        |
| ninguno   | 97         |

Tabla 3: Frecuencia de aparición de las partículas negativas estudiadas en el corpus MuchoCine.

Por último, para determinar la polaridad de las opiniones del corpus se han utilizado diferentes recursos lingüísticos: SentiWordNet, iSOL y eSOL.

SentiWordNet es un recurso léxico que asigna a cada uno de los sentidos de las palabras de WordNet tres valores que reflejan su positividad, negatividad y objetividad. WordNet es una base de datos léxica en inglés que agrupa las palabras en base a su significado. Como es ampliamente conocido por la comunidad investigadora, SentiWordNet es un recurso para AS sobre textos en inglés. Para poder aplicarlo sobre textos en español se ha recurrido a una versión de WordNet en español con el fin de desambiguar cada término y así obtener su synset. Una vez que se tiene el synset del término, se puede acudir a SentiWordNet para obtener los valores de polaridad. La versión de WordNet en español empleada es la conocida como Multilingual Central Repository (MCR) (Gonzalez-Agirre, Laparra, y Rigau, 2012). La versión de WordNet en español de MCR contiene aproximadamente 38.000 synsets, que quedan

todavía lejos de los 117.000 de WordNet.

Por su parte, iSOL es una lista de palabras indicadoras de opinión en español que es independiente del dominio (Molina-González et al., 2013). Está formada por 2509 términos positivos y 5626 negativos. Se trata de un lexicón generado a partir de la traducción automática de las palabras de la lista de Bing Liu, corregido manualmente y ampliado con nuevos términos. El último recurso empleado es eSOL. Se trata de una ampliación de la lista de palabras iSOL con términos adaptados al dominio de las críticas de cine (Molina-González et al., 2013).

#### 4.1 Experimentos

Para evaluar el sistema desarrollado se han llevado a cabo 3 experimentos. En cada uno de los experimentos se realiza un proceso común. Se analizan las frases que componen cada opinión con el objetivo de obtener el significado de cada una de las palabras y de marcar aquellas que se vean afectadas por alguna de las partículas negativas abordadas en este estudio. La diferencia entre estos experimentos radica en el recurso utilizado para calcular la polaridad de las opiniones (Figura 4).

En el primer experimento se ha utilizado SentiWordNet, de manera que la polaridad de una opinión se ha calculado como la suma de la diferencia entre los valores de positividad y negatividad de cada palabra en base a su sentido más frecuente. En el segundo experimento se ha empleado la lista de palabras iSOL. En este caso, la polaridad de la opinión se ha calculado a partir de la suma de los términos de la opinión que pertenecen a la lista. Si la palabra pertenece a la lista de términos positivos se suma con valor 1 y si la palabra pertenece a la lista de términos negativos se suma con valor -1. En el último experimento se ha utilizado la lista de palabras eSOL y se ha procedido de la misma manera que se ha descrito en el segundo experimento.

Es preciso señalar que, en todos los experimentos, para las palabras que se han detectado como pertenecientes al ámbito de la negación se ha invertido su valor, es decir, se ha multiplicado por -1. Finalmente, la opinión se ha clasificado como “positiva” si el valor de su polaridad ha sido mayor que 0 y “negativa” en caso contrario.

Como medidas de evaluación del sistema se han tomado las más utilizadas en el ámbito del AS, es decir, Precisión (Prec.), Recall, F1

y Accuracy (Acc.).

## 4.2 Resultados

Para una mejor evaluación del impacto de la identificación del ámbito de la negación en el marco de un sistema de clasificación de la polaridad, se ha ejecutado el clasificador tanto sin el módulo de identificación de la negación como con él. En las Tablas 4 y 5 se muestran los resultados obtenidos. Además, en la Tabla 5 se ha incluido una columna que refleja la mejora con respecto al valor de Accuracy.

| Recurso | Prec.  | Recall | F1     | Acc.   |
|---------|--------|--------|--------|--------|
| SWN     | 0,5896 | 0,5314 | 0,5590 | 0,5432 |
| iSOL    | 0,6283 | 0,6246 | 0,6265 | 0,6270 |
| eSOL    | 0,6365 | 0,6276 | 0,6320 | 0,6312 |

Tabla 4: Clasificación de la polaridad sin detección de la negación.

| Recurso | Prec.  | Recall | F1     | Acc.   | Mejora |
|---------|--------|--------|--------|--------|--------|
| SWN     | 0,6057 | 0,5489 | 0,5759 | 0,5596 | 3,02 % |
| iSOL    | 0,6416 | 0,6394 | 0,6405 | 0,6411 | 2,25 % |
| eSOL    | 0,6519 | 0,6430 | 0,6474 | 0,6465 | 2,42 % |

Tabla 5: Clasificación de la polaridad con detección de la negación.

En las dos tablas se puede ver la diferencia entre los tres recursos lingüísticos empleados para la clasificación de la polaridad. Los bajos resultados que obtiene SentiWordNet consideramos que están justificados porque el recurso empleado para la identificación del synset (MCR) solo cubre aproximadamente 38000 synsets de los 117000 que contiene SentiWordNet. Evidentemente, muchos sentidos no pueden ser identificados por lo que la capacidad de identificación de la orientación semántica con SentiWordNet queda bastante mermada. La diferencia entre iSOL y eSOL también es explicable. Tal y como señalan Molina-González et al. (2013) en su trabajo, la lista eSOL contiene información del dominio del cine, por lo que como es de esperar, el resultado de la clasificación es mejor.

