UNIVERSIDAD DE JAÉN

ESCUELA POLITÉCNICA SUPERIOR
DE JAÉN
DEPARTAMENTO DE INFORMÁTICA

TESIS DOCTORAL

GENERACIÓN DE RECURSOS PARA ANÁLISIS DE OPINIONES EN ESPAÑOL

PRESENTADA POR:
MARÍA DOLORES MOLINA GONZÁLEZ

DIRIGIDA POR:
DRA. DÑA. MARÍA TERESA MARTÍN VALDIVIA

JAÉN, 28 DE NOVIEMBRE DE 2014

ISBN 978-84-8439-967-4

UNIVERSIDAD DE JAÉN

Escuela Politécnica Superior de Jaén

Departamento de Informática

Generación de Recursos para Análisis de Opiniones en Español

Memoria de Tesis presentada por

María Dolores Molina González

para optar al grado de Doctor

Directora: Dra. María Teresa Martín Valdivia

Jaén, octubre de 2014

La memoria titulada *"Generación de recursos para Análisis de Opiniones en español"* que presenta D.ª María Dolores Molina González para optar al grado de doctor, ha sido realizada dentro del Programa de Doctorado en Ingeniería y Arquitectura de la Universidad de Jaén bajo la dirección de D.ª María Teresa Martín Valdivia. Para su evaluación, esta memoria se presenta como un conjunto de trabajos publicados, acogiéndose y ajustándose a lo establecido en el punto 3 del artículo 23 del *Reglamento de los Estudios de Doctorado de la Universidad de Jaén,* aprobado en Febrero de 2012.

En Jaén, Octubre de 2014

La Doctoranda

La Directora

Fdo: D.ª María Dolores Molina González

Fdo: Dra. D.ª María teresa Martín Valdivia

Agradecimientos

Dedico esta memoria a todos aquellos que han confiado en mí y me han dado su apoyo moral y logístico.

En primer lugar, quiero agradecer a mi directora de tesis, Maite Martín Valdivia, el confiar en mí y hacerme trabajar desde el primer día. Sin sus ideas brillantes, su ímpetu por avanzar y hacer las cosas bien, sus consejos para desarrollarme como investigadora y, sobretodo, su tiempo y cordialidad, esta tesis nunca hubiera llegado a su fin. No menos importante es mi compañero Eugenio Martínez Cámara (Geni) con el que he compartido trabajo desde el principio en la mejor sintonía y siempre ha estado dispuesto a ayudarme. A Alfonso Ureña López, por acogerme en su grupo de investigación 'Sistemas inteligentes de acceso a la información' (SINAI) sin poner objeciones. Muchas gracias por estar conmigo en todo momento.

También quiero agradecer su apoyo y amistad a muchos compañeros del departamento de Electrónica, Física e Informática de la Universidad de Jaén, que me han hecho la vida mucho más agradable, animándome continuamente en esta ardua tarea de la ejecución de la tesis. En especial quiero nombrar a Cati y a Lina, que me han ayudado en el terreno familiar y profesional sin descanso. A Eli, Paco, Juan Domingo, Lila y Elena ¡¡¡qué distinto hubiera sido mi paso por la cuarta planta del A3 sin vuestra compañía y experiencia!!!

Agradecer a mis hermanos, sobrinos y padres los buenos momentos que me recargan las pilas. A mi madre, agradecerle sus tupper de buenas comidas cuando mi tiempo escaseaba. Gracias.

Por último, quiero dedicar este trabajo a mis dos hijas, Eva Alejandra y Delia Mª, porque son 'el azúcar que hace que la píldora se tome mejor', como cantaba *Mary Poppins*. Espero poder dedicaros el tiempo que antes no tuve.

Índice

I. Memoria

II. Publicaciones: Trabajos Publicados, Aceptados y en Revisión

Índice de tablas

Índice de figuras

Parte I. Memoria

1. Introducción

Nuestro interés en esta memoria reside en el estudio del Análisis de Opiniones (AO) o Minería de Opiniones (MO) centrándonos en la clasificación de la polaridad basada en la Orientación Semántica (OS). Esta tarea tiene como objetivo determinar la polaridad de un documento, frase o característica (positiva o negativa) y medir el grado de dicha polaridad expresada en dichos textos por el autor. Para alcanzar dicho objetivo y contribuir a la mejora de los sistemas existentes de clasificación de la polaridad, proponemos desarrollar e integrar recursos lingüísticos adaptados al dominio y al español.

Para llevar a cabo este estudio, la presente memoria se divide en 5 secciones. Esta primera sección está dedicada al "Planteamiento" del problema, describiéndose las técnicas empleadas para resolverlo. Asimismo se exponen las deficiencias encontradas en este marco de trabajo que nos genera la "Justificación" para el desarrollo de esta tesis. Además se exponen los "Objetivos" propuestos al inicio del estudio realizado. En la segunda sección se exponen y describen brevemente los "Recursos Lingüísticos para la Minería de Opiniones" más usados para la clasificación de la polaridad y algunos métodos para la generación de los mismos. Posteriormente en la tercera sección se incluye la "Discusión de resultados" que proporciona la información resumida de los resultados obtenidos más interesantes recogidos en las distintas publicaciones. En la sección cuarta "Conclusiones" se hace una discusión general de todos los datos en su conjunto finalizándose con unos comentarios sobre futuros trabajos que quedan abiertos en la presente tesis. Por último, la quinta sección "Resultados" recoge las 7 contribuciones publicadas agrupadas en las 4 propuestas resumidas en la sección tercera.

1.1 Planteamiento

En los últimos años, el interés por el Análisis de Opiniones (AO) (conocido en inglés como S*entiment Classification, Sentiment Analysis* u *Opinion Mining*) ha crecido súbitamente debido a varios factores. Por una parte, el incremento de la creación y compartición de datos por parte de los usuarios de Internet haciendo uso de nuevos servicios emergentes. Por otra parte, empieza a ser una tarea rutinaria el consumo de datos online para la toma de decisiones a nivel personal o colectivo.

Aunque son muchas las tareas tratadas en el campo del AO, la clasificación de polaridad de las opiniones es una de las más consolidadas e importantes [Tur02, PL08], siendo las técnicas que más se han aplicado para los sistemas de clasificación de la polaridad las dos siguientes:

- **Aproximación basada en aprendizaje automático (*Machine Learning-ML*) o enfoque supervisado.** Este enfoque se basa en entrenar unos clasificadores a partir de una colección de datos previamente clasificados usando distintos algoritmos como pueden ser *Support Vector Machine* (SVM), Máxima Entropía (ME), Naïve Bayes (NB), regresión logística bayesiana (BBR), *K Nearest Neighbor* (KNN) o árbol de decisión (C4.5) [DLP03, PLV02]. Diferentes características son las empleadas por los algoritmos para hacer su entrenamiento, siendo las más usadas las siguientes:
    - la presencia de los términos (palabra simple o n-gramas) y su frecuencia;
    - la categoría gramatical (*Part of Speech, PoS*) como por ejemplo los adjetivos, buenos indicadores de subjetividad;
    - las palabras o frases de opinión como por ejemplo ‘odiar’, ‘amar’ o la frase ‘costar un ojo de la cara’;
    - la dependencia sintáctica de los términos, así pues no indica la misma intensidad ‘bello’ que ‘extremadamente bello’ o si la posición de un término es al principio o final de frase;
    - y por último, el uso de la negación, que normalmente cambia la polaridad del término que lo sigue.

- **Aproximación basada en la Orientación Semántica (*Semantic Orientation-SO*) o enfoque no supervisado**. Este enfoque no necesita un entrenamiento previo, sino que se tiene en cuenta la polaridad de una palabra o conjunto de palabras. Distintos métodos se aplican para calcular esta orientación, de la cual se obtiene un valor real que si es positivo implica opinión subjetiva favorable y si es negativo implica lo contrario. Diferentes valores de esta medida indicarán el grado de inclinación hacia un lado u otro. Los primeros estudios en este tipo de clasificación se llevaron a cabo en [HM97, KMMR04].

1.2 Justificación

Ambas aproximaciones tienen ventajas e inconvenientes aunque con la que mejores resultados se obtiene es con la aproximación basada en un enfoque supervisado. Es por ello que nuestro interés se centra precisamente en la otra alternativa, en **la aproximación basada en un enfoque no supervisado,** no solo porque nuestra principal intención es mejorar los resultados e incluso alcanzar los obtenidos con el enfoque supervisado, sino porque la mayoría de los estudios han estado orientados a utilizar aprendizaje automático.

Por otra parte, la mayor parte de la investigación en el ámbito del AO se ha realizado en textos escritos en inglés, debido principalmente a la etapa temprana en la que se encuentra y a la falta de recursos en otros idiomas. Sin embargo, actualmente la presencia de otros idiomas distintos al inglés en Internet crece cada vez más. Por lo tanto, es necesario orientar la investigación no solo a textos escritos en inglés sino a otros idiomas o a sistemas que sean capaces de funcionar independientemente de la lengua utilizada.

Según el *Internet Word State Rank*[1], el lenguaje español es el tercero más usado por los usuarios de Internet. Este hecho unido a la escasez de recursos en dicho idioma para la clasificación de polaridad son los que despiertan nuestro interés para desarrollar el trabajo de investigación que se describe en la presente memoria.

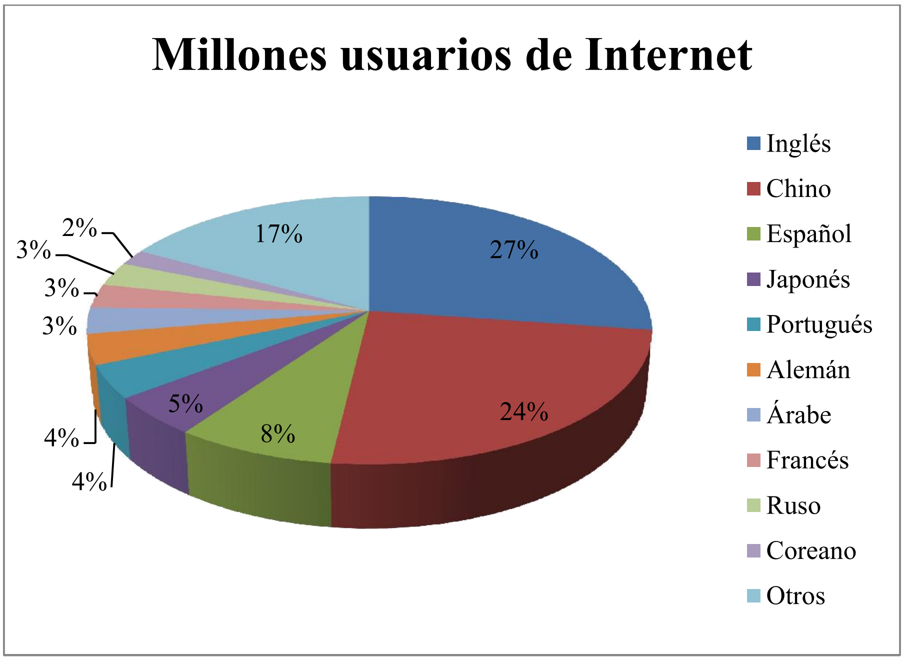


Figura 1. Millones de usuarios en los 10 idiomas más usados en Internet (Mayo 2011)

[1] www.internetworldstats.com

Tal y como se ha comentado anteriormente, la aproximación por aprendizaje automático obtiene mejores resultados que la aproximación basada en orientación semántica. Es por ello, que se necesita más investigación en este campo pero el problema que se encuentra es el escaso número de recursos lingüísticos para realizar tal investigación si el idioma de destino es distinto al inglés. Por tanto, el principal objetivo de esta tesis es **la generación y aportación de recursos enfocados al idioma español**, el tercer idioma más usado por los usuarios web para hacer posible la clasificación de opiniones textuales en la web 2.0 basada en Orientación Semántica.

1.3 Objetivos

Como se ha introducido en el punto anterior, esta tesis se centra en la clasificación de polaridad en español basada en la orientación semántica, debido principalmente al escaso número de recursos lingüísticos existentes en el lenguaje español. Las metas que se pretenden alcanzar tras la realización del trabajo de investigación descrito en este documento son las siguientes:

- Realización de un nuevo recurso en español para minería de opiniones compuesto por una lista de palabras valoradas positiva y negativamente.
- Demostración de la validez de nuestro recurso, realizando varios experimentos sobre distintos corpora en español. Concretamente, se usarán el corpus español MuchoCine MC [CTEO08] de opiniones sobre cine y el corpus SFU *(Simon Fraser University)* [BTT09] compuesto por ocho grupos de opiniones correspondientes a los siguientes dominios: hoteles, lavadoras, películas, teléfonos móviles, libros, coches, ordenadores y música. Los experimentos se compararán con otros recursos que existan para el español.
- Mejora de las listas de palabras valoradas, con la inserción de palabras positivas y negativas específicas del dominio del corpus de opiniones para el cual se esté haciendo el análisis semántico. Para ello, se diseñará una heurística basada en la aparición de palabras más frecuentes en opiniones positivas y negativas.
- Creación de un corpus de opiniones escritas en español para la comunidad investigadora con el fin de poder probar el lexicón generado sobre el dominio específico 'Hoteles' y seguir trabajando para la mejora de resultados en la orientación semántica.

Para el planteamiento de los objetivos descritos anteriormente, se parte de las siguientes hipótesis:

Hipótesis 1. La tarea de minería de opiniones trabaja con documentos, que son textos de tamaños distintos procedentes de portales web, blogs, comentarios, opiniones de productos, etc. Estos documentos están escritos en lenguaje natural y tratan sobre, al menos, un ítem o dominio. El sentimiento de estos documentos es la actitud, la opinión o la emoción que los autores de dichos documentos expresan sobre los ítems o dominios tratados.

Hipótesis 2. El sentimiento de los documentos puede ser expresado en forma numérica o simbólica acotada en un rango, en emoticonos o mediante el uso de palabras que expresan alguna de las emociones humanas. Dichas emociones, sin embargo, pueden, según muchos autores, clasificarse binariamente dependiendo de su inclinación hacia el lado positivo o negativo.

Hipótesis 3. La polaridad de sentimiento puede ser evaluada por el análisis de opiniones textuales. En las opiniones se pueden encontrar descripciones positivas y negativas. Nuestra estrategia cuenta las palabras que aparecen en la opinión y que están en las listas de palabras positivas o negativas. Las opiniones van a ser valoradas como positivas o negativas según si el número mayor de palabras encontradas es positiva o negativa.

Hipótesis 4. Debido al amplio uso del lenguaje, palabras que para un dominio resultan positivas o negativas, para otro dominio presentan una polaridad inversa. Como consecuencia, cuanto más conocimiento se tenga sobre el dominio de opiniones tratadas, mejores resultados se obtendrán y mejor será la clasificación de documentos extraídos.

2. Recursos lingüísticos para Minería de Opiniones

El término "Recursos lingüísticos" hace referencia a un conjunto de datos y sus descripciones en formato electrónico para, entre muchas otras cosas, evaluar el lenguaje natural. Ejemplos de recursos lingüísticos son los corpora escritos y hablados, bases de datos léxicas, incluso, aplicaciones software para la recopilación, preparación, gestión y uso de los mismos [CMU+96]. El uso de recursos lingüísticos (corpora y lexicones) en el Procesamiento de Lenguaje Natural (PLN) es requisito indispensable para la construcción de los clasificadores de polaridad de sentimientos. Es por ello, que a continuación se presenta un estado del arte de los dos recursos lingüísticos más utilizados para tal finalidad.

2.1 Corpora

Un corpus de lenguaje escrito es una recopilación de textos representativos de una lengua disponible en formato electrónico [FK82]. Dependiendo de sus características se pueden establecer distintos tipos de corpus. Entre dichos tipos, se pueden encontrar corpus textuales u orales, corpus de propósito general o centrado en un dominio, corpus monolingües o multilingües y corpus anotados o no anotados. Fijándonos en el último tipo de corpus, se puede decir que un corpus anotado está enriquecido con información adicional al texto en forma de marcas, puntuaciones, que aporta un conocimiento. El corpus no-anotado sólo dispone de la colección de texto sin ninguna información adicional.

Actualmente, Internet es el mayor receptor y contenedor de información sobre los temas más variados. Esta información puede contener datos objetivos y/o subjetivos, es decir, pueden ser meramente descripciones y/u opiniones personales cargadas de subjetividad, haciendo uso de emoticones o palabras que expresan alguna de las emociones humanas. Debido a la alta dependencia que se tiene del conocimiento por parte de la sociedad para la toma de decisiones principalmente, Internet es el sitio de mayor consulta por parte de millones de personas en todo el mundo. Es por estos motivos que los corpora más utilizados en la última década en la clasificación de texto se han generado a partir de información procedente de la web 2.0.

En la literatura se encuentran muchos corpora diferentes sobre opiniones principalmente escritos en inglés.

- Así en [PL04] se presenta el corpus de opiniones de películas (*Cornell movie-review*) el cual incluye diferentes conjuntos de opiniones según la polaridad de sentimientos a nivel de frase y de documento, y según la subjetividad de las frases de dichas opiniones.
- El corpus empleado en [HL04] consiste en opiniones de cinco productos electrónicos descargados de Amazon[2] y Cnet[3].
- El Corpus MPQA *(Multi-Perspective Question-Answering)* se describe en [WWH05, WWC05]. Este corpus contiene 535 artículos de noticias procedentes de una gran variedad de fuentes etiquetadas manualmente a nivel de frase y subfrase y a nivel de estados personales, como creencias, emociones, especulación, sentimientos, etc.

Aunque son minoritarios, también se dispone de algunos corpora en otros idiomas distintos al inglés.

- Así en [BFK04] se presenta un corpus consistente en 702 frases de periódicos franceses y belgas con etiquetas asignadas por diez jueces como contenidos adecuados, neutrales o no-adecuados en una escala de 7 puntos.
- En [Den08] se usan comentarios escritos en alemán sobre cine. Las opiniones son extraídas de Amazon[4] y cuenta con 100 opiniones clasificadas como positivas y otras 100 clasificadas como opiniones negativas.
- Dos corpora escritos en chino se usan en [ZZL+09]. Uno está compuesto por opiniones sobre la eutanasia recogidas de diferentes portales web, mientras que el otro es un conjunto

[2] www.amazon.com
[3] www.cnet.com
[4] www.amazon.de

de opiniones sobre 6 categorías de productos recogidos de Amazon. El corpus sobre la eutanasia fue revisado y clasificado manualmente. El corpus procedente de Amazon fue distribuido con 310.390 opiniones positivas y 29.540 opiniones negativas.

- En [GJ10] se presenta un corpus con 2.000 opiniones de películas en francés. El corpus procedente de la web[5], extrae 1.000 opiniones negativas y otras 1.000 opiniones positivas sobre 10 películas, estando el tamaño de las opiniones comprendido entre 500 a 1.000 caracteres cada una.
- En [ALT10] se presenta un corpus anotado manualmente con noticias del mercado financiero de Croacia escritas en rumano.
- OCA (*Opinion Corpus for Arabic)* [RMUP11] es un corpus de opiniones de películas de cine escritas en árabe que fueron extraídas revisando manualmente 15 portales web eligiendo 250 opiniones positivas y 250 opiniones negativas.

Pocos son también los trabajos enfocados al desarrollo de corpus en el idioma destino de nuestra tesis aunque se pueden citar los siguientes:

- En [CTEO08] se describe la generación de un corpus de críticas de cine escritas en español (corpus MC) por los propios usuarios web. Estas críticas se han extraído de la web MuchoCine[6]. El número total es de 3.878 y están puntuadas con un valor comprendido entre 1 y 5.
- En [BBMM09] se presenta el corpus EmotiBlog que incluye comentarios sobre tres temas (el protocolo Kyoto, las elecciones en Zimbabwe y las elecciones en USA) escritos en blogs en tres idiomas distintos: español, inglés e italiano. EmotiBlog contempla el etiquetado a nivel de documento, frase y elemento, distinguiéndolos entre 'objetivos' y 'subjetivos'. Cada elemento va anotado con los siguientes atributos comunes: polaridad, grado (o intensidad) y emoción.
- El corpus SFU[7] [BTT09] está compuesto por ocho grupos de opiniones correspondientes a los siguientes dominios: hoteles, lavadoras, películas, teléfonos móviles, libros, coches, ordenadores y música. Para cada dominio se tienen 25 opiniones positivas y otras 25 opiniones negativas, haciendo un total de 400 opiniones recopiladas. Las opiniones fueron extraídas del portal web Ciao[8].

[5] www.allocine.com
[6] www.muchocine.net
[7] http://www.sfu.ca/~mtaboada/download/downloadCorpusSpa.html
[8] www.ciao.es

- El corpus SHoRe *(Spanish HOtel REviews)* es una gran colección de opiniones escritas en español generada a partir de TripAdvisor[9]. Este corpus se compone de opiniones sobre más de 3.000 hoteles de los archipiélagos Canario y Balear y de los usuarios de los hoteles que aportan su opinión. La colección contiene más de 325.000 opiniones. Cabe destacar que este corpus está compuesto por opiniones procedentes de usuarios españoles como de usuarios no-españoles, por lo que las opiniones de este último grupo de usuarios son traducciones generadas automáticamente al español, lo que proporciona un decremento en la calidad de esas opiniones.
- En [MMMU14] se describe la generación del corpus COAH *(Corpus of Opinion about Andalusian Hotels)*, el cual extrae comentarios sobre 10 hoteles de cada una de las ocho provincias andaluzas, obteniendo un total de 1.816 opiniones escritas en español en los últimos años sobre los 80 hoteles elegidos en total. El portal web TripAdvisor es el origen del corpus COAH. Este corpus se compone de dos tipos de información. Una sobre el hotel (nombre, dirección) y otra sobre la opinión del huésped del hotel (valoración global entre 1 y 5, la identificación del usuario, la valoración de relación calidad/precio, la limpieza, etc.). En el siguiente fragmento se muestra un ejemplo de los datos recogidos de un hotel:

```
<ID>1</ID>
<Nombre>Alcazaba Mar Hotel</Nombre>
<Categoria>4</Categoria>
<Dirección>Juegos del Argel, Urbanizacion El Toyo | near Cabo de Gata </Dirección>
<CódigoPostal>04131</CódigoPostal>
<localidad>Retamar</localidad>
<Provincia>Almería</Provincia>
<País>España</País>
<Viajero>--------------</Viajero>
<Localidad_Viajero>------- </Localidad_Viajero>
<Valoración>3</Valoración>
<Título>"Adecuada la calidad al precio del hotel"</Título>
<Opinión>Acabamos de llegar del hotel. La verdad es que nos fuimos con mucho miedo por los
comentarios escritos aquí. Nuestra opinion es que es un hotel comodo, tiene piscina buena, animacion
excelente, y un personal muy amable. Quizas lo mas tenido en cuenta es el
buffet……………………………………………. </Opinión>
<Fecha_TipoViajero>Se alojó el Agosto de 2012, viajó con la familia</Fecha_TipoViajero>
<Relación_calidad-precio>3</Relación_calidad-precio>
<Ubicación>2</Ubicación>
<Calidad_del_sueño>3</Calidad_del_sueño>
<Habitaciones>3</Habitaciones>
<Limpieza>3</Limpieza>
<Servicio>4</Servicio>
```

Figura 2. Ejemplo de un hotel del corpus COAH

[9] www.tripadvisor.es

2.2 Lexicones

Como se ha comentado en el punto anterior, según la información lingüística existente en los corpora, se pueden distinguir dos tipos de corpora, el anotado y el no anotado. El corpus no anotado, sólo dispone de la colección de textos sin ninguna información adicional, lo cual hace difícil el conocimiento de la inclinación subjetiva si la hubiera y, por tanto, la toma de decisiones respecto al tema tratado en dicha información. Por tal motivo, surge el interés en desarrollar lexicones y otros métodos para facilitar la clasificación de opiniones en información no anotada.

Existen distintos tipos de lexicones, desde los más sencillos consistentes en listas de palabras separadas según su polaridad, hasta las más extensas colección de palabras o n-gramas que llevan asociadas una serie de características que facilitará el conocimiento gramatical y sentimental de dichas palabras o n-gramas. Tres son los principales métodos que hay para reunir estas listas de palabras con opinión: el método manual, el basado en diccionario y el basado en corpus.

- El **método manual** [DC07] consume mucho tiempo y requiere de bastantes recursos humanos. Por este motivo suele ser combinado con métodos automatizados.
- El **método basado en diccionario** [AB06] consiste en coger un conjunto de palabras manualmente con orientación conocida (semillas) e ir incrementando el número de palabras mediante el uso de algún diccionario o base de conocimiento léxico *(Lexical Knowledge Base-LKB)*. Es importante destacar que en la búsqueda de palabras con polaridad en los *LKB* se emplean las relaciones léxicas y semánticas. Por ejemplo, a partir de una palabra se buscan sus 'sinónimos' y sus 'antónimos', teniendo los primeros la misma polaridad y los segundos contraria polaridad de la palabra original. Estas dos relaciones son las básicas, pero nos podemos encontrar otras como la 'hiperonimia' (día es hiperónimo de lunes o martes), 'hiponimia' (jueves es hipónimo de día), 'meronimia' (dedo merónimo de mano), 'derivado de' (joven->juvenil), 'perteneciente a' (instrumento musical), etc. Este método tiene sus limitaciones puesto que no es capaz de encontrar palabras con orientación específica para dominios concretos.
- El **método basado en corpus** solventa esta carencia [HM97, DTCY10]. Se han empleado diferentes técnicas aunque todas ellas comienzan con una lista de palabras conocidas e intentan encontrar otras relacionadas en un corpus de un dominio específico. Así palabras con orientación positiva en un dominio ('impredecible' en el dominio cine) pueden tener orientación negativa en otro ('impredecible' en el dominio automóviles).

A continuación se describen brevemente los lexicones más usados en el ámbito de la Minería de Opiniones:

1. SWN (SentiWordNet) [ES06, BES10] es uno de los recursos más usados en MO y está construido sobre la base de datos léxica WordNet[10] [Fell98]. WordNet agrupa adjetivos, nombres, verbos y otras formas de clases gramaticales en un conjunto de sinónimos llamados **synsets**. SWN asigna a cada synset en WordNet tres propiedades (positivo, negativo y objetivo), e indica la intensidad de cada uno de estos tres atributos para cada synset. Estos valores que están comprendidos entre [0,1] son obtenidos usando un método semi-supervisado y al ser sumados deben dar 1. Las anotaciones SWN abarcan más de 117.000 synsets. Este recurso se distribuye libremente para su uso no comercial.

2. Lexicón de Bing Liu [HL04] consiste en una lista con 4.783 palabras negativas y otra de 2.006 palabras positivas que se actualizan desde el 2.004. En estas listas se pueden encontrar muchas palabras mal escritas. Realmente no son errores, sino palabras que en el contexto social suelen aparecer con frecuencia escritas de esa forma, por ejemplo, *assult, bumpping, cartoonish* y *pettifog* como palabras negativas y *goood, heros, jollify* y *lovably* como palabras positivas.

3. MPQA *(Multi-Perspective Question-Answering)* [WWC05] es un lexicón subjetivo, que fue generado de diversas maneras y de distintas fuentes. Algunas entradas fueron entresacadas manualmente de otros recursos desarrollados. Otras fueron identificadas usando datos etiquetados y no etiquetados. La mayoría de las entradas fueron reunidas como parte de un trabajo [RW03]. Hoy en día, el lexicón MPQA está compuesto por 8.222 entradas, las cuales llevan asociadas cinco características sobre la palabra o palabras existentes por cada entrada. Las características son la intensidad de la subjetividad, número de palabras en la entrada, categoría gramatical (*PoS*), si es una palabra raíz (*stem)* y su polaridad.

4. General Inquirer [SDSO66] es un lexicón que incluye información pragmática, semántica y sintáctica de una palabra o palabras etiquetadas con un *PoS*. La versión distribuida actualmente combina las categorías de análisis de contenido de los diccionarios[11] Harvard IV y Lasswell, y cinco categorías basadas en el trabajo de conocimiento social de Semin y Fiedler [SF88], haciendo un total de 182 categorías. Cada categoría es una lista de palabras con cierto sentido. Ejemplos de las categorías pueden ser "colores", "partes del cuerpo" o "pronombres demostrativos". La categoría "positiva" tiene 1.915 entradas, mientras que la categoría

[10] http://wordnet.princeton.edu/
[11] http://www.wjh.harvard.edu/~inquirer/homecat.htm

"negativa" con 2.291 entradas, es la más grande de todas. En total, General Inquirer tiene 11.788 entradas que contendrán la palabra o palabras ordenadas alfabéticamente, el diccionario del que surge, la categoría a la que pertenece y el resto es información añadida que da más conocimiento sobre esa palabra o palabras.

5. WordNet-AFFECT [SV04] es un lexicón que se desarrolló en dos etapas haciendo uso de WordNet. La primera etapa consistió en la identificación de un primer "núcleo" de synsets afectivos. El segundo paso consistió en la ampliación del núcleo con las relaciones definidas en WordNet. Para tener un juego inicial de palabras afectivas se realizó manualmente un recurso preliminar llamado AFFECT. Esta base de datos léxica contenía 1.903 términos referidos directa o indirectamente a los estados mentales (por ejemplo, emocionales). Para la recopilación de palabras se contó con la ayuda de diccionarios. Posteriormente, se seleccionó un subconjunto de WordNet que contenía todos los synsets en los que había al menos una palabra de la lista de palabras afectivas de AFFECT y se rechazó esos synsets que no fueron reconocidos como conceptos afectivos. Una vez definidas las relaciones semánticas y léxicas del subconjunto de WordNet y las palabras de AFFECT, se chequeó manualmente la posible extensión de palabras. Actualmente, WordNet-AFFECT contiene 2.874 synsets y 4.787 palabras.

6. Q-WordNet[12] [AG10]. Este lexicón está inspirado en el trabajo realizado para la clasificación de polaridad de las palabras en SentiWordNet, tomando como punto de partida WordNet. No implementa un método supervisado automático sino que se aprovecha de la información lingüística contenida en WordNet anotada por expertos lexicógrafos. Por lo tanto, en lugar de empezar por una lista de palabras (semillas) recogidas manualmente, simplemente se empieza por el synset 'Quality' contenido en WordNet. A partir de aquí y de las siguientes relaciones léxicas obtenidas con otros synsets, se va construyendo este lexicón. La última versión de Q-WordNet 3.0 (procedente de WordNet 3.0) implementa 7.402 palabras positivas y 8.108 palabras negativas.

Los lexicones nombrados anteriormente pertenecen al grupo de recursos lingüísticos en inglés, pero como se ha comentado anteriormente, al haber cada día más información escrita en otros idiomas se hace necesario el desarrollo de lexicones para otros idiomas. En [PBM12] se presenta un marco para la generación de los lexicones de subjetividad en un idioma de destino usando recursos en inglés con datos anotados manual y automáticamente. Obviamente, no será el único

[12] http://nlp.uned.es/semantics/qwn/qwn.html

trabajo que se apoya en recursos lingüísticos en inglés ni el único método para la generación de recursos en distintos idiomas. A continuación, nombramos algunos trabajos que contribuyen al avance del AO en español mediante la generación y creación de lexicones enfocados a dicho idioma.

A. SO-CAL *(Semantic Orientation-CALculator)* [BBT09]. Este lexicón es probablemente el sistema más relevante en AO en español. Este sistema al igual que SO-CAL para textos en inglés (aunque para este idioma no tuvo gran relevancia), además de resolver la orientación semántica almacenada a nivel individual en adjetivos, sustantivos, verbos y adverbios, incluye modificadores de la polaridad como son la negación o los intensificadores (muy, poco, bastante...). También es capaz de detectar y descartar el sentimiento reflejado en el contenido no fáctico del texto, representado, por ejemplo, mediante expresiones condicionales o subjuntivas.

B. iSOL[13] *(improved SOL)* [MMMP13]. Este lexicón fue generado a partir de una traducción automática del inglés al español del lexicón de Bing Liu generando el recurso SOL *(Spanish Opinion Lexicon)*. El recurso iSOL se obtuvo después de llevar a cabo una revisión manual sobre SOL con el fin de mejorar la lista final de las palabras de opinión. Por un lado, debido a la gramática española, tenemos por ejemplo que mientras un adjetivo inglés no tiene ni género ni número, y por lo general, es representado por un solo término, al adjetivo español le corresponde hasta cuatro posibles palabras traducidas al español, dos para el género (masculino o femenino) y dos para el número (singular o plural). Por otra parte, siguiendo la filosofía de Bing Liu se introdujo en las listas algunas palabras mal escritas ya que aparecen con mucha frecuencia en el contenido de los medios de comunicación social, como por ejemplo kaput, pillín o coñacete. Finalmente iSOL se compone de 2.509 palabras positivas y 5.626 palabras negativas. Por lo tanto, este lexicón español contiene 8.135 palabras de opinión.

C. SEL[14] (*Spanish Emotion Lexicon*) es un lexicón presentado en [SMV+13]. Está compuesto de 2.036 palabras que fueron analizadas y anotadas manualmente. Estas palabras fueron etiquetadas con las categorías emocionales básicas: alegría, enfado, miedo, tristeza, sorpresa y repulsión y se le asignó un porcentaje de probabilidad de ser usada con un sentido emocional. A este dato se le llamó FPA (Factor de Probabilidad de uso Afectivo). El lexicón está disponible libremente.

[13] http://sinai.ujaen.es/isol/
[14] http://www.cic.ipn.mx/~sidorov/#SEL

D. ElhPolar[15] [SS13]. Este lexicón fue creado a partir de diferentes fuentes de datos. Por un lado, se hizo una traducción automática del lexicón de polaridad existente en inglés OpinionFinder [WHS+05] mediante un diccionario bilingüe Inglés-Español. Las traducciones ambiguas se resolvieron de forma manual. Como segunda fuente del lexicón, las palabras se extraen automáticamente de un corpus generado a partir de tweets. Además, incluyen una lista de vocabulario coloquial generado mediante la recopilación de palabras a partir de dos fuentes: el diccionario "Diccionario de jerga y expresiones coloquiales"[16] y el vocabulario coloquial editado por los usuarios en una web de *crowdsourcing*[17].

E. ML-SentiCon[18] *(Multi-Layered Miltilingual Sentiment lexiCon).* En [CTPO14] se presentó un método automático para la construcción de lexicones de polaridades semánticas a nivel de lema, para inglés, español y las otras tres lenguas oficiales en España (catalán, gallego y euskera). Para generar estos lexicones se reprodujo el método original empleado para la construcción de SentiWordNet 3.0 añadiéndole unas mejoras, que refleja significativamente mejores estimaciones de la positividad y negatividad, de acuerdo con sus evaluaciones. El método se divide en dos partes claramente diferenciadas: un primer cálculo individual de la polaridad, y un segundo cálculo global de la polaridad a partir de los valores obtenidos en la primera parte. Los lexicones están estructurados en varias capas, lo que permite seleccionar distintos compromisos a priori entre la cantidad de palabras disponibles y la exactitud de las estimaciones. En total se han creado 8 capas.

Por último, vale la pena nombrar MCR *(Multilingual Central Repository)* [AVR+04; GLR12] porque combinándolo con SWN en inglés es posible conseguir un SentiWordNet para el idioma español. MCR es un recurso lingüístico a gran escala que puede ser usado en procesos semánticos que necesitan gran cantidad de conocimiento lingüístico. MCR integra en el mismo marco de trabajo de EuroWordNet, diversas versiones de WordNet para las diferentes lenguas, inglés, español, vasco, catalán y gallego. MCR se puede considerar como un inventario o repositorio de nombres, verbos, adjetivos y adverbios para estos cinco idiomas. La versión final de MCR contiene alrededor de 1,6 millones de relaciones semánticas entre los synsets, siendo la mayoría de ellos adquiridos mediante métodos automáticos. Los synsets han sido construidos siguiendo el modelo

[15] http://komunitatea.elhuyar.org/ig/files/2013/10/ElhPolar_esV1.lex
[16] http://www.ual.es/EQUAL-ARENA/Documentos/coloquio.pdf
[17] www.diccionariojerga.com
[18] http://timm.ujaen.es/recursos/ml-senticon/

propuesto por EuroWordNet, en los cuales los WordNet se enlazan mediante un índice entre lenguas (*InterLingual Index-ILI).* Por medio de este *ILI* los lenguajes están conectados, haciendo posible ir desde una palabra de un idioma a otras palabras similares traducidas a otros idiomas. Este recurso está en continuo crecimiento siendo la última versión entregada MCR 3.0.

2.3 Métodos para generación de lexicones adaptados al dominio

Los lexicones anteriormente nombrados se han generado usando principalmente el método basado en diccionario, pero desde que en [AG05] se demostró que los clasificadores de polaridad realizaban una buena estimación en los dominios en los que los clasificadores eran entrenados y que su rendimiento caía drásticamente cuando era usado para clasificar datos de diferentes dominios, trabajos para mejorar los lexicones fueron apareciendo, normalmente haciendo uso del método basado en corpus.

En [KJM08] se genera un sistema de análisis de opinión considerando el conocimiento específico en el campo de la economía. En lugar de crear una gran base de conocimiento costosa, recopilan información de distintas fuentes. Por un lado, un grupo de anotadores especialistas en economía manualmente anotaron términos económicos con su polaridad y su intensidad haciendo uso de diccionarios específicos en dicha materia. Los términos no económicos, pero influyentes en el domino, también fueron recogidos y, además, también emplearon una colección de noticias generales para adjuntar palabras con mayor número de frecuencia en alguna de las clases con una polaridad independientemente del dominio.

En [QLBC09] se propuso un método para extraer palabras de opinión de un dominio específico utilizando algunas palabras de opinión (semillas). La idea principal es explotar ciertas relaciones sintácticas de las palabras de opinión y las características del objeto para la extracción. Se demostró que las palabras de opinión se asocian casi siempre con las características del objeto en algunos aspectos. Por lo tanto, las palabras de opinión pueden ser reconocidas por características identificadas, y las características pueden ser identificadas por palabras de opinión conocidas. Las palabras y las características de opinión extraídas se utilizan para identificar las palabras nuevas de opinión y nuevas características, que se utilizan de nuevo para extraer más palabras y características de opinión. Este proceso de propagación termina cuando no hay más palabras de opinión o características que se pueden encontrar. Como el proceso consiste en la propagación a través de las palabras y las características de opinión, el método se llama propagación doble. Las reglas de extracción están diseñadas basándose en las relaciones existentes entre las palabras y las características de opinión y viceversa.

[DYTS12a] proponen un método para construir un sistema de clasificación de polaridad dependiente del dominio. El dominio seleccionado por los autores es el de hoteles. Cada comentario está representado por un conjunto de características independientes del dominio y un conjunto dependiente del dominio. Las características independientes del dominio se extraen de SWN. Para construir el conjunto de características dependientes del dominio, los autores proponen tomar el lexicón construido por Hu y Liu [HL04] y elegir esas palabras positivas/negativas que se repiten un número significativo en los comentarios positivos/negativos del corpus de entrenamiento utilizado para la experimentación. Se utilizó un corpus de 6.000 comentarios en inglés obtenidos de TripAdvisor.

[DYTS12b] proponen un método para adaptar un recurso lingüístico independiente del dominio, como SWN, a un dominio específico. La evaluación se realiza con un corpus de comentarios en inglés sobre hoteles descargado de TripAdvisor. La clave de la adaptación al dominio de SWN no es difícil. El método consiste en la actualización del valor de polaridad (positiva/negativa) de una palabra que aparece en SWN si dicha palabra es más frecuente en la otra clase (negativa/positiva).

3. Discusión de resultados

Esta sección muestra un resumen de las distintas propuestas que se recogen en esta memoria y presenta una breve discusión sobre los resultados obtenidos para cada una de ellas.

3.1 Clasificación de polaridad supervisada y no supervisada para un corpus comparable en inglés y español

Se trata del primer trabajo que se realizó y en el que se puso de manifiesto la necesidad de investigar y estudiar en detalle el AO en español.

