Un Modelo Lingüístico-Semántico Basado en Emociones para la Clasificación de Textos según su Polaridad e Intensidad

Tesis doctoral

Presentada por

Jorge Carrillo de Albornoz Cuadrado

para optar al grado de Doctor en Informática

Dirigida por el

Prof. Dr. D. Pablo Gervás Gómez-Navarro

Departamento de Ingeniería del Software e Inteligencia Artificial

Facultad de Informática

Universidad Complutense de Madrid

Madrid, septiembre de 2011

Agradecimientos

Me gustaría agradecer en primer lugar el apoyo y la confianza depositada por mi director de tesis, el Dr. Pablo Gervás, sin el cual este trabajo no habría podido realizarse.

Por otro lado, me gustaría agradecer al Departamento de Ingeniería del Software e Inteligencia Artificial de la Facultad de Informática su calurosa acogida, especialmente a su director Luis Hernández, por la inestimable ayuda prestada en esta nueva y apasionante carrera profesional.

También me gustaría agradecer a todos los componentes del grupo de investigación Nil su apoyo y ayuda, en especial a Alberto Díaz, gracias por tus ánimos y consejos.

No podría terminar estos agradecimientos sin dedicar esta tesis a mi familia, mi madre, mi padre y mi hermana, quienes siempre han estado ahí para apoyarme y ayudarme en todas las aventuras en las que me he embarcado. Os quiero, no cambiéis nunca.

Por último, quiero dedicar especialmente esta tesis a la persona más maravillosa que he conocido en mi vida, Laura. Muchas gracias por todo tu apoyo, ayuda y otras mil cosas que no cabrían en mil páginas. Sin todo ello, nunca habría podido terminar esta tesis. Gracias por estar ahí siempre, por ser como eres y por hacerme como soy. Te quiero, hoy y siempre. Finalmente, dedico esta tesis a mi pequeño, te quiero enano.

Resumen

El trabajo presentado en esta tesis describe una nueva aproximación para la clasificación de textos según su polaridad e intensidad emocional, basada en el análisis semántico del texto, y en el uso de reglas lingüísticas avanzadas. El objetivo es determinar cuando una oración o documento expresa un sentimiento positivo, negativo o neutral, así como la intensidad del mismo. El método hace uso de un algoritmo de desambiguación semántica para trabajar a nivel de conceptos en lugar de términos, y utiliza el léxico afectivo *SentiSense*, desarrollado como parte de esta tesis, para extraer el conocimiento emocional y representar cada texto de entrada como un conjunto de categorías emocionales.

Por otro lado, la negación, los cuantificadores y las formas modales son construcciones lingüísticas que pueden modificar el significado emocional del texto al que afectan, por lo que es imprescindible tener en cuenta su posible efecto a la hora de desarrollar un sistema de análisis sentimental. La negación ha sido tradicionalmente considerada como un simple modificador que invierte la polaridad del texto al que afecta, mientras que los cuantificadores son tratados como modificadores que amplifican o disminuyen la fuerza de esa polaridad. La modalidad, por el contrario, solamente ha sido estudiada desde un punto de vista superficial, considerando las formas modales como modificadores que neutralizan la polaridad del texto. Sin embargo, aunque tratar estas construcciones lingüísticas como simples modificadores de la polaridad puede ser adecuado para la tarea de clasificación de la polaridad, no es suficiente cuando se abordan tareas más complejas del análisis sentimental, como, por ejemplo, la clasificación de la intensidad. Es por ello que un modelo más preciso basado en emociones en lugar de en expresiones polares se hace necesario para capturar todo el significado emocional del texto a clasificar.

En esta tesis, se estudia en profundidad el efecto de dichos modificadores

sobre las emociones a las que afectan cuando el texto es representado como un conjunto de emociones. Así mismo, se propone un modelo emocional para la negación, los cuantificadores y la modalidad, especialmente concebido para tareas de análisis sentimental. El método presentado es comparado con la aproximación tradicional basada en expresiones polares. Los resultados obtenidos demuestran que la representación del texto como un conjunto de categorías emocionales mejora el rendimiento para diferentes tareas de clasificación, así como que dicha representación permite un modelado de los modificadores emocionales más eficiente que se traduce, a su vez, en mejoras adicionales en las tareas de clasificación. Además, se estudian los usos más comunes de los modificadores emocionales en los textos afectivos, y se cuantifica su impacto en dos tareas diferentes del análisis sentimental, la detección de la polaridad y la detección de la intensidad emocional. Por último, se analiza el efecto combinado del tratamiento de todos los modificadores, concluyendo que existen interesantes sinergias entre ellos.