En cuanto al tratamiento de la negación, la Tabla 5 evidencia que se ha producido una pequeña mejora. Esto se ha debido a una reducción del número de Falsos Positivos y Falsos Negativos, ya que se han podido tratar correctamente los términos positivos o negativos que se encuentran negados. Estos resultados nos animan a seguir investigando en el

tratamiento de la negación, y en consecuencia, en el estudio del resto de partículas que todavía quedan por incluir en nuestro sistema.

## 5 Conclusiones y trabajo futuro

En este artículo se ha presentado una primera aproximación a la identificación del ámbito de la negación que se ha aplicado sobre un sistema de clasificación de opiniones en español. La identificación de la negación se ha realizado siguiendo un enfoque basado en reglas lingüísticas, que se han definido después de realizar un estudio de cómo afectan algunas partículas negativas a los elementos que las circundan en una oración. Posteriormente, el módulo generado se ha incluido en un sistema de clasificación de la polaridad basado en el uso de recursos lingüísticos de opinión. Los resultados que se han obtenido, aunque las mejoras han sido pequeñas, certifican nuestra hipótesis inicial de que la toma en consideración del ámbito de la negación es de relevancia para la clasificación de la polaridad, lo que nos anima a seguir investigando.

Nuestro trabajo futuro se va a centrar en llevar a cabo un análisis de la significancia estadística y en el estudio del resto de partículas de negación que no se han incluido en este primer trabajo. Además, se va a realizar un análisis exhaustivo del corpus para determinar las oraciones en las que las partículas negativas afectan a expresiones polares, para así poder comprobar cómo han funcionado las reglas definidas, con objeto de poder refinarlas. También, se va a iniciar un estudio sobre la importancia de las partículas de intensificación a la hora de determinar la polaridad de una oración.

## Bibliografía

- Brooke, Julian, Milan Tofiloski, y Maite Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. En *International Conference RANLP*, 50-54.
- Cruz, Fermín L, Jose A Troyano, Fernando Enriquez, y Javier Ortega. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento de Lenguaje Natural*, 41.
- Cruz Díaz, Noa Patricia. 2014. *Detección de la Negación y la Especulación en Textos*

- Médicos y de Opinión*. Ph.D. tesis, Universidad de Huelva.
- Das, Sanjiv y Mike Chen. 2001. Yahoo! for amazon: Extracting market sentiment from stock message boards. En *Proc. of the Asia Pacific finance association annual conference (APFA)*, volumen 35, página 43. Bangkok, Thailand.
- Española, Real Academia. 2009. *Nueva gramática de la lengua española*, volumen 1. Espasa Calpe Madrid, Spain.
- Gonzalez-Agirre, Aitor, Egoitz Laparra, y German Rigau. 2012. Multilingual central repository version 3.0. En *LREC*, 2525-2529.
- Jia, Lifeng, Clement Yu, y Weiyi Meng. 2009. The effect of negation on sentiment analysis and retrieval effectiveness. En *Proc. of the 18th ACM conference on Information and knowledge management, 1827-1830*.
- Konstantinova, Natalia, Sheila CM de Sousa, Noa P Cruz Díaz, Manuel J Maña López, Maite Taboada, y Ruslan Mitkov. 2012. A review corpus annotated for negation, speculation and their scope. En *LREC*, 3190-3195.
- Liu, Bing. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1-167.
- Martín-Valdivia, María-Teresa, Eugenio Martínez-Cámara, Jose-M Perea-Ortega, y L Alfonso Ureña-López. 2013. Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*, 40(10):3934-3942.
- Molina-González, M Dolores, Eugenio Martínez-Cámara, María-Teresa Martín-Valdivia, y José M Perea-Ortega. 2013. Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, 40(18):7250-7257.
- Padró, Lluís. 2012. Analizadores multilingües en freeling. *Linguamática*, 3(2):13-20.
- Pang, Bo, Lillian Lee, y Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. En *Proc. of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, páginas 79-86. ACL.
- Polanyi, L y A Zaenen. 2004. Context valence shifters. En *Proc. of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- Rushdi-Saleh, Mohammed, M Teresa Martín-Valdivia, L Alfonso Ureña-López, y José M Perea-Ortega. 2011. Oca: Opinion corpus for arabic. *Journal of the American Society for Information Science and Technology*, 62(10):2045-2054.
- Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, y Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267-307.
- Taboada, Maite, Kimberly Voll, y Julian Brooke. 2008. Extracting sentiment as a function of discourse structure and topicality. *Simon Fraser University School of Computing Science Technical Report*.
- Vilares Calvo, David, Miguel Ángel Alonso Pardo, y Carlos Gómez Rodríguez. 2013. Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias. *Procesamiento de Lenguaje Natural*, 50:13-20.
- Wiegand, Michael, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, y Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. En *Proc. of the workshop on negation and speculation in natural language processing*, páginas 60-68. ACL.
- Wilson, Theresa, Janyce Wiebe, y Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. En *Proc. of the conference on human language technology and empirical methods in natural language processing*, páginas 347-354. ACL.
- Zhang, Changli, Daniel Zeng, Jiexun Li, Fei-Yue Wang, y Wanli Zuo. 2009. Sentiment analysis of chinese documents: From sentence to document level. *Journal of the American Society for Information Science and Technology*, 60(12):2474-2487.