En este trabajo se presentan un conjunto de experimentos para corpora de opiniones de diferentes productos escritos en los idiomas inglés y español. Los corpora son comparables y los empleados son el corpus SFU en inglés y corpus SFU en español. Se implementarán un método supervisado basado en SVM y dos métodos no supervisados. Para la clasificación de polaridad no supervisada se usa un método basado en grafos [MMMU12] y otro basado en lexicón. El lexicón de Bing Liu es usado en la clasificación para el corpus en inglés mientras que una traducción automática al español de dicho lexicón con algunas mejoras manuales es el que se usa para la clasificación del corpus español, concretamente, el lexicón usado para español es SOL (*Spanish Opinion Lexicon).* Los resultados de la clasificación de polaridad usando una aproximación no supervisada se muestran en la Tabla 1.

Corpus SFU	Método	Precision	Recall	F1	Accuracy
Inglés	**Lexicón Bing Liu**	69,56%	64,42%	66,89%	64,75%
Español	**SOL**	66,91%	61,94%	64,33%	**62,25%**
Inglés	**Basado en grafo**	68,83%	62,50%	65,51%	62,50%
Español	**Basado en grafo**	65,91%	63,50%	64,68%	**63,05%**

Tabla 1. Clasificación de polaridad no supervisada para el corpus SFU

Este artículo es el punto de partida que justifica la necesidad de generar y desarrollar lexicones en español para la clasificación de polaridad basada en orientación semántica, porque como se puede observar en la Tabla 1, los resultados son muy similares para la clasificación de polaridad no supervisada usando un método basado en grafos y otro basado en lexicón para el corpus SFU en la versión española. Esto nos hace pensar que no siempre implementar métodos tediosos y más complicados nos lleva a una mejora considerable y que con el método basado en lexicón incorporando mejoras se puede conseguir resultados superiores.

El artículo asociado a esta parte es:

a) Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. (2013). Bilingual experiments on an opinion comparable corpus. *WASSA 2013*, 87.

3.2 Generación de lexicones para la clasificación de polaridad basada en orientación semántica para corpus en español

En esta temática se han realizado varias contribuciones en las que se proponen distintas métodos para generar los lexicones que nos facilitarán la clasificación de polaridad basada en orientación semántica para corpora en español.

1. Propuesta para generar lexicones de propósito general

En este trabajo b) se presentan tres lexicones en español para hacer la clasificación de polaridad del corpus en español MuchoCine MC.

Inicialmente se explica el desarrollo y creación del recurso lingüístico SOL (*Spanish Opinion Lexicon*) basado en la traducción al español del lexicón de Bing Liu [HL04]. Este lexicón SOL contiene 1.503 palabras negativas y 523 palabras positivas menos que el lexicón de Bing Liu. Esto es debido a que muchas palabras en inglés tienen en su traducción la misma palabra en español y otro motivo es que muchas palabras son coloquialmente usadas pero no son aceptadas por el diccionario o no tienen traducción al español.

A partir de este primer recurso lingüístico SOL generado, en este trabajo se hace una mejora y se crea el lexicón iSOL (*improved* SOL). Siguiendo la filosofía del lexicón en inglés, se añaden las palabras que están escritas erróneamente o son inexistentes en el diccionario pero son usadas en la lengua española. También se agregan manualmente palabras sinónimas de aquellas traducciones al español formadas por más de una palabra, por ejemplo 'sin cerebro' o 'sin rumbo' son sustituidas por 'descerebrado' o 'desorientado', respectivamente. Otro tema que se tuvo en cuenta fue el relacionado con el género y el número presente en la gramática española, mientras en inglés sólo existe una forma para los adjetivos, en español frecuentemente existen cuatro. Así por ejemplo, la palabra inglesa '*good*' puede ser traducida al español por 'bueno', 'buena', 'buenos' y 'buenas'. Por lo tanto, el lexicón iSOL contiene 843 palabras negativas y 503 palabras positivas más que el lexicón inglés.

Estos dos lexicones en español descritos anteriormente son de propósito general. Pero como es sabido por la comunidad investigadora en AO, la orientación semántica de las palabras depende del dominio en el cual se trabaja. Así pues, en este trabajo se hace una primera aproximación generando un lexicón eSOL (*enriched* SOL) dependiente del dominio 'Cine'. Para dicho fin, se usa el método basado en corpus. El corpus utilizado es el corpus MC. Haciendo una revisión manual y subjetiva, se seleccionan 13 palabras negativas y 26 palabras positivas. En la Tabla 2 se pueden ver dichas palabras.

Palabras positivas		**Palabras negativas**
gran	sorpresa	menos
buen	sorprendente	fallida
imprescindible	increible	previsible
original	emotiva	falta
calidad	unica	minimo
redonda	sorprende	difícil
espectaculo	funciona	pretensiones
delicia	soberbia	pierde
delicias	preciosa	carece
conseguida	impactante	aburre
conseguido	gustara	predecible
trepidante	gustan	fallido
talento	transmite	olvidable

Tabla 2. Palabras incluidas en iSOL para generar eSOL en el dominio 'Cine'

La clasificación de polaridad con los distintos lexicones se muestra en la Tabla 3, obteniendo resultados que nos animan a seguir por este camino.

Lexicón	Macro-precision	Macro-recall	Macro-F1	Accuracy
SOL	56,15%	56,00%	56,07%	56,23%
iSOL	62,22%	61,47%	61,84%	61,83%
eSOL	63,93%	62,74%	63,33%	**63,16%**

Tabla 3. Clasificación del corpus MC en español con lexicones SOL, iSOL y eSOL

2. Propuesta para generar lexicones adaptados al dominio

Los resultados obtenidos con el lexicón eSOL basado en el corpus MC, en el dominio 'Cine', hacen que nuestros siguientes pasos se adentren en ese camino. Los artículos asociados a esta parte son 3. En el primer y segundo artículo (artículos c y d listados más adelante) se hace una aproximación semiautomática para generar lexicones adaptados a dominios y en el tercer artículo e) se implementa una aproximación totalmente automática.

En el primer trabajo c) se experimenta con otro corpus en español con opiniones para 8 dominios distintos y se desarrollan distintos lexicones para cada dominio. El corpus empleado es el SFU en español. Para hallar la frecuencia de las palabras se emplean dos métodos distintos, a los que llamamos 'local' y 'global'.

El método 'local' consiste en contar la frecuencia absoluta de las palabras por clase (opiniones positivas y negativas) para cada dominio.

El método 'global' consiste en contar la aparición de las palabras en cada opinión y en caso de aparecer, independientemente del número de veces que ello ocurra, solo se cuenta como 1. De esta manera, la frecuencia obtenida no podrá ser mayor que el número de opiniones por clase.

Para enriquecer el lexicón de propósito general tomando como base iSOL, y generar lexicones para cada tipo de método, se utilizan dos formas distintas. La primera, consiste en añadir el grupo de palabras con polaridad dependiente del dominio que más aparezca por clase, y la segunda manera, consiste en añadir además del grupo de palabras nombradas anteriormente, aquellos sustantivos que no tienen polaridad pero son usadas en un cierto tipo de clase de opinión y cumplen una ecuación. La ecuación que deben cumplir es:

$$lista(palabra) = \begin{cases} positiva\ \ if\ \ (f^- = 0\ and\ f^+ \geq 3) OR\ \left(\dfrac{f^+}{f^-} \geq 3\right) \\ negativa\ \ if\ \ (f^+ = 0\ and\ f^- \geq 3) OR\ \left(\dfrac{f^-}{f^+} \geq 3\right) \end{cases}$$

Siendo f^- la frecuencia con que aparecen las palabras en las opiniones negativas y f^+ la frecuencia con la que aparecen las palabras en las opiniones positivas.

En este trabajo se han generado 32 lexicones (4 lexicones por dominio), dependiendo de la búsqueda de las frecuencias en las palabras y del grupo de palabras añadidas finalmente. Para validar estos recursos se han empleado 20 opiniones por dominio. Los lexicones han mejorado la clasificación de polaridad en 4 dominios, la han mantenido igual en 2 dominios y lo han empeorado en otros 2 dominios.

En este trabajo además hemos hecho un experimento en el cual, se han desarrollado 4 lexicones dependientes del corpus completo, es decir, independientemente de los ocho distintos dominios. En este caso la clasificación de polaridad de las 160 opiniones que no han intervenido en la generación de los lexicones ha dado unos resultados que mejoran la clasificación con el lexicón de propósito general tomado como base iSOL. En la Tabla 4 se muestran los resultados.

Lexicón	Macro-P	Macro-R	Macro-F1	Accuracy
iSOL	64,52%	61,25%	62,84%	61,25%
eSOLLocal	69,50%	64,38%	66,84%	64,38%
eSOLLocal*	70,67%	61,25%	65,62%	61,25%
eSOLGlobal	69,50%	64,38%	66,84%	64,38%
eSOLGlobal*	69,53%	62,50%	65,83%	62,50%

Tabla 4. Clasificación del corpus SFU en español con lexicones adaptados

Los lexicones terminados con un asterisco *, son aquellos que se han generado añadiendo el grupo de palabras formadas por las que tienen polaridad en el dominio y los sustantivos que no tienen polaridad pero sí son usadas más a menudo en una clase (positiva o negativa) que en otra. Puede comprobarse que con este tipo de inclusión de palabras se obtienen mejores resultados que con el lexicón base iSOL, pero ofrece peores datos que aquella clasificación con los lexicones que solo incluyen palabras con polaridad.

En el segundo d) y tercer artículo e) se expone la generación de lexicones añadiendo las palabras encontradas en el corpus COAH de opiniones escritas en español en el dominio 'Hoteles' al lexicón de propósito general iSOL.

El método para encontrar las palabras en el corpus y el tipo de palabras que serán añadidas es diferente en cada uno de estos dos artículos.

En el segundo artículo d) se halla la frecuencia de las palabras por clase y, posteriormente, se hace una revisión manual para hallar las palabras con polaridad más empleadas. Para validar el lexicón generado se hace uso de la parte del corpus SFU en el dominio 'Hoteles' (50 opiniones). Los resultados obtenidos son satisfactorios y superan a los obtenidos con el lexicón base iSOL. En la Tabla 5 se pueden ver como la mejora es considerable con este método tan sencillo. Haciendo la clasificación de polaridad mediante la aproximación supervisada (SVM) comprobamos como para este corpus las distancias en la clasificación de polaridad se acortan.

Enfoque		MacroF1	Accuracy.
No Supervisado	**iSOL**	73,52%	70,0%
No Supervisado	**eSOLHotel**	81,22%	78,0%
Supervisado	**SVM**	82,71%	82,0%

Tabla 5. Clasificación del dominio 'Hoteles' del corpus SFU en español

En el tercer artículo e), la frecuencia de las palabras halladas se hace automáticamente. Por lo tanto crearemos lexicones eSOLHotel^n añadiendo al lexicón iSOL *n* grupos de palabras que deben cumplir la siguiente ecuación.

$$lista(palabra) = \begin{cases} positiva\ if\ \ (f^- = 0\ and\ f^+ \geq n)OR\ \left(\frac{f^+}{f^-} \geq n\right) \\ negativa\ if\ \ (f^+ = 0\ and\ f^- \geq n)OR\ \left(\frac{f^-}{f^+} \geq n\right) \\ \qquad n = 2,3,4,5,6,7 \end{cases}$$

En la Tabla 6 se muestran algunas de las palabras que se han añadido a todos los lexicones generados automáticamente. Estas palabras no tienen por qué tener ninguna información subjetiva en el dominio 'Hoteles', aunque queda demostrado que tienden a ser más utilizadas cuando se opina positiva o negativamente sobre dicho asunto.

Palabras positivas	Palabras negativas
Desconectar	Pelos
Decoradas	Dinero
Tapas	Muelles
Repetir	Moho
Serviciales	Sabana
Climatizada	Pelusas
Azotea	Moqueta
Silencio	Voces
Grata	Desconchones
Disposición	Pagado

Tabla 6. Ejemplo de palabras añadidas a los lexicons eSOLHotel^n

Como puede observarse la palabra 'grata' es una palabra con polaridad positiva independientemente del dominio. Por otro lado, las palabras 'climatizada' o 'desconchones' son palabras dependientes del dominio 'Hoteles' con polaridad positiva y negativa, respectivamente. Por último, las palabras 'azotea' o 'voces' son palabras sin información subjetiva que tienden a usarse más cuando se opina positiva o negativamente en el dominio 'Hoteles'.

Para la validación usamos el corpus SHoRe de opiniones de hoteles. Solamente las opiniones procedentes de los usuarios de España (*Spain*) son tomadas en cuenta. El número de estas opiniones son 32.920. Los resultados demuestran cómo la inclusión de palabras empleadas en el dominio mejora la resolución de la clasificación de polaridad, siendo con el grupo n=4 donde los resultados alcanzan las mayores mejoras. Dichos datos se muestran en la Figura 3.

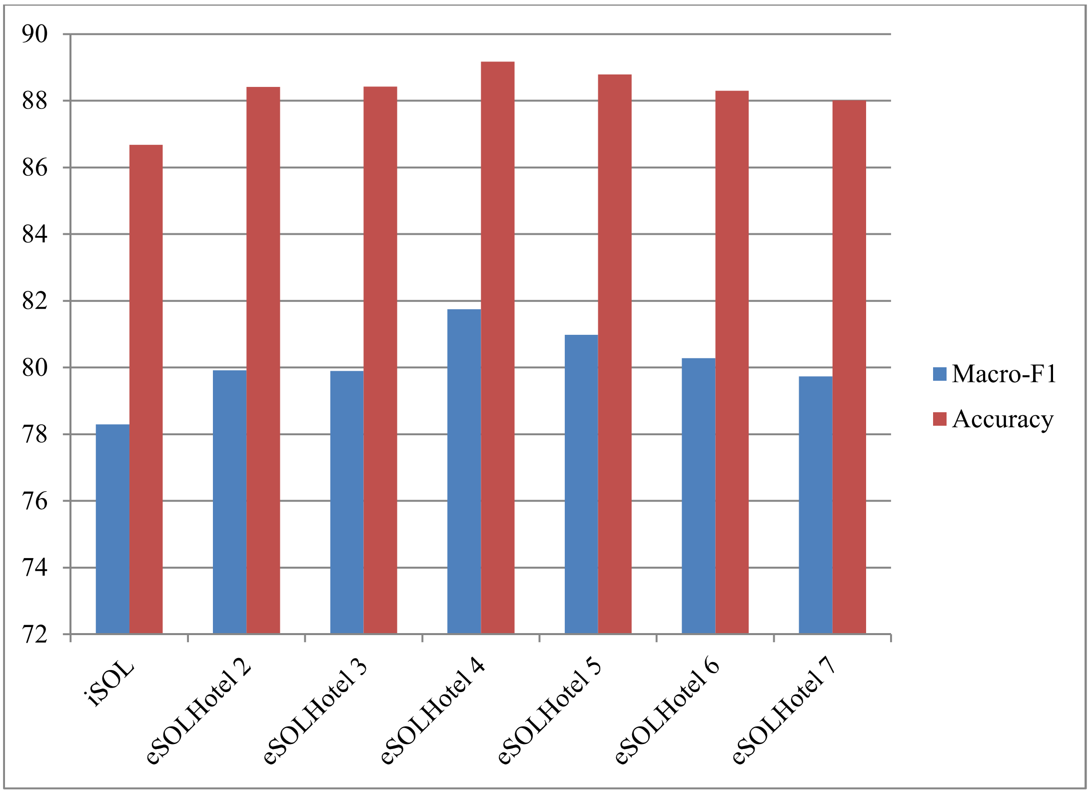


Figura 3. Clasificación de polaridad del corpus SHoRe para distintas con diferentes lexicones.

Los artículos asociados a esta parte son:

b) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, *40*(18), 7250-7257.

c) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014). A Spanish Semantic Orientation Approach to Domain Adaptation for Polarity classification. *Information Processing and Management*

d) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Jiménez Zafra, S. (2014). eSOLHotel: Generación de un lexicón de opinión en español adaptado al dominio turístico. *Procesamiento del Lenguaje Natural*, nº 54. En revisión

e) Molina-González, M. D., Martín-Valdivia, M. T., Martínez-Cámara, E., Ortega F. J., & Cruz F. L. (2014). Automatic generation of subjective lexicons adapted to a specific domain using linguistic resources. *Knowledge-Based Systems.* En revisión

3.3 Generación de nuevo corpus en español

Debido a la falta de corpus etiquetados en español con gran cantidad de opiniones, en este artículo se presenta un nuevo corpus de opiniones en español en el dominio 'Hoteles' llamado COAH (*Corpus of Opinion about Andalucian Hotels*). En la Tabla 7 se recogen algunas de las características del corpus.

#Opiniones	1.816
#Hoteles	80
Media de opiniones por hotel	22,7
#Palabras	264,303
#Frases	9.952
#Adjetivos	17.800
#Adverbios	15.219
#Verbos	38.590
#Sustantivos	53.640
Media de palabras por frase	26,55
Media de palabras por opinión	145,54
Media de adjetivos por opinión	9,80
Media de adverbios por opinión	8,38
Media de verbos por opinión	21,25
Media de sustantivos por opinión	29,54

Tabla 7. Estadísticas de COAH

Como puede verse en la Tabla 7, el corpus COAH tiene 1.816 opiniones. La distribución de las opiniones según su valoración puede verse en la Tabla 8, donde el valor 1 corresponde a un hotel considerado pésimo por el usuario y el valor 5 corresponde a un hotel considerado excelente por el usuario.

Valoración	**Número de opiniones**
1	312
2	199
3	285
4	489
5	531
Total	**1.816**

Tabla 8. Distribución por valoración

Además de presentar el corpus en este trabajo también se ha realizado la clasificación de polaridad. Para ello, se ha reutilizado el lexicón adaptado al dominio 'Hoteles' generado a partir del método 'local', en el cual solo hemos añadido las palabras con polaridad más frecuentemente usadas en cada clase. Este lexicón empleado está basado en corpus, habiéndose empleado la parte del corpus SFU en español perteneciente al dominio 'Hoteles'.

El artículo asociado a esta parte es:

f) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014). Cross-domain Sentiment Analysis using Spanish Opinionated Words. *The 19th international Conference on Application of Natural Language to Information Systems,* NLDB'14. DOI: 10.1007/978-3-319-07983-7_28

3.4 Meta-clasificadores para clasificación de polaridad integrando distintos recursos léxicos en español

En este trabajo hemos querido avanzar en el campo del bilingüismo, usando recursos en inglés para mejorar la clasificación de polaridad para un corpus en español. Para ello realizamos experimentos individuales en ambos idiomas y, posteriormente, haremos uso de meta-clasificadores para obtener mejoras en la clasificación de polaridad en español. El esquema empleado es el que se muestra en la Figura 4.

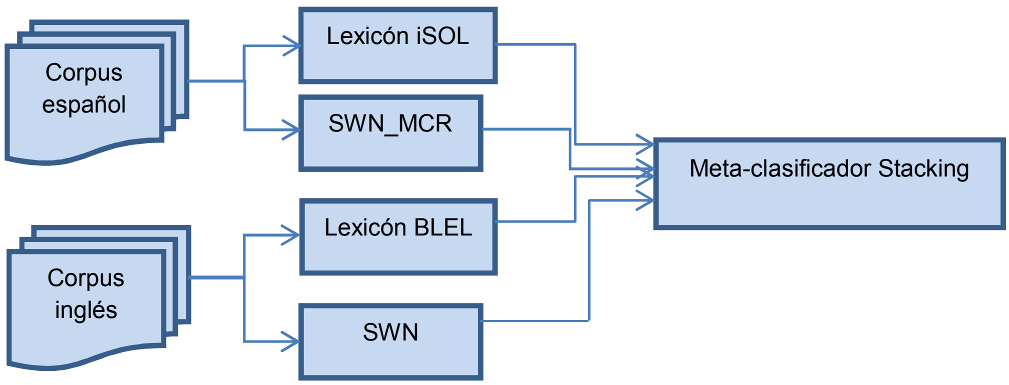


Figura 4. Esquema del meta-clasificador

El corpus empleado es el corpus MC y los recursos lingüísticos empleados son SWN, MCR, el lexicón de Bing Liu y el lexicón iSOL. Inicialmente se realizan dos experimentos individuales dependientes del lenguaje, es decir, se obtiene la clasificación de polaridad para el corpus español MC usando el lexicón iSOL y por otro lado, con el recurso lingüístico SWN enlazando el synset inglés con su correspondiente en español de MCR. Este procedimiento es novedoso y hasta el momento no se ha contemplado en la literatura del AO en español. Al igual que se hace para el corpus en español MC, se realiza una traducción de dicho corpus al inglés y se obtiene la clasificación de polaridad usando el lexicón de Bing Liu y SWN

Según se ve en la Tabla 9, se consiguen mejores resultados en español que en inglés si usamos los lexicones. Sin embargo, se obtienen mejores resultados si usamos el recurso lingüístico SWN sobre el corpus en inglés, que combinado con MCR para el corpus en español (Tabla 10).

Clasificación	Macro-P	Macro-R	Macro-F1	Acc
iSOL sobre MC Español	62,22%	61,47%	61,84%	61,83%
BLEL sobre MC Inglés	61,92%	56,58%	59,13%	57,56%

Tabla 9. Clasificación de polaridad para corpora usando lexicones

SWN	Macro-P	Macro-R	Macro-F1	Acc
(Adj+adv) sobre MC Español	63,68%	58,79%	61,14%	59,66%
(Adj+nouns+verb+adv) sobre MC Inglés	65,13%	64,72%	64,92%	64,95%

Tabla 10. Clasificación de polaridad para corpora usando SWN (mejor resultado)

Finalmente, con el uso del meta-clasificador, usando la combinación de los resultados conseguidos con los recursos en español exclusivamente aumentamos el Acurracy en la clasificación de polaridad hasta un 63,85%, sin embargo, sí en la combinación, además de los resultados en español, usamos los conseguidos en la clasificación de polaridad para el corpus traducido al inglés empleando el recurso SWN, aún mejoramos más los resultados finales. En la Tabla 11 se puede ver un resumen.

Meta-clasificador (NB) combinando distintos recursos	Macro-F1	Accuracy
iSOL_SWN_SP_pred	63,80%	63,85%
iSOL_SWN_SP_pred_polar_MCE_SWN_EN_pred_polar	64,79%	64,68%

Tabla 11. Clasificación de polaridad para corpus MC en español

El artículo asociado a esta parte es:

g) Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Perea-Ortega, J. M. (2014). Integrating Spanish Lexical Resources by Meta-classifiers for polarity classification. *Journal of Information Science*. DOI: 10.1177/0165551514535710

4. Conclusiones

Como acabamos de describir, hemos seguido una línea de trabajo totalmente encadenada que comienza con una visión general de la clasificación de polaridad supervisada y no supervisada para corpora comparables en dos idiomas, el inglés y el español. Centrándonos en la clasificación basada en la aproximación no supervisada sobre corpus en español usando dos métodos distintos, comprobamos que los resultados del método basado en lexicón son comparables a los obtenidos con el basado en grafos, método más complicado y tedioso de implementar. Este será nuestro punto de partida para decantarnos por la creación de recursos lingüísticos en español para la clasificación de la polaridad en nuestro idioma destino.

Los primeros recursos lingüísticos que generamos son lexicones que están traducidos automáticamente del lexicón de Bing Liu y mejorados manualmente. Estos lexicones son de propósito general y comparando la clasificación de polaridad basada en orientación semántica para corpora de opiniones en español aplicando nuestro lexicón iSOL con la obtenida con otro recurso léxico, como puede ser el lexicón SEL, concluimos que son acertadas las decisiones tomadas.

Posteriormente, damos un paso adelante generando lexicones basados en corpus. Con la ayuda de dos corpora existentes en español, creamos lexicones adaptados a varios dominios. Comprobamos que tanto para la generación como para la validación de los lexicones es necesario el uso de corpus con bastantes opiniones (del orden del centenar como mínimo), siendo por tanto el corpus SFU limitado para sacar conclusiones aceptables, aunque en la mayor parte de los experimentos realizados con tal corpus, los resultados han mejorado los hallados con el lexicón tomando como base iSOL.

La mayor parte de los artículos que se han discutido en la sección 3 han utilizado solo dos corpora en español, debido a la dificultad de encontrar otros disponibles en este idioma. Por tal motivo y ante la necesidad de experimentar con distintos corpora la clasificación de polaridad basada en orientación semántica con los lexicones creados, nos vemos animados a realizar otro corpus de un dominio distinto al de Cine, ya que el corpus MC es sobre dicho dominio. Además, aunque el corpus SFU es variado en dominios, en número de opiniones por dominio es escaso. Concretamente, hemos creado un corpus en el dominio ‘Hoteles’. La elección de dicho dominio viene tomada por el interés que despierta en todas las clases sociales el conocimiento sobre un posible hospedaje en cualquier lugar del mundo. Este dominio además dispone de portales web confiables y con gran cantidad de opiniones etiquetadas aceptablemente para generar un corpus con un número de opiniones considerable.

Este nuevo corpus de opiniones en el dominio ‘Hoteles’ es usado para la generación de otro lexicón adaptado al dominio, llegando a la conclusión general que la inclusión de palabras al lexicón base iSOL, hace mejorar la clasificación de polaridad.

Ampliando el estudio al uso de recursos lingüísticos en inglés para mejorar la clasificación de polaridad para corpora en español, llegamos a la conclusión de que la integración de recursos semánticos ayuda al proceso de clasificación de polaridad. Por otra parte, y tras comprobar los resultados obtenidos, no solo en nuestros experimentos, sino también en los de otros investigadores que lo han utilizado, llegamos a la conclusión de que el lexicón iSOL es un recurso para análisis de sentimientos en español sobradamente efectivo. Esto hace reafirmarnos en la elección de la línea de investigación.

En la Tabla 12 se muestran los recursos lingüísticos para análisis de opiniones en español, que han sido el principal aporte que ha suscitado la realización de esta tesis, junto a la ubicación donde poder descargarlos para posteriores fines de investigación.

Recurso Lingüístico	URL
iSOL	http://sinai.ujaen.es/isol/
eSOL (enriquecido con corpus MC)	http://sinai.ujaen.es/esol/
eSOLDomainGlobal	http://sinai.ujaen.es/esoldomainglobal
COAH	http://sinai.ujaen.es/coah
eSOLHotel^{n}(enriquecido con COAH)	http://sinai.ujaen.es/esolhoteln

Tabla 12. Recursos lingüísticos para análisis de opiniones en español

5. Perspectivas futuras

A continuación, se presentan algunas líneas futuras que se plantean a partir de los métodos propuestos en esta memoria y ante la necesidad de acotar distancias entre la clasificación de polaridad basada en la aproximación supervisada y no supervisada.

1. Desarrollo de varios lexicones en español adaptados al dominio para AO siguiendo diversos métodos para la inclusión de palabras. No todos los dominios son iguales, por tanto, la inclusión de palabras siguiendo una frecuencia determinada y fija, como vimos en el artículo c) no tiene porqué mejorar la clasificación en todos los dominios por igual. Por tal motivo, sería necesario hacer estudios independientes para dominios diferentes.

2. En los trabajos discutidos se ha visto como el uso de la ironía y la negación de palabras cambian la polaridad de las opiniones. La ironía es una tarea bastante compleja de abordar pero en la negación de las palabras, se pueden seguir las reglas gramáticas básicas en la construcción de frases para mejorar la eficiencia en los clasificadores de opiniones. Por tanto, es un campo abierto para la investigación futura.

3. También sería interesante hacer un estudio para el análisis de sentimientos implementando en las listas del lexicón iSOL, otra serie de características como por ejemplo el *PoS*, utilizando los distintos grupos semánticos individualmente y las posibles combinaciones de ellos.

4. No todas las palabras tienen la misma carga de subjetividad positiva y negativa, así, apoyándonos en algún recurso existente (español o inglés) podríamos generar las listas del lexicón iSOL con información que nos permitiera saber qué grado de positividad o negatividad existe en una opinión, sumando los datos de cada una de las palabras encontradas.

5. Nuestra investigación ha estado orientada al español, sin embargo, los métodos desarrollados aquí, podrían extenderse a otras lenguas con similares características lingüísticas, como por ejemplo, el catalán, gallego, portugués o francés, todo ellos procedentes del latín.

Parte II. Publicaciones: Trabajos Publicados, Aceptados y en Revisión

1. Clasificación de polaridad supervisada y no supervisada para un corpus comparable en inglés y español

Las publicaciones en revista asociadas a esta parte son:

a) Martínez-Cámara, E., Martın-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. (2013). Bilingual experiments on an opinion comparable corpus. *WASSA 2013*, 87

Estado: Publicado
Área de conocimiento: Computer Science
Citas: 2

Bilingual Experiments on an Opinion Comparable Corpus

E. Martínez-Cámara
SINAI research group
University of Jaén
E-23071, Jaén (Spain)
emcamara@ujaen.es

M. T. Martín-Valdivia
SINAI research group
University of Jaén
E-23071, Jaén (Spain)
maite@ujaen.es

M. D. Molina-González
SINAI research group
University of Jaén
E-23071, Jaén (Spain)
mdmolina@ujaen.es

L. A. Ureña-López
SINAI research group
University of Jaén
E-23071, Jaén (Spain)
laurena@ujaen.es

Abstract

Up until now most of the methods published for polarity classification are applied to English texts. However, other languages on the Internet are becoming increasingly important. This paper presents a set of experiments on English and Spanish product reviews. Using a comparable corpus, a supervised method and two unsupervised methods have been assessed. Furthermore, a list of Spanish opinion words is presented as a valuable resource.

1 Introduction

Opinion Mining (OM) is defined as the computational treatment of opinion, sentiment, and subjectivity in text. The OM discipline combines Natural Language Processing (NLP) with data mining techniques and includes a large number of tasks (Pang and Lee, 2008). One of the most studied tasks is polarity classification of reviews. This task focuses on determining which is the overall sentiment-orientation (positive or negative) of the opinions contained within a given document.

Two main appraoches are followed by researches to tackle the OM task. On the one hand, the Machine Learning (ML) approach (also known as the supervised approach) is based on using a collection of data to train the classifiers (Pang et al., 2002). On the other hand, (Turney, 2002) proposed an unsupervised method based on the semantic orientation of the words and phrases in the reviews. Both methodologies have their advantages and drawbacks. For example, the ML approach depends on the availability of labelled data sets (training data), which in many cases are impossible or difficult to achieve, partially due to the novelty of the task. On the contrary, the unsupervised method requires a large amount of linguistic resources which generally depend on the language, and often this approach obtains lower recall because it depends on the presence of the words comprising the lexicon in the document in order to determine the polarity of opinion.

Although opinions and comments on the Internet are expressed in any language, most of research in OM is focused on English texts. However, languages such as Chinese, Spanish or Arabic, are ever more present on the web. Thus, it is important to develop resources for these languages. The work presented herein is mainly motivated by the need to develop polarity classification systems and resources in languages other than English. We present an experimental study over the SFU Review Corpus (Taboada, 2008), a comparable corpus that includes opinions of several topics in English and in Spanish. We have followed this line of work: Firstly, we have taken as baseline a supervised experiment using Support Vector Machine (SVM). Then we have tried different unsupervised strategies. The first one uses the method presented in (Montejo-Ráez et al., 2012). This approach combines SentiWordNet scores with a random walk analysis of the concepts found in the text over the WordNet graph in order to determine the polarity of a tweet. This method obtained very good results in short texts (tweets) and so, we want to try it using larger document. Although we have carried out several experiments using different parameters and modifications, the results are not as good as we hoped. For this, we have

Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 87–93,
Atlanta, Georgia, 14 June 2013.

tried a very simple experiment using a list of opinionated words in order to classify the polarity of the reviews. For English we have used the Bin Liu English lexicon (BLEL) (Hu and Liu, 2004) and for Spanish we have automatically translated the BLEL lexicon into Spanish. In addition, we have also checked manually and improved the Spanish list.

The paper is organized as follows: Section 2 briefly describes papers that study non-English sentiment polarity classification and, specifically work related to Spanish OM. In Section 3 we explain the resources used in the unsupervised methods assessed. Section 4 presents the experiments carried out and discusses the main results obtained. Finally, we outline conclusions and further work.

2 Related Work

There are some interesting papers that have studied the problem using non-English collections. Denecke (2008) worked on German comments collected from Amazon. These reviews were translated into English using standard machine translation software. Then the translated reviews were classified as positive or negative, using three different classifiers: LingPipe7, SentiWordNet (Baccianella et al., 2010) with classification rule, and SentiWordNet with machine learning. Ghorbel and Jacot (2011) used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determine the polarity of the reviews. In (Rushdi-Saleh et al., 2011a) a corpus of movies reviews in Arabic annotated with polarity was presented and several supervised experiments were performed. Subsequently, they generated the parallel EVOCA corpus (English version of OCA) by translating the OCA corpus automatically into English. The results showed that they are comparable to other English experiments, since the loss of precision due to the translation process is very slight, as can be seen in (Rushdi-Saleh et al., 2011b).

Regarding Spanish, there are also some interesting studies. Banea et al. (2008) showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. In (Brooke et al., 2009) several experiments are presented dealing with Spanish and English resources. They conclude that although the ML techniques can provide a good baseline performance, it is necessary to integrate language-specific knowledge and resources in order to achieve an improvement. Cruz et al. (2008) manually recollected the MuchoCine (MC) corpus to develop a sentiment polarity classifier based on the semantic orientation of the phrases and words. The corpus contains annotated Spanish movie reviews from the MuchoCine website. The MC corpus was also used in (Martínez-Cámara et al., 2011) to carry out several experiments with a supervised approach applying different ML algorithms. Finally, (Martín-Valdivia et al., 2012) also dealt with the MC corpus to present an experimental study of supervised and unsupervised approaches over a Spanish-English parallel corpus.

3 Resources for the unsupervised methods

In order to tackle the unsupervised experiments we have chosen several well-known resources in the OM research community. In addition, we have also generated a new Spanish linguistic resource.

Comparable corpora are those consisted of texts in two or more languages about the same topic, but they are not the translated version of the texts in the source language. For the experiments, we chose the comparable corpus SFU Review Corpus. The SFU Review Corpus is composed of reviews of products in English and Spanish. The English version (Taboada and Grieve, 2004) has 400 reviews (200 positive and 200 negative) of commercial products downloaded in 2004 from the Epinions web which are divided into eight categories: books, cars, computers, cookware, hotels, movies, music and phones. Each category includes 25 positive reviews and 25 negative reviews. Recently, the authors of SFU Review Corpus have made available the Spanish version of the corpus[1]. The Spanish reviews are divided into the same eight categories, and also each category has 25 positive and 25 negative reviews.

In the unsupervised experiments we have analysed the performance of two approaches, the first one is based on lexicon and the other one in a graph-based method. We have selected the BLEL lexicon (Hu and Liu, 2004) to carry out the experiment based

[1] http://www.sfu.ca/~mtaboada/download/downloadCorpusSpa.html

on lexicon on the English version of the corpus. The lexicon is composed by 6,787 opinion words that indicate positive or negative opinions, which 2,005 are positive and 4,782 are negative. With the aim of following the same approach over the Spanish version, firstly we have translated the BLEL lexicon with the Reverso machine translator, and them we have checked manually the resultant list. The Spanish Opinion Lexicon[2] (SOL) is composed by 2,509 positive and 5,627 negative words, thus in total SOL has 8,136 opinion words. If a review has more or the same positive words than negative the polarity is positive, otherwise negative.

The graph-based method is a modular system which is made up of different components and technologies. The method was first presented in (Montejo-Ráez et al., 2012) with a good performance over a corpus of English tweets. The main idea of the algorithm is to represent each review as a vector of polarity scores of the senses in the text and senses related to the context of the first ones. Besides, the polarity score is weighted with a measure of importance. Taking a review as input, the workflow of the algorithm is the following:

1. Disambiguation: To get the corresponding sense of the words that are in the text is required to disambiguate them. Thus, the output of this first step is one unique synset from WordNet[3] (Miller, 1995) for each term. The input of the algorithm is the set of words with a POS-Tag allowed in WordNet. The graph nature of the WordNet structure is the basis for the UKB disambiguation method proposed by (Agirre and Soroa, 2009). The UKB disambiguation algorithm apply PageRank (Page et al., 1999) on the WordNet graph starting from term nodes, where each term node points to all its possible senses or synsets. The output of the process is a ranked list of synsets for each input word, and the highest rank synset is chosen as candidate sense.

   For the Spanish disambiguation process we have chosen the Spanish WordNet version offered by the project Multilingual Central Repository (MCR) (Gonzalez-Agirre et al., 2012). The Spanish WordNet of MCR has 38,702 synsets while WordNet has 117,659, i.e. the MCR covers the 32.89% of WordNet.

2. PPV: Once the synsets for the reviews are computed, the following step performs a second run of PageRank described in (Agirre and Soroa, 2009). Using the *Personalized PageRank*, a set of Personalized PageRank Vectors (PPVs) is obtained. This vector is a list of synsets with their ranked values. The key of this approach is to take from this vector additional synsets not related directly to the set of synsets disambiguated in the first step. The result is a longer list of pair <*synset, weight*> where the weight is the rank value obtained by the propagation of the weights of original synsets across the WordNet graph.

3. Polarity: The following step is to calculate the polarity score. For this purpose it is necessary a semantic resource to take the polarity score for each retrieved synset in the two previous steps. The semantic resource selected is SentiWordNet (Baccianella et al., 2010). According to these values, the three following equations have been applied to obtain the final polarity value:

$$p(r) = \frac{1}{|r|} \sum_{s \in r} \frac{1}{|s|} \sum_{i \in s} (p_i^+ - p_i^-) w_i \tag{1}$$

$$\begin{aligned} p(r) &= \frac{1}{|r|} \sum_{s \in r} \frac{1}{|s|} \sum_{i \in s} f(p_i) \\ f(p_i) &= \begin{cases} p_i^+ & \text{if } p_i^+ > p_i^- \\ p_i^- & \text{if } p_i^+ <= p_i^- \end{cases} \end{aligned} \tag{2}$$

$$\begin{aligned} p(r) &= \frac{1}{|r|} \sum_{s \in r} \frac{1}{|s|} \sum_{i \in s} f(p_i) \\ f(p_i) &= \begin{cases} 1 & \text{if } i \in \text{ [positive words]} \\ -1 & \text{if } i \in \text{ [negative words]} \\ p_i^+ & \text{if } p_i^+ > p_i^- \\ p_i^- & \text{if } p_i^+ <= p_i^- \end{cases} \end{aligned} \tag{3}$$

where $p(r)$ is the polarity of the review; $|r|$ is the number of sentences in the review r; s is a sentence in r, being itself a set of synsets; i is a synset in s; p_i^+ is the positive polarity of synset i; p_i^- is the negative polarity of synset i and w_i is the weight of synset i.

[2] `http://sinai.ujaen.es/wiki/index.php/SOL`

[3] We have used the 3.0 release of WordNet.

4 Experiments and Results

Systems based on supervised approach are the most successfully in the OM literature. Therefore, we began the set of experiments applying a machine learning algorithm to the SFU corpus. Also, we have carried out a set of unsupervised experiments following a lexicon-based approach and a graph-based algorithm. For all the experiments the evaluation measures have been: precision, recall, F1 and Accuracy (Acc.). The validation approach followed for the supervised approach has been the well-known 10-cross-validation.

The algorithm chose for the supervised experiments is SVM (Cortes and Vapnik, 1995) because is one of the most successfully used in OM. LibSVM[4] (Chang and Lin, 2011) was the implementation selected to carry out several experiments using SVM. We have evaluated unigrams and bigrams as minimum unit of information. Also, the influence of stemmer have been assessed. The weight scheme for representing each unit of information is TF-IDF. The same configuration has been applied to English and Spanish version of SFU corpus. Table 1 and Table 2 show the results for English version and Spanish version respectively.