Finalmente, se ha realizado una evaluación exhaustiva sobre textos de diferentes dominios (titulares de noticias y opiniones de diferentes productos) para determinar la factibilidad de la aproximación presentada en diferentes tareas de clasificación de la polaridad y la intensidad. Así mismo, se ha analizado y cuantificado el efecto de la ambigüedad léxica y la presencia de negaciones, cuantificadores y modales en textos afectivos. Los resultados obtenidos muestran que el enfoque propuesto se comporta significativamente mejor que otros sistemas evaluados bajo las mismas condiciones. Estos resultados demuestran, además, que el uso de técnicas de desambiguación semántica, así como la representación del texto mediante un conjunto de emociones y el tratamiento del efecto de los modificadores emocionales sobre dichas emociones, pueden mejorar considerablemente el rendimiento de los sistemas que abordan tareas de análisis sentimental.

Abstract

The present work describes a novel approach to text polarity and intensity classification based on semantic analysis and linguistic rules. The aim is to determine whether a sentence or document expresses a positive, negative or neutral sentiment, as well as its intensity. The method performs word sense disambiguation in order to work with concepts rather than terms, and makes use of the knowledge in an affective lexicon, *SentiSense*, to label these concepts with emotional categories.

Negation, quantification and modality are common linguistic constructions that may modify the emotional meaning of the text affected by them, and therefore need to be taken into consideration in sentiment analysis. Negation is usually considered as a simple polarity shifter, while quantifiers are regarded as amplifiers or diminishers of the strength of such polarity. Modality, in turn, has only been addressed in a very naive fashion, so that modal forms are believed to neutralize the polarity of the terms within their scope. However, processing these constructions as mere polarity modifiers may be adequate for polarity classification, but it is clearly not enough for more complex sentiment analysis tasks (e.g., intensity classification), for which a more fine-grained model based on emotions instead of polar expressions is needed.

In this work, we also study the effect of modifiers on the emotions affected by them, and propose a model of negation, quantification and modality especially conceived for sentiment analysis tasks. We compare our emotion-based strategy with the traditional approach based on polar expressions and found that representing the text as a set of emotions increases accuracy in different classification tasks, and that this representation allows a more accurate modeling of modifiers that results in further classification improvements. We also study the most common uses of modifiers in affective texts and quantify their impact in two different sentiment analysis tasks, polarity

recognition and intensity classification. Finally, we analyze the combined effect of emotional modifiers and found that there exist interesting synergies between them.

An extensive evaluation on texts from different domains (news headlines and various product reviews) is performed in order to determine the feasibility of our approach in different polarity and intensity classification tasks. The effect of word ambiguity and the presence of negations, quantifiers and modality in opinionated text is also analyzed and quantified. The results obtained compare favorably with those achieved by other systems addressing similar evaluations. They also demonstrate that using word sense disambiguation techniques and dealing with emotional modifiers may considerably improve the performance of sentiment analysis related systems.

Índice general

1. Introducción **1**
1.1. Motivación 1
1.2. Retos del análisis sentimental 3
1.3. Propuesta y objetivos 10
1.4. Estructura del documento 12

2. Estado del arte **15**
2.1. La emoción y los sentimientos 15
2.2. Análisis sentimental 18
2.2.1. Breve historia del análisis sentimental 18
2.2.2. Clasificación de la subjetividad 24
2.2.3. Clasificación de la polaridad 28
2.2.4. Clasificación de la intensidad 35
2.3. La negación en el análisis sentimental 40
2.4. Los cuantificadores en el análisis sentimental 43
2.5. La modalidad en el análisis sentimental 45

3. Herramientas y recursos **49**
3.1. Herramientas lingüísticas 49
3.1.1. Stanford Parser 50
3.1.2. Gate 53
3.2. Weka 56
3.3. WordNet 62
3.4. WordNet::SenseRelate 66
3.5. Léxicos afectivos 68
3.5.1. WordNet Affect 70
3.5.2. SentiWordNet 73

4. SentiSense: un léxico afectivo basado en conceptos 77
4.1. Motivación 77
4.2. Diseño de SentiSense 80
4.3. Herramienta de etiquetado manual de SentiSense 82
4.4. Herramienta de expansión de SentiSense 84
4.5. Formato de SentiSense 85

5. Modelo lingüístico-semántico basado en emociones para la clasificación de la polaridad e intensidad 87
5.1. Preprocesado: análisis sintáctico 89
5.2. Identificación de conceptos 91
5.3. Identificación de categorías emocionales 92
5.4. Detección de negación y su ámbito 93
5.5. Detección de cuantificadores y su ámbito 98
5.6. Detección de modales y su ámbito 101
5.7. Clasificación automática 107

6. Evaluación 109
6.1. Metodología de evaluación 109
6.1.1. Métricas de evaluación 109
6.1.2. Colecciones de evaluación 112
6.1.3. Algoritmos de aprendizaje máquina 118
6.1.4. Líneas base 118
6.1.5. Experimentos 121
6.2. Clasificación de la polaridad 121
6.2.1. Clasificación de la polaridad a nivel de oración 121
6.2.2. Clasificación de la polaridad a nivel de documento . . 123
6.3. Clasificación de la intensidad 124
6.3.1. Clasificación de la intensidad a nivel de oración 124
6.3.2. Clasificación de la intensidad a nivel de documento . . 125
6.4. El efecto de los modificadores emocionales 126
6.4.1. Efecto de la negación 129
6.4.2. Efecto de los cuantificadores 131
6.4.3. Efecto de la modalidad 133
6.4.4. Efecto de todos los modificadores emocionales 134
6.5. El efecto de la ambigüedad 136