	Precision	Recall	F1	Acc.
Unigrams	79.07%	78.50%	78.78%	78.50%
Unigrams & stemmer	79.82%	79.50%	79.66%	79.50%
Bigrams	78.77%	78.25%	78.51%	78.25%
Bigrams & stemmer	**80.64%**	**80.25%**	**80.44%**	**80.25%**

Table 1: SVM results for English SFU corpus

	Precision	Recall	F1	Acc.
Unigrams	73.65%	73.25%	73.45%	73.25%
Unigrams & stemmer	**74.10%**	**73.75%**	**73.92%**	**73.75%**
Bigrams	74.02%	73.50%	73.76%	73.50%
Bigrams & stemmer	73.90%	73.50%	73.70%	73.50%

Table 2: SVM results for Spanish SFU corpus

The results show one of the differences between the works published in SA, the use of unigrams or bigrams. In (Pang et al., 2002) is asserted that the reviews should be represented with unigrams, but in (Dave et al., 2003) bigrams and trigrams outperformed the unigrams features. In our case, regarding the results reached without using a stemmer, the use of unigrams as minium unit of information achieves better result than the use of bigrams when the language is English, but bigrams outperform unigrams when the texts are in Spanish. On the other hand, the best result both in English and Spanish is reached when a stemmer algorithm is applied. So, one conclusion of the supervised experiments is that the use of stemmer enhances the polarity classification in reviews. The following conclusion is that in English the presence of pair of words separate better the positive and negative classes, while in Spanish the use of unigrams is enough to classify the polarity when a stemmer algorithm is used.

The set of unsupervised experiments begins with a lexicon-based method. The method consists of find the presence in the reviews of opinion words which are included in a lexicon of opinion words. BLEL has been used for the English reviews, and SOL for the Spanish reviews. The results are presented in Table 3.

	Precision	Recall	F1	Acc.
BLEL lexicon	69.56%	64.42%	66.89%	64.75%
SOL	66.91%	61.94%	64.33%	62.25%

Table 3: Lexicon-based approch results

The differences in the results between the English and Spanish version of SFU Review Corpus are lower when a lexicon is used instead of a machine learning algorithm is applied. In a lexicon-based method is very important the recall value, because it indicates whether the set of words covers the vocabulary of the corpus. The recall value is upper 60% regarding English and Spanish, although is not an excellent value, is good for the two small and independent-domain lexicons. In the case of Spanish the supervised method is only 15.59% better regarding Accuracy. The results show that may be considered the use of a lexicon-based method for Spanish due to the few computer resources needed. Moreover, it must be highlighted the performance of SOL, so it is the first time that this resource is used to resolve a polarity classification problem.

[4] `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

The graph-based method has been described as a modular and flexible algorithm. Due to its modular nature we have carried out several experiments:

1. **wnet_ant+_eq1_[en|es]**: As baseline, we have run the algorithm with the same configuration as is described in (Montejo-Ráez et al., 2012), i.e. using the equation 1.
2. **wnet_ant-_eq1_[en|es]**: We have assessed the algorithm with a version of WordNet without the antonym relation.
3. **wnet_ant+_eq2_[en|es]**: The equation to calculate the polarity is 2
4. **wnet_ant-_eq2_[en|es]**: The same as wnet_ant+_eq2_[en|es] but the antonym relation is not considered.
5. **wnet_ant+_eq3_[en|es]**: The same as wnet_ant+_eq2_[en|es] but the equation 3 is used to calculate the polarity.
6. **wnet_ant-_eq3_[en|es]**: The same as wnet_ant+_eq3_[en|es] but the antonym relation is not considered.

Furthermore, one of the key elements of the algorithm is the possibility of setting the number of related synsets to get from WordNet. In all of the experiments we have evaluated from an expansion of 0 sysnsets to 100 synsets. In Table 4 are the best results obtained with the English and the Spanish version of SFU corpus.

Regarding the results, only for English is evident that the selection of the right equation to calculate the polarity score is important. On the other hand, the initial assumption that the relation of antonym could complicate the calculation of the final polarity, and the use of a graph of WordNet without antonym could enhance the results cannot be demonstrated because these experiments have reached the same results as the obtained ones using the graph with the relation of antonym. The equation 3, which includes additional information (in this case the BLEL lexicon) to calculate the final polarity score, outperforms the original way to get the polarity score (equation 1). The equation 3 for the English version of the corpus reaches 5.84% and 8.4% better results than equation 1 regarding F1 and Accuracy respectively.

The results obtained with the Spanish reviews are a bit different. In this case, the results are always improved when the antonym relation is not taking into account. So the first conclusion is the relation of antonym is not convenient for the calculation of the polarity value on Spanish texts. The process of expansion with related senses has not been relevant for the final results on the English reviews, but when the language of the reviews is Spanish the expansion is more decisive. For the *wnet_ant-_eq3_es* experiment the best result has been reached considering 71 related senses, so we can conclude that for Spanish the context should be considered. Although the best results is obtained with the configuration *wnet_ant+_eq3_es*, it must be highlighted the precision value of 68.03% reached by the configuration *wnet_ant+_eq2_es*. In some OM experiments is more important the precision of the system than the recall or other evaluation measures, so for Spanish reviews should be taken account this configuration too.

Regarding English and Spanish results, Table 4 shows similar performance, i.e. the graph-based algorithm obtained better results when the antonym is not considered and the use of a lexicon of opinion words enhances considerably the results.

The supervised approach clearly outperforms the two unsupervised approaches. The results obtained by the two unsupervised approaches are closer. The lexicon based method has a better performance on English reviews regarding the four different evaluation measures considered. Thus, the lexicon method not only has better results but also it is simpler, faster and more efficient than the graph-based method. Nevertheless, the graph-based method on Spanish reviews outperforms in precision regarding the configuration *wnet_ant+_eq2_es* and in the other three measures take into account the configuration *wnet_ant+_eq3_es*. However, the graph-based method is only 1.64% better regarding the precision value, and 0.54% better regarding F1. Therefore, we also considered the lexicon-based approach as the more suitable approach than the graph-based one.

	Expansion	Precision	Recall	F1	Accuracy
wnet_ant+_eq1_en	2	66.86%	57.25%	61.68%	57.25%
wnet_ant-_eq1_en	2	66.86%	57.25%	61.68%	57.25%
wnet_ant+_eq2_en	0	65.27%	55.5%	59.99%	55.50%
wnet_ant-_eq2_en	0	65.27%	55.5%	59.99%	55.50%
wnet_ant+_eq3_en	3	68.83%	62.50%	65.51%	62.50%
wnet_ant-_eq3_en	**3**	**68.83%**	**62.50%**	**65.51%**	**62.50%**
wnet_ant+_eq1_es	0	65.42%	54.5%	59.46%	54.5%
wnet_ant-_eq1_es	19	64.39%	57.75%	60.89%	57.75%
wnet_ant+_eq2_es	0	68.03%	52.75%	59.42%	52.75%
wnet_ant-_eq2_es	70	64.62%	58.00%	61.13%	58.00%
wnet_ant+_eq3_es	71	65.91%	63.50%	64.68%	63.05%
wnet_ant-_eq3_es	**71**	**65.91%**	**63.50%**	**64.68%**	**63.05%**

Table 4: Results of the graph-based algorithm

5 Conclusion and future work

In this work, we have presented a set of experiments with a comparable corpora in English and Spanish. As it is usual, the supervised experiment has outperforms the unsupervised ones. The unsupervised experiments have included the evaluation of two different approaches: lexicon-based and graph-based. In the lexicon-based approach we have presented a new resource for the Spanish OM research community, being an important contribution of this paper. The results reached with SOL are very closed to the ones obtained with graph-based methods. Although, for short texts the graph-based method performed well, for the kind of reviews used in these experiments is not as good. Due to the fact that for English the BLEL lexicon has reached better results, for Spanish the results of SOL are nearly the same ones obtained by the graph method, and the use of a lexicon is more efficient, we conclude that the lexicon-based method is most suitable.

Currently we are improving the SOL lexicon, and also we are adding domain information to the words in SOL. Furthermore, one of our main objectives is the treatment of the negation because we considered that is essential for OM.

Acknowledgments

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), TEXT-COOL 2.0 project (TIN2009-13391-C04-02) and ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. Also, this paper is partially funded by the European Commission under the Seventh (FP7 - 2007-2013) Framework Programme for Research and Technological Development through the FIRST project (FP7-287607). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

Eneko Agirre and Aitor Soroa. 2009. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, pages 33–41, Stroudsburg, PA, USA. Association for Computational Linguistics.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Carmen Banea, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. 2008. Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 127–135, Stroudsburg, PA, USA. Association for Computational Linguistics.

Julian Brooke, Milan Tofiloski, and Maite Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the International Conference RANLP-2009*, pages 50–54, Borovets, Bulgaria, September. Association for Computational Linguistics.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20:273–297.

Fermín L. Cruz, Jose A. Troyano, Fernando Enriquez, and Javier Ortega. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural*, 41:73–80.

Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, WWW '03, pages 519–528, New York, NY, USA. ACM.

Kerstin Denecke. 2008. Using sentiwordnet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, pages 507–512. IEEE.

Hatem Ghorbel and David Jacot. 2011. Sentiment analysis of french movie reviews. *Advances in Distributed Agent-Based Retrieval Tools*, pages 97–108.

Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012. Multilingual central repository version 3.0. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña López. 2011. Opinion classification techniques applied to a spanish corpus. In *Proceedings of the 16th international conference on Natural language processing and information systems*, NLDB'11, pages 169–176, Berlin, Heidelberg. Springer-Verlag.

M. Teresa Martín-Valdivia, Eugenio Martínez-Cámara, Jose M. Perea-Ortega, and L. Alfonso Ureña López. 2012. Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*. In press.

George A. Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Arturo Montejo-Ráez, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña López. 2012. Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 3–10, Jeju, Korea, July. Association for Computational Linguistics.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November. Previous number = SIDL-WP-1999-0120.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mohammed Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña López, and José M. Perea-Ortega. 2011a. OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology*, 62(10):2045–2054, October.

Mohammed Rushdi-Saleh, Maria Teresa Martn-Valdivia, Luis Alfonso Urea-Lpez, and Jos M. Perea-Ortega. 2011b. Bilingual Experiments with an Arabic-English Corpus for Opinion Mining. In Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, and Nicolas Nicolov, editors, *RANLP*, pages 740–745. RANLP 2011 Organising Committee.

Maite Taboada and Jack Grieve. 2004. Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Re# port SS# 04# 07), Stanford University, CA, pp. 158q161. AAAI Press*.

Maite Taboada. 2008. Sfu review corpus. `http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html`.

Peter D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

2. Generación de lexicones para la clasificación de polaridad basada en orientación semántica para corpus en español

2.1 Propuesta para generar lexicones de propósito general

Las publicaciones en revista asociadas a esta parte son:

b) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, *40*(18), 7250-7257.

Estado: Publicado
Índice de Impacto (JCR 2013): 1,965
Área de conocimiento: Computer Science, Artificial Intelligence, Ranking: 30/121
Cuartil: Q1
Citas: 12
ISSN: 0957-4174
DOI: 10.1016/j.eswa.2013.06.076

Expert Systems with Applications 40 (2013) 7250–7257

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Semantic orientation for polarity classification in Spanish reviews

M. Dolores Molina-González [a], Eugenio Martínez-Cámara [a,*], María-Teresa Martín-Valdivia [a], José M. Perea-Ortega [b]

[a] SINAI Research Group, University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain
[b] European Commission, Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), Via Fermi, 2749 21027 Ispra (VA), Italy

ARTICLE INFO

Keywords:
Sentiment polarity detection
Multilingual opinion mining
Spanish resources for sentiment analysis

ABSTRACT

Until now most of the published methods for polarity classification have been applied to English texts, but other languages are becoming increasingly important. This paper presents a new resource for the Spanish sentiment analysis research community. We have generated a new lexicon by translating into Spanish the Bin Liu English Lexicon. In order to assess the validity of the proposed lexicon a set of experiments on a Spanish review corpus are presented. In addition, the resource presented is compared with another existing Spanish lexicon. The results show that our resource outperforms the currently available Spanish lexicon for sentiment analysis.

1. Introduction

Recently, interest in Opinion Mining (OM) has grown significantly due to different factors. On the one hand, the rapid evolution of the World Wide Web has changed our view of the Internet. It has turned into a collaborative framework where technological and social trends come together, resulting in the over exploited term Web 2.0. On the other hand, the tremendous use of e-commerce services has been accompanied by an increase in freely available online reviews, comments and opinions about products and services. Web sites such as Amazon,[1] Epinions[2] or IMDb,[3] are queried everyday by customers who want to buy a product and are interested in other buyer's opinions. However, the huge amount of information makes it necessary to develop new methods and strategies to tackle the problem.

Sentiment analysis (SA) systems can be both helpful and influential not only for individual customers but also for any company or institution. These systems automatically accumulate feedback and comments originating from multiple sources, effectively aggregate this information, and present the results in an appropriate way to the user. Thus, SA is becoming one of the main research areas that combines Natural Language Processing (NLP) and Text Mining (TM) to automatically identify and analyze opinions and emotions in documents (Tsytsarau & Palpanas, 2012).

Several subtasks related to SA have been studied such as subjectivity detection (Wiebe, Wilson, & Bell, 2001), review summarization (Somprasertsri & Lalitrojwong, 2010), humor detection (Mihalcea & Strapparava, 2006) and emotion classification (Strapparava & Mihalcea, 2008). One of the most widely studied tasks is sentiment classification, which focuses on determining the polarity of a document, sentence or feature (positive or negative) and on measuring the degree of the polarity expressed in the document (Pang & Lee, 2008). Polarity classification aims to classify a subjective text as positive or negative, according to the overall sentiment expressed by the author. Thus, given a subjective text a sentiment classifier must determine whether the opinion is positive or negative. Although different approaches have been applied to the field of polarity classification, the mainstream basically consists of two major methodologies. On the one hand, the Machine Learning (ML) approach is based on using a collection of data to train the classifiers (Pang, Lee, & Vaithyanathan, 2002). On the other hand, the approach based on computing the semantic orientation (SO) of the words in the texts does not need prior training, but takes into account the orientation of words, positive or negative (Turney, 2002). Both methodologies have their advantages and drawbacks. For example, the ML approach requires training data, which in many cases are difficult or impossible to obtain, partially due to the novelty of the task. On the contrary, the SO approach requires a large amount of linguistic resources which generally depend on the language. In order to take advantage of both methods, some studies apply a hybrid approach (Prabowo & Thelwall, 2009) (Martín-Valdivia, Martínez-Cámara, Perea-Ortega, & Ureña-López, 2013). Usually, the ML approaches obtain better results and currently we can find very good systems working over different domains (Rushdi-Saleh, Martín-Valdivia, Ureña-López, &

* Corresponding author. Tel.: +34 953 21 19 56.
E-mail addresses: mdmolina@ujaen.es (M.D. Molina-González), emcamara@ujaen.es (E. Martínez-Cámara), maite@ujaen.es (M.-T. Martín-Valdivia), jose-manuel.perea-ortega@jrc.ec.europa.eu (J.M. Perea-Ortega).

[1] http://www.amazon.com.
[2] http://www.epinions.com.
[3] http://www.imdb.com.

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.06.076

Perea-Ortega, 2011c). However, the SO method needs more research in order to obtain similar ML results. This is one of the reasons why this paper is focused on semantic orientation for polarity classification.

Another reason is concerned with language. Although opinions and comments in the Internet are expressed in any language, most of the research in OM, and specifically in polarity classification, only deals with English documents. However, languages such as Chinese (Tan & Zhang, 2008), Spanish (Martín-Valdivia et al., 2013) or Arabic (Rushdi-Saleh, Martín-Valdivia, Ureña-López, & Perea-Ortega, 2011a), are ever more present on the web. Therefore, it is important to develop resources to help researchers to work with these languages.

There are two main ways of addressing the problem of applying SA to non-English languages: on the one hand, we can generate resources for the target language, for example corpora, dictionaries, and lists of opinion words. These resources are then used in order to carry out the classification process. On the other hand, we can extract information in the target language, for example in Spanish or Arabic, and translate it into English. This information can then be managed using the available English resources like SentiWordNet (Esuli & Sebastiani, 2006) or WordNet Affect (Strapparava & Valitutti, 2004). This second approach has been successfully applied in several studies, for example translating into German (Denecke, 2008), Arabic (Rushdi-Saleh, Martín-Valdivia, Ureña-López, & Perea-Ortega, 2011b) or Spanish (Martín-Valdivia et al., 2013). However, the generation of resources for the target language is a more difficult and time-consuming task that requires deeper research. Some corpora have been created in other languages than English in order to apply them to a polarity classification system, for example in Arabic (Rushdi-Saleh et al., 2011a) and Chinese (Zhang, Zeng, Li, Wang, & Zuo, 2009). Although we can find some lexicons in several languages, it is noteworthy that there are very few resources for Spanish. Therefore, another motivation of this work is to investigate the effect of using a Spanish lexicon over a corpus of reviews.

In this paper we present a new Spanish resource for OM composed of a list of opinion words; SOL (Spanish Opinion Lexicon). Our main goal is to develop a Spanish lexicon based on one of the most widely-used English lexicons for polarity classification (we will call it BLEL: the Bing Liu English Lexicon). Specifically, we focus on the use of opinion words. In the research literature opinion words are also known as polar words, opinion-bearing words, and sentiment words. Positive opinion words are used to express desired states while negative opinion words are used to express undesired states. Apart from individual words, there are also opinion phrases and idioms. Collectively, they are called the opinion lexicon.

Thus, we have taken the BLEL[4] (Hu & Liu, 2004) and have automatically translated it into Spanish, obtaining the SOL resource. Then we have manually reviewed the lexicon in order to improve the final list of words obtaining iSOL (improved SOL). In order to demonstrate the validity of this resource we have carried out several experiments over a Spanish corpus of movie reviews called MuchoCine (Cruz, Troyano, Enríquez, & Ortega, 2008). The results obtained show that the use of an improved list of sentiment words from the same language can be considered a good strategy for unsupervised polarity classification. Moreover, we have generated another list by integrating the positive and negative words present in the MuchoCine corpus. In this way, we attempt to integrate domain knowledge in the lexicon. Experiments with this enriched eSOL (enriched SOL) show the advantages of integrating external knowledge. Furthermore, we provide a comparative study between our eSOL and other recently published Spanish lexicon, which is known as Spanish Emotion Lexicon (SEL), with the aim of showing the relevance for the research community of the lexicons introduced in this paper. SEL is a resource provided by Sidorov (Sidorov et al., 2012) and it has two implementation details that worth pointing out. Firstly, SEL is composed of 2036 words. Secondly, these SEL words are associated to the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion: joy, anger, sadness, surprise and disgust. The higher the value of the PFA, the more probable the association of the word with the emotion is.

Due to our resource is focused on opinion words, our classification is binary and therefore in order to establish a feasible comparison, we have had to consider the joy and surprise categories as positive and the other as negative words. Thus, we notice that our polarity lexicon is significantly larger than SEL and the experiments show that eSOL has improved accuracy on a reviews polarity classification task opposed to SEL.

The remainder of the paper is organized as follows: Section 2 briefly describes previous related work on semantic orientation for polarity classification and papers that study the problem regarding non-English texts. In Section 3 we explain the methodology used to build the Spanish lexicon as well as different improvements achieved. Section 4 describes the different resources used in our experiments. Section 5 presents the experiments carried out and discusses the main results obtained. Finally, we outline conclusions and further work.

2. Related work

Two main approaches can be distinguished in the field of polarity classification. On the one hand, ML techniques are more extensively used for the classification of reviews. In this approach, the document is represented by different features that may include the use of n-grams or defined grammatical roles like, for instance, adjectives or other linguistic feature combinations. Then a machine learning algorithm is applied. Commonly used machine learning algorithms are Support Vector Machines (SVM), Maximum Entropy (ME) and Naïve Bayes (NB). A survey of studies using ML can be found in Pang and Lee (2008), Liu (2012) or Tsytsarau and Palpanas (2012).

On the other hand there is a lot of work based on the semantic orientation approach, which represents the document as a collection of words. Then the sentiment of each word can be determined by different methods, for example using a web search (Hatzivassiloglou & Wiebe, 2000) or consulting a lexical database like WordNet[5] (Kamps, Marx, Mokken, & de Rijke, 2004). Regarding methods that consider some linguistic features such as adjectives and adverbs, we can find many studies in the literature (Ding & Liu, 2007; Hatzivassiloglou & McKeown, 1997; Kamps et al., 2004; Turney, 2002; Wiebe, 2000). Specifically, our paper is based on the paper by Hu and Liu (2004).

Regarding polarity classification using non-English languages, we can find some interesting studies that apply a semantic orientation approach based on sentiment words. Kim and Hovy (2006) compared opinion expression between an aligned corpus of emails in German and English. One of their experiments translates English opinion-bearing words into German and then analyzes German emails using the German opinion-bearing words. Zhang et al. (2009) applied Chinese sentiment analysis to two datasets. In the first one euthanasia reviews were collected from different web sites, while the second dataset was about six product categories collected from Amazon (Chinese reviews). They proposed a rule-based approach including two phases: firstly, determining each

[4] Available in http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon.

[5] http://wordnet.princeton.edu.

sentence's sentiment based on word dependency, and secondly, aggregating sentences in order to predict the document sentiment. Wan (2009) studied how to reduce the need of using Chinese linguistic resources for SA in Chinese. The author followed a supervised approach and proposed a co-training system based on the use of an English corpus for polarity classification of Chinese products reviews and the usage of a machine translation system.

Finally, there are also some remarkable studies regarding polarity classification focused on Spanish using SO based on bearing words lists. For example, Baena et al. (2008) proposed several approaches to cross-lingual subjectivity analysis by directly applying the translations of opinion corpus in English to training an opinion classifier in Romanian and Spanish. This work showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. Cruz et al. (2008) gathered a corpus of Spanish movie reviews from the MuchoCine website[6]. The MuchoCine (MC) corpus was manually annotated and used to develop a polarity classifier based on the semantic orientation of the words. Brooke, Tofiloski, and Taboada (2009) presented several experiments dealing with Spanish and English resources. They conclude that although the ML techniques can provide a good baseline performance, it is necessary to integrate language-specific knowledge and resources in order to achieve an improvement. They proposed three approaches: the first one uses Spanish resources generated manually and automatically. The second one applies ML to a Spanish corpus. The last one translates the Spanish corpus into English and then applies the SO-CAL (Semantic Orientation CALculator), a tool developed by themselves (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011).

3. Sentiment word lists for Spanish

Three main approaches exist for the compilation of a set of polar words: the manual approach, dictionary-based approach and corpus-based approach. The manual approach is tedious and time consuming, so it is not usually used. However, the manual method is used combined with automated approaches as the final check, because automated ones may make mistakes.

The dictionary-based approach consists of taking manually a small set of sentiment words as seeds with known positive or negative orientations. The following step is enlarging the initial set of seeds by searching in a lexical knowledge base such as WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) for their synonyms and antonyms. The newly found words are included in the seed list. It is an iterative process which ends when no more new words can be found. An example of this method is the paper (Hu & Liu, 2004) where BLEL is presented.

The corpus-based approach is usually applied in two different situations:

1. Given a list of polar words, encounter other opinion words and their polarity from a domain sentiment label corpus.
2. To adapt a general-purpose sentiment lexicon to a new one using a domain corpus for SA applications in the domain.

A good representative of this method is the work of Kanayama and Nasukawa (2006).

Each method has its advantages and drawbacks. The dictionary-based approach is more suitable for the compilation of general-purpose lexicons, while the corpus-based method is better for the generation of domain-dependent sentiment lexicons.

Most of the studies on sentiment words only deal with English documents, perhaps due to the lack of resources in other languages. However, according to the Internet World Stats[7], the number of Internet users with Spanish as their source language is 8%, third after English and Chinese. For this reason, we consider the need to develop a resource as complete as possible composed of sentiment words for Spanish that would be useful for further research activities. This resource was developed in incremental versions which are described in detail below.

Table 1
Examples of English words with same meaning in Spanish.

Several English words	Spanish meaning
Bogus disingenuous dud false phony spurious untrue	*Falso*
Castigate chasten chastise penalize punish	*Castigar*
Crabby glum ill-tempered moody peevish sullen	*Malhumorado*
Absurd absurdness farcical ludicrous preposterous	*Absurdo*
Gaily jolly joyfully joyously merrily	*Alegremente*
Beautifully gloriously marvelously splendidly wonderfully	*Maravillosamente*
Bright lustrous shiny sparkling twinkly	*Brillante*
Affordable economical low-cost low-priced thrifty	*Económico*

3.1. Original list SOL (Spanish Opinion Lexicon)

In a first version we generated a parallel list of sentiment words in Spanish from the opinion lexicon in English provided by Liu[4] (BLEL). This resource was generated by applying automatic machine translation techniques and is composed of approximately 6800 positive and negative opinion words. Reverso[8] was used as an automatic machine translation system, taking into account the first translated word that the system returned for each original word from BLEL. In the process, 1068 negative and 364 positive opinion words were eliminated because their meanings were the same. Table 1 shows some examples of these English words that shared the same first translations.

During the process, we found words that do not have a translation in Spanish with Reverso. We noticed that some of these words were misspelled in the BLEL lists, but they should not be considered mistakes, because they appear frequently in social media content. The rest of words simply have not been recognized by Reverso. Due to both reasons, 435 negative and 159 positive words have been discarded in our resource. Table 2 shows some examples of misspelled and not recognized words.

Other conventions followed when we generated the list were related to writing criteria. For example, each word was written by using non capital letters, without special characters and without accented vowels.

Finally, our first lexicon is composed of approximately 4800 positive and negative opinion words. This resource was called SOL[9] (Spanish Opinion Lexicon).

3.2. Improved SOL (iSOL)

After generating the SOL list, we decided to improve it by addressing some issues that were raised. The resulting list was called iSOL[10] (improved SOL).

At first, we included misspelled Spanish words frequently used in sentiment opinion following the philosophy of the BLEL. Table 3 shows some examples of these misspelled Spanish words.

One of the problems we had to take into consideration was related to the fact that the translation of an English word returned two or more words. For these cases we had to assign manually

[6] http://www.muchocine.net

[7] . Estimation of the number of Internet users by language as of 31 May 2011.
[8] http://www.reverso.net
[9] http://sinai.ujaen.es/?p=1224.
[10] http://sinai.ujaen.es/?p=1202.

Table 2
Examples of discarded words.

Misspelled words	Not recognized words
Assult	Bonny
Goood	Fav
Prospros	Pettifog
Slooow	Bumpping
2-faces	Jollily
Danken	Brainiest
Jutter	Prik

Table 3
Examples of possible translations of misspelled Spanish words in English.

Misspelled Spanish word	Possible translated English word
Cool	*Cool*
Kaput	*Thumbs down*
Pillin	*Naughty*
Coñacete	*Pain in the neck*
Top	*Number one*

Table 4
Examples of some manually reviewed translations.

English	Automatic Spanish translation	Manual assignment
Brainless	*Sin cerebro*	*Descerebrado*
Aimless	*Sin rumbo*	*Desorientado*
Arrogantly	*Con arrogancia*	*Arrogantemente*
Deadlock	*Punto muerto*	*Estancado*
Worthless	*Sin valor*	*Devaluado*
Fashionable	*A la moda*	*Moderno*

the best synonym (composed of only one term) for the translated word. Table 4 shows some examples of these translations performed manually.

Another issue was related to the repetition of words in both lists. For these cases we decided to discard them. A total of 36 words were discarded. Some examples of these words were: *ansioso*, *presumido*, *aturdido*, *increible*, exaltado, *asombrar*, etc.

Finally, the last issue was related to the genre and number present in Spanish grammar. While an English adjective has neither genre nor number and is usually represented by a single term, a Spanish adjective can have four possible translated words, two for the genre (male or female) and two for the number (singular or plural). Table 5 shows some examples of possible translations of English adjectives in Spanish.

3.3. Enriched SOL (eSOL)

The two lexicons described are general-purpose sentiment lexicons. As is well-known in the SA research community, the semantic orientation of a word is domain-dependent. Within the approaches followed by research into the compilation of a set of polar words, the most suitable for obtaining domain-dependent opinion words is that known as the corpus-based approach. Hatzivassiloglou and McKeown (1997) take some adjectives as seeds to find additional sentiment adjectives in the corpus. Their method took advantage of a set of conventions on connectives with the aim of identifying more polar words and their orientation from a sentiment label corpus.

Taking as baseline the lexicon iSOL, we generated a list of opinion words for the cinema domain. We followed the corpus-based approach. The key element of the corpus-based approach is the use of a sentiment labeled corpus. The Spanish corpus selected

Table 5
Examples of possible translations of English adjectives in Spanish.

English	Spanish
Good	*Bueno, buena, buenos, buenas*
Famous	*Famoso, famosa, famosos, famosas*
Pretty	*Guapo, guapa, guapos, guapas*
Ugly	*Feo, fea, feos, feas*
Aching	*Dolido, dolida, dolidos, dolidas*
Bad	*Malo, mala, malos, malas*

Table 6
Number of sentiment words in resources.

Resourses	Number of negative words	Number of positive words
BLEL	4783	2006
SOL	3280	1483
iSOL	5626	2509
eSOL	5639	2536

for the process was MuchoCine (MC), which is described in detail in Section 4.

We followed the same assumption as (Du, Tan, Cheng, Yun, 2010), i.e. a word should be positive (or negative) if it appears in many positive (or negative) documents. Thus, we calculated the word frequency in each class of documents (positive and negative). We found about 15 negative words and 25 positive words. Therefore, these 40 most frequent words that were not yet contained in the iSOL list were added to the final list. This new list integrating information from the corpus was called eSOL[11] (enriched SOL).

Finally these lists have been freely made available as a lexical resource of positive and negative opinion words[9] for use in sentiment analysis for Spanish.

As can be seen in Table 6, we have increased the size of the iSOL and eSOL lists for both negative and positive lists of words with regard to the original list provided by Liu (BLEL). Specifically, for the iSOL list 843 negative and 503 positive words were added, while for the eSOL list 856 negative and 530 positive words were added.

4. Experimental framework

This section presents the measures employed for evaluating the experiments carried out in this paper. Moreover, the main features of the MuchoCine corpus are also shown.

4.1. Evaluation measures

In order to evaluate the different approaches, we have used the traditional measures employed in text classification: precision (P), recall (R), F1 and Accuracy:

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN}$$
$$F1 = \frac{2PR}{P + R}$$
$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP (True Positives) are those assessments where the system and a human expert agree on a label, FP (False Positives) are those labels assigned by the system that do not agree with the expert assignment, and FN (False Negatives) are those labels that the sys-

[11] http://sinai.ujaen.es/?p=1188.

Table 7
Rating distribution.

Rating	#Reviews
1	351
2	923
3	1253
4	890
5	461
Total	3875

Table 8
Binary classification of the MC corpus.

Classes	#Reviews
Positive	1274
Negative	1351
Total	2625

tem failed to assign as they were given by the human expert. F1 is a measure that combines both precision and recall, calculating the proportion of true results (both true positives and true negatives) (Sebastiani, 2002). For ease of comparison, we summarize the F1 scores over the different categories (positive and negative) using the macro-averages of F1 scores:

$$\text{Macro-F1} = \text{Average of within-category F1 values}$$

In the same way, we can obtain the Macro-Recall and Macro-Precision as follows:

$$\text{Macro-Recall} = \text{Average of within-category Recall values}$$
$$\text{Macro-Precision} = \text{Average of within-category Precision values}$$

4.2. The MC corpus

In order to demonstrate the effectiveness of our approach we selected the MuchoCine corpus (MC), available for the SA research community in Spanish (Cruz et al., 2008). The corpus consists of 3878 movie reviews collected from the MuchoCine website. The reviews are written by web users instead of professional film critics. This increases the difficulty of the task because the sentences found in the documents may not always be grammatically correct, or they may include spelling mistakes or informal expressions. The corpus contains about 2 million words and an average of 546 words per review.

The opinions are rated on a scale from 1 to 5. One point means that the movie is very bad and 5 means very good. Films with a rating of 3 can be considered as "neutral", which means that the user considers the film is neither bad nor good. Table 7 shows the number of reviews per rating. This corpus has been widely used in different studies such as (del-Hoyo, Hupont, Lacueva, & Abadía, 2009), (Barreiro & Gonçalo, 2011), (Malvar-Fernández & Pichel-Campos, 2011), (Martínez-Cámara, Martín-Valdivia, & Ureña-López, 2011) and (Martín-Valdivia et al., 2013).

In our experiments we discarded the neutral examples. In this way, opinions rated with 3 were not considered, the opinions with ratings of 1 or 2 were considered as positive and those with ratings of 4 or 5 were considered as negative. Table 8 shows the class distribution of the binary classification of MC.

In MC corpus a movie review consists of four fields: the identifier of the review, the review rating, the summary and the description. Fig. 1 shows an excerpt of a review from MC.

id|rating|summary|body

1000|-1|*Silicona, esteroides, pactos demoniacos y otras basuras habituales son la base que sustentan esta aberración de vergüenza.*| *Una fiesta llena de excesos, rubias despampanantes, musculitos por doquier, algún que otro muerto. Nada nuevo. La alianza del mal es el nombre de este thriller sobrenatural que narra las peripecias de unos jóvenes...*

Fig. 1. Excerpt of a review from the MuchoCine corpus.

5. Experiments and results

Several experiments were carried out in order to verify the utility of the three lists of sentiment words generated for Spanish: SOL, iSOL and eSOL.

Before carrying out the experiments we performed a preprocessing step to the MC corpus in order to apply the same criteria followed during the generation of the lists. For example, for both summary and body we had to change capital letters to non-capital letters, accented letters to non-accented letters and special characters had to be deleted from the opinions. Moreover, the stop words and proper nouns were discarded. The named entity recognition was carried out using the Freeling tool[12].

In order to decide whether a review was positive or negative we followed a simple approach based on counting the number of words included in the lists of sentiment words for Spanish. Therefore, our system assesses the review as positive if the number of positive words is greater than or equal to the negative ones, and as negative in the opposite case. Using this approach, we carried out the binary classification of the MC corpus by using these three lists of sentiment words. Table 9 shows the comparison between the different results obtained.

As shown in Table 9, using the last versions of the lexicon (improved and enriched) we obtained the best results for all the measures employed. Employing the Macro-F1 as the evaluation measure and the approach applying SOL as base case, the system using the iSOL achieves an improvement of +10.29%, while using eSOL the improvement achieved is +12.95%. Therefore, we can conclude that the most completed list of sentiment words provided (eSOL) can be considered an interesting resource for use in sentiment analysis tasks related to Spanish language, particularly for unsupervised approaches.

5.1. Comparison with other related work

In the literature we can find an interesting resource called the Spanish Emotion Lexicon (SEL[13]) provided by Sidorov et al. (2012). This resource is freely available[11] for research purposes. SEL is composed of 2036 words that are associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion or category: *joy*, *anger*, *fear*, *sadness*, *surprise*, and *disgust*. It was marked manually by 19 annotators by using a scale with four values: null, low, medium and high.

In order to establish a feasible comparison by using the SEL resource for binary classification of MC, we considered the *joy* and *surprise* categories as positive and the others as negative. We carried out two different experiments: taking into account all the words provided by SEL and considering only those words whose PFA value was greater than or equal to 0.2. Table 10 shows the comparison between the results obtained by using the SEL resource and those obtained by using the eSOL resource provided in this paper.

Regarding the results in Table 10 the words included in SEL are

[12] http://nlp.lsi.upc.edu/freeling.
[13] http://www.cic.ipn.mx/~sidorov/#SEL.

Table 9
Results obtained for the binary classification of the MC corpus by using the three lists of sentiment words generated.

	Macro-precision	Macro-recall	Macro-F1	Accuracy
SOL	0.5615	0.5600	0.5607	0.5623
iSOL	0.6222	0.6147	0.6184	0.6183
eSOL	**0.6393**	**0.6274**	**0.6333**	**0.6316**

Table 10
Comparison for binary classification of MC by using the SEL and eSOL resources.

	Macro-precision	Macro-recall	Macro-F1	Accuracy
SEL (all words)	0.5240	0.5162	0.5200	0.5249
SEL (PFA > 0,2)	0.5256	0.5181	0.5218	0.5264
eSOL	**0.6393**	**0.6274**	**0.6333**	**0.6316**

Table 11
Some samples of MC corpus classified with eSOL and SEL resources.

Id	Original rating	#positive words in eSOL	#negative words in eSOL	eSOL rating	#positive words in SEL	#negative words in SEL	SEL rating
1002	−1	8	9	−1	4	0	1
1008	1	23	13	1	5	6	−1
1011	−1	8	13	−1	7	5	−1
1023	1	13	11	1	4	5	−1
1036	1	49	26	1	1	16	−1
1045	1	20	13	1	4	10	−1

Table 12
Negative and positive words classified with eSOL and SEL.

Id	Negative words found with eSOL	Negative words found with SEL	Positive words found with eSOL	Positive words found with SEL
1008	Falta Prision Impotencia Pequeña Ansiedad Intenso Nerviosa Dolor Muerte Soledad Irregular Sobria Falta	Impotencia Dolor Muerte Soledad Decaer Solo	Maestra Mayor Ventaja Prometedor Emociones Intimas Unica Profunda Mejor Estabilidad Amor Emotiva Afecto Amor Impecable Sabio Humano Maestra Mayor Ventaja Prometedor Emociones Intimas	Amor Afecto Amor Expresion Suave

not as discriminative as eSOL. Due to the low performance of SEL, we revised SEL and noticed that words with a PFA value closed to zero should not be taken into account, because those terms could introduce noise. Thus we only considered terms with a PFA value over 0.2. This subset of SEL achieved slightly better results, +0.3456%, +0.2854% regarding Macro-F1 and Accuracy respectively. However, the results achieved by SEL over 0.2 are unsatisfactory. The difference between SEL over 0.2 is noteworthy, +19.3057% and +18.1693% considering Macro-F1 and Accuracy, respectively.

Table 13
Polarity classification results over MC corpus.

	Approach	Precision	Recall	F1	Accuracy
Cruz et al. (2008)	Unsupervised	N/A	N/A	N/A	0.6950
	Supervised	N/A	N/A	N/A	0.7750
del-Hoyo et bal. (2009)	Hybrid	N/A	N/A	N/A	0.8086
Malvar-Fernández & Pichel-Campos (2011)	Supervised	0.77	0.77	N/A	N/A
Martínez-Cámara et al. (2011)	Supervised	0.8684	0.8667	0.8675	0.8674
Martín-Valdivia et al. (2013)	Hybrid	0.8858	0.8857	0.88575	0.8857
eSOL	Unsupervised	**0.6393**	**0.6274**	**0.6333**	**0.6316**

Table 11 shows some samples of MC corpus classified with our resource eSOL and the resource SEL with PFA > 0.2 for the reason that we have explained before. The results show that the resource eSOL is more suitable for polarity classification of Spanish texts than the SEL resource.

In next table we show the negative and positive words that eSOL and SEL found in the review with identifier 1008.

As we can see in Table 12, negative words such as *ansiedad* (*anxiety*), *nerviosa* (*nervous*), *irregular* (*irregular*), and positive words such as *sabio* (*wise*), *prometedor* (*promising*), *maestra* (*masterpiece*) are not included in the resource SEL. This is the main reason why this review was classified correctly by using our resource eSOL and incorrectly by using SEL.

On the other hand, it is noteworthy that the MC corpus has been widely used by the SA Spanish research community. Some authors assessed different methods over MC corpus, so in Table 13 we show a comparison of our proposed method with other works.

As usual in data mining the approaches based on supervised methods achieve better results than those based on unsupervised methods. The authors of the MC corpus followed the unsupervised method proposed by Turney (2002), which takes advantage of the search engine AltaVista[14]. That study is the only which describes an unsupervised approach over the MC Corpus. As the authors indicate in their paper, they did not use the whole corpus (3878 reviews), neither the 2625 reviews resultant of getting rid of the opinions tagged with a polarity value of 3, as we do in our experiments. They only used 400 reviews (200 positive and 200 negative) for the experimentation which had been randomly selected from the subset of 2625 reviews, i.e. the reviews labeled with a value of 1–2 (negative) or a value of 4–5 (positive). With this subset of 400 reviews the authors achieved 0.6950 of Accuracy. On the other hand, we used a subset of the original MC corpus to assess the lexicons which concerns 2625 reviews. With a larger set of data our unsupervised method achieved 0.6316 of Accuracy, only 0.0634 lower than the method proposed by the authors of the corpus. Taking into account the simplicity of our lexicon-based method, the results achieved by eSOL can be considered as good. Also, we highlighted the fact that, as far as we know, this is the first work that has used all the positive and negative reviews of the MC corpus for a polarity classification experimentation following an unsupervised method.

[14] http://www.altavista.com/

6. Conclusions and further work

In this paper we have presented an experimental study of polarity classification over a corpus of film reviews written in Spanish, the MuchoCine corpus (MC). Firstly, we translated the lexicon of BLEL in order to generate sentiment word lexicons for Spanish. Several improvements were carried out in order to build an improved sentiment words list, and finally, an enriched lexicon for Spanish.

The results show the validity of the two lexicons presented in this paper, iSOL and eSOL, for polarity classification of Spanish reviews. In addition, the results show that a lexicon-based method is suitable for solving the task of polarity classification of Spanish texts. The experiments carried out in this paper encourage us to continue working along this line. A lexicon such as iSOL and eSOL can be used as the sole semantic resource or can be used as another element within the workflow of a polarity classification system. Therefore, we consider that the lexicons developed, which are freely available, are valuable resources for the Spanish SA research community.

Currently we are working on the development of several Spanish lexicons for domain-dependent SA following the method proposed here, i.e. selecting the words with a higher frequency in a corpus. In addition, we are interested in another novelty method combining a random walk algorithm for building domain-oriented sentiment lexicons (Tan & Wu, 2011). Finally, we are studying the treatment of negation in SA, which we think is essential for the resolution of the polarity classification task.

Acknowledgements

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), TEXT-COOL 2.0 project (TIN2009-13391-C04-02) and ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. Also, this paper is partially funded by the European Commission under the Seventh (FP7 – 2007-2013) Framework Programme for Research and Technological Development through the FIRST project (FP7-287607). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP '08)* (pp. 127–135). Stroudsburg, PA, USA: Association for Computational Linguistics.

Barreiro, A., & Gonçalo, O. (2011). Cross-language semantic relations between English and Portuguese. In *Proceedings of the workshop on Iberian cross-language natural language processing tasks (ICL 2011)* (pp. 49–58).

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the RANLP Conference* (pp. 50–54).

Cruz, F., Troyano, J. A., Enríquez, F., & Ortega, J. (2008). Experiments in sentiment classification of movie reviews in Spanish. *Procesamiento de Lenguaje Natural, 41*, 73–80.

del-Hoyo, R., Hupont, I., Lacueva, F. J. & Abadía, D. (2009). Hybrid text affect sensing system for emotional language analysis. In *Proceedings of the international workshop on affective-aware virtual agents and social robots*. Boston, Massachusetts.

Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008* (pp. 507–512). IEEE 24th International Conference on IEEE.

Ding, X., & Liu, B. (2007). The utility of linguistic rules in opinion mining. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 811–812). ACM.

Du, W., Tan, S., Cheng, X., Yun, X. (2010). Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. In *Proceedings of ACM international conference on web search and data mining (WSDM-2010)* (pp. 111–120). ACM, New York.

Esuli, A. & Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th conference on language resources and evaluation (LREC'06)* (pp. 417–422).

Hatzivassiloglou, V. & McKeown, K. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the association for computational linguistics* (pp. 174–181).

Hatzivassiloglou, V. & Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the international conference on computational linguistics (COLING)* (pp. 299–305).

Hu, M. Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, KDD'04* (pp. 168–177). New York, NY, USA. ACM.

Kamps, J., Marx, M., Mokken, R. J. & de Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. In *LREC, European Language Resources Association*.

Kanayam, H., Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of conference on empirical methods in natural language processing (EMNLP-2006)* (pp. 355–363).

Kim, S. M., & Hovy, E. (2006). Identifying and analyzing judgment opinions. In *Proceedings of the joint human language technology/North American chapter of the ACL conference (HLT-NAACL)* (pp. 200–207). Stroudsburg, PA, USA: Association for Computational Linguistics.

Liu, B. (2012). *Sentiment analysis and opinion mining. synthesis lectures on human language technologies*. Morgan & Claypool Publishers.

Malvar-Fernández, P., & Pichel-Campos, J. R. (2011). Generación semiautomática de recursos de Opinion Mining para el gallego a partir del portugués y el español. In *Proceedings of the workshop on Iberian cross-language natural language processing tasks (ICL 2011)* (pp. 59–63).

Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2011). Opinion classification techniques applied to a Spanish corpus. In *Proceedings of the 16th international conference on Natural language processing and information systems, NLDB'11* (pp. 169–176). Springer-Verlag.

Martín-Valdivia, M. T., Martínez-Cámara, E., Perea-Ortega, J. M., & Ureña-López, L. A. (2013). Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications, 40*(10), 3934–3942. http://dx.doi.org/10.1016/j.eswa.2012.12.084.

Mihalcea, R., & Strapparava, C. (2006). Learning to laugh (automatically): Computational models for humor recognition. *Journal of Computational Intelligence, 22*, 126–142.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). *WordNet: An on-line lexical database*. Oxford University Press.

Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing 10*, (pp. 79–86).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics, 3*(2), 143–157.

Rushdi Saleh, M., Martín-Valdivia, M. T., Montejo-Ráez, A., & Ureña-López, L. A. (2011c). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications, 38*(12), 14799–14804.

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A. & Perea-Ortega, J. M. (2011b). Bilingual Experiments with an Arabic-English Corpus for Opinion Mining. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)* (pp. 740–745).

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Perea-Ortega, J. M. (2011a). OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology, 62*(10), 2045–2054.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1–47.

Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2012). Empirical study of opinion mining in Spanish tweets. *LNAI, 7629–7630*, 14.

Somprasertsri, G., & Lalitrojwong, P. (2010). Mining feature-opinion in online customer reviews for opinion summarization. *Journal of Universal Computer Science, 16*, 938–955.

Strapparava, C. & Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on applied computing (SAC '08)* (pp. 1556–1560). ACM.

Strapparava, C. & Valitutti, A. (2004). WordNet-Affect: An affective extension of WordNet. In *Proceedings of the 4th international conference on language resources and evaluation* (pp. 1083–1086).

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. D., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267–307.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Systems with Applications, 38*(10), 12094–12100.

Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. *Expert Systems with Applications, 34*(4), 2622–2629.

Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery, 24*(3), 478–514. http://dx.doi.org/10.1007/s10618-011-0238-6.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424).

Wan, X. (2009). Co-training for cross-lingual sentiment classification. In Keh-Yih Su, Jian Su, & Janyce Wiebe (Eds.), *'ACL/AFNLP', The Association for Computer, Linguistics* (pp. 235–243).

Wiebe, J. (2000). Learning subjective adjectives from corpora. In *Proceedings of AAAI*.

Wiebe, J., Wilson, T. & Bell, M. (2001). Identifying collocations for recognizing opinions. In *Proceedings of the ACL'01 association for computational linguistics workshop on collocation: computational extraction, analysis, and exploitation* (pp. 24–31).

Zhang, C., Zeng, D., Li, J., Wang, F. Y., & Zuo, W. (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.

2.2 Propuesta para generar lexicones adaptados al dominio

Las publicaciones en revista asociadas a esta parte son:

c) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014).A Spanish Semantic Orientation Approach to Domain Adaptation for Polarity classification. *Information Processing and Management*

Estado: Aceptado y pendiente de publicación
Índice de Impacto (JCR 2013): 1,069
Área de conocimiento: Computer Science, Information System, Ranking 68/135
Cuartil: Q3
ISSN: 0306-4573
DOI: 10.1016/j.ipm.2014.10.002

d) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Jiménez Zafra, S. (2014). eSOLHotel: Generación de un lexicón de opinión en español adaptado al dominio turístico. Procesamiento del Lenguaje Natural nº 54.

Estado: en Revisión
Incluida en RECyT (Repositorio Español de Ciencia y Tecnología).
Sello de calidad (ISO9001) que la acredita como excelente durante un periodo de tres años (14 de marzo de 2012 al 14 de marzo de 2015).
ISSN: 1135-5948

e) Molina-González, M. D., Martín-Valdivia, M. T Martínez-Cámara, E., & Ortega J. F., Cruz F. (2014). Automatic generation of subjective lexicons adapted to a specific domain using linguistic resources. *Knowledge-Based Systems.*

Estado: en Revisión
Índice de Impacto (JCR 2013): 3,058
Área de conocimiento: Computer Science, Artificial Intelligence, Ranking 15/121
Cuartil: Q1
ISSN: 0950-7051

A Spanish Semantic Orientation Approach to Domain Adaptation for Polarity Classification

M. Dolores Molina-González[a], Eugenio Martínez-Cámara[a,*], M. Teresa Martín-Valdivia[a], L. Alfonso Ureña-López[a]

[a]*SINAI Research Group,*
University of Jaén
Campus Las Lagunillas E-23071, Jaén (Spain)

Abstract

One of the problems of Opinion Mining is the domain adaptation of the sentiment classifiers. There are several approaches to tackling this problem. One of these is the integration of a list of opinion bearing words for the specific domain. This paper presents the generation of several resources for domain adaptation to polarity detection. On the other hand, the lack of resources in languages different from English has orientated our work towards developing sentiment lexicons for polarity classifiers in Spanish. The results show the validity of the new sentiment lexicons, which can be used as part of a polarity classifier.

Keywords: Spanish Opinion Mining, Sentiment lexicon, domain adaptation

1. Introduction

Opinion Mining (OM) is defined as the computational treatment of opinion, sentiment, and subjectivity in text. This new area of research is becoming more and more important mainly due to the growth of social media where users continually post contents on the web in the form of comments, opinions, emotions, etc. The OM discipline combines Natural Language Processing (NLP) with data mining techniques and includes a large number of tasks (Pang & Lee, 2008). One of the most widely studied tasks is the polarity classification of reviews. This task

*Corresponding author
Email addresses: mdmolina@ujaen.es (M. Dolores Molina-González), emcamara@ujaen.es (Eugenio Martínez-Cámara), maite@ujaen.es (M. Teresa Martín-Valdivia), laurena@ujaen.es (L. Alfonso Ureña-López)

focuses on determining the overall sentiment-orientation (positive or negative) of the opinions contained within a given document.

Although different approaches have been applied to polarity classification, the mainstream basically consists of two major methodologies. On the one hand, the Machine Learning (ML) approach (also known as the supervised approach) is based on using a collection of data to train the classifiers (Pang et al., 2002). On the other hand, the approach based on Semantic Orientation (SO) does not need prior training, but takes into account the positive or negative orientation of words (Turney, 2002). This method, also known as the unsupervised approach, makes use of lexical resources like lists of opinion words, lexicons, dictionaries, etc. Both methodologies have their advantages and drawbacks. For example, the ML approach depends on the availability of labelled data sets (training data), which in many cases are impossible or difficult to achieve. On the contrary, the SO strategy requires a large amount of linguistic resources which generally depend on the language, and often this approach obtains lower recall because it depends on the presence of the words comprising the lexicon in the document in order to determine the orientation of opinion. In this paper we focus on the generation of linguistic resources to tackle the problem of polarity classification using an unsupervised approach.

While opinions and comments on the Internet are expressed in any language, most research in OM is focused on English texts. However, languages such as Chinese, Spanish or Arabic, are even more present on the web. Thus it is important to develop resources to help researchers to work with these languages. The work presented herein is mainly motivated by the need to develop polarity classification systems and resources in languages other than English. Specifically, in this paper we deal with Spanish reviews. We present an experimental study over the SFU Review Corpus[1] (Brooke et al., 2009), which is a comparable corpus that includes opinions of several topics in English and in Spanish in different domains.

One of the open problems in OM is that of domain adaptation. Although movie reviews have been the most studied domain in sentiment analysis, a wide range of areas are being investigated such as political debates, hotels or music. However, when we train a classifier using a specific domain we need to adapt it in order to apply it to another domain. For example, the sentence “Definitively, you should read the book” most likely refers to positive polarity for Book reviews but negative sentiment for Movie reviews.

[1] http://www.sfu.ca/œmtaboada/research/SFU_Review_Corpus.html

Thus the problem of domain adaptation is attracting more and more attention in OM. In this paper we carry out an experimental study of domain adaptation of linguistic resources for Spanish reviews in different domains. We have used the Spanish version of SFU, which includes 400 reviews for 8 different domains. We have generated several lists of opinionated words integrating knowledge from the different domains and we have compared the results obtained. A corpus-based approach is followed with the aim of adapting a general-purpose sentiment lexicon to a specific domain by integrating lists of opinion bearing words. iSOL[2] (Molina-González et al., 2013) is the general-purpose sentiment lexicon chosen. The Spanish version of the SFU corpus was the corpus selected for the adaptation process due mainly to the fact that it covers 8 different domains. Following different heuristics, which will be described later, the most frequent opinion bearing words are appended to iSOL. Several experiments were carried out with the goal of assessing the new domain-specific sentiment lexicons. The analysis of the results shows the validity of the new lists.

The paper is organized as follows: Section 2 briefly describes other papers that study non-English sentiment polarity classification and, specifically work related to Spanish OM. In addition, we include some papers studying the domain adaptation problem. In Section 3 we explain the different resources used. Sections 4 and 5 present the experiments carried out and discusses the main results obtained. Finally, we outline conclusions and further work.

2. Related work

In this study we focus on two open problems in opinion mining: Non-English polarity classification and the domain adaptation problem. Next, we will comment on some papers that have inspired our work.

2.1. Non-English polarity classification

There are some interesting papers that have studied the problem using non-English collections. For example, Denecke (2008) worked on German comments collected from Amazon. These reviews were translated into English using standard machine translation software. Then the translated reviews were classified as positive or negative, using three different classifiers: LingPipe3, SentiWordNet (Esuli & Sebastiani, 2006) with classification rule, and SentiWordNet with

[2]The iSOL resource is freely available for research purpose at `http://sinai.ujaen.es/?p=1202`

machine learning. In (Zhang et al., 2009) Chinese sentiment analysis is applied on two datasets. In the first one euthanasia reviews were collected from different web sites, while the second dataset was about six product categories collected from Amazon (Chinese reviews). Ghorbel & Jacot (2011) used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determinate the polarity of the reviews. In (Agić et al., 2010) a manually annotated corpus is presented with news on the financial market in Croatia. In (Rushdi-Saleh et al., 2011) a corpus of movies reviews in Arabic annotated with polarity was presented and several experiments using machine learning techniques were performed.

Regarding Spanish, there are also some interesting studies. For example, Banea et al. (2008) proposed several approaches to cross lingual subjectivity analysis by directly applying the translations of opinion corpus in English to training an opinion classifier in Romanian and Spanish. This study showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. In (Brooke et al., 2009) several experiments dealing with Spanish and English resources are presented. They conclude that although the ML techniques can provide a good baseline performance, it is necessary to integrate language-specific knowledge and resources in order to achieve an improvement. They proposed three approaches: the first one uses Spanish resources generated manually and automatically. The second one applies ML to a Spanish corpus. The last one translates the Spanish corpus into English and then applies the SO-CAL (Semantic Orientation CALculator), a tool developed by themselves (Taboada et al., 2011). Cruz et al. (2008) manually recollected the MuchoCine (MC) corpus in order to develop a sentiment polarity classifier based on semantic orientation. The corpus contains annotated Spanish movie reviews from the MuchoCine website[3]. The MC corpus was also used in (Martínez-Cámara et al., 2011) to carry out several experiments with a supervised approach applying different ML algorithms (SVM, NB,BBR, KNN, C4.5). The results are much better than those obtained with the unsupervised approach proposed by Cruz et al. (2008).

One of the barriers in the study of how to resolve the problem of polarity classification on Spanish reviews is the lack of Spanish linguistic resources for opinion mining. However, some new sentiment linguistic resources, mainly lists of opinion bearing words, have been made available in the last years. Sidorov

[3]http://www.muchocine.net/

et al. (2013) provided the Spanish Emotion Lexicon (SEL). SEL is composed of 2,036 words that are associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion or category: joy, anger, fear, sadness, surprise, and disgust. Molina-González et al. (2013) describe a new Spanish sentiment lexicon. The authors translated the Bing Liu English Opinion Lexicon (Hu & Liu, 2004) into Spanish. Subsequently, the translated version was manually corrected and improved with Spanish opinion bearing words. The result is the lexicon iSOL, which is composed of 8,135 words. iSOL has been also used in (Martínez-Cámara et al., 2013) with promising results.

2.2. Domain adaptation for sentiment analysis

Different methods have been proposed for tackling the domain adaptation problem. One of the primary studies in sentiment analysis is (Blitzer et al., 2007). They use Structural Correspondence Learning (SCL) to find correspondences between features from source and target domains through modelling their correlations with pivot features. The proposed approach was successfully tested on review data from 4 domains (DVDs, books, kitchen appliances and electronics). Following the same idea, Pan et al. (2010) present the Spectral Feature Alignment (SFA) that uses spectral clustering to align domain-specific and domain- independent words into a set of feature-clusters. The results obtained surpass the SCL. Jiang & Zhai (2007) describe two distinct needs. On the one hand, instance adaptation takes into account the change of instance probability, e.g., the change of vocabulary or the change of words' frequency from one domain to another; On the other hand, labelling adaptation models the changes of the labelling function, since one feature that is positive in the source domain may express the opposite meaning in the target domain. Most studies tackle the instance adaptation problem, while Xia et al. (2013) propose a combination taking into account both kinds of adaptation, obtaining good results. In Ponomareva & Thelwall (2012) graph-based approaches are applied. They model the data as a graph of documents, taking into account not only the document content but also document connectivity, which is modelled as document sentiment similarity rather than content similarity.

3. Domain adaptation method

Like some of the studies mentioned in the previous section, herein we propose a domain adaptation method for sentiment analysis. However, we focus our study on reviews written in Spanish. In addition, in contrast to the aforementioned methods which mainly focus on machine learning algorithms, we propose a lexicon-based approach to the domain adaptation problem. We follow a very simple strategy by generating lists of opinionated words for each domain in an automatic way. The Spanish version of Taboada corpus SFU is used in our experiments. Firstly we apply a general lexicon to the corpus, taking into account the different domains. Then, four different opinionated word lists are generated for each of the eight different domains and four different word lists for all domains of the corpus. Following a corpus based method, two heuristics are assessed with the aim of integrating into each list the most frequent words used for positive and negative reviews. A subset of the corpus is used to build the lists and the other part to test the new resources. The results obtained show an improvement over the experiments using the general lexicon.

3.1. Corpus

In order to carry out the experiment we chose the Spanish part of the comparable SFU Review Corpus. The SFU Review Corpus is composed of reviews of products in English and Spanish. The English version (Taboada & Grieve, 2004) has 400 reviews (200 positive and 200 negative) of commercial products downloaded in 2004 from the Epinions[4] web which are divided into eight categories. Each category includes 25 positive reviews and 25 negative reviews. Subsequently, the authors of the SFU Review Corpus have made available the Spanish version of the corpus[5] with the aim of offering a comparable corpus for the research community. The Spanish reviews are divided into eight similar categories: books, cars, computers, washing machines, hotels, movies, music and phones. Each category also has 25 positive and 25 negative reviews, which are defined as positive or negative based on the number of stars given by the reviewer (1-2=negative; 4-5=positive; 3-star reviews are not included). In this case, the reviews are downloaded from the Ciao.es[6] website.

3.2. Opinion lists generation

We followed a lexicon-based approach to tackle the problem. The iSOL lexicon (Molina-González et al., 2013) was selected to carry out the experiments. This resource was generated from the BLEL lexicon (Hu & Liu, 2004) by automatically translating it into Spanish and obtaining the SOL (Spanish Opinion Lexicon) resource. Then, this resource was manually reviewed in order to improve the final

[4]http://www.epinions.com
[5]http://www.sfu.ca/~mtaboada/download/downloadCorpusSpa.html
[6]http://www.ciao.es/

list of words obtaining iSOL (improved SOL). This resource has been successfully evaluated in (Molina-González et al., 2013) using a Spanish corpus of movie reviews called MuchoCine (Cruz et al., 2008). The results showed that the use of an improved list of sentiment words from the same language could be considered as a good strategy for unsupervised polarity classification. Moreover, another list was generated appending the positive and negative words of the MuchoCine corpus. In this way, domain knowledge was added in the lexicon. The result of the process was a new lexicon which is called eSOL (enriched SOL). The experiments with eSOL showed the advantages of using domain knowledge. Thus, the main motivation of this paper is the integration of knowledge from the domain in order to improve the final polarity classification system.

The improved Spanish Opinion Lexicon (iSOL) is composed of 2,509 positive and 5,626 negative words, thus in total the Spanish lexicon has 8,135 opinion words.

As is well-known in the SA research community, the semantic orientation of a word is domain-dependent. Within the approaches followed by research into the compilation of a set of polar words, the most suitable for obtaining domain-dependent opinion words is that known as the corpus-based approach. Hatzivassiloglou & McKeown (1997) take some adjectives as seeds in order to find additional sentiment adjectives in the corpus. Their method takes advantage of a set of conventions on connectives with the aim of identifying more polar words and their orientation from a sentiment label corpus. On the other hand, Du et al. (2010) follow a similar assumption, and they consider that a word should be positive (or negative) if it appears in many positive (or negative) documents.

We follow a more straightforward method which consists of enlarging iSOL with the most frequent words of a sample of the SFU Spanish Review Corpus. The key point of the method is to automatically find domain sentiment words in the different domains of the corpus with the goal of developing a domain specific sentiment lexicon for each domain covered by the SFU Spanish Review Corpus. Four different word lists for each domain of the corpus and four different word lists for the categories (positive and negative) of the corpus are generated. Then, the new resources are assessed over the reviews of the corpus which are not utilised for building the lists. To build the first four lists, we split the 50 reviews for each category into two random groups of 15 and 10 positive reviews and 15 and 10 negative reviews. We used the group of 15 reviews of both polarities (30 reviews in total) to seek the words and integrate them into the new resources. Then we used the group of 20 reviews (10 positive, 10 negative) to test the validity of the new domain specific lexicons.

Taking the general-purpose sentiment lexicon iSOL, we generated our first list of opinion words for each domain of the SFU Spanish Review Corpus. After removing the stop words from the documents, the selection of the polarity domain-dependent words consists of calculating the absolute frequency of each word per class (positive/negative), and then the most-used positive and negative words were appended to iSOL. The new list with domain information is called eSOLdomainLocal (enriched SOL Local) where domain = cars, hotels, washing machine, books, phones, music, computers, movies.

The second way to enrich the iSOL lexicon consists of adding not only the sentiment words but also the most frequently used domain-dependent words. Commercial names and proper nouns were discarded from this selection. Some examples of these discarded words were: Fagor, BMW, Almudena Grandes, Quijote, Siemens, Citroen, Acer, Nokia, Bon Jovi, AC/DC, Bosh, Hannibal Lecter, George Lucas, etc. The most frequent domain-dependent words are selected using the following formula:

$$\text{list(word)} = \begin{cases} positive & if(f^- = 0 \wedge f^+ \geq 3) \vee \left(\dfrac{f^+}{f^-} \geq 3\right) \\ negative & if(f^+ = 0 \wedge f^- \geq 3) \vee \left(\dfrac{f^-}{f^+} \geq 3\right) \end{cases} \tag{1}$$

where f^+ is the frequency of the word in positive reviews and f^- in negative reviews. Thus, those words that satisfy Equation 1 are appended to the positive or negative list of eSOLdomainLocal. These new resources are called eSOLdomainLocal*. Tables 1 and 2 show some examples of domain-dependent words that have been appended to the lists.

The third and fourth lists generated are similar to the first and second ones. The difference is how to find the most used sentiment and domain-dependent words. In these lists, if one word is used one or more times in one positive or negative review we considered that its frequency is one. That is, although the word appears several times in a specific review, its frequency is one. Therefore in these lists, the highest possible frequency of a word is 15. The new resources are called eSOLdomainGlobal with only sentiment words and eSOLdomainGlobal* including the most frequently used sentiment words and domain-dependent words.

In order to generate the last four lists, we considered all the domains together. Again, we split the corpus into two groups, one for integrating opinion words into the lists and another one for testing the new resources. Thus, we used 120 positive reviews (the same 15 positive reviews per domain used before multiplied by 8 domains) and 120 negative reviews to generate new resources from eSOL, and

Word	Domain	Freq. in Positive Reviews	Freq. in Negative Reviews
Consumo (consumption)	Cars	10	1
Maletero (boot)	Cars	6	0
Menú (menu)	Hotels	4	0
Minibar (minibar)	Hotels	5	0
Temperatura (temperature)	Washing machine	12	1
Capacidad (capacity)	Washing machine	7	2
Recuerdos (memories)	Books	10	1
Introducción (introduction)	Books	5	1
Conectividad (connectivity)	Phones	6	1
Navegación (navigation)	Phones	6	0
Ritmos (rhythms)	Music	8	1
Sonidos (sounds)	Music	8	1
Rendimiento (performance)	Computer	13	1
Plataforma (platform)	Computer	11	0
Escena (scene)	Movies	19	2
Estreno (premiere)	Movies	5	0

Table 1: Some positive words included in eSOLdoaminLocal*

we carried out the experiment with the rest of the corpus, that is 160 reviews, 80 positive (10 positive reviews for each 8 different domains) and 80 negative.

On the one hand, we generated the new eSOLLocal resource taking into account only the most frequent sentiment words. Then, we generated the new eSOLLocal* taking into account the sentiment words and also the most frequent domain words, which were obtained following the Equation 1.

On the other hand, in the compilation of the latter two lists the difference is how to find the most used sentiment and domain-dependent words. If one word is used one or more times in one positive or negative review we have considered that its frequency is one. Therefore, in these lists the highest frequency is 120, and this only happens if the word is in all the reviews. The new resource, with only sentiment words added to iSOL, is called eSOLGlobal, and the resource with not only the sentiment words but also including the most frequent domain words is called eSOLGlobal*.

Regarding the original lexicon iSOL, we increased the size of the generated eSOLdomainLocal and eSOLdomainLocal* lists for both negative and positive lists of words. Tables 3 and 4 show the number of words added to iSOL in each resource respectively.

Word	Domain	Freq. in Positive Reviews	Freq. in Negative Reviews
Taller (workshop)	Cars	2	19
Sensor (sensor)	Cars	0	5
Manchas (spots)	Hotels	0	4
Moqueta (fitted carpet)	Hotels	1	6
Acero (steel)	Washing machine	0	3
Cocina (kitchen)	Washing machine	1	8
Serie (series)	Books	2	9
Ritmo (rhythm)	Books	0	5
Covertura (coverage)	Phones	0	8
Carga (charge)	Phones	0	5
Remix (remix)	Music	1	6
Versiones (versions)	Music	0	4
Pantalla (screen)	Computer	0	4
Computadora (computer)	Computer	0	8
Trailer (trailer)	Movies	1	6
Saga (saga)	Movies	0	9

Table 2: Some negative words included in eSOLdoaminLocal*

Concerning the eSOLdomainGlobal and eSOLdomainGlobal*, we also increased the size of the original iSOL lexicon. Tables 5 and 6 show the number of words added to iSOL and also the final size of each new list.

Regarding the eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal*, the size also increased compared to the original iSOL lexicon, and the number of positive and negative words integrated in the new lists is shown in Table 7.

4. Experiments and results

In order to evaluate the different approaches, we used the traditional measures employed in text classification: precision (P), recall (R), F1 and Accuracy:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

eSOLdomainLocal	#positive words	#negative words
Cars	18 (2527)	23 (5649)
Hotels	9 (2518)	10 (5636)
Washing machines	11 (2520)	13 (5639)
Books	19 (2528)	26 (5652)
Cell phones	20 (2529)	33 (5659)
Music	27 (2536)	19 (5645)
Computers	17 (2526)	19 (5645)
Movies	32 (2541)	22 (5648)

Table 3: Number of words included in the new eSOLdomainLOCAL lexicon and final size of the lists

eSOLdomainLocal*	#positive words	#negative words
Cars	28 (2537)	36 (5662)
Hotels	24 (2533)	15 (5641)
Washing machines	18 (2527)	22 (5648)
Books	29 (2538)	36 (5662)
Cell phones	42 (2551)	36 (5662)
Music	43 (2552)	26 (5652)
Computers	51 (2560)	25 (5651)
Movies	58 (2567)	29 (5655)

Table 4: Number of words included in the new eSOLdomainLocal* lexicon and final size of the lists

$$F1 = \frac{2PR}{P+R} \tag{4}$$

$$Acc. = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

where TP (True Positives) and TN (True Negatives) are those assessments where the system and a human expert agree on a label (in this case, TP and TN are those positive or negative reviews rightly classified), FP (False Positives) and FN (False Negatives) are those labels assigned by the system that do not agree with the expert assignment, in plain English, the positive and negatives reviews misclassified. F1 is a measure that combines both precision and recall, calculating the proportion of true results (both true positives and true negatives) Sebastiani (2002). Due to the fact that the system classifies two classes, the P, R and F1 of

eSOLdomainGlobal	#positive words	#negative words
Cars	19 (2528)	22 (5648)
Hotels	9 (2518)	10 (5636)
Washing machines	11 (2520)	13 (5639)
Books	20 (2529)	25 (5651)
Cell phones	20 (2529)	31 (5657)
Music	29 (2538)	19 (5645)
Computers	18 (2527)	19 (5645)
Movies	26 (2535)	22 (5648)

Table 5: Number of words included in the new eSOL*domain*Global lexicon and final size of the lists

eSOLdomainGlobal*	#positive words	#negative words
Cars	28 (2537)	34 (5660)
Hotels	21 (2530)	15 (5641)
Washing machines	18 (2527)	17 (5643)
Books	27 (2536)	29 (5655)
Cell phones	35 (2544)	33 (5659)
Music	37 (2548)	21 (5647)
Computers	30 (2539)	21 (5647)
Movies	39 (2548)	25 (5651)

Table 6: Number of words included in the new eSOL*domain*Global* lexicon and final size of the lists

each class have been calculated. Then, the overall P, R and F1 of the system have been obtained following the macro-averaged evaluation measures. The macro-avergaed evaluation measures formulae for P, R and F1 are the following:

$$\text{Macro-F1} = \frac{2 * \text{Macro-Precision} * \text{Macro-Recall}}{\text{Macro-Precision} + \text{Macro-Recall}} \tag{6}$$

Where Macro-Recall and Macro-Precision are obtained as follows:

$$\text{Macro-Recall} = \frac{\sum_{i=1}^{c} R_i}{c} \tag{7}$$

$$\text{Macro-Precision} = \frac{\sum_{i=1}^{c} P_i}{c} \tag{8}$$

Where c is the number of classes (c=2).

All domains	#positive words	#negative words
eSOLLocal	113 (2622)	399 (6025)
eSOLLocal*	118 (2627)	150 (5776)
eSOLGlobal	113 (2622)	278 (5904)
eSOLGlobal*	118 (2627)	141 (5767)

Table 7: Number of words included in the new resources considering all the domains and final size of the lists

Several experiments were carried out in order to verify the utility of the new resources generated for Spanish from iSOL: eSOLdomainLocal(*), eSOLdomainGlobal(*), eSOLLocal(*) and eSOLGlobal(*) where domain = cars, hotels, washine machine, books, phones, music, computers, movies. The general method consists of finding the presence in the reviews of opinion words which are included in a lexicon of opinion words. If a review has more positive words than negative ones, the document polarity is positive, otherwise negative (Equation 9).

$$p(r) = \begin{cases} 1 & \text{if } |positive| > |negative| \\ -1 & \text{if } |positive| \leq |negative| \end{cases} \tag{9}$$

where p(r) is the polarity of the review, |positive| is the number of positive words and |negative| is the number of negative words.

Before carrying out the experiments we performed a pre-processing step on the SFC corpus in order to apply the same criteria followed in the generation of the enriched iSOL lists. For example, we changed capital letters to non-capital ones, accented letters to non-accented ones and special characters were separated from words in the reviews. Moreover, the stop words were discarded.

For the baseline experiment we took the same 20 reviews used to test each domain separately and applied the iSOL lexicon. The results are presented in Table 8.

The next experiments were carried out over the 160 reviews used for testing purposes (10 positive reviews and 10 negative reviews per domain chosen randomly and not used to generate the lexicons). Thus the results obtained with the eSOLdomainLocal are shown in Table 9. This resource was built by adding the most used sentiment words in positive or negative reviews of each domain to iSOL. Table 9 also includes the percentage of improvement over the baseline experiment (Table 8) using the following equation:

	Macro-P	Macro-R	Macro-F1	Accuracy
Cars	0.8125	0.7000	0.7521	0.7000
Hotels	0.8571	0.8000	0.8276	0.8000
Washing machines	0.5667	0.5500	0.5582	0.5500
Books	0.7083	0.7000	0.7041	0.7000
Cell phones	0.7778	0.6000	0.6774	0.6000
Music	0.4333	0.4500	0.4415	0.4500
Computers	0.5667	0.5500	0.5582	0.5500
Movies	0.5549	0.5500	0.5525	0.5500

Table 8: Polarity classification over the SFU corpus using iSOL

$$\text{Improvement} = \frac{\text{Macro-F1eSOLdomain[Local}|\text{Global]} - \text{MacroF1iSOLdomain}}{\text{MacroF1iSOLdomain}} * 100 \tag{10}$$

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement
Cars	0.8571	0.8000	0.8276	0.8000	10.04%
Hotels	0.8125	0.8000	0.8062	0.8000	-2.58%
Washing machines	0.8000	0.8000	0.8000	0.8000	43.31%
Books	0.7083	0.7000	0.7041	0.7000	0.00%
Cell phones	0.7941	0.6500	0.7149	0.6500	5.53%
Music	0.5000	0.5000	0.5000	0.5000	13.24%
Computers	0.5000	0.5000	0.5000	0.5000	-14.42%
Movies	0.5000	0.5000	0.5000	0.5000	-9.49%

Table 9: Polarity classification over the SFU corpus using eSOLdomainLocal

The second eSOLdomainLocal* resource enriched the iSOL lexicon with sentiment words, but also the most frequent domain-dependent words (Equation 1). Table 10 shows the results obtained with this lexicon.

The next two experiments are similar to the two previous ones but using eSOLdomainGlobal and eSOLdomainGlobal*. The difference between Local and Global is how to find the most used sentiment and domain-dependent words. If one word is used one or more times in a positive or negative review we considered that its frequency is one. Therefore, in this experiment the highest possible

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement
Cars	0.8571	0.8000	0.8276	0.8000	10.04%
Hotels	0.8571	0.8000	0.8276	0.8000	0.00%
Washing machines	0.8846	0.8500	0.8670	0.8500	55.31%
Books	0.7083	0.7000	0.7041	0.7000	0.00%
Cell phones	0.7778	0.6000	0.6774	0.6000	0.00%
Music	0.5980	0.5500	0.5730	0.5500	29.78%
Computers	0.2368	0.4500	0.3103	0.4500	-44.41%
Movies	0.5000	0.5000	0.5000	0.5000	-9.49%

Table 10: Polarity classification over the SFU corpus using eSOLdomainLocal*

frequency of a word is 15. Tables 11 and 12 show the results obtained using the eSOLdomainGlobal and eSOLdomainGlobal* resources respectively.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement
Cars	0.8571	0.8000	0.8276	0.8000	10.04%
Hotels	0.8125	0.8000	0.8062	0.8000	-2.58%
Washing machines	0.8000	0.8000	0.8000	0.8000	43.31%
Books	0.7083	0.7000	0.7041	0.7000	0.00%
Cell phones	0.7941	0.6500	0.7149	0.6500	5.53%
Music	0.5000	0.5000	0.5000	0.5000	13.24%
Computers	0.5000	0.5000	0.5000	0.5000	-10.42%
Movies	0.5549	0.5500	0.5525	0.5500	0.00%

Table 11: Polarity classification over the SFU corpus using eSOLdomainGlobal

Taking as an example the cars domain to simplify, Table 13 shows how many new words were found in the reviews when we use the new lists generated. As we can see, in reviews “coches_no_2_12” and “coches_no_2_20” whose rank is -1, with iSOL the review is classified as Positive (FP), and with the new lists as Negative, which means an improvement of the system.

For the last experiments we did not take into account the different domains individually, and so we did not separate the domains of the SFU Reviews Corpus. Therefore, we have two new groups of reviews, one for generating the lists and another one for testing the new resources. The group for generating new resources eSOLLocal(*) and eSOLGlobal(*) includes a total of 240 documents: 120 pos-

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement
Cars	0.8571	0.8000	0.8276	0.8000	10.04%
Hotels	0.8333	0.7500	0.7895	0.7500	-4.6%
Washing machines	0.8846	0.8500	0.8670	0.8500	55.31
Books	0.7083	0.7000	0.7041	0.7000	0.00%
Cell phones	0.7778	0.6000	0.6774	0.6000	0.00%
Music	0.5980	0.5500	0.5730	0.5500	29.78%
Computers	0.5980	0.5500	0.5730	0.5500	2.64%
Movies	0.5000	0.5000	0.5000	0.5000	-9.49%

Table 12: Polarity classification over the SFU corpus using eSOLdomainGlobal*

itive reviews (15 positive reviews per domain multiplied by 8 domains) and 120 negative reviews. The group of reviews used for evaluating the generated lists is composed of 160 documents (10 positive and 10 negative reviews for each of the 8 domains). Thus, for the eSOLLocal we added only the most frequent sentiment words to iSOL lists and for the eSOLLocal* resource we also included the most frequent domain words if their frequency in positive (or negative) reviews was three or more times as much as negative (or positive) reviews. A similar process was followed to generate the eSOLGlobal and eSOLGlobal* resources. The difference is how to find the most used sentiment and domain words. If one word is used one or more times in one positive or negative review we considered that its frequency is one. Therefore, in this experiment the highest possible frequency of a word is 120. Table 14 shows the results obtained with these new resources, including the baseline experiment using the original iSOL list.

5. Analysis of results

In Table 14 we can see that the results obtained are improved for all the cases when we integrate domain information without taking into account the domains individually but considering all them together.

As we can see by comparing the different tables of results, the baseline experiment is improved upon for five domains (Cars, Washing machines, Music and Mobile phones) when we apply domain sentiment lexicons. Nevertheless, the results obtained with the new resources over the domains Hotels, Books, Computers and Movies are worse or equal than the original iSOL list.