7. Discusión **139**
7.1. Clasificación de la polaridad 139
7.2. Clasificación de la intensidad 144
7.3. El efecto de los modificadores emocionales 148
7.4. El efecto de la ambigüedad 154

8. Conclusiones y trabajo futuro **157**
8.1. Conclusiones 157
8.2. Trabajo futuro 160

Bibliografía **165**

A. Publicaciones **179**
A.1. Clasificación de la polaridad y la intensidad 179
A.2. Aplicación del análisis sentimental a otros dominios de la inteligencia artificial 180
A.3. Aplicación de procesos de análisis de textos a otros dominios 180
A.4. Aplicación de otras técnicas de inteligencia artificial 180

Índice de Figuras

1.1. Fragmento de un comentario sobre un teléfono 4
1.2. Diferente orientación semántica de una misma palabra dependiendo del contexto 5
1.3. Fragmento de un comentario sobre un hotel 5
1.4. Problema de co-referencia en un comentario sobre un hotel . 7
1.5. Importancia de la estructura del discurso en el análisis sentimental 8
1.6. Ejemplo de errores ortográficos y gramaticales 9

2.1. Mapa bidimensional del espacio emocional de Watson y Tellegen (1985) 17

3.1. Etiquetado gramatical del Stanford Parser 51
3.2. Árbol sintáctico del Stanford Parser 52
3.3. Dependencias del Stanford Parser 52
3.4. Interfaz gráfica de Gate 55
3.5. Pantalla principal de Weka 57
3.6. Ejemplo de conjunto de datos en formato *arff* 59
3.7. Pestaña *Preprocess* del *Explorer* de Weka 60
3.8. Pestaña *Classify* del *Explorer* de Weka 61
3.9. Resultados de la búsqueda de la palabra *love* en WordNet . . 63
3.10. Jerarquía del primer significado de *love* en WordNet 63
3.11. Salida del paquete WordNet::SenseRelate::AllWords 68
3.12. Jerarquía de las categorías emocionales de WordNet Affect . . 72
3.13. Entrada en SentiWordNet para el término *good* 74
3.14. Entrada en SentiWordNet para el término *cancer* 75

4.1. Pantalla de la aplicación de etiquetado de SentiSense 83

4.2. Pantalla de la aplicación de visualización de SentiSense 84
4.3. Extracto del archivo de categorías emocionales de SentiSense 85
4.4. Extracto del archivo de *synsets* etiquetados en SentiSense . . 86

5.1. Diagrama de ejecución del sistema 88
5.2. Árbol sintáctico de una oración simple con negación 95
5.3. Árbol sintáctico de una oración subordinada con negación . . 97

6.1. Términos más frecuentes de negación, cuantificación y modalidad 128
6.2. Distribución de los modificadores en clases de polaridad . . . 129

Índice de Tablas

2.1. Patrones para la extracción de sintagmas de dos términos (Turney, 2002) 32

3.1. Algoritmo de WordNet::SenseRelate::AllWords 68

4.1. Categorías emocionales y sus antónimos en SentiSense 81
4.2. Distribución de los *synsets* en SentiSense 82

5.1. Términos incluidos en la lista de negación 94
5.2. Lista de delimitadores del ámbito de la negación 96
5.3. Ejemplos de oraciones con expresiones negadas 98
5.4. Términos incluidos en la lista de cuantificadores 99
5.5. Ejemplos de oraciones con cuantificadores 101
5.6. Lista de verbos modales 102
5.7. Verbos modales y su efecto 106
5.8. Ejemplos de oraciones con formas modales 106

6.1. Posibles combinaciones de resultados en la clasificación en dos clases 110
6.2. Posibles combinaciones de resultados en la clasificación en tres clases 111
6.3. Comentario extraído de la colección HotelReview corpus . . . 118
6.4. Categorías emocionales y sus valores de polaridad 119
6.5. Resultados promedio para diferentes clasificadores en la tarea de clasificación de la polaridad a nivel de oración 122
6.6. Resultados promedio para diferentes clasificadores en la tarea de clasificación de la polaridad a nivel de documento 123
6.7. Resultados promedios para diferentes clasificadores en la tarea de clasificación de la intensidad a nivel de oración 125