On the other hand, taking the means of the Macro-F1 and accuracy of results,

		iSOL		eSOLcarsLocal		eSOLcarsLocal*		eSOLcarsGlobal		eSOLcarsGlobal*	
id. Review	Rank	Pos.	Neg	Pos.	Neg	Pos.	Neg	Pos.	Neg	Pos.	Neg
coches_yes_5_10	1	4	0	6	0	8	0	6	0	8	0
coches_yes_5_12	1	28	14	31	17	38	18	31	16	38	18
coches_yes_5_15	1	17	3	17	4	24	4	17	4	24	4
coches_yes_5_17	1	7	3	10	3	10	3	10	3	10	3
coches_yes_5_21	1	16	12	20	12	24	15	22	12	24	15
coches_yes_5_25	1	8	7	8	7	10	7	8	7	10	7
coches_yes_5_4	1	16	3	16	3	22	3	16	3	22	3
coches_yes_5_5	1	3	2	8	3	9	5	8	3	9	5
coches_yes_5_7	1	10	5	10	5	15	5	10	5	15	5
coches_yes_5_8	1	20	3	22	4	29	6	23	4	29	6
coches_no_2_10	-1	8	16	9	17	15	17	12	17	15	17
coches_no_2_12	-1	3	2	4	4	4	4	4	4	4	4
coches_no_2_14	-1	3	4	3	6	3	6	3	6	3	6
coches_no_2_16	-1	7	3	7	6	8	3	7	3	8	3
coches_no_2_17	-1	14	5	18	8	20	8	18	8	20	8
coches_no_2_20	-1	7	2	8	8	9	9	8	8	9	9
coches_no_2_22	-1	2	6	3	6	3	8	3	6	3	8
coches_no_2_24	-1	13	10	13	12	17	12	13	11	17	12
coches_no_2_7	-1	8	7	11	7	13	8	11	7	13	8
coches_no_2_9	-1	3	6	3	11	6	11	3	11	6	11

Table 13: Number of words of the different lists in the reviews

	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement
iSOL	0.6452	0.6125	0.6284	0.6125	-
eSOLLocal	0.6950	0.6438	0.6684	0.6438	6.36%
eSOLLocal*	0.7067	0.6125	0.6562	0.6125	4.42%
eSOLGlobal	0.6950	0.6438	0.6684	0.6438	6.36%
eSOLGlobal*	0.6953	0.6250	0.6583	0.6250	4.75%

Table 14: Polarity classification over the SFU corpus using lexicons without taking into account the domain (eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal*)

we can see that the eSOLdomainGlobal and eSOLdomainGlobal* lists obtained results a little better than eSOLdomainLocal and eSOLdomainLocal* lexicons, respectively. This means that in order to generate domain adapted opinion bearing word lists it is advisable to measure the frequency of the words as the number of documents of the corpus where the word appears. So, if a word is in most of the documents of the corpus, it is more representative than the word which is repeated a lot of times in a single document but does not appear in the others. In our case, if a word is in most of the positive documents it is very likely that the word expresses a positive opinion or sentiment, but if that word is only repeated several times in a positive document it does not mean that it expresses a positive meaning.

However, the differences between the eSOLdomainGlobal lexicons and eS-

...Espero que algún día alguien en Hollywood lea "Casa de muñecas", de Ibsen, y sepan que hay mujeres que no se rinden al mandato de la vida cómoda y standar, que quizás alguna sepa dar portazo a tanta tonteria. **No me ha gustado**. Se nota, ¿no?

```
...I hope one day somebody in Hollywood reads ``House of dolls'' by
Ibsen, and they will know that there are women who do not surrender
to the command of comfortable living standard, and also who will know
slamming to such foolishness.  I did not like it.  You notice, don't you?
```

Figure 1: Excerpt of the review "no_2_4.txt" from the Movies domain

OLdomainGlobal* are not significant because they achieved very similar results, so we consider that the eSOLdomainGlobal resource has the most suitable list of opinion bearing words, because it adds less words than eSOLdomainGlobal*. Although we should emphasise that the performance with the domains computers and movies is not good.

As we have said previously the domain adaptation process has not worked as we expected for some domains. One of the problematic domains is "Computers". After reading some of the reviews we have noticed that in the some reviews the author expresses his disagreement and also advice the purchase of distinct computer. Thus in the same review there are positive and negative expressions that could have driven the domain adaptation method to introduce unsuitable words to the lists.

5.1. Negation and irony

A deep analysis of the results shows that the number of FP is quite high. After reading some of the test reviews we can see that some possible causes of the misclassification are associated with the poor treatment of some issues in opinion mining: negation and irony. For example, some reviews that belong to the negative class use negative expressions with positive words to state a negative opinion. An example can be read in file "no_2_4.txt" (Figure 1) of the Movies domain: *no me ha gustado* (I did not like it). The word *gustado* (liked) is in the positive lists of eSOLLocal and eSOLGlobal, so the system considers the word as positive, but it must be negative because the word *no* (not) changes the polarity of *gustado* (liked). Another example can be found in file "no_2_12.txt" (Figure 2) of the Music domain. This file includes the sentence *sin ideas geniales* (without brilliant ideas). The word *geniales* is also in the positive lists of all resources, so the system

... Tras disolver Smashing Pumpkins e intentar volar por si sólo Billy ha tenido que volver al grupo que le dió el éxito en el pasado (eso sí, sólo se mantienen 2 miembros originales) pero ha vuelto **sin ideas geniales** y se ha dedicado a recrear un Revival (bastante mediocre) de lo que el grupo...

`... After dissolving Smashing Pumpkins and trying to go in his own business Billy had to return to the group that gave him success in the past (only two original members remain) but he has returned` **`without brilliant`** `ideas and has dedicated to recreate a Revival (pretty average) of the group...`

Figure 2: Excerpt of the review “no_2_12.txt” from the Music domain

...pero ni acero ni nada, baquelita o malquita o no sé qué historias, que se parte por la mitad. Y eso sí, hasta que acabó de romperse, me pasé un par de meses recogiendo del suelo el agua que se salía del aparato **¡una maravilla!**

`...but not steel or anything, bakelite or malquita or what stuff to be split in half isn’t known . And yes, until finally they break, I spent a couple of months collecting water out of the appliance,` **`it is wonderful!`**

Figure 3: Excerpt of the review “no_1_20.txt” from the Washing machines domain

considers it as positive. As in the previous example the word *sin* (without) changes the polarity of the word *geniales* (brilliant).

However, this kind of error could not be associated to the lexicon because the lexicon only includes bearing words. On the contrary, it is necessary to perform a deeper analysis of the content and develop strategies for dealing with negation.

On the other hand, one of the features of irony is the use of positive words to express a negative point of view about something or somebody. After reading some reviews of the corpus the use of irony is very common in some domains. The expression *¡una maravilla!* (it is wonderful!) in the review “no_1_20.txt” 3 of the Washing machines domain is a clear example.

These are the main reasons for the low performance in some domains, so the errors are not caused by a low quality of the lexicons. Thus the main problem is that the classifier built for domain lexicons assessment only takes into account the words that are on the lists and does not consider other issues of OM. The classifier does not consider negation or irony because the main goal of the paper is the description of the new domain specific sentiment lexicons.

5.2. Evaluating the lexicons over other corpus

To finish our analysis of results, we would like to evaluate the validity of the generated lexicons by comparing the system with other corpora. However, the availability of Spanish corpora is very sparse, so this evaluation is very difficult to carry out. Also, for a complete evaluation of all the lists we need eight different corpora, one per each domain. The only Spanish corpus available is a corpus of movie reviews. The corpus is called MuchoCine (Cruz et al., 2008). Thus, only the list that achieved better results in the Movie domain (eSOLMovieGlobal) has been evaluated with the corpus MuchoCine. The results achieved by iSOL and eSOLMovieGlobal are shown in Table 15. The evaluation of eSOLMovieGlobal with the corpus MuchoCine has shown that the domain adaptation method presented in this paper is also valid for other corpus.

	Macro-P	Macro-R	Macro-F1	Accuracy	Improv.
iSOL	0.6222	0.6147	0.6184	0.6183	–
eSOLMovieGlobal	0.6253	0.6151	0.6206	0.6198	0.35%

Table 15: Polarity classification over the MC corpus using iSOL and eSOLMovieGlobal lexicons

6. Conclusions and further work

In this paper we study the integration of domain information for a Spanish polarity classification system. We have carried out several experiments in order to test the different resources generated from the original Spanish lexicon iSOL: a polarity classification of each domain using iSOL; a polarity classification with the sentiment lexicons eSOLdomainLocal and eSOLdomainLocal*; the same experiment but with the lexicons eSOLdomainGlobal and eSOLdomainGlobal*; and the last experiments with the lists iSOL, eSOLLocal, eSOLLocal*, eSOLGlobal and eSOLGlobal*. All these resources are freely available for research purposes[7].

The results obtained in the polarity classification of the entire corpus independently for the domain, the lexicons eSOLLocal, eSOLLocal*, eSOLGlobal and

[7] `http://sinai.ujaen.es/?p=1264`

eSOLGlobal* are very similar, although we highlight that eSOLLocal and eSOLGlobal achieve better results than eSOLLocal* and eSOLGlobal*. The four lists surpass the results achieved by iSOL, but the differences between them are not significant.

However, according to the domain polarity classification the results over four domains surpass the baseline, while the other four domains seem to be harder to classify. An analysis of the errors shows that the possible cause of the misclassification could be the use of irony and negation in these reviews. Thus, our future work would be focused on the development of techniques for the treatment of negation in OM with the goal of improving the polarity classification systems. Another research line for the future is the analysis whether the application of a homogeneous factor for all the domains is a good strategy, because the analysis of the results shows up that it is very likely that each domain needs its own factor in equation 1.

Acknowledgements

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. The project AORESCU (P11-TIC-7684 MO) from the regional government of Junta de Andalucía partially supports this manuscript, and the project CEATIC-2013-01 from the University of Jaén.

References

Agić, Ž., Ljubešić, N., & Tadić, M. (2010). Towards sentiment analysis of financial texts in croatian. In N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta; ELRA: European Language Resources Association. 19-21.

Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* EMNLP '08 (pp. 127–135). Stroudsburg, PA, USA: Association for Computational Linguistics.

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of* *the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 440–447). Prague, Czech Republic: Association for Computational Linguistics.

Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the International Conference RANLP-2009* (pp. 50–54). Borovets, Bulgaria: Association for Computational Linguistics.

Cruz, F., Troyano, J. A., Enriquez, F., & Ortega, J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural*, *41*, 73–80.

Denecke, K. (2008). Using sentiwordnet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on* (pp. 507–512).

Du, W., Tan, S., Cheng, X., & Yun, X. (2010). Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. In *Proceedings of the third ACM international conference on Web search and data mining* WSDM '10 (pp. 111–120). New York, NY, USA: ACM.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)* (pp. 417–422).

Ghorbel, H., & Jacot, D. (2011). Sentiment analysis of french movie reviews. In V. Pallotta, A. Soro, & E. Vargiu (Eds.), *Advances in Distributed Agent-Based Retrieval Tools* (pp. 97–108). Springer Berlin Heidelberg volume 361 of *Studies in Computational Intelligence*.

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics* EACL '97 (pp. 174–181). Stroudsburg, PA, USA: Association for Computational Linguistics.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* KDD '04 (pp. 168–177). New York, NY, USA: ACM.

Jiang, J., & Zhai, C. (2007). Instance weighting for domain adaptation in nlp. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 264–271). Prague, Czech Republic: Association for Computational Linguistics.

Martínez-Cámara, E., Martín-Valdivia, M., & Ureña-López, L. (2011). Opinion classification techniques applied to a spanish corpus. (pp. 169–176). Springer Berlin / Heidelberg volume 6716 of *Natural Language Processing and Information Systems*. 10.1007/978-3-642-22327-3_17.

Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Ureña López, L. A. (2013). Bilingual experiments on an opinion comparable corpus. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 87–93). Atlanta, Georgia: Association for Computational Linguistics.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, *40*, 7250–7257.

Pan, S. J., Ni, X., Sun, J.-T., Yang, Q., & Chen, Z. (2010). Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th International Conference on World Wide Web* WWW '10 (pp. 751–760). New York, NY, USA: ACM.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (pp. 79–86). Stroudsburg, PA, USA: Association for Computational Linguistics volume 10 of *EMNLP '02*.

Ponomareva, N., & Thelwall, M. (2012). Do neighbours help?: An exploration of graph-based algorithms for cross-domain sentiment classification. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* EMNLP-CoNLL '12 (pp. 655–665). Stroudsburg, PA, USA: Association for Computational Linguistics.

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña López, L. A., & Perea-Ortega, J. M. (2011). Oca: Opinion corpus for arabic. *Journal of the American Society for Information Science and Technology*, *62*, 2045–2054.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*, 1–47.

Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Veláquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2013). Empirical study of machine learning based approach for opinion mining in tweets. In I. Batyrshin, & M. González Mendoza (Eds.), *Advances in Artificial Intelligence* (pp. 1–14). Springer Berlin Heidelberg volume 7629 of *Lecture Notes in Computer Science*.

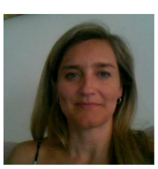
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*, 267–307.

Taboada, M., & Grieve, J. (2004). *Analyzing appraisal automatically*. Technical Report Stanford University.

Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* ACL '02 (pp. 417–424). Stroudsburg, PA, USA: Association for Computational Linguistics.

Xia, R., Zong, C., Hu, X., & Cambria, E. (2013). Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *Intelligent Systems, IEEE*, *28*, 10–18.

Zhang, C., Zeng, D., Li, J., Wang, F.-Y., & Zuo, W. (2009). Sentiment analysis of chinese documents: From sentence to document level. *Journal of the American Society for Information Science and Technology*, *60*, 2474–2487.



M. Dolores Molina-González received the B.S degree in Telecommunication Engineering and an M.S degree in Telecommunication Engineering from the University of Alcalá de Henares, Spain and the University of Valencia, Spain in 1992 and 1995, respectively. She is an assistant professor in the Department of Telecommunication Engineering in the University of Jaén, Spain and she is currently pursuing her Ph. D. in Natural Language Processing from the Computer Science department of the University of Jaén. Her current research interests include sentiment analysis, opinion mining, domain adaptation and information retrieval.

Eugenio Martínez-Cámara received a B.S degree in Computer Science and Management and an M.S degree in Computer Science from the University of Jaén, Spain, in 2008 and 2010, respectively. He is currently pursuing his Ph. D. in Natural Language Processing from the Computer Science department of the University of Jaén. His current research interests include sentiment analysis, opinion mining, text mining, and information retrieval.

M. Teresa Martín-Valdivia is Reader in the Department of Computer Science at Jaén University (Spain). She received her M.S. degree in Computer Science from the University of Granada, and Ph. D. in Computer Science from the Software Engineering Department of Málaga. Her current research interests include natural language processing, tools for text mining, neural networks, machine learning algorithms, text categorization, and multilingual and multimodal information retrieval. In addition, she is Treasurer of the SEPLN Committee since 2007 and editor of the Procesamiento de Lenguaje Natural Journal. She is author or co-author of more than 50 scientific publications. She is a technical reviewer in several journals and in the Program Committee of some major conferences.

L. Alfonso Ureña-López is a Professor in the Department of Computer Science at Jaén University (Spain). He received his M.S. degree in Computer Science from the University of Granada, and Ph. D. in Computer Science from the Software Engineering Department of Granada. His current main research interest is Intelligent Search, including Search Engines, Search Assistants, tools for Text Mining and Retrieval and Human Computer Interaction. Dr. Ureña is Director and founder of the Research Group of Intelligent Systems of Information Access. Also, he is President of the Spanish NLP Society since 2007 and editor of the Procesamiento de Lenguaje Natural Journal.

eSOLHotel: Generación de un lexicón de opinión en español adaptado al dominio turístico

eSOLHotel: Building an Spanish opinion lexicon adapted to the tourism domain

M. Dolores Molina González, Eugenio Martínez Cámara, M.Teresa Martín Valdivia, Salud Jiménez Zafra
Departamento de Informática, Escuela Politécnica Superior de Jaén
Universidad de Jaén, E-23071 - Jaén
{mdmolina, emcamara, maite, sjzafra}@ujaen.es

Resumen: Desde que la web 2.0 es el mayor contenedor de opiniones en todos los idiomas sobre distintos temas o asuntos, el estudio del Análisis de Sentimientos ha crecido exponencialmente. En este trabajo nos centramos en la clasificación de polaridad de opiniones en español y se presenta un nuevo recurso léxico dependiente del dominio (eSOLHotel). Este nuevo lexicón usa el enfoque basado en corpus. Hemos realizado varios experimentos usando la aproximación no supervisada para la clasificación de polaridad de las opiniones en la categoría de hoteles del corpus SFU. Los resultados obtenidos con el nuevo lexicón eSOLHotel superan los resultados obtenidos con otro lexicón de propósito general.
Palabras clave: Clasificación de polaridad, corpus de opiniones en español, lexicón dependiente del dominio.

Abstract: Since Web 2.0 is the largest container for subjective expressions about different topics or issues expressed in all languages, the study of Sentiment Analysis has grown exponentially. In this work, we focus on Spanish polarity classification of hotel reviews and a new domain-dependent lexical resource (eSOLHotel) is presented. This new lexicon has been compiled following a corpus-based approach. We have carried out several experiments using an unsupervised approach for the polarity classification over the category of hotels from corpus SFU. The results obtained with the new lexicon eSOLHotel outperform the results with other general purpose lexicon.
Keywords: polarity classification, Spanish reviews corpus, dependent-domain lexicon.

1 *Introducción*

En los últimos años, el interés por el Análisis de Sentimientos (AS) (conocido en inglés como sentiment analysis u opinion mining) ha crecido significativamente debido a diferentes factores (Pang y Lee, 2008) (Liu, 2012) (Tsytsarau y Palpanas, 2012). Por una parte, el incremento de la creación y compartición de datos por parte de los usuarios de Internet haciendo uso de las nuevas plataformas y servicios que están emergiendo continua y expeditamente. Por otra parte, el consumo de datos online comienza a ser una tarea imprescindible y rutinaria para la toma de decisiones a nivel individual o colectivo.

Muchas son las tareas estudiadas en AS, siendo una de las más consolidas la clasificación de la polaridad. En esta tarea se han seguido distintas aproximaciones, aunque son dos las líneas principales. Por una parte, la aproximación basada en técnicas de aprendizaje automático (Machine Learning ML), la cual se basa en entrenar unos modelos a partir de una colección de datos etiquetada a priori con el objetivo de predecir el valor de salida correspondiente a cualquier dato de entrada válido. Los clasificadores pueden estar basados en distintos algoritmos, entre los más utilizados están Support Vector Machines (SVM), Maximum Entropy (ME) o Naïve Bayes (NB). Los clasificadores tienen el inconveniente de necesitar gran cantidad de datos de entrada para un entrenamiento previo y poder obtener buenos resultados. Trabajos como (Pang, Lee y Vaithyanathan, 2002) usan este enfoque supervisado para resolver el problema de la clasificación de polaridad.

La segunda línea conocida como aproximación basada en Orientación Semántica (Semantic Orientation SO) obtiene la polaridad de cada documento como la agregación de la inclinación positiva o negativa de sus palabras. La polaridad de las palabras puede ser determinada por diferentes métodos, por ejemplo usando una lista de palabras de opinión (Hu y Liu, 2004), utilizando búsquedas en la web (Hatzivassiloglou y Wiebe, 2000), consultando en una base de datos léxica como WordNet (Kamps et al., 2004) o considerando alguna característica lingüística para determinar el sentimiento a nivel de palabra (Ding y Liu, 2007) (Hatzivassiloglou y Mckeown, 1997) (Turney, 2002). Esta aproximación no necesita de una colección de datos etiquetada a priori para un entrenamiento previo aunque sí de recursos léxicos normalmente dependientes del idioma para determinar la polaridad de las palabras. Aunque ambas aproximaciones tienen ventajas e inconvenientes, nuestro trabajo se engloba en la aproximación basada en SO. Muchos investigadores han guiado sus pasos intentando resolver estos problemas pero aún quedan otros retos que afrontar y abordar como es la adaptación de la clasificación de opiniones al dominio tratado (Aue y Gamon, 2005). Es en este reto donde centraremos el esfuerzo de este artículo.

Por otra parte, la mayoría de los trabajos en SA tratan con documentos escritos en inglés a pesar de que cada vez es mayor la cantidad de información subjetiva que publican los usuarios de internet en su propio idioma. Es por esta razón, que la generación y uso de recursos propios en el idioma de los documentos a tratar se esté convirtiendo en un tema crucial para realizar la clasificación de opiniones mediante orientación semántica. Así pues, nuestro artículo está enfocado al AS en español de manera que los recursos que utilizaremos estarán en este idioma, tanto corpora como lexicones.

Resumiendo, el desarrollo de recursos lingüísticos nuevos es muy importante para seguir progresando en AS. Además, se hace necesario que esos nuevos recursos se implementen en otros idiomas distintos al inglés, como el español por ejemplo. Así, la descripción de un corpus nuevo de opiniones en el dominio turístico, la descripción de un lexicón de palabras con sentimientos dependiente del dominio y unos experimentos que certifiquen la validez de dichos recursos son la principal contribución de este artículo.

El presente artículo se estructura de la siguiente manera: en la sección 2 se describen brevemente otros trabajos relacionados con la clasificación de polaridad en opiniones escritas en español, trabajos que generan nuevos recursos léxicos y algunos trabajos relacionados con la adaptación al dominio en AS. En la sección 3 se explican los diferentes recursos utilizados, así como la metodología utilizada para la generación del nuevo lexicón adaptado al dominio. En la sección 4 se muestran los experimentos realizados y se discute los resultados obtenidos. Por último, se exponen las conclusiones y el trabajo futuro.

2 *Trabajos relacionados*

Centrándonos en los trabajos realizados sobre AS a continuación se presentan los más relevantes en un idioma distinto del inglés. Como primer trabajo se tiene el de Banea et al. (2008), el cual propone varios enfoques para el análisis de la subjetividad en varios idiomas mediante la aplicación directa de las traducciones de un corpus de opiniones etiquetadas en inglés para el entrenamiento de un clasificador de opiniones en rumano y español. Este trabajo muestra que la traducción automática es una alternativa viable para la construcción de recursos y herramientas para el análisis de la subjetividad en un idioma distinto al inglés. Brooke et al. (2009) presentan varios experimentos relacionados con los recursos en español e inglés. Llegan a la conclusión de que, aunque las técnicas de aprendizaje automático pueden proporcionar un buen rendimiento, es necesario integrar el conocimiento y los recursos específicos del idioma con el fin de lograr una mejora notable. Se proponen tres enfoques: el primero utiliza los recursos de forma manual y automáticamente generados para el español. El segundo aplica aprendizaje automático sobre un corpus español y el último traduce los corpus del español al inglés y luego aplica SO-CAL, (Semantic Orientation

Calculator), una herramienta desarrollada por ellos mismos (Taboada et al., 2011). Martínez-Cámara et al. (2011) emplean un corpus de críticas de cine llamado MuchoCine (Cruz et al., 2008) para clasificar opiniones escritas en español usando un enfoque supervisado, y Martín-Valdivia et al. (2013) empleando el mismo corpus de cine en español y generando el corpus paralelo en inglés MCE realiza una combinación de la clasificación supervisada realizada sobre ambos corpus y una clasificación no supervisada integrando SentiWordNet (Esuli and Sebastiani, 2006) sobre el corpus en inglés.

Para realizar la clasificación de la polaridad siguiendo un enfoque basado en orientación semántica, muchos autores usan o generan recursos léxicos en el idioma en el que están escritas las opiniones. Así, Taboada et al. (2008) ponen a disposición de los investigadores el corpus SFU en inglés con 400 opiniones distribuidas en 8 categorías con 25 opiniones positivas y otras 25 negativas cada categoría. Al poco tiempo, generan otro corpus en español siguiendo la misma filosofía con 8 categorías similares, el corpus SFU en español. En Cruz et al. (2008) se describe la generación de un corpus MC de críticas de cine escritas en español a partir de la página web MuchoCine.com[1]. El corpus cuenta con 1274 opiniones clasificadas como negativas y 1351 opiniones clasificadas como positivas. Boldrini et al. (2009) presenta el corpus EmotiBlog que incluye comentarios sobre varios temas en tres idiomas: español, inglés e italiano. En Molina-González et al (2013) se presenta un nuevo recurso para la comunidad investigadora en Análisis de Sentimientos en español. El recurso, llamado iSOL y que será utilizado en este artículo, es un lexicón de palabras de opinión generado a partir del conocido y ampliamente usado lexicón existente en inglés de Bing Liu (Hu and Liu, 2004). En Díaz-Rangel et al (2014) se proporciona un lexicón de emociones en español compuesto de 2036 palabras que llevan asociada un factor de probabilidad de uso afectivo (PFA) con respecto al menos una de las emociones básicas: alegría, enfado, tristeza, sorpresa y disgusto.

Por otra parte, como es bien sabido, la orientación semántica de muchas palabras son dependientes del dominio que se trate, existiendo diversos documentos que corroboran este hecho como son (Engström, 2004) (Owsley et al., 2006), (Blitzer et al., 2007). Existen trabajos más actuales como (Dehkharghani et al., 2012) el que se propone un método para construir un sistema de clasificación de la polaridad dependiente del dominio. El dominio seleccionado por los autores son comentarios de los huéspedes de hoteles. Cada opinión se representada por un conjunto de características independientes del dominio y un conjunto de características dependientes del dominio. En (Demiroz et al., 2012) se propone un método para adaptar un recurso lingüístico de sentimientos independiente del dominio, como SentiWordNet, a un dominio específico. En (Molina-González et al., 2013) se detalla la generación de un recurso léxico basado en listas de palabras de opinión adaptado al dominio de cine. Nuestra propuesta sigue un enfoque basado en corpus, pero en este caso el dominio utilizado es el turístico y concretamente, usaremos un corpus con opiniones extraídas de TripAdvisor para diferentes hoteles de Andalucía. Los buenos resultados obtenidos en los experimentos demuestran que nuestra propuesta es válida independientemente del dominio elegido.

3 *Recursos: corpora y lexicones*

En esta sección se describe en primer lugar el corpus de opiniones sobre hoteles. Este corpus se llama COAH (Corpus of Opinions from Andalusian Hotels) y está disponible libremente para fines de investigación bajo petición a los autores. Los lexicones usados para la experimentación son el lexicón iSOL independiente del dominio usado en varios trabajos como (Molina-González et al., 2013) y el nuevo lexicón eSOLHotel (iSOL enriquecido para el dominio de hoteles) generado a partir del corpus COAH. El corpus usado para probar la bondad del lexicón generado eSOLHotel es el corpus SFU en español (Taboada et al., 2004) en particular de las opiniones en la categoría de hoteles.

3.1 Corpus COAH

Para compilar un corpus de opiniones es muy importante saber elegir la fuente de dichos datos. En nuestro caso, hemos intentado satisfacer los siguientes requisitos:

- Debe haber gran cantidad de opiniones y éstas deben ser escritas por usuarios de los hoteles.

[1] http://www.muchocine.net/

- Cada opinión debe estar valorada por el propietario de dicha opinión.
- El portal web debe ser un portal confiable en el dominio de hoteles.
- Debe ser un portal prestigioso internacionalmente en la búsqueda de información sobre hoteles.

Después de estudiar varios portales web, nuestra elección final fue TripAdvisor[2]. El corpus generado consiste en una colección que contiene 1816 opiniones escritas por usuarios no profesionales. Este hecho incrementa la dificultad de la tarea, porque los textos pueden no ser gramaticalmente correctos, incluso contener palabras mal escritas o expresiones informales. Se han seleccionado solo hoteles andaluces. Por cada provincia de Andalucía (Almería, Cádiz, Córdoba, Granada, Jaén, Huelva, Málaga and Sevilla), se han elegido 10 hoteles, siendo 5 de ellos de valoración muy alta y los otros 5 con las peores valoraciones. Todos los hoteles seleccionados deben tener al menos 20 opiniones escritas en español en los últimos años. Finalmente, se han obtenido 1.816 opiniones.

Las opiniones están valoradas en una escala de 1 a 5. El valor 1 significa que el autor manifiesta una opinión muy negativa sobre el hotel, mientras que una puntuación de 5 quiere decir que el autor tiene muy buena opinión sobre el hotel. Los hoteles con valor 3 se pueden catalogar como hoteles neutros, ni buenos ni malos. En la Tabla 1 se muestra el número de opiniones por valoración.

Valoración	Número de opiniones
1	312
2	199
3	285
4	489
5	531
Total	1.816

Tabla 1 Distribución por valoración

#Opiniones	1.816
#Hoteles	80
Media de opiniones por hotel	22,7
#Palabras	264.303
#Frases	9.952
#Adjectivos	17.800
#Adverbios	15.219
#Verbos	38.590
#Sustantivos	53.640
Media de palabras por frase	26,55
Media de palabras por opinión	145,54
Media de adjetivos por opinión	9,80
Media de adverbios por opinión	8,38
Media de verbos por opinión	21,25
Media de sustantivos por opinión	29,54

Tabla 2 Estadísticas de COAH

En la Tabla 2 se muestran algunas características del corpus. De los metadatos mostrados en la Tabla 2 se puede resaltar que las opiniones tienen una media de 145 palabras suficientes para dar la opinión subjetiva sin implicarse en descripciones objetivas fuera de nuestro estudio. Las páginas web extraídas fueron transformadas en ficheros xml (uno por hotel). Cada fichero xml tiene 20 opiniones. Cada opinión tiene dos tipos de información, una sobre el hotel y otra sobre la opinión del huésped del hotel.

A partir de los ficheros xml se genera un documento que solo alberga la valoración de un hotel específico, el título y la opinión. Para los experimentos se descartan aquellas opiniones neutras, es decir, con valoración 3. El resto de opiniones son catalogadas como positivas si su valoración es 4 ó 5, y negativas si su valoración es 1 ó 2. Por tanto la clasificación binaria de las opiniones sobre hoteles del corpus COAH es la que se muestra en la Tabla 3.

Clases	**Número de opiniones**
Positiva	1.020
Negativa	511
Total	1.531

Tabla 3 Clasificación binaria del corpus COAH

[2] http://www.tripadvisor.es

En las Figuras 1 y 2 se muestra un ejemplo de un hotel, en XML y en formato texto.

```
<ID>1</ID>
<Nombre>Alcazaba Mar Hotel</Nombre>
<Categoria>4</Categoria>
<Dirección>Juegos del Argel, Urbanizacion El Toyo |Cabo de Gata </Dirección>
<CódigoPostal>04131</CódigoPostal>
<localidad>Retamar</localidad>
<Provincia>Almería</Provincia>
<País>España</País>
<Viajero>-------------</Viajero>
<Localidad_Viajero>-------
</Localidad_Viajero>
<Valoración>3</Valoración>
<Título>"Adecuada la calidad al precio del hotel"</Título>
<Opinión>Acabamos de llegar del hotel. La verdad es que nos fuimos con mucho miedo por los comentarios escritos aquí. Nuestra opinion es que es un hotel comodo, tiene piscina buena, animacion excelente, y un personal muy amable. Quizas lo mas tenido en cuenta es el buffet…………………………………… </Opinión>
<Fecha_TipoViajero>Se alojó el Agosto de 2012, viajó con la familia</Fecha_TipoViajero>
<Relación_calidad-precio>3</Relación_calidad-precio>
<Ubicación>2</Ubicación>
<Calidad_del_sueño>3</Calidad_del_sueño>
<Habitaciones>3</Habitaciones>
<Limpieza>3</Limpieza>
<Servicio>4</Servicio>
```

Figura 1 Ejemplo de un hotel en el corpus COAH

Valoración| Título| Opinión

1 | "Un hotel digno de mención!" | Como bien les comenté a los propietarios a la hora de abandonar el hotel, no dudaré un momento en recomendar una y otra vez el Hotel Albero de Granada. Su situación respecto del centro de Granada no es la mejor, pero para nuestros propósitos era perfecto (escapada de fin de semana con visita a la Alhambra). Se encuentra en la carretera de……………………………… Si vuelvo a Granada no dudaré en hospedarme en el mismo hotel. Muchas gracias por todo!!

Figura 2 Fragmento de una opinión del corpus COAH

3.2 Corpus SFU

Para realizar los experimentos, se elige parte del corpus SFU Corpus. El Corpus SFU se compone de opiniones de productos en inglés y español. La versión en inglés (Taboada y Grieve, 2004) tiene 400 opiniones (200 positivas y 200 negativas) de productos comerciales descargados de la web Epinions[3] en el año 2004. Se divide en ocho categorías: libros, coches, ordenadores, utensilios de cocina, hoteles, películas, música y teléfono. Cada categoría incluye 25 opiniones positivas y 25 de opiniones negativas. Posteriormente, los autores de SFU Corpus hacen disponible la versión española del corpus[4], con el objetivo de ofrecer un corpus comparable para las siguientes investigaciones. Las opiniones en español se dividen en ocho categorías similares, y también cada categoría tiene 25 opiniones positivas y 25 de opiniones negativas. En este caso, las opiniones se descargan desde la web Ciao.es[5]. Para realizar nuestros experimentos se eligen las opiniones de la categoría hoteles.

3.3 Lexicón iSOL

Este recurso fue generado a partir del lexicón en inglés de Bing Liu (Hu & Liu, 2004) traduciéndolo automáticamente al español obteniendo el recurso SOL (Spanish Opinion Lexicon). Posteriormente, la lista fue revisada manualmente. La lista final de palabras de opinión se llama iSOL (improved SOL). El lexicón iSOL se compone de 2.509 palabras positivas y 5.626 palabras negativas, en total, el lexicón español tiene 8.135 palabras polarizadas. Este recurso ha sido evaluado satisfactoriamente en (Molina-González et al., 2013) usando el corpus MuchoCine (Cruz et al., 2008). Los resultados mostraron que el uso de la lista mejorada de palabras polarizadas puede ser una buena estrategia para la clasificación de polaridad no supervisada.

3.4 Lexicón eSOLHotel

El lexicón iSOL es de propósito general, sin embargo en las aplicaciones reales, la orientación semántica de una palabra, frase o documento depende del dominio tratado. Dentro de los enfoques seguidos para la compilación de un conjunto de palabras de opinión, el más adecuado para obtener términos con carga semántica dependientes del dominio es el que se conoce como el enfoque basado en corpus (Kanayama y Nasukawa, 2006).

Tomando como referencia el lexicón iSOL, se ha generado una lista de palabras de opinión para el dominio de hoteles. Para la generación

[3] http://www.epinions.com/
[4] https://www.sfu.ca/~mtaboada/download/downloadCorpusSpa.html
[5] http://www.ciao.es/

de la lista de palabras de opinión se ha seguido el enfoque basado en corpus. El elemento clave del enfoque basado en el corpus es el uso de un corpus etiquetado según su polaridad. El corpus español seleccionado para el proceso es COAH. Hemos seguido el mismo supuesto que (Du et al., 2010), es decir, una palabra debe ser positiva (o negativa) si aparece en muchos documentos positivos (o negativos). Por lo tanto, hemos calculado la frecuencia de la palabra en cada clase de documentos (positivos y negativos). Se han encontrado unas 166 palabras positivas y 131 palabras negativas. Por lo tanto, se añadieron las 297 palabras más frecuentes que aún no figuraban en la lista iSOL a la lista final obteniendo un total de 8.432 palabras polarizadas (2.675 positivas y 5.757 negativas). Esta nueva lista de integración de la información del corpus ha sido llamada eSOLHotel (SOL enriquecido y adaptado al dominio Hotel). En la siguiente Tabla 4 se muestran algunas de las palabras que han sido añadidas.

Palabras positivas	Palabras negativas
ensueño	asqueroso
luminoso	cucaracha
coqueto	desconchones
comodisima	humedades
intachable	mejorable
remodelado	reclamaciones
pasada	tugurio
supercentrico	zulo

Tabla 4 Palabras positivas y negativas añadidas al lexicón eSOLHotel

4 *Experimentos y resultados*

Antes de llevar a cabo los experimentos, a las opiniones de hoteles del corpus SFU se le ha realizado un *preprocesamiento* con el fin de tener los mismos criterios que se han tenido en cuenta en la generación de los lexicones iSOL y eSOLHotel. Por ejemplo, las letras mayúsculas se han cambiado a minúsculas, a las vocales acentuadas se les ha quitado el acento y caracteres especiales han sido separado de las palabras, para aislar dichas palabras.

Para decidir si una opinión se considera positiva o negativa, seguimos un simple método basado en la cuenta del número de palabras incluidas en las listas iSOL y eSOLHotel encontradas en las opiniones de hoteles del corpus SFU etiquetado en español. Así nuestro método clasifica la opinión como positiva si el número de palabras positivas encontradas es igual o mayor que el número de palabras negativas encontradas o como negativa en el resto de casos.

En la Tabla 5 se muestran los resultados obtenidos en la categoría de hoteles del corpus SFU en español usando los lexicones iSOL (independiente del dominio) y eSOLHotel (adaptado al dominio de hoteles).

Lexicón	Precisión	Recall	F1	Acc.
iSOL	77,41%	70,0%	73,52%	70,0%
eSOLHotel	84,72%	78,0%	81,22%	78,0%

Tabla 5 Resultados obtenidos en la clasificación binaria de corpus SFU usando iSOL y eSOLHotel

Los resultados que se muestran en la Tabla 5 confirman nuestra hipótesis de partida, es decir, que la inclusión de información del dominio en una lista de palabras de opinión genérica mejora los resultados de la clasificación de la polaridad. El porcentaje de mejora que se ha obtenido con la inclusión de información del dominio ha sido de un 11,43%. Siguiendo una metodología muy simple, como la que se ha descrito, se ha obtenido una mejora muy importante.

Con el fin de profundizar en el estudio de la bondad de la metodología seguida para la inclusión de información del dominio, se ha construido un clasificador supervisado. Para ello, se ha aplicado un *stemmer* a los documentos, se han representado como vectores de unigramas pesados con su valor TF-IDF. Por último se ha realizado una validación cruzada con el algoritmo Support Vector Machine (SVM). Los resultados que se han obtenido son 82% y un 82,71% de Accuracy y F1-score respectivamente. De nuevos los resultados de la Tabla 5 indican la bondad la metodología presentada en el artículo, dado que la diferencia de F1-score entre SVM y eSOLHotel es solo de un 0,96%. Por lo tanto, la perdida de exactitud es tan mínima que puede considerarse aconsejable el uso de la lista en lugar del método supervisado, ya que en este caso no se necesitaría de un modelo de aprendizaje automático previamente entrenado.

5 Conclusiones y trabajos futuros

En este artículo se ha presentado una metodología de adaptación de un lexicón de palabras de opinión a un dominio concreto. Para ello se ha tomado un corpus de opiniones de hoteles como referencia (COAH), se han calculado la frecuencia de los términos que componen el corpus y se han seleccionado las palabras de opinión más representativas del corpus. La metodología se ha evaluado con las opiniones de hoteles del corpus SFU en español. Los resultados que se han obtenido (Tabla 5) ponen de manifiesto la bondad de la metodología y nos anima a seguir perfeccionando la metodología de adaptación al dominio.

El sistema de clasificación se puede todavía mejorar aún más. Como trabajo futuro se va a incluir un tratamiento de la negación basado en reglas lingüísticas específico para español. Este nuevo elemento del sistema nos va a permitir clasificar correctamente las opiniones negativas expresadas con términos positivos negados.

Agradecimientos

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. The project AORESCU (P11-TIC-7684 MO) from the regional government of Junta de Andalucía partially supports this manuscript. Also the project CEATIC (CEATIC-2013-01) from the University of Jaén partially support this paper.

Bibliografía

Aue, A. y M. Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*

Banea, C., R. Mihalcea, J. Wiebe, y S. Hassan, S. 2008. Multilingual subjectivity analysis using machine translation. *In Proceedings of the conference on empirical methods in natural language processing (EMNLP '08)* (pp. 127–135). Stroudsburg, PA, USA: Association for Computational Linguistics.

Blitzer, J., M. Dredze, y F. Pereira. 2007. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the Association for Computational Linguistics (ACL)*.

Boldrini, E., A. Balahur, P. Martínez-Barco, y A. Montoyo. 2009. "Emotiblog: an annotation scheme for emotion detection and analysis in non-traditional textual genres." *In DMIN, pages 491–497. CSREA Press.*

Brooke, J., M. Tofiloski, y M. Taboada. (2009). Cross-linguistic sentiment analysis: From English to Spanish. *In Proceedings of the International Conference RANLP-2009 (pp. 50–54).* Borovets, Bulgaria: Association for Computational Linguistics.

Cruz, F.L., J.A. Troyano, F. Enriquez, y J. Ortega. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento de Lenguaje Natural, vol. 41, pp 73-80.*

Dave, K., S. Lawrence, y D. M. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of WWW, pp. 519–528.

Demiroz, G., B. Yanikoglu, D. Tapucu, y Y. Saygin. 2012. Learning domain-specific polarity lexicons. *In: Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on. 674-679*

Díaz Rangel, I., G. Sidorov y S. Suárez-Guerra. 2014 Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. *Onomazein, 29, 23 p*

Ding, X., y B. Liu. 2007. The utility of linguistic rules in opinion mining. *In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 811–812). ACM.

Du, W.T., S. Cheng, y X. Yun. 2010. Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. *In Proceedings of ACM International Conference on Web search and data mining (WSDM-2010).*

Engström, C. 2004. Topic dependence in sentiment classification. *Master's thesis*, University of Cambridge.

Esuli, A. y F. Sebastiani. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *In Proceedings of Language Resources and Evaluation (LREC).*

Hatzivassiloglou, V., y K. McKeown. 1997. Predicting the semantic orientation of adjectives. *In Proceedings of the eighth conference on European chapter of the association for computational linguistics* (pp. 174–181).

Hatzivassiloglou, V., y J. Wiebe, J. 2000. Effects of adjective orientation and gradability on sentence subjectivity. *In Proceedings of the international conference on computational linguistics (COLING)* (pp. 299–305).

Hu, M. y B. Liu. 2004. Mining and summarizing customer reviews. *In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* Seattle, Washington, USA, pp. 168-177.

Kamps, J., M. Marx, R.J. Mokken, y M. de Rijke. 2004. Using WordNet to measure semantic orientations of adjectives. *In LREC, European Language Resources Association*.

Kanayama, H. y T. Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 355–363, Sydney, Australia, July 2006. Association for Computational Linguistics.

Liu, B. 2012. Sentiment analysis and opinion mining. synthesis lectures on human language technologies. *Morgan & Claypool Publishers.*

Martín-Valdivia, M.T., E. Martínez-Cámara, J.M. Perea-Ortega, y L.A. Ureña-López. 2013. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications, 40 (10)* pp. 3934–3942

Martínez-Cámara, E., M.T. Martín-Valdivia, y L.A. Ureña-López. 2011 “Opinion classification techniques applied to a Spanish corpus”. *Proceedings of the 16th international conference on Natural language processing and information systems, NLDB’11*, Springer-Verlag, pp. 169–176

Molina-González, M. D., E. Martínez-Cámara, M.T. Martín-Valdivia,y J.M. Perea-Ortega. 2013. Semantic Orientation for Polarity Classification in Spanish Reviews. *Expert Systems with Applications; 40(18)*, pp. 7250-7257.

Owsley, S., S. Sood, y K.J. Hammond. 2006. Domain specific affective classification of documents. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pages 181–183,.

Pang, B., y L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Pang, B., L. Lee, y S. Vaithyanathan. 2002. Thums up?: Sentiment Analysis classification using machine learning techniques. *In Proceedings of the ACL-02 Conference on Empirical methods in natural language processing. Stroudsburg, PA, USA: Association for Computational Linguistics; 10:79-86.*

Read, J. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. *In Proceedings of the ACL Student Research Workshop.*

Taboada, M., J. Brooke, M. Tofiloski, K.D. Voll, y M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.

Taboada, M. 2008. SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University.

Tsytsarau, M., y T. Palpanas. 2012. Survey on mining subjective data on the web. *Data Mining and Knowledge. Discovery. 24, 3* 478-514.

Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 417-424. Association for Computational Linguistics.*

Automatic generation of subjective lexicons adapted to a specific domain using linguistic resources

M. Dolores Molina-González[a], María-Teresa Martín-Valdivia[a], Eugenio Martínez Cámara[a,*], F. Javier Ortega[b], Fermín L. Cruz[b]

[a]SINAI Research Group, University of Jaén, Campus Las Lagunillas, E-23071, Jaén (Spain)
[b]ITALICA Research Group, University of Seville, Avda. Reina Mercedes s/n, 41012, Seville (Spain)

Abstract

In this paper we study the generation of lexicons for polarity classification, one of the most popular areas in Sentiment Analysis. Lexicons have been widely used for sentiment and subjectivity analysis because of their simplicity and effectiveness when it comes to building Semantic Orientation (SO) based polarity classifiers. The use and integration of linguistic resources such as lexicons or corpora is paramount in polarity classification systems based on SO. However, the generation of lexicons must take into account not only the language of the classifier but also the specific domain where it will be applied. We will work with Spanish documents and our main goal is to define a method to automatically create domain-specific sentiment lexicons. Our effort will be focused on investigating the adaptation to a specific domain in order to ease and improve the polarity classification system using a corpus-based methodology to build lexicons. We propose a simple strategy to integrate the knowledge extracted from a Spanish corpus of hotel reviews. We carry out a study to determine how many polar words should be included in the new lexicon. The results obtained are very good, showing an improvement for all the cases studied and promoting the idea that including frequent words in positive/negative reviews is a promising methodology for generating lexicons adapted to a specific domain.

Keywords:
domain adaptation, sentiment analysis, lexicon generation, semantic orientation, Spanish opinion mining, Spanish linguistic resources

1. Introduction

In recent years, interest in Sentiment Analysis (SA) has grown quickly due to various factors [21] [5]. The generation and sharing of data by Internet users making use of new platforms and services that are emerging is increasing. Moreover, consumption of online data for decision-making at individual or collective level is becoming an essential and common task. Knowledge of the degree of satisfaction of customers is very important for the industrial, commercial, political

*Corresponding author

Email addresses: mdmolina@ujaen.es (M. Dolores Molina-González), maite@ujaen.es (María-Teresa Martín-Valdivia), emcamara@ujaen.es (Eugenio Martínez Cámara), fjortega@lsi.us.es (F. Javier Ortega), fermin@lsi.us.es (Fermín L. Cruz)

Preprint submitted to Knowledge-Based Systems *October 10, 2014*

and social environment. The reviews posted on web sites by users are the best source of feedback about a specific issue. The area that analyses these opinion data is Sentiment Analysis (SA), also known as Opinion Mining (OM).

Among the different tasks which are encompassed in SA, one of the most relevant is polarity classification. In this task, two main approaches can be followed. On the one hand, the approach based on Machine Learning (ML) techniques is based on training a model from an annotated dataset in order to predict the output value (positive or negative) corresponding to any valid input data [29]. The classifiers can be based on different algorithms such as Support Vector Machines (SVM), Maximum Entropy (ME) or Naïve Bayes (NB). This methodology has the disadvantage of requiring large amounts of input data for prior training in order to obtain good results. The second approach, known as Semantic Orientation (SO), computes the polarity of each document as the aggregation of the negative and positive polarities of its words. The polarity of words can be determined by different methods, for example using a list of words [15], web searches [34], lexical databases such as WordNet [18], sentiment databases as WeFeelFine[1] [26], or considering any linguistic feature to determine the sentiment at word level [11] [14]. This approach does not require an annotated dataset for training, although normally dependent lexical resources of the language are required in order to determine the polarity of words. Usually, SO obtains worse results than ML and it has been less studied. Moreover, the SO approach presents several issues, such as language dependence, domain adaptation and the absence of lexical resources in a specific language. Our work is centered on solving some of these problems, mainly the last one.

Most of the work in SA deals with documents written in English, even though the amount of opinions published on the Internet in other languages is increasing. For this reason the generation and use of resources in the language of the documents under consideration is becoming a crucial task for the classification of reviews by their semantic orientation. Moreover, there are well-known lexicons for English, such as SentiWordNet[3], General Inquirer [33], MPQA [37] and Bing Liu's Lexicon [15]. However, it is difficult to find these kinds of linguistic resources for other languages like Spanish or Chinese. Our main goal is to build a Spanish lexicon adapted to the hotel domain. To this end, we take as a baseline a list of polar words (the iSOL lexicon) [23] and then we automatically add positive and negative words to this list, extracting them from a Spanish corpus of hotel reviews.

Several studies show performance differences between different domains in polarity classification. In [2] different methodologies for the adaptation of a sentiment classification system to a new domain in the absence of large amounts of annotated data are presented. [27] highlight the importance of building a domain-specific sentiment classifier. The creation of lists of polar words is normally addressed by means of dictionary-based or corpus-based methods. This last method is usually applied in order to create domain-specific sentiment lexicons. Our effort will be focused on building new resources adapted to specific domains in order to ease and improve the polarity classification system using corpus-based methods. We propose a simple strategy to integrate the knowledge extracted from a Spanish corpus of hotel reviews. We carry out a complete study to determine how many polar words should be included in the new lexicon. The results obtained are very good showing an improvement for all the cases and revealing interesting features in the domain adaptation.

The research related to SA is focused on discovering features that represent the intention of a review. These features are usually starred by linguistic elements with a sentiment sense. Thus,

[1]WeFeelFine is described in [19]

most of the linguistic resources utilised and also being developed are constituted by words or idioms that have at least one subjective meaning. But opinions are not only manifested through subjective statements. In [21] it is asserted that there are two kinds of opinions, explicit and implicit opinions. Explicit opinions are subjective statements that give a regular or comparative opinion. Due to the fact that they are subjective clauses, subjective linguistic elements occur in those sentences, so the sentiment linguistic resource can be used to detect them. For example, in the sentence "My teacher is great", the word "great" is an adjective with a subjective meaning. Meanwhile, implicit opinions are objective statements that imply a regular or comparative opinion. For example, in the sentence "in my hotel room there are hairs" there are no words that can trigger a polarity classifier based on linguistic resources, but the sentence does express a fact that is mostly considered as negative, because people do not like hotel rooms with the hairs of previous guests. The methodology herein described of adapting a general domain opinion lexicon to a specific domain takes into account the existence of implicit opinions, because it not only appends adjectives or adverbs, which are the morphological categories most related to subjectivity, to the list but it adds frequent words without taking into account their morphological category, with the aim of covering the implicit opinions.

The rest of the paper is organized as follows: The next section presents related studies that apply a semantic orientation approach focusing mainly on the generation and use of lexical resources. We also comment on some research studies which deal with domain-dependent SA. Section 3 introduces the main resources used to carry out our experiments and to generate the new lexicons. So in Section 4 we describe the construction process and the features of a lexicon adapted to a specific domain that we used in our experiments. Section 5 presents the results obtained in the experiments we performed. Finally, conclusions and future work are presented.

2. Related Work

The use and integration of linguistic resources is paramount in polarity classification systems based on SO. Perhaps one of the most well-known lexicons is SentiWordNet (SWN) [3]. SWN is a resource that is widely used in opinion mining, and is based on the English lexical database WordNet [22]. SWN attaches three sentiment scores: positivity, negativity and objectivity to each WordNet synset. The scores, which are in the range of $[0, 1]$ and add up to 1, are obtained using a semi-supervised machine learning method. SWN is composed of about $117,000$ synsets. It is freely distributed for non-commercial use, and licenses are available for commercial applications. The MPQA (Multi-Perspective Question Answering) is another subjectivity lexicon. It was collected from a number of sources: some were chosen from manually developed resources, others were automatically identified using both annotated and unannotated data. A majority of the words were collected as part of the work reported in [31]. Nowadays, the MPQA lexicon is composed of 8222 lines and each line has five features (*type_strength*, *length*, *partofspeech*, *stemmed* and *prior_polarity*) concerning a given word. In [1] a very simple method was presented of extracting polarity information by means of the quality synset in WordNet. They developed the resource Q-WordNet. Another widely used resource is the English lexicon of Bing Liu [15]. The Bing Liu English Lexicon (BLEL) consists of a polar list of $4,783$ negative words and $2,006$ positive words. We can find many misspelled words in the list, but they are not mistakes. They are included due to their frequent appearances in social media content. [36] propose a method to identify and contextualize ambiguous sentiment terms using a specific domain to enrich lexical resources as SenticNet [6].

Regarding the resources for Spanish, we can find some interesting lexicons like the Spanish Emotion Lexicon (SEL) developed in [32]. SEL is composed of 2,036 words that are associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion or category out of this list: joy, anger, fear, sadness, surprise, and disgust. It was manually marked by 19 annotators using a scale of four values: null, low, medium and high. [30] proposed a new method to build a subjectivity and sentiment lexicon for Spanish, which they employ in performing sentence level sentiment classification, as well as seeking to enrich it through a bootstrapping process in the target language. Their work explores the WordNet structure to extract parallelism across languages, and does not make use of the embedded additional relations such as hypernymy, hyponymy, meronymy, antonymy, etc., and to a limited extent synonymy. Thus they use WordNet for cross-language expansion. In [8] an automatic method was presented for building lemma-level sentiment lexicons, which has been applied to obtain lexicons for English, Spanish and the other three official languages in Spain (Basque, Catalan and Galician). The lexicons are multi-layered, allowing applications to trade off between the amount of available words and the accuracy of the estimations. As a previous step to the lemma-level lexicons, they have built a synset-level lexicon for English similar to SentiWordNet 3.0. The resource, called ML-SentiCon, is publicly available[2]. Finally, in [23], a new Spanish semantic lexicon (iSOL) was generated by taking the Bing Liu English Lexicon. This English lexicon was translated into Spanish applying automatic machine translation techniques and manually reviewed in order to improve the final lists of sentiment words. The resource iSOL has been tested successfully in several studies, obtaining very good results in different domains [24]. In this paper, we take this lexicon as a baseline and then we improve it with a corpus-based strategy in order to increase the performance of our final polarity classification system.

The previously highlighted lexicons are formed mostly by adjectives, adverbs and some verbs. They do not pay much attention to nouns or noun phrases that can express an opinion in an objective sentence. [39] studied how to detect nouns and noun phrases that have a subjective sense inside a factual clause. The authors apply a method divided into two steps: First a set of candidates are selected following a syntactic approach, which consists of selecting those nouns and noun phrases that are in a positive or a negative context, in other words, that the noun or the noun phrase is modified by a polar word. The second step is mainly a pruning phase, because from the set of candidates those ones that have the same frequent of positive and negative contexts are erased. As will be explained later, our method also includes nouns with a polar meaning in an objective sentence, because in the domain adaptation process the morphological category of the words in the corpus is not taken into account.

Concerning the domain adaptation issue, different methods have been proposed for tackling the problem in SA. One of the primary studies in sentiment analysis is [4]. They use Structural Correspondence Learning (SCL) to find correspondences between features from source and target domains through modelling their correlations with pivot features. The proposed approach was successfully tested on review data from 4 domains (DVDs, books, kitchen appliances and electronics). Following the same idea, [28] present the Spectral Feature Alignment (SFA) that uses spectral clustering to align domain-specific and domain-independent words into a set of feature-clusters. The results obtained surpass those of SCL. [17] describe two distinct needs. On the one hand, instance adaptation takes into account the change of instance probability, e.g., the change of vocabulary or the change of words frequency from one domain to another. On the other

[2] `http://timm.ujaen.es/recursos/ml-senticon/`

hand, labelling adaptation models the changes of the labelling function, since one feature that is positive in the source domain may express the opposite meaning in the target domain. Recently, [16] propose a strategy based on a semi supervised methodology to automatically construct sentiment lexicons adapted to a specific domain. This method requires some sentiment seeds to trigger the propagation learning process. The approach get a slight improvement over the car and hotel domains although the whole process is quite complex.Demiroz et al. [10] propose a method to adapt a domain-independent sentiment linguistic resource, like SWN, to a specific domain. The assessment is carried out with a corpus of English hotel reviews downloaded from TripAdvisor [35]. The key of the domain adaptation of SWN is not difficult. The method consists of updating the polarity value (positive/negative) of a word in SWN if the word is more frequent in a different class (positive/negative) of its polarity value in SWN. [38] propose a new method called Feature Ensemble plus Sample Selection (SS-FE) that combines labelling adaptation and instance adaptation in order to improve the final system that uses these features independently. They apply the method over a multi-domain corpus extracted from Amazon and the results show a slight improvement on the baseline. In [9] a method to build a domain-dependent polarity classification system is proposed. The domain selected by the authors is hotel reviews. Each review is represented by a set of domain-independent features and a set of domain-dependent ones. The domain-independent features are extracted from SWN. To build the set of domain-dependent features the authors propose taking the lexicon built by [15] and choosing those positive/negative words that occur in a significant number of positive/negative reviews of the training corpus used for the experimentation. They used a corpus of 6,000 English reviews gathered from TripAdvisor. [13] develop a polarity classification system based on the use of a list of opinion words developed by the authors. The corpus used for the evaluation is a set of hotel reviews written in Spanish and gathered from TripAdvisor. The method developed consists of counting the number of positive and negative words that appear in the text.

The method proposed in this paper is similar to the two last papers. We have taken as a base the iSOL lexicon and then we have enriched it with the most frequent words in positive/negative reviews from a Spanish corpus compiled from TripAdvisor. We have implemented an automatic method to determine the best ratio between positive and negative words in order to integrate them into the new adapted lexicon. In addition, we have tested our method over a different corpus of Spanish hotel reviews composed of more than $32,000$ opinions.

3. Resources

In this section we describe the Spanish linguistic resources used for the experiments. Firstly, the iSOL lexicon and the COAH corpus are briefly introduced. We only comment on the main features because these resources have been used and described in previous work. However, the SHoRe corpus used in the experiments to prove our approach is described in more detail because it is presented for the first time. All these resources have been generated by the authors and they are freely available for research purposes.

3.1. iSOL Lexicon

The iSOL resource was generated from the English Lexicon of Bing Liu [15], translated into Spanish automatically, obtaining the resource SOL (Spanish Opinion Lexicon). Subsequently, the list was reviewed manually. The final list of opinion words was called iSOL (improved SOL). The iSOL is a generic lexicon that consists of 8,135 polarized words, 2,509 positive words and

5,626 negative words. This resource has been successfully evaluated in [23] using the corpus MuchoCine [7]. The results showed that the use of the refined list of polarized words can be a good strategy for unsupervised polarity classification. The iSOL is freely available for research purposes[3].

3.2. COAH Corpus

The Corpus of Opinions of Andalusian Hotels (COAH) has been generated recently using reviews from the TripAdvisor site. The corpus is described in [25] and a preliminary experiment was presented. The collection contains 1,816 reviews which were written by non-professional reviewers. The texts for hotel reviews may not be grammatically correct, or they can include spelling mistakes or informal expressions. The corpus contains only reviews from Andalusian Hotels: ten hotels for each province of Andalusia (Almería, Cádiz, Córdoba, Granada, Jaén, Huelva, Málaga and Sevilla). So the corpus contains reviews for 80 hotels with an average of 23 reviews per hotel. In Table 1 some interesting features of the COAH are shown.

#Opinions	1,816
#Hotels	80
#Words	264,303
#Sentences	9,952
#Adjectives	17,800
#Adverbs	15,219
#Verbs	38,590
#Nouns	53,640
Mean of opinions per hotel	22.7
Mean of words per sentence	26.55
Mean of words per opinion	145.54
Mean of adjectives per opinion	9.80
Mean of adverbs per opinion	8.38
Mean of verbs per opinion	21.25
Mean of names per opinion	29.54

Table 1: Some features of the corpus COAH.

3.3. SHoRe Corpus

The SHoRe (Spanish Hotel Reviews) Corpus is an enormous annotated dataset generated from the TripAdvisor website. SHoRe is composed of information about hotels in Canary Islands and Balearic Islands (Spain) and user reviews of those hotels. These users are not professional writers and therefore the opinions may have misspellings or grammatical mistakes. The collection contains more than 324K reviews from more than 3K hotels. In Table 2 we show more details of the corpus.

The compilation of the dataset is driven by the hotels, so SHoRe contains all the information and opinions in TripAdvisor about the crawled hotels, until January 2013. Since the crawling was centered on the hotels, we do not have all the opinions written by the users appearing in the corpus; in other words, we have hotel profiles instead of user profiles.

[3] `http://sinai.ujaen.es/isol/`

Island	#Hotels	#Opinions	Opinions per Hotel
El Hierro	13	110	8.46
Fuerteventura	181	33,517	185.18
La Gomera	51	849	16.65
Gran Canaria	438	39,657	90.54
La Palma	96	719	7.49
Lanzarote	215	17,627	81.99
Tenerife	400	73,342	183.36
Formentera	28	1,832	65.43
Ibiza	332	33,066	99.60
Mallorca	1,091	104,834	96.09
Menorca	163	18,548	113.79
Total	3,008	324,101	107.75

Table 2: Hotels and opinions in corpus SHORE per island

Specifically, the structured data contained in SHoRe for each hotel consists of: name of the hotel, category (in the range of 0-5 stars), location (the island where the hotel is located) and the average of the scores provided by the users. About the textual opinions, we have gathered the following information: the user who wrote the opinion, the origin of the user, the profile (whether the user has traveled "solo", i.e. alone, with friends or with family), the textual opinion itself, and a set of detailed scores given by the users regarding six specific features: location, service, comfort, cleanliness, rooms and quality of the hotel. TripAdvisor allows their users to assign a numerical value to each of these features, computing the overall score of a hotel as the average of the feature scores. This information is publicly available through the web interface of TripAdvisor.

Although the size of the corpus is remarkable, we have detected some minor problems in some opinions that prevent us from using all of them for the experiments in this study. On the one hand there are a number of reviews with an absence of textual opinion but with numerical scores. Since we intend to build lexicons, these reviews are not included in the experiments. On the other hand, TripAdvisor detects the origin of the visitor and it automatically translates the textual opinions to his native language. Since we have compiled the corpus in Spain, all the reviews written in other languages different from Spanish have suffered an automatic translation process, decreasing the quality of those texts. For this reason we have included in our experiments only the opinions written by Spanish users, ensuring that the texts have not suffered any automatic preprocessing. In Table 3 we show some metrics of the subset of SHoRe used for the experiments.

4. Generating eSOLHotel Lexicons

The iSOL is a general purpose lexicon. However, in real applications the semantic orientation of a word, phrase, or domain document depends on the specific domain. Among the approaches followed for the compilation of a set of words of opinion, the most appropriate method for semantically dependent domain is known as the corpus-based approach [20].

The method proposed in this study follows the same scheme proposed in [12]. We take as input the lexicon to be enriched and a corpus labeled with the polarity of the texts that will be taken as a reference, whose terms will be added to the lexicon when they satisfy certain

#Opinions	32,920
#Hotels	1,888
#Words	5,539,070
#Sentences	364,348
#Adjectives	294,799
#Adverbs	306,894
#Verbs	570,624
#Nouns	795,139
Mean of opinions per hotel	17.43
Mean of words per sentence	17.66
Mean of words per opinion	195.54
Mean of adjectives per opinion	8.95
Mean of adverbs per opinion	9.32
Mean of verbs per opinion	17.33
Mean of nouns per opinion	24.15

Table 3: Some features of the corpus SHoRe

conditions. In our case, the lexicon to be enriched is the domain independent iSOL lexicon and the reference corpus is COAH. The basic method works as follows: a word, w, is added to the new domain dependent lexicon, eSOLHotel, if it appears in COAH in a number of positive or negative labeled opinions. Therefore, we have implemented an automatic method of determining the best ratio between the occurrences of the words in positive and in negative reviews in order to integrate them into the new adapted lexicons. The groups of words are selected using the following equation:

$$\text{list(word)} = \begin{cases} positive \quad if(f^- = 0 \wedge f^+ \geq n) \vee \left(\dfrac{f^+}{f^-} \geq n\right) \\ negative \quad if(f^+ = 0 \wedge f^- \geq n) \vee \left(\dfrac{f^-}{f^+} \geq n\right) \end{cases} \tag{1}$$

$$n = 2, 3, 4, 5, 6, 7$$

where f^- and f^+ correspond to the absolute frequency of the occurrences of a given word in negative and positive reviews, respectively. In the generation of the lexicon, the different groups of words that satisfy the equation are added to iSOL. These new lists have been noted as $eSOLHotel^n$ (enriched SOL adapted to domain Hotel), where $n = 2, 3, 4, 5, 6, 7$ is the ratio between the amount of positive and negative words. The lower the ratio is, the greater the number of words found. In Table 4, the number of added words in each lexicon is shown.

Table 5 shows a set of words that have been added to all the lexicons. It is worth pointing out that these words could have positive or negative orientation in the hotel domain. Also, the fact must be highlighted that some of the words added are nouns that a have a subjective interpretation in the hotel domain. The presence of those words in the lexicon makes them triggers of polarity for a polarity classifier, so that the sentiment of even objective sentences can be extracted. For example, when we mention “Silencio” (silence) in the hotel domain it usually will be to comment on something positive (read Figure 1). However, if we are in the film domains, the same word could appear in positive, negative or neutral reviews. The automatic method does not imply that

Lexicon	Positive Words	Negative Words	Total Words
$eSOLHotel^2$	$2,058$	$1,793$	$3,851$
$eSOLHotel^3$	$1,097$	770	$1,867$
$eSOLHotel^4$	675	421	$1,096$
$eSOLHotel^5$	471	246	717
$eSOLHotel^6$	316	156	472
$eSOLHotel^7$	230	109	339

Table 4: Number of added words in the $eSOLHotel^n$ lexicons

> Tuve la suerte de recibir el regalo de permanecer en la suite del Cala Grande en Las Negras. La habitación era increíble y la terraza con spa propio hicieron el resto. Fue una estancia para repetir, una estancia para relajarse, para disfrutar del silencio y de las estrellas.
>
> `I had the fortune of receiving the gift of staying in the Cala Grande's suite room at Las Negras. The room was incredible and the roof terrace with private spa did the rest. It was a stay to repeat, a stay for relaxing, for enjoying the silence and the stars.`

Figure 1: Excerpt of the review no_1_20.txt from the Washing machines domain

the words added to the lexicons contain subjectivity information, but it assures that those words appear frequently in positive or negative opinions.

Positive Words	Negative Words
Desconectar (break away)	Pelos (hair)
Decoradas (decorated)	Dinero (money)
Tapas (tapas)	Muelles (bedsprings)
Repetir (to repeat)	Moho (mould)
Silencio (silence)	Sábana (bedsheet)
Acierto (success)	Limpiado (cleaned)
Azotea (roof terrace)	Moqueta (carpet)
Disposición (arrangement)	Pagado (paid)

Table 5: Some positive and negative words added to create eSOLHotel lexicons.

5. Experimental Framework and Results

In order to evaluate the performance of the different lexicons generated we have used the ShoRe Corpus. As we show in 3, the subset of SHoRe Corpus used for our experiments is composed of $32,920$ reviews. Before carrying out the experiments we performed a pre-processing step on the subset of SHoRe Corpus in order to apply the same criteria followed during the generation of the iSOL and eSOLHotel lists. For example, for both summary and body fields we had to change capital letters for non-capital letters, accented letters for non-accented letters, and all special characters had to be deleted from the opinions.

For the experiments we used the traditional measures employed in test classification: precision (P), recall (R), F_1 and Accuracy (Acc):

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP (True Positives) are those assessments where the system and a human expert agree on a label, FP (False Positives) are those positive labels assigned by the system that do not agree with the expert assignment and FN (False Negatives) are those negative labels that the system assigns that do not agree with the human expert. F_1 is a measure that combines both precision and recall, calculating the proportion of true results (both true positives and true negatives). For a feasible comparison, we summarize the F_1 scores over the different categories (positive and negative) using the macro-averages[4] of F_1 scores:

$$Macro - F_1 = \frac{2 \cdot Macro - Precision \cdot Macro - Recall}{Macro - Precision + Macro - Recall}$$

where $Macro - Recall$ and $Macro - Precision$ are obtained as follows:

$$Macro - Recall = \frac{\sum_{j=1}^{c} r_j}{c}$$

$$Macro - Precision = \frac{\sum_{j=1}^{c} p_j}{c}$$

where r is the recall value, p is the precision value, and c is the number of classes ($c = 2$).

Finally, in order to calculate the polarity (*pol*) of a review (*rev*) with each lexicon, we take into account the total number of positive words (#*positive*) and the total number of negative words (#*negative*) within the review, according to the following strategy:

$$pol(rev) = 1 \leftrightarrow \#positive > \#negative$$

$$pol(rev) = -1 \leftrightarrow \#positive \leq \#negative$$

Table 6 shows the results obtained over the part of the SHoRe corpus concerning Spanish users by using the iSOL (domain independent) and eSOLHoteln lexicons (adapted to the domain of hotels).

As we can see in Table 6, the results obtained are very promising. All the lexicons including words of the specific domain surpass the performance of the baseline iSOL. In fact, iSOL is a very good general purpose lexicon for Spanish, as demonstrated in [23]. However, the results assess the intuition that adding knowledge from a domain-dependent source like SHoRe improves

[4]Macro-Precision (Macro-P); Macro-Recall (Macro-R)

	Macro-P	Macro-R	Macro-F1	Acc	%Impr. Macro-F1	%Impr. Acc.
iSOL	78.96%	77.64%	78.30%	86.68%	-	-
$eSOLHotel^2$	84.28%	75.98%	79.92%	88.41%	2.07%	2.00%
$eSOLHotel^3$	84.56%	75.71%	79.90%	88.43%	2.04%	2.02%
$eSOLHotel^4$	84.23%	79.40%	81.75%	89.18%	4.41%	2.88%
$eSOLHotel^5$	83.72%	78.40%	80.98%	88.78%	3.42%	2.42%
$eSOLHotel^6$	82.50%	78.18%	80.28%	88.30%	2.53%	1.87%
$eSOLHotel^7$	82.08%	77.51%	79.73%	88.01%	1.83%	1.53%

Table 6: Results obtained using the different lexicons

the system, even when the classification algorithm being applied is as simple as counting the occurrences of positive and negative words from the lexicon.

Specifically we obtain the best result with $eSOLHotel^4$, which is the lexicon resulting from the inclusion of positive or negative words when the proportion of their appearances in positive or negative reviews is greater than 4. In this case we have included a little more than $1,000$ words in the iSOL baseline (see Table 4). The improvement over the general purpose iSOL is 3.5 and 2.5 points for F1 and Accuracy, respectively. However, it is interesting to note that all the lexicons that apply the method for adaptation have achieved better accuracy than the iSOL. It is worth mentioning that iSOL does not contain nouns that could express polarity in objective clauses, so the fact that all the new lists have better results than iSOL could be due to the inclusion of not just polarity words.

Table 6 shows that $eSOLHotel^4$ is the cut-off point of the different list assessed. From $eSOLHotel^2$ to $eSOLHotel^4$ the system improves the results gradually, so it is a sign that the tightening of the requirements for including words in the lexicon benefits the classification. But from $eSOLHotel^4$ to $eSOLHotel^7$ the results are poorer. So a first conclusion is that the number of words to include in the lexicon has a limit. A first thought is that with the appending of additional words to the lexicon, the results will be better or at least the recall will be better because with more words the system will cover a wider domain-specific vocabulary. But the results show that this assertion is false. With our experiments we have demonstrated that there exists a limit from which the new words that we are appending are not valid words. In plain English, an excessive number of new words introduces noise into the system.

6. Conclusions and future work

The main goal of this work is to generate automatically lexicons for a specific domain in Spanish for the polarity classification task. To this end, an automatic method is proposed consisting of adding to the independent domain lexicon iSOL those words that appear in the labeled corpus COAH. In order to be added to the lexicon, the words must be found in positive reviews n-times more than in negative ones, or vice versa. We have tested the approach over the SHoRe corpus. Our hypothesis has two main keys:

- The integration of semantic resources helps the process of polarity classification.
- Human beings tend to use the same words (verbs, nouns, adverbs and adjectives) when describing positive or negative characteristics of a topic. Sometimes, these words have no opinion information.

In the previous sections we have described the experiments that assess the method proposed. As baseline, the COAH corpus was selected to seek six different bags of words that are more frequently used in one class of reviews (positive/negative) than the other class (negative/positive). These bags of words have being added to the general purpose lexicon iSOL. The inclusion of these words in the general purpose lexicon improves the polarity classification. This polarity classification achieves the best results when the ratio between positive and negative words is four.

As future work we plan to apply the proposed method to other languages and domains. We also plan to continue researching methodologies for the classification of implicit opinions, because a great amount of users express their intention through objective clauses. The treatment of negation is another challenge in SA, so we are going to make efforts to include a negation scope identification module in our method.

Acknowledgements

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. The project AORESCU (P11-TIC-7684 MO) from the regional government of Junta de Andalucía partially supports this manuscript, and the CEATIC-2013-01 project from the University of Jaén.

References

[1] R. Agerri, A. García-Serrano, Q-wordnet: Extracting polarity from wordnet senses, in: N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (eds.), Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association, Valletta, Malta; ELRA, 2010, 19-21.

[2] A. Aue, M. Gamon, Customizing sentiment classifiers to new domains: A case study, in: Recent Advances in Natural Language Processing - RANLP, 2005.
URL `http://research.microsoft.com/char126anthaue/new_domain_sentiment.pdf`

[3] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (eds.), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, 2010, pp. 2200–2204.

[4] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 440–447.
URL `http://www.aclweb.org/anthology/P/P07/P07-1056`

[5] E. Cambria, B. Schuller, B. Liu, H. Wang, C. Havasi, Statistical approaches to concept-level sentiment analysis, Intelligent Systems, IEEE 28 (3) (2013) 6–9.

[6] E. Cambria, R. Speer, C. Havasi, A. Hussain, Senticnet: A publicly available semantic resource for opinion mining.
URL `http://www.aaai.org/ocs/index.php/FSS/FSS10/paper/view/2216`

[7] F. Cruz, J. A. Troyano, F. Enriquez, J. Ortega, Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español, Procesamiento del Lenguaje Natural 41 (2008) 73–80.
URL `http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/2551`

[8] F. L. Cruz, J. A. Troyano, B. Pontes, F. J. Ortega, Ml-senticon: Un lexicón multilingüe de polaridades semnticas a nivel de lemas, Procesamiento del Lenguaje Natural 53 (0) (2014) 113–120.
URL `http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5041`

[9] R. Dehkharghani, B. Yanikoglu, D. Tapucu, Y. Saygin, Adaptation and use of subjectivity lexicons for domain dependent sentiment classification, in: Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on, 2012, pp. 669–673.

[10] G. Demiroz, B. Yanikoglu, D. Tapucu, Y. Saygin, Learning domain-specific polarity lexicons, in: Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on, 2012, pp. 674–679.

[11] X. Ding, B. Liu, The utility of linguistic rules in opinion mining, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '07, ACM, New York, NY, USA, 2007, pp. 811–812.
URL http://doi.acm.org/10.1145/1277741.1277921

[12] W. Du, S. Tan, X. Cheng, X. Yun, Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon, in: Proceedings of the third ACM international conference on Web search and data mining, WSDM '10, ACM, New York, NY, USA, 2010, pp. 111–120.
URL http://doi.acm.org/10.1145/1718487.1718502

[13] A. García, S. Gaines, M. T. Linaza, A lexicon based sentiment analysis retrieval system for tourism domain, in: ENTER 2012 Idea Exchange. 19th International Conference on Information Technology and Travel & Tourism ENTER 2012'eTourism Present and Future Services and Applications', Helsingborg, Sweden, 24-27 January 2012., vol. 10, Texas A & M University Press, 2012, pp. 35–38.

[14] V. Hatzivassiloglou, K. R. McKeown, Predicting the semantic orientation of adjectives, in: Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, EACL '97, Association for Computational Linguistics, Stroudsburg, PA, USA, 1997, pp. 174–181.
URL http://dx.doi.org/10.3115/979617.979640

[15] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04, ACM, New York, NY, USA, 2004, pp. 168–177.
URL http://doi.acm.org/10.1145/1014052.1014073

[16] S. Huang, Z. Niu, C. Shi, Automatic construction of domain-specific sentiment lexicon based on constrained label propagation, Knowledge-Based Systems 56 (0) (2014) 191 – 200.
URL http://www.sciencedirect.com/science/article/pii/S0950705113003596

[17] J. Jiang, C. Zhai, Instance weighting for domain adaptation in nlp, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 264–271.
URL http://www.aclweb.org/anthology/P/P07/P07-1034

[18] J. Kamps, M. Marx, R. J. Mokken, M. de Rijke, Using wordnet to measure semantic orientations of adjectives, in: Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal, 2004.
URL http://www.lrec-conf.org/proceedings/lrec2004/pdf/734.pdf

[19] S. D. Kamvar, J. Harris, We feel fine and searching the emotional web, in: Proceedings of the fourth ACM international conference on Web search and data mining, ACM, 2011, pp. 117–126.

[20] H. Kanayama, T. Nasukawa, Fully automatic lexicon expansion for domain-oriented sentiment analysis, in: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP '06, Association for Computational Linguistics, Stroudsburg, PA, USA, 2006, pp. 355–363.
URL http://portal.acm.org/citation.cfm?id=1610075.1610125

[21] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5 (1) (2012) 1–167.
URL http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016

[22] G. A. Miller, Wordnet: A lexical database for english, Commun. ACM 38 (11) (1995) 39–41.
URL http://doi.acm.org/10.1145/219717.219748

[23] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia, J. M. Perea-Ortega, Semantic orientation for polarity classification in spanish reviews, Expert Systems with Applications 40 (18) (2013) 7250–7257.
URL http://www.sciencedirect.com/science/article/pii/S0957417413004752

[24] M. D. Molina-González, E. Martínez Cámara, M. T. Martín Valdivia, L. A. Ureña López, A spanish semantic orientation approach to domain adaptation for polarity classification, Infomration Processing & Management In press.

[25] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, Cross-domain sentiment analysis using spanish opinionated words, in: E. Métais, M. Roche, M. Teisseire (eds.), Natural Language Processing and Information Systems, vol. 8455 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 214–219.
URL http://dx.doi.org/10.1007/978-3-319-07983-7_28

[26] A. Montejo-Ráez, WeFeelFine as resource for unsupervised polarity classification, Procesamiento del Lenguaje Natural 50.
URL http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4656

[27] S. Owsley, S. Sood, K. J. Hammond, Domain specific affective classification of documents, in: AAI Spring Symposium: Computational Approaches to Analyzing Weblogs, American Association for Artificial Intelligence, 2006, pp. 181–183.

[28] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, Z. Chen, Cross-domain sentiment classification via spectral feature alignment, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 751–760.
URL http://doi.acm.org/10.1145/1772690.1772767

[29] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: Sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10 of EMNLP '02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 79–86.
URL http://dx.doi.org/10.3115/1118693.1118704

[30] V. Perez-Rosas, C. Banea, R. Mihalcea, Learning sentiment lexicons in spanish, in: N. C. C. Chair), K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey, 2012.

[31] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: Proceedings of the 2003 conference on Empirical methods in natural language processing, EMNLP '03, Association for Computational Linguistics, Stroudsburg, PA, USA, 2003, pp. 105–112.
URL http://dx.doi.org/10.3115/1119355.1119369

[32] G. Sidorov, S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Veláquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño, J. Gordon, Empirical study of machine learning based approach for opinion mining in tweets, in: I. Batyrshin, M. González Mendoza (eds.), Advances in Artificial Intelligence, vol. 7629 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 1–14.
URL http://dx.doi.org/10.1007/978-3-642-37807-2_1

[33] P. J. Stone, D. C. Dunphy, M. S. Smith, The General Inquirer: A Computer Approach to Content Analysis., M.I.T. press, 1966.

[34] P. D. Turney, Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 417–424.
URL http://dx.doi.org/10.3115/1073083.1073153

[35] H. Wang, Y. Lu, C. Zhai, Latent aspect rating analysis on review text data: A rating regression approach, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10, ACM, New York, NY, USA, 2010, pp. 783–792.
URL http://doi.acm.org/10.1145/1835804.1835903

[36] A. Weichselbraun, S. Gindl, A. Scharl, Enriching semantic knowledge bases for opinion mining in big data applications, Knowledge-Based Systems 69 (0) (2014) 78 – 85.
URL http://www.sciencedirect.com/science/article/pii/S0950705114001695

[37] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, Language Resources and Evaluation 39 (2-3) (2005) 165–210.

[38] R. Xia, C. Zong, X. Hu, E. Cambria, Feature ensemble plus sample selection: Domain adaptation for sentiment classification, Intelligent Systems, IEEE 28 (3) (2013) 10–18.