6.8. Resultados promedio para diferentes clasificadores en la tarea de clasificación de la intensidad a nivel de documentos 126
6.9. Distribución de la negación, los cuantificadores y los modales en oraciones y documentos 128
6.10. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de la negación 129
6.11. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de la negación 130
6.12. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de los cuantificadores 132
6.13. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de los cuantificadores 133
6.14. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de la modalidad 134
6.15. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de la modalidad 134
6.16. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de los modificadores 135
6.17. Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de los modificadores 136
6.18. Resultados promedio para diferentes clasificadores en las tareas de clasificación de la polaridad e intensidad a nivel de oración utilizando diferentes estrategias de desambiguación semántica 137
6.19. Resultados promedio para diferentes clasificadores en las tareas de clasificación de la polaridad e intensidad a nivel de documento utilizando diferentes estrategias de desambiguación semántica 137

7.1. Porcentajes de incrementos en exactitud y medida-F respecto de los obtenidos en la línea base *Sin modificadores* del método propuesto, para diferentes combinaciones de los modificadores emocionales evaluados 149
7.2. Exactitud y medida-F media en la tarea de clasificación de la polaridad e intensidad a nivel de documento, cuando solo son evaluados los documentos que contienen negaciones, cuantificadores y modales 152
7.3. Precisión y cobertura media por clase en la tarea de clasificación de la polaridad e intensidad para 3 clases 153

Capítulo 1

Introducción

1.1. Motivación

Las opiniones y comentarios de los consumidores sobre productos y servicios son de gran importancia para las empresas. Por un lado, dichas opiniones permiten realizar estudios de satisfacción de los clientes sobre los productos o servicios ofertados, así como determinar cuáles son las características y atributos de los mismos a los que más importancia conceden, sus fortalezas y sus debilidades. Por otro lado, las opiniones son de gran utilidad para delimitar los distintos "segmentos de mercado", que definen las características generales de los diferentes tipos de consumidores de un producto o servicio, así como sus gustos y preferencias. Finalmente, un análisis detallado de las opiniones sobre los productos y servicios que ofertan los competidores directos permite evaluar la posición y las posibilidades de los productos y servicios propios.

Estas prácticas no son nuevas, las empresas llevan años recopilando información sobre el grado de satisfacción de los usuarios. El proceso de recopilación y análisis de las opiniones ha sido tradicionalmente una tarea difícil, con un alto coste tanto económico como temporal. En la mayoría de los casos, estos análisis suponen la contratación de empresas externas especializadas que realizan encuestas a grupos de usuarios representativos de los distintos segmentos, y cuyos resultados tienen que ser extrapolados al total de los consumidores. Además, los resultados obtenidos mediante este proceso suelen ser válidos durante un período de tiempo limitado, puesto que los gustos y preferencias de los usuarios cambian rápidamente hoy en día.

Tradicionalmente, este tipo de encuestas tenían su razón de ser, puesto que el boca a boca era la manera normal para un usuario de informarse acerca de un producto o servicio. Sin embargo, con el crecimiento de Internet y la rápida proliferación de páginas web de intercambio de opiniones, foros, blogs y grupos de discusión, donde los usuarios intercambian opiniones y comentarios acerca de infinidad de productos y servicios, el coste que supone a un usuario el proceso de recopilación de información a la hora de adquirir un producto ha disminuido sustancialmente. Más concretamente, de acuerdo a dos encuestas (Hitlin y Rainie, 2004; ComScore, 2007), el 81 % de los usuarios de Internet de Estados Unidos afirman haber realizado búsquedas de opiniones sobre productos al menos una vez, y entre el 73 % y el 87 % de usuarios que han realizado búsquedas sobre hoteles, restaurantes o agencias de viajes han resaltado que las opiniones encontradas han influido significativamente en su decisión de compra (Pang y Lee, 2008). Queda clara, por tanto, la influencia de las opiniones y recomendaciones de otras personas sobre los consumidores potenciales de un producto o servicio, así como su importancia para la reputación de las empresas encargadas de su producción o comercialización.

Sin embargo, la ingente cantidad de opiniones vertidas en Internet cada día hace cualquier tipo de tratamiento manual imposible. Es por esto que el diseño y desarrollo de nuevas tecnologías capaces de procesar automáticamente esta información de una forma eficiente, generando información útil para las empresas, ha despertado el interés de la comunidad científica, surgiendo numerosas líneas de investigación relacionadas. Pese a ello, como expone Liu (2010), monitorizar las opiniones de la Web es una tarea difícil y los modelos actuales del problema, así como de la información necesaria o de sus posibles soluciones, son todavía limitados.

La información suscrita en las páginas de opiniones o foros, comúnmente conocida como *contenido generado por el usuario* (*user-generated content*), se presenta, por regla general, como texto libre no estructurado, por lo que el procesamiento automático y el análisis de las opiniones requiere de técnicas avanzadas de procesamiento del lenguaje natural. Es por ello que, en los últimos años, ha surgido una nueva disciplina dentro del procesamiento del lenguaje natural encargada de tratar el problema del análisis computacional de las opiniones, sentimientos y subjetividad en el texto, llamada *análisis sentimental* y algunas veces referida como *minería de opiniones*. El análisis

sentimental, a su vez, engloba diferentes tareas como la *detección de subjetividad*, la *detección de polaridad*, la *detección de polaridad e intensidad* o el *análisis sentimental basado en tópicos*. Aunque la finalidad de las distintas tareas es diferente (determinar si un texto es subjetivo u objetivo, o si un texto expresa un sentimiento positivo o negativo, o incluso la intensidad de dicha polaridad) todas abordan la tarea de analizar el texto desde el punto de vista de las emociones y opiniones expresadas en él.

Restringir el alcance del análisis sentimental al estudio de opiniones de usuarios para servir objetivos empresariales es no hacer justicia a las posibilidades de la disciplina. De hecho, el análisis sentimental de texto ha sido utilizado recientemente para predecir los resultados de las últimas elecciones al senado de los Estados Unidos de America mediante el análisis de las conversaciones en redes sociales y blogs, o incluso para analizar la opinión pública respecto al vertido de petróleo del golfo de Méjico (Brustein, 2010).

Resulta pertinente, por tanto, afirmar que las posibilidades y ventajas que proporcionaría un sistema capaz de determinar la polaridad e intensidad de un texto de manera automática supondrían un beneficio sustancial para los usuarios y consumidores de bienes y servicios, a quienes ayudaría en su proceso de decisión de compra; para empresas y compañías, que podrían utilizar estas opiniones para mejorar su marketing-mix; e incluso para gobiernos y organizaciones, que podrían así conocer la opinión pública sobre temas importantes y controvertidos.

1.2. Retos del análisis sentimental

La identificación/clasificación automática de textos no estructurados no es un problema trivial, como la comunidad científica ha descubierto en los últimos años. Muchos son los retos que hay que superar para obtener una clasificación sentimental adecuada de una revisión u opinión de un producto, de una noticia, o de cualquier otro tipo de texto. Sin embargo, actualmente estos sistemas han logrado un considerable porcentaje de éxito, principalmente motivado por el avance de los algoritmos de aprendizaje máquina y la disponibilidad pública de recursos para entrenar dichos algoritmos, lo que ha hecho muy atractiva esta disciplina tanto para compañías y gobiernos, como para la comunidad científica.

La evaluación emocional de un texto debe resolver diferentes problemas

que se engloban dentro del procesamiento del lenguaje natural, para poder clasificar y determinar correctamente las opiniones del usuario o emisor. A lo largo de esta sección utilizaremos distintos fragmentos de texto extraídos del corpus presentado en (Taboada y Grieve, 2004) sobre comentarios de teléfonos y del corpus presentado en (Carrillo de Albornoz et al., 2011) sobre comentarios de hoteles, para ilustrar dichos problemas y dificultades.

Para poder evaluar correctamente un texto desde el punto de vista emocional, el primer paso necesario es determinar qué oraciones son las que expresan las opiniones del usuario. Antes de continuar, es preciso diferenciar entre dos grupos de información textual: los *hechos* y las *opiniones*. Los *hechos* son fragmentos de texto donde el usuario expresa sentencias objetivas sobre el producto o servicio, mientras que las *opiniones* exponen los sentimientos y vivencias de los usuarios respecto a él. Las expresiones subjetivas pueden presentarse de muchas formas, como deseos, opiniones, creencias o especulaciones, lo que hace de esta tarea un difícil reto. Por ejemplo, en la figura 1.1 las oraciones (1) y (2) representan hechos expresados por el usuario al comentar las prestaciones de un teléfono, mientras que las oraciones (3) y (4) expresan las opiniones del usuario sobre diferentes características del mismo.

```
(1) We bought this phone to replace a cordless phone we
had for about 9 years.
(2) The phone we replaced was very cheap but worked
perfectly for us until I dropped it into a bucket of water.
(3) We like Panasonic products so decided to give
this phone a try.
(4) The sound quality wasn't good at all for us.
```

Figura 1.1: Fragmento de un comentario sobre un teléfono

Para poder identificar las expresiones subjetivas, es necesario capturar de alguna manera la semántica de las palabras así como su orientación (i.e., objetiva *vs* subjetiva). Es por esto que se deben usar léxicos especiales donde cada palabra está etiquetada de acuerdo a su orientación. Sin embargo, las palabras normalmente poseen más de un significado, por lo que un mismo término puede actuar como subjetivo u objetivo en una oración dependiendo del contexto. La figura 1.2 ilustra claramente este problema. La palabra *disaster* en la primera oración denota una opinión o experiencia subjetiva sobre el servicio de un hotel, mientras que la misma palabra en la segunda

oración tiene claramente una orientación objetiva. Es por esto que, a la hora de trabajar con la orientación de las palabras, es imprescindible desambiguar el significado de las mismas, utilizando para ello técnicas de desambiguación del significado (del inglés *Word Sense Disambiguation*) (Ide y Véronis, 1998; Agirre y Edmonds, 2006).

```
(1) The hotel service is a disaster.
(2) The disaster has caused 10 victims.
```

Figura 1.2: Diferente orientación semántica de una misma palabra dependiendo del contexto