[39] L. Zhang, B. Liu, Identifying noun product features that imply opinions, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2, HLT '11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 575–580.
URL http://dl.acm.org/citation.cfm?id=2002736.2002849

3. Generación de nuevo corpus en español

Las publicaciones en revista asociadas a esta parte son:

f) Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T. y Ureña-López, L.A., (2014). Cross-domain Sentiment Analysis using Spanish Opinionated Words. *The 19th international Conference on Application of Natural Language to Information Systems,* NLDB'14

   Estado: Publicado online
   ISBN: 978-3-319-07983-7
   DOI: 10.1007/978-3-319-07983-7_28

Cross-Domain Sentiment Analysis Using Spanish Opinionated Words*

M. Dolores Molina-González, Eugenio Martínez-Cámara,
M. Teresa Martín-Valdivia, and L. Alfonso Ureña-López

Computer Science Department,
University of Jaén,
Campus Las Lagunillas, 23071, Jaén, Spain
{mdmolina,emcamara,maite,laurena}@ujaen.es

Abstract. A common issue of most of NLP tasks is the lack of linguistic resources in languages different from English. In this paper is described a new corpus for Sentiment Analysis composed by hotel reviews written in Spanish. We use the corpus to carry out a set of experiments for unsupervised polarity detection using different lexicons. But, in addition, we want to check the adaptability to a domain for the lists of opinionated words. The obtained results are very promising and encourage us to continue investigating in this line.

Keywords: Sentiment Polarity Detection, Spanish Opinion Mining, Spanish hotel review corpus, domain adaptation.

1 Introduction

Sentiment Analysis (SA), also known as Opinion Mining (OM) is a challenging task that combines Natural Language Processing (NLP) and Text Mining (TM). Polarity classification is one of the most studied tasks of OM that is focused on determining which is the overall sentiment-orientation of the opinions contained within a given document. The document is supposed to contain subjective information such as product reviews or opinionated posts in blogs. In this paper, we focus on semantic orientation for polarity classification in reviews over a tourism domain. We want to analysis the goodness of some lexicons and the adaptability to a specific domain. Specifically we have chosen the tourism domain to carry out our experimental study. We have generated a corpus with hotel reviews from the Tripadvisor website[1]. The results obtained are very promising being even comparable with machine learning approach.

On the other hand, although SA is a very recent area, the number of papers, applications and resources dedicated to this task is impressive. However, most

* This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. The project AORESCU (TIC - 07684) from the regional government of Junta de Andaluca partially supports this manuscript.

[1] http://www.tripadvisor.es

E. Métais, M. Roche, and M. Teisseire (Eds.): NLDB 2014, LNCS 8455, pp. 214–219, 2014.

of works related to opinion mining only deal with English texts. In this paper we focus on Spanish SA. Our main interest is to check the behaviour of different Spanish lexicons generated using as a base the SOL (Spanish Opinion Lexicon) resource [1]. This lexicon was manually checked obtaining the iSOL (improved SOL) resource.

The rest of the paper is organised as follows: The next section presents related studies that apply a semantic orientation approach focusing mainly on the use of lexical resources. Section 3 introduces the main resources used and describes the construction process and the features of a corpus of hotel reviews (COAH) that we used in our experiments. Section 4 presents the results obtained in the experiments we performed. Finally, the conclusions and future work are presented.

2 Related Works

Several studies have been published related to extract the opinion from the users reviews posted in touristic web sites. In [2] a supervised polarity classification system is described. The authors compiled a corpus from the travel column of Yahoo![2] , which is composed by 1191 reviews. The authors study the performance of two machine learning classifiers, Naïve Bayes and SVM, and the N-gram based character language model for SA. In this case SVM was the algorithm that reached the best results. Chinese hotel reviews have also been utilised by the SA research community. In [3] the authors classify Chinese hotel reviews. As in the former work, the authors develop a supervised classifier based on the use of the algorithm SVM.

A considerable number of papers have also been published proposing methodologies to adapt resources to the hotel domain. In [4] is proposed a method to build a domain-dependent polarity classification system. The domain selected by the authors is hotel reviews. Each review is represented by a set of domain-independent features and a set of domain-dependent ones. The domain-independent features are extracted from SentiWordNet. To build the set of domain-dependent features the authors propose to take the lexicon built by Hu and Liu [5] and choose those positive/negative words that occur in a significant number of positive/negative reviews of the training corpus used for the experimentation. In [6] is proposed a method to adapt a domain-independent sentiment linguistic resource, like SentiWordNet, to a specific domain. The assessment is done with a corpus of English hotel reviews downloaded from Tripadvisor [7].

Our proposal for the domain adaptation problem is the use of specific lists of opinion words per each domain. So, in the following sections a corpus of Spanish hotel reviews and an opinion list for tourism domain are described. Also, a set of experiments are shown with the aim of illustrating the value of these new linguistic resources.

[2] `http://travel.yahoo.com/`

3 Resources: Corpora and Word Lists

The development of new linguistic resources is very important to make progress in solving the problem of cross-domain SA. Also, the need of new linguistic resources is higher in languages other than English, like Spanish. So, the main contribution of this paper is the description of a new corpus of reviews in the tourism domain, and also new experiments that certificate the goodness of two sentiment lexicons developed by us.

Firstly, a Spanish corpus of hotel reviews has been compiled. The corpus is called COAH, which means Corpus of Opinion about Andalusian Hotels. Using COAH, an unsupervised polarity classification system has been developed with the aim of assessing two domain-independent opinion lexicons and a sentiment lexicon adapted to the tourism domain. The domain-independent lexicons are SOL and iSOL lexicons [1], and the tourism lexicon is eSOLHotel.

3.1 Corpus: COAH

For our experiments we have created the Corpus of Opinion about Andalusian Hotels COAH from the TripAdvisor site. The collection contains 1,835 reviews not written by professional writers, but rather by the web users. This may appear anecdotal, though it increases the difficulty of the task, because the texts may not be grammatically correct, or they can include spelling mistakes or informal expressions. We have selected only Andalusian hotels, ten hotels per each province of Andalusian (Almería, Cádiz, Córdoba, Granada, Jaén, Huelva, Málaga and Sevilla). We have selected ten hotels, five of them with higher rating and the other five with worse rating. All the hotels must have at least twenty opinions in the latter years written in Spanish. Finally, we have obtained 1,835 reviews.

The opinions are rated on a scale from 1 to 5. A rank of 1 means that the hotel is very bad, and 5 means very good. Rated 3 hotels can be categorised as "neutral" which means the user consider the hotel is neither bad nor good. Table 1 shows the number of reviews per rating. In our experiments we discarded the neutral reviews. In this way, opinions rated with 3 were not considered, the opinions with ratings of 1 or 2 were considered as negative reviews (519 in total) and those with ratings of 4 or 5 were considered as positives (1,025 in total).

Table 1. Rating distribution

Rating	**1**	**2**	**3**	**4**	**5**	**Total**
#Reviews	316	203	291	493	532	1835

In Table 2 is shown some interesting features of the corpus.

Table 2. COAH statistics

#Reviews	1,835
#Hotels	80
Mean of reviews per hotel	22.93
#Tokens	266,410
#Sentences	10,042
#Adjectives	17,910
#Adverbs	15,357
#Verbs	38,889
#Names	53,924
Mean of tokens per sentence	26.52
Mean of tokens per review	145.18
Mean of adjectives per review	9.76
Mean of adverbs per review	8.36
Mean of verbs per review	21.19
Mean of names per review	29.38

3.2 iSOL and eSOLHotel Lexicons

The iSOL resource was generated from the BLEL lexicon [5] by automatically translating it into Spanish and obtaining the SOL (Spanish Opinion Lexicon) resource. Then this resource was manually reviewed in order to improve the final list of words obtaining iSOL (improved SOL). The iSOL is composed of 2,509 positive and 5,626 negative words, thus in total the Spanish lexicon has 8,135 opinion words.

On the other hand, the eSOLHotel List is a resource generated from the iSOL lexicon by using domain knowledge. In this respect, we chose the Spanish part of the comparable SFU Reviews Corpus [8]. The SFU Reviews Corpus is composed of reviews of products in English and Spanish. The Spanish reviews are divided into eight categories: books, cars, computers, washing machine, hotels, movies, music and phones. In order to generate the enriched Spanish Lexicon for hotels (eSOLHotel), we search the most frequent words into the hotel category in the SFU Spanish Review Corpus.

4 Experiments and Results

In order to evaluate the experiments, we used the traditional measures employed in text classification: precision (P), recall (R), F1 and Accuracy (Acc.). The polarity of a review is calculated by taking into account the total number of positive words (#positive) and the total number of negative words (#negative) within the review. Table 3 shows the results obtained by using the three lists of opinionated words over the COAH corpus.

Table 3. Results obtained by using the three lists of opinionated words

	Macro-P	Macro-R	Macro-F1	Acc.
SOL	84.70%	75.22%	79.68%	82.24%
iSOL	91.61%	83.25%	87.23%	88.46%
eSOLHotel	91.59%	84.31%	87.80%	89.05%

4.1 Comparison with Other Related Work

In the literature we can find an interesting resource called the Spanish Emotion Lexicon (SEL) provided by Sidorov [9]. This resource is freely available for research purposes[3]. SEL is composed of 2,036 words that are associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion or category: *joy*, *anger*, *fear*, *sadness*, *surprise*, and *disgust*. In order to establish a feasible comparison by using the SEL resource for binary classification of COAH, we considered the joy and surprise categories as positive and the others as negative.

As is widely known, supervised learning overcomes unsupervised learning, so an unsupervised system is better when its results will be closer to the ones reached by a supervised system. Thus, a supervised experiment has been carried out with the aim of comparing the lexicon-based classifier described previously. The supervised system used is the same that is described in [10], i.e. the SVM algorithm has been used as classifier, and the reviews have been represented as a set of vectors of tokens weighted by TF-IDF. The results achived with ML and SEL are shown in Table 4.

Table 4. Comparison for binary classification of COAH by using ML (SVM algorithm) and eSOLHotel

	Macro-P	Macro-R	Macro-F1	Acc.
SEL	81.72%	69.00%	74.82%	78.16%
SVM	95.22%	93.14%	94.17%	94.82%
eSOLHotel	91.59%	84.31%	87.80%	89.05%

The high results in Table 3 show the validity of the two lexicons presented in this paper, iSOL and eSOLHotel, for polarity classification of Spanish reviews in hotel domain. Besides, in the same table we can observe that accuracy is better with eSOLHotel, lexicon where we are implement the domain knowledge.

Therefore, we consider that the lexicons developed and the new corpus COAH, which are freely available, are valuable resources for the Spanish SA research community.

Regarding Accuracy and F1 of Table 3 and Table 4, the percentage difference between eSOLHotel and SVM is only 6.27% and 7% respectively. The reduced

[3] `http://www.cic.ipn.mx/~sidorov/#SEL`

difference shows the goodness of eSOLHotel, which is also certificated by the good results reached with the hotel reviews section of the corpus SFU. This good performance also shows that eSOLHotel covers correctly vocabulary related to tourism specially the vocabulary related to hotels.

5 Further Work

Currently we are working on the development of several Spanish lexicons for domain-dependent SA following the method proposed here, i.e. selecting the words with a higher frequency in a corpus. Also, we want to deep in the evaluation of the domain-dependent lists of opinion words using larger sets of reviews.

References

1. Molina-González, M.D., Martínez-Cámara, E., Martín-Valdivia, M.T., Perea-Ortega, J.M.: Semantic orientation for polarity classification in spanish reviews. Expert Systems with Applications 40(18), 7250–7257 (2013)
2. Ye, Q., Zhang, Z., Law, R.: Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst. Appl. 36(3), 6527–6535 (2009)
3. Shi, H.X., Li, X.J.: A sentiment analysis model for hotel reviews based on supervised learning. In: 2011 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 3, pp. 950–954 (July 2011)
4. Dehkharghani, R., Yanikoglu, B., Tapucu, D., Saygin, Y.: Adaptation and use of subjectivity lexicons for domain dependent sentiment classification. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), pp. 669–673 (December 2012)
5. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004, pp. 168–177. ACM, New York (2004)
6. Demiroz, G., Yanikoglu, B., Tapucu, D., Saygin, Y.: Learning domain-specific polarity lexicons. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), pp. 674–679 (December 2012)
7. Wang, H., Lu, Y., Zhai, C.: Latent aspect rating analysis on review text data: A rating regression approach. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2010, pp. 783–792. ACM, New York (2010)
8. Taboada, M., Grieve, J.: Analyzing appraisal automatically. Technical report, Stanford University (March 2004)
9. Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., Gordon, J.: Empirical study of machine learning based approach for opinion mining in tweets. In: Batyrshin, I., González Mendoza, M. (eds.) MICAI 2012, Part I. LNCS, vol. 7629, pp. 1–14. Springer, Heidelberg (2013)
10. Martínez-Cámara, E., Martín-Valdivia, M.T., Perea-Ortega, J.M., Ureña-López, L.A.: Opinion classification techniques applied to a spanish corpus. Procesamiento del Lenguaje Natural 47, 163–170 (2011)

4. Meta-clasificadores para clasificación de polaridad integrando distintos recursos léxicos en español

Las publicaciones en revista asociadas a esta parte son:

g) Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., y Perea-Ortega, J. M. (2014). Integrating Spanish Lexical Resources by Meta-classifiers for polarity classification. *Journal of Information Science*

Estado: Publicado online (19 Mayo 2014)
Índice de Impacto (JCR 2013): 1,087
Área de conocimiento: Computer Science, Information Systems, Ranking 65/135
Cuartil: Q2
Citas: 0
ISSN: 0165-5515
DOI: 10.1177/0165551514535710

JIS

Article

Integrating Spanish Lexical Resources by Meta-classifiers for polarity classification

Journal of Information Science
1–19

DOI: 10.1177/0165551510000000
jis.sagepub.com
SAGE

Eugenio Martínez-Cámara
Computer Science Department, University of Jaén, Spain

M. Teresa Martín-Valdivia
Computer Science Department, University of Jaén, Spain

M. Dolores Molina-González
Computer Science Department, University of Jaén, Spain

José M. Perea-Ortega
European Commission, Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), Ispra, Italy.

Abstract
In this paper we focus on unsupervised Sentiment Analysis in Spanish. The lack of resources for languages other than English, as for example Spanish, adds more complexity to the task. However, we should take advantage of some good already existing lexical resources. We have carried out several experiments using different unsupervised approaches in order to compare the different methodologies for solving the problem of the Spanish polarity classification in a corpus of movie reviews. Among all these approaches, perhaps the newest one integrates SentiWordNet with the Multilingual Central Repository to tackle the polarity detection directly over the Spanish corpus. However, the results obtained are not as promising as we expected, and so we have carried out another group of experiments combining all the methods by using meta-classifiers. The results obtained with stacking outperform the individual experiments and encourage us to continue in this way.

Keywords
Unsupervised polarity detection, SentiWordNet (SWN), Multilingual Central Repository (MCR), Stacking algorithm, Meta-classifiers, Spanish lexical resources, Lexical resources for opinion mining

1. Introduction

Sentiment Analysis (SA), also known as Opinion Mining (OM), is an area of Natural Language Processing (NLP) that refers to the treatment of the subjective information in texts, mainly product reviews, comments on blogs or personal opinions. One of the basic tasks in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature level, i.e., whether the opinion expressed in a document, a sentence or an entity feature is positive, negative, or neutral. Many studies have investigated the polarity classification problem but most only consider documents written in English. However, nowadays more and more people express their comments, opinions or points of view in their own language, making languages like Spanish, Chinese or Arabic increasingly important in OM. For this reason it is necessary to develop systems than can extract and analyse all this information in different languages. In this work we focus on polarity detection for Spanish reviews. We are mainly concerned with linguistic resources for Spanish sentiment analysis because, in addition to the lack of resources for this language in this area, it is currently the third most used language in the Web according to the Internet World Stats[1].

Corresponding author:
Eugenio Martínez Cámara, Computer Science Department. University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain.
Email: emcamara@ujaen.es

Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

On the other hand, polarity classification has usually been tackled following two main approaches. The first one applies Machine Learning (ML) algorithms in order to train a polarity classifier using a labelled corpus [1]. This approach is also known as the supervised approach. The second one is known as Semantic Orientation (SO), or the unsupervised approach, and it integrates linguistic resources in a model in order to detect the polarity [2]. Both approaches have advantages and drawbacks. For example, ML methods require annotated corpora to train the model that are normally difficult to achieve. However, this strategy usually obtains better performances. On the contrary, the SO methodology requires a large amount of linguistic resources which generally depend on the language, although the model does not require labelled corpora for learning. Until now the results obtained with unsupervised models do not outperform the ML classifiers. However, there are several semantic resources that we believe must be analysed and integrated in order to improve these systems.

In this study we use one of these interesting semantic resources: SentiWordNet (SWN) [3]. Specifically, our proposal focuses on adapting this resource to the Spanish language in order to be applied directly over a Spanish movie review corpus. As a main novelty we make use of the Multilingual Central Repository (MCR) [4] [5] by linking each synset of SWN to their equivalent Spanish semantic words. The MCR integrates wordnets from five different languages (English, Spanish, Catalan, Basque and Galician), allowing connections from words in one language to equivalent translations in any of the other languages thanks to the automatically generated mappings among WordNet versions. To our knowledge this is the first time that MCR has been integrated with SWN in order to classify the opinion polarity in a Spanish review corpus.

According to [6] there are two main approaches in the context of multilingual SA: The first one is the corpus-based approach, where a subjectivity-annotated corpus for the target language is built through projection, and then a statistical classifier is trained on the resulting corpus (Figure 1). The second one is the lexicon-based approach, where a target-language subjectivity classifier is generated by translating an existing lexicon into another language (Figure 2).

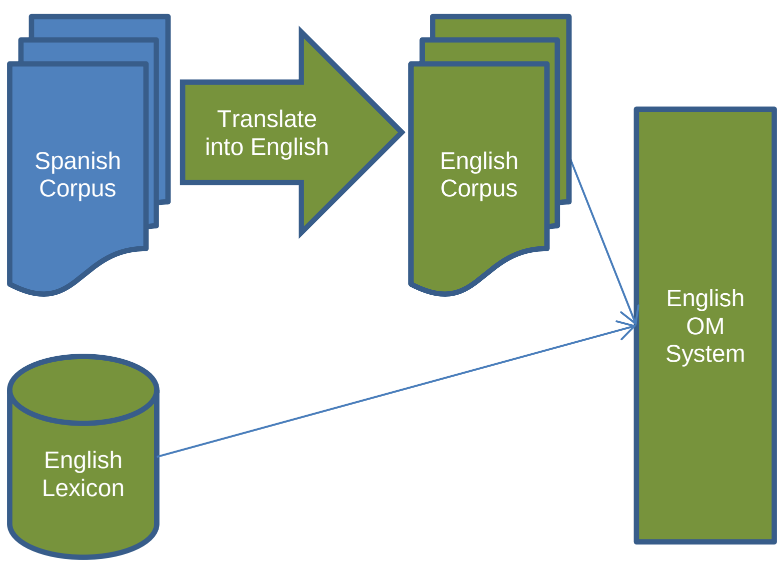


Figure 1. Corpus-based approach

In this paper we combine both approaches, using meta-classifiers in order to improve the final system. For the corpus-based approach, we translate a Spanish corpus of movie reviews called MuchoCine (MC) into English and then we apply different English resources (SWN and opinionated lists of words). For the lexicon-based approach, we use the MC corpus directly in Spanish. Therefore we have used two different semantic resources. First, we use a list of opinionated words translated into Spanish, and secondly we apply the MCR in Spanish linked with SWN in order to integrate a Spanish lexicon over the MC corpus.

Finally, we propose to take advantage of the combination of different linguistic resources and the certainty that subjectivity tends to be preserved between languages. Several combinations of classifiers were studied with the goal of improving the performance of the Spanish polarity classification. The results show that the combination of different linguistic resources and also the use of meta-classifiers enhance the performance of a polarity classification system for Spanish texts.

The rest of the article is organized as follows: The next section presents related studies that apply a semantic orientation approach focusing mainly on the use of SWN and other lexical resources. We also comment on some research which deals with languages other than English and multilingual OM. In addition, some interesting papers about meta-classifiers are also referred. Section 3 outlines the method proposed. Section 4 introduces the main resources used in our experiments. Then the individual and combined systems are described and the results obtained analysed. Finally, the conclusions and future work are presented.

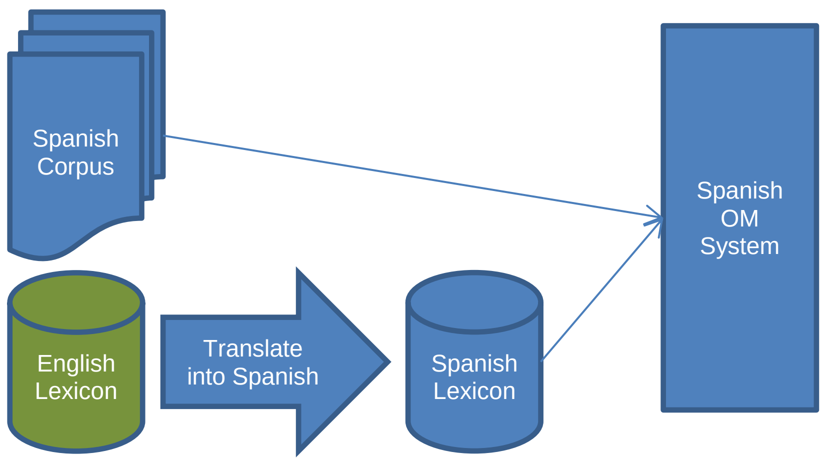


Figure 2. Lexicon-based approach

2. Related work

Research into opinion mining has experimented an exponential growing in recent years. Some recent surveys can be found in [7] and [8]. Concerning Semantic Orientation (SO) there are also several studies. In the SO approach the document is represented as a collection of words and manual rules and lexicons are applied. The sentiment of each word can be determined by different methods, for example using a list of opinionated words [9], applying web search [10], making use of annotated terms in dictionaries [11], or lexical resources such as General Inquirer [12] and WordNet [13]. Moreover, there are other studies that apply specific sentiment analysis resources like SentiSense [14], WordNet-Affect [15] or SentiWordNet [16].

2.1. Non-English Sentiment Analysis

Regarding opinion mining focused on languages other than English some studies can be highlighted. For example, Zhang et al. [17] applied Chinese sentiment analysis on two datasets. In the first one euthanasia reviews were collected from different web sites, while the second dataset was about six product categories collected from Amazon (Chinese reviews). Agić et al. [18] presented a manually annotated corpus with news on the financial market in Croatia. In [19] a corpus of movie reviews in Arabic annotated with polarity was presented and several experiments using machine learning techniques were performed. Regarding Spanish, there are also some interesting studies. For example, Banea et al. [20] proposed several approaches to cross lingual subjectivity analysis by directly applying the translations of opinion corpus in English to training an opinion classifier in Romanian and Spanish. This work showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. Brooke et al. [21] presented several experiments dealing with Spanish and English resources. They conclude that although the ML techniques can provide a good baseline performance, it is necessary to integrate language-specific knowledge and resources in order to achieve an improvement. They proposed three approaches: the first one uses Spanish resources generated manually and automatically. The second one applies ML to a Spanish corpus. The last one translates the Spanish corpus into English and then applies the SO-CAL (Semantic Orientation CALculator), a tool developed by themselves [11]. Cruz et al. [22] manually recollected the MuchoCine (MC) corpus to develop a sentiment polarity classifier based on semantic orientation. The corpus contains annotated Spanish movie reviews from the MuchoCine website. This MC corpus is used in this paper for our experimental study.

On the other hand, although SentiWordNet has been used in several studies most of them only deal with English documents. A few studies try to apply SWN to languages other than English. For example, Denecke [23] worked on German comments collected from Amazon. These reviews were translated into English using standard machine translation software. Then the translated reviews were classified as positive or negative, using three different classifiers: LingPipe, SentiWordNet with classification rule, and SWN with machine learning. Ghorbel and Jacot [24] used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determine the polarity of the reviews. Martín-Valdivia et al. [25] presented an experimental study of supervised and unsupervised approaches over a Spanish-English parallel corpus, by integrating SWN in different ways over the translated English corpus. Perea-Ortega et al. [26] carried out several experiments by combining both machine learning and semantic orientation approaches over the Opinion Corpus for Arabic (OCA) [19] and its parallel English version named EVOCA. They applied a voting system based on majority rule showing a slight improvement when both approaches were combined. In all of these examples the original opinion corpus is translated into English and then SWN is applied over the translated English text.

In this paper we study the use of SWN directly over the original corpus. In order to apply SWN over a non-English corpus it is necessary to use another resource to link the synset in English to its corresponding synset in Spanish. For this case we have integrated the Multilingual Central Repository. MCR has been applied in several studies, but for sentiment analysis we can only find one [27]. The idea emerges after several studies related to Spanish polarity detection over the MC corpus. In our first paper [25] we followed the corpus-based approach and we generated an English parallel corpus, called MCE (MuchoCine English version)[2]. The MCE corpus was built by applying automatic machine translation techniques to the Spanish MC corpus. Then we combined supervised and unsupervised approaches using meta-classifiers. First we generated two individual models using these two corpora (MC Spanish and English corpus) and applying Support Vector Machines (SVM) algorithms. Then we integrated SentiWordNet into the English corpus, generating a new unsupervised model. Finally, the three systems were combined using a meta-classifier that allows us to apply several combination algorithms such as voting system or Stacking [28].

On the other hand, our second study [29] was oriented to a lexicon-based approach dealing with the Spanish MC corpus and using the semantic orientation strategy. The paper presented a new resource for the Spanish sentiment analysis research community (iSOL, improved Spanish Opinion Lexicon). We generated the new lexicon iSOL by translating into Spanish the Bing Liu English Lexicon (BLEL) [9], and then the resource was manually revised and improved.

2.2. Using Meta-classifiers for Sentiment Analysis

As mentioned before, there are several methods for generating classifiers for polarity detection. Each one has advantages and drawbacks. Thus some researchers have tried to take advantage of the ensemble methods or meta-classifiers theory. The main idea of ensemble methodology is to combine a set of classifiers in order to obtain a composite of combine learners, with more accurate estimations that can be achieved by using a single classifier [30]. Broadly speaking, the ensemble methodology tries to learn from the errors of the base classifiers with the aim of achieving a more accurate final classifier. A wide range of methodologies of combining classifiers are described in the literature due to their potential usefulness. Several factors differentiate the various ensemble or combined methods. The main factors are:

(1) Inter-classifier relationship: Depending on whether each classifier is affected by the other ones the ensemble methods can be divided into two types: sequential and concurrent. The sequential ensemble methods are those where the final model is built in an iterative process of model generation, in which the model of *i*-th iteration depends on the previous model. An example of a sequential combined classifier is AdaBoost [31]. On the other hand, the concurrent ensemble methods are those where dependency between the models that concern the

classification process is minimal, and they are built concurrently. The most representative concurrent ensemble method is Stacking [28].

(2) The combined method. Each ensemble method has to choose between several ways to combine the output of different classifiers. Voting schemes, Bayesian combinations, distribution summation, likelihood combination or statistical methods can be used.

(3) The diversity generator. Some ensemble methods require that the classifiers concerned in the process generate a diversity output. Loosely speaking, the combined classification will be more successful when the outputs of the classifiers are more different.

(4) Ensemble size. The number of classifiers involve in the classification process is another important factor.

In [32] the authors carried out a broad experimentation. The authors built two systems; one employing a lexicon-based method and the other one based on the use of the machine learning algorithm Support Vector Machines (SVM). Although the machine learning system achieved good results, the authors wanted to enhance the overall performance. They tried to combine the two systems. Firstly the authors developed a weighted voting method, and then they built a meta-set with the output of the two base learners. The meta-classifier based on SVM outperformed the weighted voting scheme and also all the variations of the two base learners. They concluded that the combination of several classification models helps to enhance the results of a polarity classification system.

Wan [33] worked with a corpus of Chinese product reviews. The author designed a framework focused on the combination of two unsupervised classifiers. One of them was used for classifying the original Chinese products reviews and the other one for its English translated version. The two base learners consisted of counting the number of positive and negative terms. For the combination, the author assessed three ensemble methods: average, weighted average and voting scheme. As in the previous study the results demonstrated that the combination of sentiment classification models enhance the final performance of the system.

In [6] it is asserted that subjectivity tends to be preserved across languages, but in [20] it is hypothesized that subjectivity is expressed differently in various languages due to lexicalization, formal versus informal markers, etc. Thus, in [34] the authors tried to demonstrate that several parallel corpora in different languages can complement each other in polarity classification. The authors took the MPQA corpus [35] and translated it into Spanish, Arabic, French, German and Romanian. Then several individual polarity classification experiments were carried out using Naïve Bayes, and they also combined the individual classifiers with a majority vote meta-classifier. The authors concluded that more languages are better for multilingual sentiment classification as they are able to complement each other, and together they provide better classification results.

Balahur and Tuchi [36] studied the manner in which Sentiment Analysis can be performed for languages other than English using machine translation. The authors studied the issue in three different languages (Spanish, French and German), taking an English sentiment corpus as the original source. As baseline they classified the English corpus with the SVM SMO machine learning algorithm and then the concurrent ensemble method Bagging [37] was applied in order to study the way in which noise in the training data could be removed. Finally they translated the English corpus into Spanish, French and German, and repeated the same experimentation. The results obtained were very similar, although the German classification with Bagging methodology outperformed the baseline. The authors concluded that the use of machine translation systems is a good strategy for tackling the problem of multilingual sentiment classification, mainly due to the noteworthy current performance of the machine translation systems. In addition, they mentioned that meta-classifiers can be used to reduce the noise that the translation process may include.

3. Proposed approach

Our latest studies have followed the hypothesis proposed by [6] which states that subjectivity tends to be preserved across languages. This affirmation is very valuable for research in those languages whose available sentiment linguistic resources are scarce. Spanish is one of these languages with limited resources for opinion mining. For this reason we are currently focusing on adapting English linguistic resources for polarity classification to Spanish, and on generating Spanish sentiment resources.

In our first study in this area [25] ensemble methods and meta-classifiers were explored with the aim of including English sentiment resources in a Spanish polarity classification system. The combined system enhanced the results obtained by the supervised system that only took into account the lexical information of the Spanish sentiment corpus, and encouraged us to continue researching into the application of ensemble methods in sentiment analysis. As a result, due to the lack of lexicons of opinion bearing words in Spanish, and following the lexicon-based approach proposed by [6], a new Spanish sentiment lexicon called iSOL was generated [29].

After our previous experiments we focused on using a greater number of linguistic resources in Spanish. Since SentiWordNet (SWN) is a widely used sentiment resource for opinion mining tasks, but is only available for English, we tried to apply some Spanish linguistic resource that would be able to link WordNet synsets (and therefore SentiWordNet words) with Spanish words. In this sense the Multilingual Central Repository (MCR) provides a reduced version of WordNet in Spanish and allows users to make connections between equivalent translations in different languages such as English and Spanish. A brief description of MCR is shown below in Section 4.3.

Therefore, focusing on an unsupervised approach, our proposal in this study is to combine different semantic resources available for different languages (English and Spanish) in order to address the polarity classification task with Spanish documents and try to improve the individual performance of these resources for the task. On the one hand, we combine the semantic resources (iSOL and SWN_SP) for the original Spanish corpus, where SWN_SP is the adapted version of SentiWordNet for Spanish by using the MCR resource. On the other hand, we combine the semantic resources (BLEL and SWN) for the English translated version of the corpus. Finally, the results of both classifications are combined again by following a stacking strategy. Figures 3, 4 and 5 show the overview of our proposed approach.

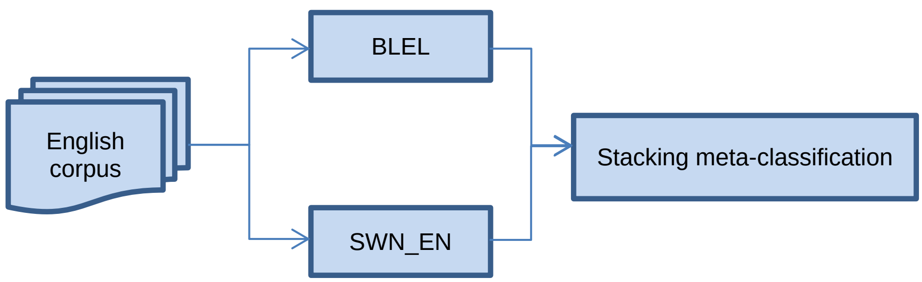


Figure 3. English Stacking Scheme

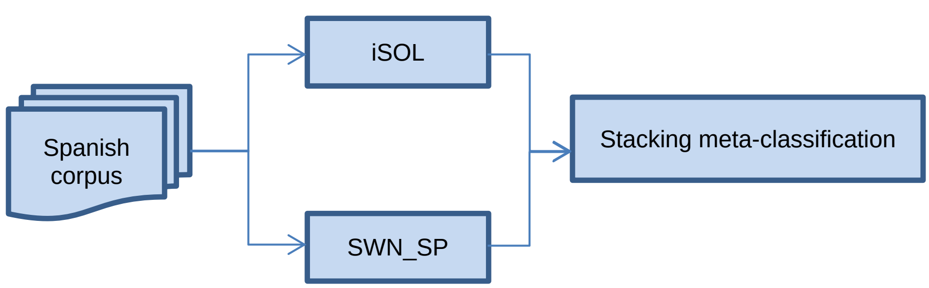


Figure 4. Spanish Stacking Scheme

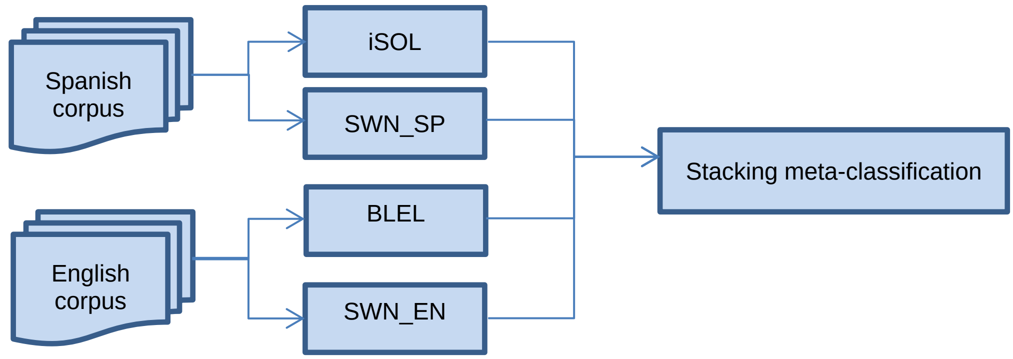


Figure 5. Bilingual Stacking Scheme

In order to carry out the stacking approach different machine learning algorithms have been applied: Support Vector Machines (SVM), Naïve Bayes (NB) and Bayesian Logistic Regression (BLR). SVM and NB are broadly known by the research community in Natural Language Processing, but BLR not so much. BLR [38] is a Bayesian implementation of the logistic regression that avoids over fitting the training data. The algorithm is based on the calculation of the following conditional likelihood:

$$P(y|\beta, x_i) = \omega(\beta^T, x_i) = \omega\left(\sum_i \beta_j x_{i,j}\right) \quad (1)$$

Where $y \in \{+1, -1\}$ are the classes. Each document is represented by a vector (xi) of values, βj are the predictors variables, and ω is a logistic link function.

$$\omega(r) = \frac{e^r}{1+e^r} \quad (2)$$

The use of a Bayesian approach to avoid over fitting involves a prior distribution on β specifying that each β_j is likely to be near 0. The prior distribution selected was a Gaussian distribution.

When the regression model is built an iterative optimization process starts. It begins by setting all variables to some initial value. It then sets the first variable to a value than minimizes the objective function holding all other variables constant. When all variables have been traversed, the algorithm begins again. Multiple passes are made over the variables until some convergence criterion is met. BLR has achieved good results in text classification problems [38] [39] and sentiment analysis [40].

4. Semantic resources

This section describes the semantic resources used for the experiments carried out. Firstly, two main lists of opinionated words were used: the improved Spanish Opinion Lexicon (iSOL) for the experiments using the Spanish corpus, and the Bing Liu English Lexicon (BLEL) for the English parallel corpus. Secondly, SentiWordNet was also applied for both corpora, but making use previously of the MCR resource when the Spanish corpus was processed.

4.1. Lists of opinionated words

The improved Spanish Opinion Lexicon[3] (iSOL) [29] was generated from the Bing Liu English Lexicon[4] (BLEL) [9] by automatically translating it into Spanish and obtaining the SOL (Spanish Opinion Lexicon) resource. Finally, the iSOL resource was obtained after carrying out a manual revision over SOL in order to improve the final list of opinion words. iSOL is composed of 2,509 positive and 5,626 negative words. Therefore this Spanish lexicon contains 8,135 opinion words.

On the other hand, the BLEL resource is composed of 2,006 positive and 4,783 negative words, resulting in a total of 6,789 opinion words. Both resources contain a higher proportion of adjectives, adverbs, nouns and verbs. Moreover, some misspelled words are included in both lists because they appear frequently in social media content.

The difference in the number of words in these two lexicons is due to the Spanish grammar. For instance, while an English adjective has neither genre nor number and is usually represented by a single term, a Spanish adjective can have four possible translated words, two for the genre (male or female) and two for the number (singular or plural). Table 1 shows some examples of possible translations of English adjectives in Spanish.

Table 1. Examples of possible translations of English adjectives in Spanish

English	Spanish
good	*bueno, buena, buenos, buenas*
famous	*famoso, famosa, famosos, famosas*
attractive	*guapo, guapa, guapos, guapas*
ugly	*feo, fea, feos, feas*
aching	*dolido, dolida, dolidos, dolidas*
bad	*malo,mala,malos,malas*

4.2. SentiWordNet

SentiWordNet [3] is a publicly available lexical resource for opinion mining which assigns three sentiment scores to each synset of WordNet[5]: positivity (how positive the word is), negativity (how negative the word is) and objectivity (how objective the word is). In other words, the sentiment scores of SentiWordNet mean the probability of a synset of being positive, negative and neutral. Each of the scores ranges from 0 to 1, and their sum equals 1. SentiWordNet scores have been semi-automatically computed based on the use of weakly supervised classification algorithms.

In SentiWordNet (SWN), each entry contains the pair Part Of Speech (POS) category and ID, which uniquely identifies a WordNet (3.0) synset, the PosScore and NegScore, which are the positivity and negativity scores assigned by SentiWordNet to the synset, and the terms with sense number belonging to the synset . The objectivity score can be calculated as 1 - (PosScore + NegScore).

Table 2 shows an excerpt of the subjectivity scores found in SWN for some synsets related to the words “good” and “bad”. There are 4 senses of the noun (POS ‘n’) “good”, 21 senses of the adjective (POS ‘a’) “good”, and 2 senses of the adverb (POS ‘r’) “good” in WordNet. There is one sense of the noun “bad”, 14 senses of the adjective “bad”, and 2 senses of the adverb “bad” in WordNet (3.0) synset.

Table 2. Fragment of SentiWordNet

POS	ID	PosScore	NegScore	SynsetTerms Gloss
a	00064787	0.625	0	good#5 beneficial#1
n	03076708	0	0	trade_good#1 good#4 commodity#1
r	00011093	0.375	0	well#1 good#1
a	01174222	0	1	unsound#5 unfit#3 bad#10
n	05144079	0	0.875	badness#1 bad#1
r	00016240	0.125	0.25	badly#6 bad#2

4.3. The Multilingual Central Repository

The Multilingual Central Repository[6] (MCR) [4] [5] constitutes a large-scale natural multilingual linguistic resource that can be used for semantic processes that need large amounts of linguistic knowledge. The MCR integrates into the same EuroWordNet framework wordnets from five different languages (including Spanish) together with four English WordNet versions. The final version of the MCR contains 1,642,389 semantic relations between synsets, most of them acquired by automatic means.