Sin embargo, incluso para los seres humanos, no siempre es una tarea sencilla clasificar oraciones en hechos u opiniones. Por ejemplo, la figura 1.3 muestra un fragmento de un comentario sobre un hotel extraído de la página web *booking.com*, donde se puede corroborar esta afirmación. En este texto, las oraciones (1), (5), (6), (7) y (10) son una buena representación de oraciones que expresan los sentimientos del usuario, mientras que las oraciones (2), (3), (4), (8) y (9) expresan hechos sobre el hotel. No obstante, estos hechos no deberían ser considerados como neutrales por un sistema de análisis sentimental, puesto que exponen una opinión implícita sobre el hotel, como es, por ejemplo, su buena ubicación frente a la Torre Eiffel.

```
(1) The view of Eiffel Tower was very good.
(2) We were able to see it every day and
night from our room. (3) Cooking facilities
are sufficient with Microwave and hobs.
(4) Cutleries are provided too. (5) Everything
is quite clean. (6) Sofa bed is comfortable.
(7) Location is quite far from shopping centers
metro stations. (8) No breakfast included.
(9)  Kettle did not work. (10) View of Eiffel
Tower can not be guaranteed so you have to try your luck.
```

Figura 1.3: Fragmento de un comentario sobre un hotel

Por otro lado, las distintas oraciones, y los textos en general, pueden expresar hechos o sentimientos positivos, negativos o neutrales. En concreto, si expresan positividad o negatividad, se dice que la oración tiene una orientación polar, que puede ser, a su vez, positiva o negativa. Así, por ejemplo, se puede ver que la oración (3) del ejemplo 1.1 claramente denota una punto

de vista positivo del usuario sobre los productos de Panasonic, mientras que la oración (4) expresa una opinión negativa acerca del teléfono descrito en el comentario. Para poder determinar la polaridad de un texto; es decir, clasificarlo como positivo o negativo, el proceso a seguir es muy similar al expuesto para la detección de subjetividad. Puede parecer que determinar la polaridad es un problema sencillo, ya que muchas palabras tienen por sí mismas cierta orientación sentimental. Sin embargo, por regla general, las palabras no siempre expresan el mismo sentimiento, debido a que normalmente poseen distintos significados, y a que, aún atendiendo a un mismo significado, no siempre expresan la misma orientación sentimental, por lo que en la mayoría de los casos dicha orientación depende del contexto de la oración que incluye la palabra. Así, palabras que por si mismas denotan una clara orientación negativa pueden ser neutrales, o incluso positivas, en el contexto adecuado. El grado o fuerza de esta polaridad es también un punto interesante a tener en cuenta cuando se realiza el análisis emocional de una oración. Toda oración que exprese una opinión puede clasificarse como positiva o negativa, pero es lógico pensar que no todas las oraciones expresan el mismo grado de polaridad negativa o positiva. Por ejemplo, la oración *This phone is AWFUL* es claramente más negativa que la oración *The sound quality wasn't good at all for us.*

Otro aspecto importante a la hora de evaluar una opinión es determinar correctamente el objeto sobre el que recae dicha opinión, ya que no es lo mismo una oración que exprese una opinión negativa sobre la habitación de un hotel que otra reprochando el color utilizado para pintar las paredes de la recepción. Al abordar este problema, hay que dar respuesta a tres preguntas: primero, cuáles son los atributos o características importantes para los usuarios cuando evalúan un determinado servicio o producto; segundo, qué oraciones del texto se refieren a los atributos o características detectadas; y tercero, cómo ponderar la polaridad de esas oraciones con respecto a las características identificadas. El amplio vocabulario utilizado por los usuarios para referirse a los diferentes atributos es otro punto interesante a tener en cuenta, ya que el uso de unos u otros términos para identificar los atributos puede decrementar considerablemente el porcentaje de acierto en el paso de identificación. Es también muy importante determinar correctamente el objeto sobre el que trata la oración, puesto que en muchos casos no resulta evidente. Por ejemplo, un sistema que analice el texto de la figura 1.4 debe

ser capaz de determinar que el *staff* (personal) del hotel es el objeto de la oración (2); de otra manera la oración se asignaría incorrectamente a otra característica del hotel, o incluso a ninguna. Estos casos se consideran un tipo de problema de co-referencia que suele ser abordado mediante el uso de técnicas de resolución de anáforas y co-referencias (Elango, 2005; Nicolov, Salvetti, y Ivanova, 2008).

```
(1) Staff is nice except for one receptionist
(a man) at night, he was not helpful at all,
I asked him for directions and he said there's
nothing I can do if you don't know Paris.
(2) Anyways, everybody else was nice.
```

Figura 1.4: Problema de co-referencia en un comentario sobre un hotel

El análisis a nivel de oración es una tarea crucial y muy importante para la identificación emocional de textos. Sin embargo, el principal objetivo normalmente es clasificar o asignar una puntuación a todo el comentario o documento dentro de un conjunto de grados o intensidades de polaridad (e.g., *fuertemente negativo*, *negativo*, *neutral*, *positivo* y *fuertemente positivo*). En general, los textos suelen ser una combinación de oraciones positivas, neutrales y negativas, por lo que la dificultad es aún mayor puesto que hay que tener en cuenta el peso de cada oración en el comentario o documento, los atributos o características a las que las oraciones hacen referencia, y su influencia en el conjunto de la opinión. La estructura del discurso es otro factor a tener en cuenta cuando se trabaja a nivel de documento, puesto que puede afectar a la detección de su polaridad e intensidad. Por ejemplo, el texto de la figura 1.5 empieza con dos oraciones positivas sobre el hotel, pero termina con una oración negativa que cambia por completo la polaridad global del comentario. Esta es una estructura típica del discurso en debates y comentarios de productos o servicios, donde la introducción del documento hace algunas concesiones y la conclusión termina con un juicio que expresa el punto de vista contrario.

Hay otros factores lingüísticos a tener en cuenta a la hora de evaluar automáticamente opiniones, como son el uso de negaciones, de cuantificadores, de modales, o de otras construcciones más complejas, como la ironía o el sarcasmo, que afectan a la interpretación semántica de dichas opiniones.

El efecto de la negación puede afectar al sujeto de una oración (***No one***

```
(1) The hotel is in a nice location. (2) Also, the staff
is nice and polite. (3) However the rooms were horrible
and the building is more than old, so definitely I will
never stay again.
```

Figura 1.5: Importancia de la estructura del discurso en el análisis sentimental

would like to do this), a la acción de la oración (*Beckham* ***does not want to play again for Real***), o a solo una palabra (*There is no* ***no reason*** *to do that*), lo que hace de su identificación y tratamiento tareas complejas. El análisis del efecto de la negación no solo comprende la tarea de identificar en el texto las palabras o términos que producen esa negación, sino que también es imprescindible determinar su ámbito de acción sobre el resto de palabras (i.e., las palabras afectadas por la negación). Por ejemplo, la negación en la oración *Very comfortable bed and* <u>*not*</u> *too short* solo afecta a las palabras *not too short*, mientras que en *Kettle did* <u>*not*</u> *work* la negación afecta a la acción de la oración (el verbo *work*). El efecto de esta negación también puede variar dependiendo del contexto, pudiendo cambiar la orientación de la polaridad, o simplemente disminuyendo dicha polaridad. Así, por ejemplo, la negación en la oración *I do* <u>*not*</u> *love you* claramente cambia la polaridad de la misma, mientras que la oración *The neighborhood is* <u>*not*</u> *bad* no significa necesariamente que el vecindario sea bueno, simplemente que es aceptable.

El uso de cuantificadores en las oraciones puede también afectar a la polaridad y su intensidad, puesto que no es lo mismo decir *This is a good party* que *This is a* <u>*really*</u> *good party*. Mientras que la negación típicamente invierte o disminuye la carga emocional de las palabras afectadas, los cuantificadores pueden amplificar o disminuir la intensidad de sus significados, y a su vez, con diferentes intensidades (Quirk et al., 1985). El análisis de los cuantificadores, como en el caso de la negación, también implica la detección de las palabras o términos que actúan como cuantificadores así como su ámbito de acción. Existen gran cantidad de listas de cuantificadores en inglés, lo que hace que la identificación de los mismos sea sencilla. Por el contrario, la resolución de su efecto sobre el resto de palabras de su ámbito de acción conlleva mayor complejidad, puesto que no todos los cuantificadores potencian, o disminuyen, la polaridad con la misma intensidad.

La presencia de modales es otro aspecto importante a tener en cuenta cuando se realiza el análisis emocional de un texto u oración, puesto que los

sistemas deberían ser capaces de distinguir entre situaciones que podrían, deberían, o es posible que pasen, frente a situaciones que están pasando, han pasado o van a pasar. Sin embargo, el efecto de los modales sobre la polaridad de la oración no es, por así decirlo, tan claro como el efecto de los cuantificadores, o incluso la negación, ya que los modales no siempre tienen el mismo efecto. En algunos casos, un modal puede cambiar la polaridad de la oración, como en *I should eat healthier*, o no modificar en absoluto su polaridad, como en *I would like to kill you*.

Por último, la ironía y el sarcasmo son probablemente las características más difíciles a detectar y tratar en estos sistemas, puesto que no existe un modelo conceptual sobre cómo se presentan estos fenómenos. Es por esto que muy pocas investigaciones, ya no solo en el área de análisis sentimental sino en la disciplina completa del procesamiento del lenguaje natural, abordan este fenómeno.