Describing more deeply how MCR works, it can be seen as a sense inventory for nouns, verbs, adjectives and adverbs for the languages involved (Basque, Catalan, English, Italian and Spanish). The wordnets in MCR have been constructed following the model proposed by EuroWordNet[7], i.e. the wordnets are linked to an Inter-Lingual-Index (ILI). Via this index the languages are interconnected, making it possible to go from the words in one language to similar words in any other language connected. The ILI is a set of meanings, mainly taken from WordNet[8]. The only purpose of the ILI is to mediate between the synsets of the local wordnets. Each synset in the local wordnets has at least one equivalent relation with a record in this ILI, either directly or indirectly via other related synsets. Language-specific synsets linked to the same ILI-record should thus be equivalent across the languages.

For the experiments carried out in this study we have used the ILI version that corresponds to WordNet 3.0. This resource allows us to obtain the synset id for each Spanish word and then look for the positive and negative scores in SWN.

5. Experimental Framework

5.1. Corpora

In this section, the two corpora used for the experiments carried out in this study are described. Firstly, the main features of the MuchoCine (MC) corpus are described. This corpus is composed of film reviews in Spanish. Then we explain briefly how the parallel English version of MC (MCE) was generated by applying machine translation techniques.

5.1.1. The MC corpus

MuchoCine [22] is a corpus of movie reviews in Spanish available for the SA research community[9]. The corpus consists of 3,878 movie reviews collected from the MuchoCine website[10]. The reviews are written by web users instead of professional film critics. This increases the difficulty of the task because the sentences found in the documents may not always be grammatically correct, or they may include spelling mistakes or informal expressions. The corpus contains about two million words and an average of 546 words per review.

In the MC corpus a movie review consists of four fields: the identifier of the review, the review rating, the summary and the description. Figure 6 shows an excerpt of a review from MC.

id|rank|summary|body
1000|-1|*Silicona, esteroides, pactos demoniacos y otras basuras habituales son la base que sustentan esta aberración de vergüenza.*| *Una fiesta llena de excesos, rubias despampanantes, musculitos por doquier, algún que otro muerto. Nada nuevo. La alianza del mal es el nombre de este thriller sobrenatural que narra las peripecias de unos jóvenes…*

Figure 6. Excerpt of a review from the MuchoCine corpus

The opinions are rated on a scale from 1 to 5. One point means that the movie is very bad and 5 means very good. Films with a rating of 3 can be considered as "neutral", which means that the user considers the film neither bad nor good. Table 3 shows the number of reviews per rating. This corpus has been widely used in different studies such as [41], [42], [43], [40] and [25].

In our experiments we discarded the neutral examples. In this way, opinions rated with 3 were not considered, the opinions with ratings of 1 or 2 were considered as negatives and those with ratings of 4 or 5 were considered as positives. Table 4 shows the class distribution of the binary classification of MC.

Table 3. Rating distribution

Rating	#Reviews
1	351
2	923
3	1,253
4	890
5	461
Total	3,875

Table 4. Binary classification of the MC corpus

Classes	#Reviews
Positive	1,274
Negative	1,351
Total	2,625

5.1.2. The MCE corpus

The MuchoCine English corpus (MCE) [25] is the version of MC translated into English and it is also available for the research community[11]. It was generated by applying a machine translation process in which different automatic translation tools were tested. According to the authors some difficulties were encountered during this process, but finally the Microsoft Translator Java API was selected as the automatic translation tool. As for the MC corpus, in MCE a movie review consists of four fields: the identifier of the review, the review rating, the summary and the description. Figure 7[12] shows an excerpt of a review from MCE.

id|rank|summary|body
1000|-1|*Silicone, steroids, demonic pacts and other usual garbage are the basis underpinning this aberration.*| *A party filled with excesses, stunning blondes, musculitos everywhere, some other dead. Nothing new. The Alliance of evil is the name of this supernatural thriller which tells the adventures of a few young people ………*

Figure 7. Excerpt of a review from the MuchoCine English corpus

For the MCE corpus we followed the same criteria that we used for the MC corpus, i.e. we discarded the neutral examples. In this way opinions rated with 3 are not considered, the opinions with ratings of 1 or 2 are considered as negative and those with ratings of 4 or 5 are considered as positive.

5.2. Evaluation framework

In order to assess the proposal, a *k-fold* cross-validation process was carried out. *K-fold* cross-validation consists of dividing the dataset in *k* bins or folders. The algorithm is run *k* times with *k* different training and test sets. In each iteration, *k-1* bins are considered to build the training set, and the other one is employed to test the classification model. For the experiments of this study we applied *10-fold* cross-validation (*k=10*).

For evaluating the classification accuracy, we employed the traditional measures used in text classification tasks: precision (P), recall (R), F1 and accuracy (Acc). For a feasible comparison, we summarize the F1 scores over the different categories (positive and negative) using the macro-averages of F1 scores:

$$Macro - F1 = \frac{2*Macro-Precision*Macro-Recall}{Macro-Precision+Macro-Recall} \quad (7)$$

In the same way we can obtain the Macro-Recall and Macro-Precision as follows:

$$Macro - Recall = \frac{\sum_{i=1}^{c} r_i}{c} \quad (8)$$

$$Macro - Precision = \frac{\sum_{i=1}^{c} p_i}{c} \quad (9)$$

Where r is the recall value, p is the precision value, and c is the number of classes.

6. Experiments and results

This section describes the experiments carried out and shows the results obtained. Firstly, we present the individual experiments that only make use of the two semantic resources explained in Section 4. Secondly, the experiments related to the proposed approach by combining the individual classifiers are described.

6.1. Individual experiments

6.1.1. Using lists of opinionated words

Before carrying out the experiments we performed a pre-processing step to the MC corpus in order to apply the same criteria followed during the generation of the iSOL list. For example, for both summary and body fields we had to change capital letters for non-capital letters, accented letters for non-accented letters, and all special characters had to be deleted from the opinions. Moreover, stop words and proper nouns were discarded.

For the MCE corpus we performed a simpler pre-processing step. We only had to change capital letters for non-capital letters and commas, semicolons, question marks and periods characters were deleted from both the summary and body fields of each opinion.

In order to calculate the polarity (p) of a review (r), we take into account the total number of positive words (#positive) and the total number of negative words (#negative) within the review, according to the following strategy:

$$p(r) = 1 \leftrightarrow \#positive \geq \#negative \quad (10)$$

$$p(r) = -1 \leftrightarrow \#positive < \#negative \quad (11)$$

Table 5 shows the results obtained by using the two lists of opinionated words over the corpora.

Table 5. Results obtained by using the two lists of opinionated words

	Macro-P	Macro-R	Macro-F1	Acc
iSOL over MC	62.22%	61.47%	61.84%	61.83%
BLEL over MCE	61.92%	56.58%	59.13%	57.56%

6.1.2. Using SentiWordNet

The semantic orientation approach using SentiWordNet has been also applied to both corpora, MC in Spanish and MCE in English. For both corpora we have followed the same procedure:

(1) **Part Of Speech tagging (POS tagging)**. The documents were processed by applying a POS tagger like TreeTagger[13] [44]. The aim of this process was to obtain all the nouns, adjectives, verbs and adverbs of each review.

(2) **Linguistic feature extraction**. This process extracts linguistic features detected in the previous step in order to generate different sub-corpora. A total of 15 sub-corpora from MC and MCE were provided by making a combination among the four linguistic features (nouns, adjectives, verbs and adverbs): *only-noun, only-adj, only-verb, only-adv, adj+noun, adj+verb, adj+adv, noun+verb, noun+adv, verb+adv, adj+noun+verb, adj+noun+adv, noun+verb+adv, adj+verb+adv, adj+noun+verb+adv*.

(3) **SWN score calculation**. The SWN score for each document was calculated in order to classify them as positive or negative. The SWN score or polarity score of a document was obtained by following the procedure described by Denecke [23] based on the calculation of a triplet of positivity, negativity and objectivity scores:

(3.1) For each token *A* with *n* synsets found in SWN, we calculate the average of its positivity score ($score_{pos}$) and the average of its negativity score ($score_{neg}$):

$$score_{pos}(A) = \frac{1}{n}\sum_{i=1}^{n} score_{pos}(i) \quad (12)$$

$$score_{neg}(A) = \frac{1}{n}\sum_{i=1}^{n} score_{neg}(i) \quad (13)$$

(3.2) The objectivity score ($score_{obj}$) is obtained for each token:

$$score_{obj}(A) = 1 - (score_{pos}(A) + score_{neg}(A)) \quad (14)$$

(3.3) The score-triplet for each document is determined by summing up the score-triplet of each term and dividing each score by the number of terms considered in such document.

In order to classify a review as "positive" or "negative" using SWN, we followed a similar strategy to that applied for the lists of opinionated words, i.e., we considered a review as "positive" if its positivity score ($score_{pos}$) is larger than or equal to the negativity score ($score_{neg}$) and as "negative" otherwise.

As described above, SentiWordNet is a semantic lexical resource available only in English. In order to deploy our SWN approach for the Spanish corpus (MC), we made use of the Multilingual Central Repository (see Section 4.3). The results obtained by using SWN over the corpora are shown in Table 6 for the MC corpus and Table 7 for the MCE corpus.

From the results shown in Table 6 and Table 7 we can observe that the highest accuracy results for both MC and MCE are obtained from those sub-corpora that include adjectives as linguistic features, as expected. Nevertheless, it is noteworthy the difference obtained (+8.87%) between the best results of both corpora (*adj+nouns+verb+adv* from MCE versus *adj+adv* from MC). We think that the main reason for this behavior is the significant difference between the number of synsets included in SentiWordNet (around 117,000 synsets) and those covered by the Multilingual Central Repository (around 38,000 synsets) used for applying SentiWordNet to the Spanish corpus.

Table 6. Results obtained by using SentiWordNet over the MC corpus (SWN_SP)

	Macro-P	Macro-R	Macro-F1	Acc
Only-nouns	51.27%	51.06%	51.16%	51.66%
Only-verb	51.48%	50.51%	50.99%	51.70%
Only-adj	62.06%	58.74%	60.36%	59.50%
Only-adv	55.20%	54.11%	54.65%	54.78%
Nouns+adj	60.41%	56.60%	58.45%	57.49%
Nouns+verb	48.56%	49.37%	48.96%	50.48%

Nouns+adv	53.87%	52.91%	53.39%	53.64%
Adj+verb	59.46%	54.18%	56.70%	55.28%
Adj+adv	**63.68%**	**58.79%**	**61.14%**	**59.66%**
Verb+adv	55.43%	51.79%	53.55%	52.99%
Nouns+adj+verb	58.93%	53.87%	56.29%	54.97%
Nouns+adj+adv	61.81%	56.93%	59.27%	57.87%
Nouns+verb+adv	50.13%	50.05%	50.09%	51.20%
Adj+verb+adv	61.16%	54.41%	57.59%	55.54%
Adj+nouns+verb+adv	59.19%	53.47%	56.19%	54.63%

Table 7. Results obtained by using SentiWordNet over the MCE corpus (SWN_EN)

	Macro-P	Macro-R	Macro-F1	Acc
Only-nouns	55.58%	54.31%	54.94%	55.01%
Only-verb	57.59%	54.95%	56.24%	55.81%
Only-adj	61.69%	60.78%	61.23%	61.18%
Only-adv	58.26%	55.09%	56.63%	54.17%
Nouns+adj	62.90%	60.48%	61.66%	61.10%
Nouns+verb	56.11%	53.76%	54.91%	54.67%
Nouns+adv	58.79%	58.41%	58.60%	58.10%
Adj+verb	63.42%	60.85%	62.11%	61.49%
Adj+adv	63.20%	62.82%	63.01%	62.55%
Verb+adv	58.14%	57.13%	57.63%	56.61%
Nouns+adj+verb	62.45%	58.96%	60.66%	59.73%
Nouns+adj+adv	64.27%	64.25%	64.26%	64.30%
Nouns+verb+adv	60.39%	60.39%	60.39%	60.34%
Adj+verb+adv	64.32%	64.33%	64.33%	64.30%
Adj+nouns+verb+adv	**65.13%**	**64.72%**	**64.92%**	**64.95%**

6.2. Combined experiments: stacking

We applied a stacking strategy in order to improve the results obtained with the individual experiments. The main purpose of the stacking method is to achieve the highest generalization accuracy by creating a meta-dataset which contains one tuple for each input example. The dimensions or number of features of those tuples are the outputs of the individual classifiers. Loosely speaking, the stacking technique consists of building a new classifier whose inputs are the outputs of the individual classifiers.

First we carried out the combination of the classifiers for each language. Three different machine learning algorithms were evaluated as stacking classifiers: Support Vector Machines (SVM), Naïve Bayes (NB) and Bayesian Logistic Regression (BLR).

As explained in the previous section, the formula used to calculate the polarity of a document for the SentiWordNet experiments was the one proposed by Denecke. This formula returns three scores (positivity, negativity and objectivity) for each document. The predicted class or output class is calculated taking into account those polarity scores. Thus the system based on the use of SentiWordNet returns four values: the predicted class and three polarity scores (positivity, negativity and objectivity). Taking into account these values, six meta-datasets for each language were built:

(1) MC_iSOL_SWN_SP_pred: It is composed of the predicted class of the base learners (SWN_SP and iSOL).
(2) MCE_BLEL_SWN_EN_pred: It is composed of the predicted class of the base learners (SWN_EN and BLEL).
(3) MC_iSOL_SWN_SP_polar: The dataset includes the predicted class of the classifier based on iSOL and the three polarity scores returned by the classifier based on SWN_SP.

(4) MCE_BLEL_SWN_EN_polar: The dataset includes the predicted class of the classifier based on BLEL and the three polarity scores returned by the classifier based on SWN_EN.
(5) MC_iSOL_SWN_SP_pred_polar: It is formed by the predicted classes returned by the two classifiers (iSOL and SWN_SP) and the polarity scores of SWN_SP.
(6) MCE_BLEL_SWN_EN_pred_polar: It is formed by the predicted classes returned by the two classifiers (BLEL and SWN_EN) and the polarity scores of SWN_EN.

The results achieved with these combinations of monolingual learners are shown in Table 8.

Table 8. Results obtained for the monolingual stacking experiments

Meta-dataset	Stacking learner	Macro-P	Macro-R	Macro-F1	Accuracy
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_pred	NB	63.94%	63.67%	63.80%	63.85%
	BLR	63.94%	69.67%	63.80%	63.85%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_polar	NB	63.33%	62.64%	62.93%	62.93%
	BLR	62.75%	62.08%	62.41%	62.40%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_pred_polar	NB	63.84%	61.90%	62.86%	62.44%
	BLR	63.77%	63.46%	63.61%	63.65%
	SVM	63.60%	60.85%	62.19%	61.48%
MCE_BLEL_SWN_EN_pred	NB	63.68%	62.26%	62.96%	62.70%
	BLR	63.68%	62.26%	62.96%	62.70%
	SVM	61.96%	60.07%	61.03%	60.65%
MCE_BLEL_SWN_EN_polar	NB	60.02%	60.07%	61.03%	60.65%
	BLR	62.76%	57.80%	60.18%	58.70%
	SVM	63.60%	60.85%	62.19%	61.48%
MCE_BLEL_SWN_EN_pred_polar	NB	63.45%	62.08%	62.76%	62.51%
	BLR	63.68%	62.26%	62.96%	62.70%

The results achieved with the two corpora follow the same pattern. The meta-datasets composed by the predicted classes of the base learners are those that achieved higher results. The performance of the system is usually worse when the polarity scores of the base learners are included in the meta-dataset. According to the results shown in Table 8, it is evident that the polarity scores do not offer valuable information for the stacking classification with Spanish texts (MC corpus). The same behaviour is repeated when the corpus is MCE. Therefore, we can conclude that the most suitable combination to enhance the performance of the monolingual polarity classification systems is the one that only takes into account the predicted classes of the base learners.

Taking into account that several classifiers have been developed for monolingual polarity classification and with the aim of improving the Spanish polarity classification, we studied the combination of the Spanish and English polarity classifiers. The meta-datasets built were:

(1) MC_iSOL_SWN_SP_pred_MCE_BLEL: Combination of the outputs of iSOL, SWN_SP and BLEL.
(2) MC_iSOL_SWN_SP_polar_MCE_BLEL: Combination of the output of iSOL, the probabilities calculated by SWN_SP, and the output of BLEL.
(3) MC_iSOL_SWN_SP_pred_polar_MCE_BLEL: Combination of the output of iSOL, the class and the probabilities calculated by SWN_SP and the output of BLEL.
(4) MC_iSOL_SWN_SP_pred_MCE_SWN_EN_pred: Combination of the outputs of iSOL, the class calculated by SWN_SP and the classes returned by SWN_EN.
(5) MC_iSOL_SWN_SP_polar_MCE_SWN_EN_polar: Combination of the output of iSOL, the probabilities calculated by SWN_SP, and the likelihoods returned by SWN_EN.

(6) MC_iSOL_SWN_SP_pred_polar_MCE_SWN_EN_pred_polar: Combination of the output of iSOL, the class and the probabilities calculated by SWN_SP and the class and the likelihoods returned by SWN_EN.
(7) MC_MCE_pred: The combination iSOL, BLEL, and the classes returned by SWN_SP and SWN_EN.
(8) MC_MCE_polar: The combination iSOL, BLEL, and the probabilities returned by SWN_SP and SWN_EN.
(9) MC_MCE_pred_polar: The combination iSOL, BLEL, and the classes and probabilities returned by SWN_SP and SWN_EN.

The results of the experiments above are shown in Table 9. This table shows different behaviours which depend on the combined classifiers and the stacking algorithm used. When the BLEL lexicon is used the results are very similar, but when the BLR is the stacking classifier the results usually increase slightly. Therefore, at least in this case, the BRL algorithm learns from the diversity of the base classifiers with better performance than SVM and NB. When SentiWordNet in English is used, the results are higher than those obtained when BLEL is combined to the Spanish base learners. Focusing only on the combinations between iSOL, SWN_SP and SWN_EN, it is interesting to highlight the fact that NB always enhances the results when the polar scores are added to the input of the stacking classifier. In the case of the monolingual stacking experiments, the polarity scores usually worsen the overall performance of the system. In this group of experiments, the polarity scores are the features which provided more information. The meta-dataset composed by the predicted class of the system based on iSOL, the predicted classes and the polarity scores of the classifiers based on the use of SWN_EN and SWN_SP is the one that achieves the highest results. NB is the algorithm with the best performance with that meta-dataset. At the beginning we expected that when more information was combined the results would be better, but as Table 9 shows, we were wrong. In this case, one of the classifiers introduces noise in the meta-classification, and taking into account the previous results we can conclude that BLEL does not provide valuable information to the classification process.

7. Analysis of Results

The main goal of this article is the improvement of Spanish polarity classification. To reach that purpose we propose a method which consists of the combination of two sentiment resources and the use of meta-classifiers. Our hypothesis has two main keys:

- The integration of semantic resources always helps the process of polarity classification.
- As Mihalcea et al. (2007) propose in their work, we also think that subjectivity tends to be preserved across languages.

In the previous sections we have described a number of experiments that assess the method proposed. As baseline, the MC corpus was classified by a system based on the use of a bag of opinion bearing words (iSOL). Then the Spanish corpus was also classified using a Spanish projection of the sentiment base-knowledge SentiWordNet. The next step was the evaluation of whether a meta-classifier method like staking can enhance the results. Following the second hypothesis the Spanish corpus was translated into English, with the aim of taking advantage of some English semantic resources. The same systems were built for English texts but using English semantic resources. The last step to evaluate our hypothesis was the combination of all the systems developed. The results obtained for Spanish sentiment classification are summed up in Table 10.

The results shown in Table 10 demonstrate that our hypothesis is correct. Firstly, the individual systems (1) and (2) achieved lower performance than the combined systems (3) and (4). Also, we highlight the fact that the system which uses iSOL obtained better results than that based on the use of the Spanish projection of SentiWordNet using MCR. Specifically, individual Spanish experiments using the lexicon based-approach and the corpus based-approach achieved an accuracy of 61.83% and 59.66% respectively (Table 5 and Table 6). However, if we compare the individual experiments for the English corpus MCE (Table 5 and Table 7), the use of the semantic resource SWN obtains better results than those applying the list of opinionated word BLEL (64.95 % and 57.56% of accuracy, respectively). We think that the main reason for this behavior is the low number of synsets managed in the MCR compared to those covered by the original SWN resource for English.

Regarding the combined systems, we can affirm that our hypothesis is valid because when we increase the number of semantic resources combined the overall performance is higher. System (4) achieves better results than (3) because system (4) takes advantage of Spanish and English semantic resources, while system (3) only uses Spanish semantic resources. On the other hand, we consider that subjectivity tends to be preserved across languages because the systems in Spanish and English obtain very similar results, and we think the loss of accuracy is only due to the reasonable noise included in the translation process.

Table 9. Bilingual Stacking results

Meta-dataset	Stacking learner	Macro-P	Macro-R	Macro-F1	Accuracy
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_pred_MCE_BLEL	NB	63.55%	61.06%	62.28%	61.68%
	BLR	63.94%	63.67%	63.80%	63.85%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_polar_MCE_BLEL	NB	63.53%	62.08%	62.80%	62.55%
	BLR	63.09%	62.43%	62.76%	62.74%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_pred_polar_MCE_BLEL	NB	63.99%	62.10%	63.03%	62.63%
	BLR	64.13%	63.74%	63.93%	63.96%
	SVM	64.40%	63.98%	64.19%	64.07%
MC_iSOL_SWN_SP_pred_MCE_SWN_EN_pred	NB	64.89%	63.63%	64.25%	64.04%
	BLR	64.70%	64.50%	64.60%	64.53%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_iSOL_SWN_SP_polar_MCE_SWN_EN_polar	NB	64.30%	63.89%	64.09%	64.11%
	BLR	63.27%	62.63%	62.95%	62.93%
	SVM	63.93%	62.88%	63.40%	63.23%
MC_iSOL_SWN_SP_pred_polar_**MCE_SWN_EN**_pred_polar	**NB**	**65.25%**	**64.34%**	**64.79%**	**64.68%**
	BLR	64.12%	63.70%	63.91%	63.92%
	SVM	63.55%	62.70%	63.12%	63.01%
MC_MCE_pred	NB	64.97%	63.35%	64.15%	63.81%
	BLR	64.57%	63.42%	63.99%	63.77%
	SVM	62.26%	61.47%	61.86%	61.83%
MC_MCE_polar	NB	64.55%	63.56%	64.05%	63.92%
	BLR	63.47%	62.76%	63.11%	63.08%
	SVM	63.37%	62.55%	62.96%	62.89%
MC_MCE_pred_polar	NB	65.12%	64.34%	64.73%	64.65%
	BLR	64.39%	63.70%	64.04%	64.00%

Table 10. Summary of Spanish polarity classification results

	Macro-F1	Accuracy
(1) iSOL over MC	61.84%	61.83%
(2) SWN_SP (adj+verb)	61.14%	59.66%
(3) MC_iSOL_SWN_SP_pred (NB)	63.80%	63.85%
(4) MC_iSOL_SWN_SP_pred_polar_MCE_SWN_EN_pred_polar (NB)	**64.79%**	**64.68%**

8. Conclusions and future work

In this paper we have combined the corpus-based and lexicon-based approch using meta-classifiers in order to improve the final system. For the corpus-based approach, we translate a Spanish corpus of movie reviews called MuchoCine (MC) into English and then we apply different English resources (SWN and opinionated lists of words). For the lexicon-based approach, we use MC corpus directly in Spanish. Therefore we have used two different semantic resources. First, we use a list of opinionated words translated into Spanish, and secondly we apply the MCR in Spanish linked with SWN in order to integrate a Spanish lexicon over the MC corpus. Finally, several combinations of classifiers were studied with the goal of improving the performance of the Spanish polarity classification. The results show that the combination of different linguistic resources, and also the use of meta-classifiers enhance the performance of a polarity classification system for Spanish texts. These results encourage us to continue working along this line.

On the other hand, for sentiment analysis the study of the influence of contextual valence shifters is very interesting. In English there are several publications such as [45] that study the influence of this linguistic element for the polarity classification task. The Spanish sentiment analysis research community has studied the contextual valence shifters in Spanish, but not in great depth. So currently we are carrying out a study of these elements, because we think that it is essential to the polarity classification task. Another of our future steps in the improvement of polarity classification systems in Spanish is the study of the calculation of negation scope and the use of linguistic heuristics to calculate the sentiment orientation of a sentence.

Notes

1. http://www.internetworldstats.com
2. MCE is freely available in http://sinai.ujaen.es/wiki/index.php/MCE_Corpus_(English_version)
3. http://sinai.ujaen.es/?p=1202
4. http://www.cs.uic.edu/˜liub/FBS/sentiment-analysis.html
5. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. It is available in http://wordnet.princeton.edu
6. http://adimen.si.ehu.es/web/MCR
7. http://www.illc.uva.nl/EuroWordNet
8. http://wordnet.princeton.edu
9. http://www.lsi.us.es/~fermin/corpusCine.zip
10. http://www.muchocine.net
11. http://sinai.ujaen.es/?p=1208
12. Figure 6 is the English translation of Figure 7
13. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger

Funding

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), TEXT-COOL 2.0 project (TIN2009-13391-C04-02) and ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. The project AORESCU (TIC - 07684) from the regional government of Junta de Andalucía partially supports this manuscript. Also, this paper is partially funded by the European Commission under the Seventh (FP7 - 2007-2013) Framework Program for Research and Technological Development through the FIRST project (FP7-287607). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

[1] Pang B and Lee L. Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing. Stroudsburg, PA, USA: Association for Computational Linguistics 2002; 10:79-86.

[2] Turney P D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics. 2002, pp. 417-424.

[3] Baccianella S, Esuli A and Sebastiani F. Sentiwordnet 3.0: An en-hanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valleta, Malta: European Language Resources Association (ELRA) 2010.

[4] Atserias J, et al. The MEANING Multilingual Central Repository. In: Proceedings of the Second International Global WordNet Conference (GWC'04). Brno, Czech Republic 2004.

[5] Gonzalez-Agirre A, Laparra E and Rigau G. Multilingual Central Repository version 3.0. In: Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'12), Istambul, Turkey, 2012.

[6] Mihalcea R, Banea C, and Wiebe J. Learning multilingual subjective language via cross-lingual projections. In: Proceedings of Annual meeting-association for computational linguistics 2007; 45(1): 976-983

[7] Liu B. Sentiment analysis and opinion mining. In: Synthesis Lectures on Human Language Technologies series 2012; 5(1): 1-167.

[8] Tsytsarau M and Palpanas T. Survey on mining subjective data on the web. Data Mining and Knowledge Discovery 2012; 24(3): 478-514.

[9] Hu M and Liu B. (2004). Mining and summarizing customer reviews. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Seattle, Washington, USA, 2004; pp. 168-177.

[10] Hatzivassiloglou V and Wiebe J. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In: Proceedings of the International Conference on Computational Linguistics 2000; pp. 299-305.

[11] Taboada M, Brooke J, Tofiloski M, Voll K, and Stede M. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2) (2011) 267-307.

[12] Wilson T, Wiebe J and Hoffmann P. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing 2005, pp. 347-354.

[13] Kamps J, Marx M, Mokken R J and de Rijke M. Using WordNet to measure semantic orientation of adjectives. In: The fourth international conference on Language Resources and Evaluation 2004; pp. 1115-1118.

[14] De Albornoz J C, Plaza L, and Gervás, P. SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) 2012; pp. 3562-3567.

[15] Chaumartin F R. UPAR7: a knowledge-based system for headline sentiment tagging. In: Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval'07, Association for Computational Linguistics 2007, pp. 422-425.

[16] Thet T T, Na, J. C, and. Khoo C. S. G. Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science 2010; 36(6): 823-848.

[17] Zhang C, Zeng D, Li J, Wang F-Y and Zuo W. Sentiment analysis of Chinese docu-ments: From sentence to document level. Journal of the American Society for Information Science and Technology 2009; 60(12): 2474–2487.

[18] Agić Z, Ljubešić N and Tadić M. Towards Sentiment Analysis of Financial Texts in Croatian. In: Proceedings of Language Resources and Evaluation, European Language Resources Association 2010; pp. 1164-1167.

[19] Rushdi-Saleh M, Martín-Valdivia M T, Ureña-López L A and Perea-Ortega J M. OCA: Opinion corpus for Arabic. Journal of the American Society for Information Science and Technology 2011; 62: 2045-2054.

[20] Banea C, Mihalcea R, Wiebe J and Hassan, S. Multilingual subjectivity analysis using machine translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, Association for Computational Linguistics 2008; pp. 127-135.

[21] Brooke J, Tofloski M and Taboada M. Cross-linguistic sentiment analysis: From English to Spanish. In: Proceedings of the International Conference RANLP-2009 2009; pp. 50-54.

[22] Cruz F L, Troyano J A, Enriquez F and Ortega J. Experiments in sentiment classification of movie reviews in Spanish. Procesamiento del Lenguaje Natural 2008; 41:79-80.

[23] Denecke K. Using SentiWordNet for multilingual sentiment analysis. In: Proceedings of Data engineering Workshop, ICDEW, IEEE 24th International Conference on, 2008; pp. 507-512.

[24] Ghorbel H and Jacot D. Sentiment analysis of French movie reviews. Advances in Distributed Agent-Based Retrieval Tools 2011; 361: 97-108.

[25] Martín-Valdivia M T, Martínez-Cámara E, Perea-Ortega J M and Ureña-López L A. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. Expert Systems with Applications 2013; 40(10), 3934–3942.

[26] Perea-Ortega J M, Martín-Valdivia M T, Ureña-López L A and Martínez-Cámara E. Improving polarity classification of bilingual parallel corpora combining machine learning and semantic orientation approaches. Journal of the American Society for Information Science and Technology 2013; 64(9):1864-1877.
[27] Martínez-Cámara E Martín-Validivia M T, Molina-González M. D and Ureña-López, L. A. Bilingual Experiments on an Opinion Comparable Corpus. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, Georgia. ACL 2013; pp 87-93.
[28] Wolpert d H. Stacked Generalization. Neural Networks 1992; 5: 241-259.
[29] Molina-González M D, Martínez-Cámara E, Martín-Valdivia M T and Perea-Ortega J M. Semantic Orientation for Polarity Classification in Spanish Reviews. Expert Systems with Applications 2013; 40(18):7250-7257.
[30] Rokach L. Ensemble Methods for Classifiers. In: Malmon O and Rokach L (EDS) Data Mining and Knowledge Discovery Handbook. Springer US, 2005, pp. 957-980.
[31] Freund Y and Schapire R E. Experiments with a new boosting algorithm. In: Machine Learning: Proceedings of the Thirteenth International Conference, 1996, pp. 325-332.
[32] Kennedy A and Inkpen D. Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence 2006, 22(2): 110-125.
[33] Wan X. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2008, pp. 553-561.
[34] Banea C, Mihalcea R. and Wiebe J. Multilingual subjectivity: are more languages better? In: Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp. 28-36.
[35] Wiebe J, Wilson T and Cardie C. Annotating expressions of opinions and emotions in language. Language Resources and evaluation 2005; 39(2-3):165-210.
[36] Balahur A and Turchi M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech & Language 2014; 28(1):56-75.
[37] Breiman L. Bagging predictors. Machine learning 1996; 24(2):123-140.
[38] Genkin H, Lewis D and Madigan D. Large-Scale Bayesian Logistic Regression for Text Categorization. Technometrics 2004. 49(3):291-384.
[39] Ortiz-Martos A, Martín-Valdivia M T, Ureña-López L A and Cumbreras-García, M. A. Detección automática de Spam utilizando Regresión Logística Bayesiana. Procesamiento del Lenguaje Natural 2005; 35:127-133.
[40] Martínez-Cámara E, Martín-Valdivia M T, Ureña López L A. Opinion classification techniques applied to a Spanish corpus. In: Proceedings of the 16th international conference on Natural language processing and information systems, 2011, pp. 169-176.
[41] Del-Hoyo R, Hupont I, Lacueva F J and Abadía, David. Hybrid Text Affect Sensing System for Emotional Language Analysis. In: Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, 2009.
[42] Barreiro A and Gonçalo-Oliveira H. Cross-language semantic relations between English and Portuguese. In: Proceedings of the workshop on iberian cross-language natural language processing tasks, 2011, pp. 49-58.
[43] Malvar-Fernández P and Pichel-Campos J R. Generación semiautomática de recursos de Opinion Mining para el gallego a partir del portugués y el español. In: Proceedings of the workshop on Iberian cross-language natural language processing tasks, 2011, pp. 59-63.
[44] Schmid H. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of the International Conference on New Methods in Language Processing 1994; 12:44-49.
[45] Polanyi L. and Zaenen A. Contextual valence shifters. In Janyce Wiebe (eds) Computing Attitude and Affect in Text: Theory and Applications. Springer, Dordrecht, 2006, pp 1-10.

Bibliografía

[AB06] Andreevskaia, A., & Bergler, S. (2006). Mining WordNet for a fuzzy sentiment: Sentiment tag extraction from WordNet glosses. Paper presented at the *EACL, , 6* 209-215.

[AG05] Aue, A., & Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Proceedings of recent advances in natural language processing (RANLP)* (Vol. 1, No. 3.1, pp. 2-1).

[AG10] Agerri, R., & García-Serrano, A. (2010). Q-WordNet: Extracting Polarity from WordNet Senses. In *LREC*.

[ALT10] Agić, Ž., Ljubešić, N., & Tadić, M. (2010). Towards sentiment analysis of financial texts in croatian. *Bull market*, *143*(45), 69.

[AVR+04] Atserias, J., Villarejo, L., Rigau, G., Agirre, E., Carroll, J., Magnini, B., & Vossen, P. (2004). The meaning multilingual central repository.

[BBMM09] Boldrini, E., Balahur, A., Martínez-Barco, P., & Montoyo, A. (2009). EmotiBlog: an Annotation Scheme for Emotion Detection and Analysis in Non-traditional Textual Genres. In *DMIN* (pp. 491-497).

[BES10] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *LREC* (Vol. 10, pp. 2200-2204).

[BFK04] Bestgen, Y., Fairon, C., & Kerves, L. (2004). Un barometre affectif effectif: Corpus de référence et méthode pour déterminer la valence affective de phrases. *Journées internationales d'analyse statistique des donnés textuelles (JADT)*, 182-191.

[BTT09] Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-Linguistic Sentiment Analysis: From English to Spanish. In *RANLP* (pp. 50-54).

[CMU+96] Cole, R. A., Mariani, J., Uszkoreit, H., Zaenen, A., & Zue, V. (1996). Survey of the state of the art in human language technology, 1996. *A study sponsored by the US National Science Foundation, Directorate XIIIE of the Commission of the European Communities, and the Center for Spoken Language Understanding, Oregon Graduate Institute*.

[CTEO08] Cruz, F. L., Troyano, J. A., Enríquez, F., & Ortega, J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de crıticas de cine en español. *Procesamiento de Lenguaje Natural*, *41*. (pp. 73-80).

[CTPO14] Cruz, F. L., Troyano, J. A., Pontes, B., & Ortega, F. J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, *41*(13), 5984-5994.

[DC07] Das, S. R., & Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science, 53*(9), 1375-1388.

[Den08] Denecke, K. (2008). Using sentiwordnet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on* (pp. 507-512). IEEE.

[DLP03] Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528). ACM.

[DTCY10] Du, W., Tan, S., Cheng, X., & Yun, X. (2010). Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. Paper presented at the *Proceedings of the Third ACM International Conference on Web Search and Data Mining,* 111-120.

[DYTS12a] Dehkharghani, R., Yanikoglu, B., Tapucu, D., & Saygin, Y. (2012). Adaptation and use of subjectivity lexicons for domain dependent sentiment classification. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on* (pp. 669-673). IEEE.

[DYTS12b] Demiroz, G., Yanikoglu, B., Tapucu, D., & Saygin, Y. (2012). Learning domain-specific polarity lexicons. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on* (pp. 674-679). IEEE.

[ES06] Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of LREC (Vol. 6, pp. 417-422).

[Fell98] Fellbaum, C. (1998). WordNet: An Electronic Database. Cambridge, MA: MIT Press

[FK82] Francis, W., & Kucera, H. (1982). Frequency analysis of English usage.

[GJ10] Ghorbel, H., & Jacot, D. (2011). Sentiment analysis of French movie reviews. In *Advances in Distributed Agent-Based Retrieval Tools* (pp. 97-108). Springer Berlin Heidelberg.

[GLR12] Gonzalez-Agirre, A., Laparra, E., & Rigau, G. (2012). Multilingual Central Repository version 3.0. In *LREC* (pp. 2525-2529).

[HM97] Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics* (pp. 174-181). Association for Computational Linguistics.

[HL04] Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

[KJM08] Kim, Y., Jung, Y., & Myaeng, S. H. (2008). An opinion analysis system using domain-specific lexical knowledge. In Information Retrieval Technology (pp. 466-471). Springer Berlin Heidelberg.

[KMMR04] Kamps, J., Marx, M. J., Mokken, R. J., & De Rijke, M. (2004). Using wordnet to measure semantic orientations of adjectives.

[MMMP14] Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Perea-Ortega, J. M. (2014). Integrating Spanish Lexical Resources by Meta-classifiers for polarity classification. *Journal of Information Science*.

[MMMU13] Martínez-Cámara, E., Martín-Valdivia, M. T., Molina-González, M. D., & Ureña-López, L. A. (2013). Bilingual experiments on an opinion comparable corpus. *WASSA 2013*, 87.

[MMMP13] Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, *40*(18), 7250-7257.

[MMMU14] Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014). Cross-Domain Sentiment Analysis Using Spanish Opinionated Words. In *Natural Language Processing and Information Systems* (pp. 214-219). Springer International Publishing.

[MMMU12] Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M., & Ureña-López, L. (2012). Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. Paper presented at *the Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis,* 3-10.

[PBM12] Perez-Rosas, V., Banea, C., & Mihalcea, R. (2012). Learning Sentiment Lexicons in Spanish. In *LREC* (pp. 3077-3081).

[PL04] Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics* (p. 271). Association for Computational Linguistics.

[PL08] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, *2*(1-2), 1-135.

[PLV02] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.

[QLBC09] Qiu, G., Liu, B., Bu, J., & Chen, C. (2009). Expanding Domain Sentiment Lexicon through Double Propagation. In *IJCAI* (Vol. 9, pp. 1199-1204).

[RMUP11] Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Perea-Ortega, J. M. (2011). OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology*, *62*(10), 2045-2054.

[RW03] Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In Proceedings of the 2003 conference on Empirical methods in natural language processing (pp. 105-112). Association for Computational Linguistics.

[SDSO66] Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The General Inquirer: A Computer Approach to Content Analysis.

[SF88] Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, *54*(4), 558.

[SMV+13] Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., ... & Gordon, J. (2013). Empirical study of machine learning based approach for opinion mining in tweets. In *Advances in Artificial Intelligence* (pp. 1-14). Springer Berlin Heidelberg.

[SS13] Saralegi X., & San Vicente I.. (2013). Elhuyar at TASS 2013. In *Proceedings of XXIX Congreso de la Sociedad Española de Procesamiento de lenguaje natural*. (pp. 143-150)

[SV04] Strapparava, C., & Valitutti, A. (2004). WordNet Affect: an Affective Extension of WordNet. In *LREC* (Vol. 4, pp. 1083-1086).

[Tur02] Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.

[WHS+05] Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., ... & Patwardhan, S. (2005). OpinionFinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations* (pp. 34-35). Association for Computational Linguistics.

[WWC05] Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, *39*(2-3), 165-210.

[WWH05] Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. Paper presented at the *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing,* 347-354.

[ZZL+09] Zhang, C., Zeng, D., Li, J., Wang, F. Y., & Zuo, W. (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.