Otros factores menos relevantes, pero que también pueden influir considerablemente en el rendimiento de los sistemas, y que hay que tener en cuenta a la hora de evaluar texto libre no estructurado de foros y blogs on-line, son las faltas ortográficas y gramaticales, y el uso de un vocabulario o jerga específica. Centrándonos en el idioma inglés, un gran porcentaje de usuarios que utilizan estos sistemas no tienen el inglés como lengua materna, por lo que en muchos casos escriben las palabras como suenan, lo que supone frecuentes fallos como, por ejemplo, confundir la palabra *where* con *were*. También cabe destacar errores como la omisión de espacios entre palabras, el uso incorrecto de las reglas de puntuación o la omisión de los sujetos, tal y como se puede ver en la figura 1.6, donde se muestra un fragmento de un comentario sobre un hotel extraído de *booking.com*.

```
i thought the breakfast will be provided likealways
when u are in a hotel,unfortunately there is none
for breakfast. bed wasn.t made no free drink on
arrival no continental breakfast as advertised,
noisy broken tap in kitchen health n safety
non existent, dirty showers and toilets .
```

Figura 1.6: Ejemplo de errores ortográficos y gramaticales

Todos estos retos que hay que afrontar a la hora de diseñar un sistema capaz de analizar emocionalmente un texto, han hecho esta disciplina atrac-

tiva para la comunidad científica. Aunque los primeros trabajos en el área estaban basados en aproximaciones simples de frecuencias de términos o en el uso de características lingüísticas simples del texto analizado, actualmente se está empezando a realizar análisis más profundos, fundamentalmente a nivel lingüístico, consiguiendo sistemas más escalables y con un mayor porcentaje de acierto.

1.3. Propuesta y objetivos

El trabajo realizado en esta tesis presenta un método híbrido, basado en el análisis semántico y la generación de reglas lingüísticas, para la clasificación de textos en base a su polaridad e intensidad emocional. Las aproximaciones tradicionales que encontramos en la literatura para afrontar estas tareas están generalmente formuladas como métodos supervisados de aprendizaje máquina. Las variables utilizadas para las instancias de aprendizaje son muy variadas, dependiendo de la aproximación, incluyendo las palabras o términos que se encuentran en el texto, su frecuencia o presencia, su categoría gramatical, el árbol de dependencias generado para representar el texto, o incluso la polaridad de las palabras que lo componen (extraída de un léxico afectivo). Para el correcto funcionamiento de estos sistemas, se necesitan ejemplos previamente etiquetados con los que entrenar el algoritmo (es decir, textos clasificados como positivos o negativos, y en distintos grados de intensidad), que por regla general son recopilados de páginas de opiniones de Internet. La principal debilidad de estos enfoques se encuentra en la escasa importancia que atribuyen al aspecto lingüístico del problema, y al nulo papel que juega la semántica en las soluciones adoptadas, aspectos que, desde el punto de vista del autor de esta tesis, son imprescindibles para resolver problemas como, por ejemplo, la ambigüedad del significado de las palabras. Como ya se ha dicho, las palabras pueden tener diferentes significados en función del contexto en el que se utilizan, por lo que capturar apropiadamente el significado de cada palabra resulta fundamental a la hora de analizar un texto y clasificarlo de acuerdo a su polaridad e intensidad. Esta problemática puede ser resuelta representando el texto de entrada mediante conceptos en lugar de términos o palabras, utilizando para ello un algoritmo de desambiguación semántica que asocie palabras con significados. De esta manera, la primera aportación importante de esta tesis es el

uso de una representación basada en conceptos, y la utilización de diferentes técnicas de desambiguación semántica para traducir las palabras del texto a sus correspondientes significados en la base de datos léxica WordNet. Además, se ha estudiado y evaluado el efecto de la ambigüedad en las tareas de análisis sentimental para corroborar la hipótesis expuesta.

Las aproximaciones más recientes en detección de polaridad e intensidad abordan el problema mediante las llamadas *expresiones polares*, que no son más que palabras que tienen una polaridad en sí mismas (i.e., un significado negativo o positivo). Por ejemplo, las palabras *like* o *good* son expresiones polares positivas, mientras que las palabras *hate* o *bad* son expresiones polares negativas. En estos trabajos, el texto de entrada es transformado en un valor final calculado a partir de las expresiones polares que lo componen, y que determinará la polaridad e intensidad global del texto. Sin embargo, estas aproximaciones presentan el inconveniente principal de asumir que la detección de la polaridad e intensidad depende exclusivamente de la orientación polar de las palabras que componen el texto, sin tener en cuenta los sentimientos o emociones concretas que dichas palabras expresan. Es evidente, por ejemplo, que las palabras *cancer* y *cold*, aunque teniendo ambas una orientación negativa, expresan emociones o sentimientos muy diferentes: un cáncer es generalmente asociado con *miedo* y *tristeza*, mientras que un resfriado se asocia más bien con *malestar* o *disgusto*. Esta información más específica del significado emocional de las palabras puede ser necesaria cuando se pretende clasificar un texto en positivo o negativo. Consideremos, por ejemplo, la oración *Superb location, terrible room* y la oración *Superb location, mediocre room* extraídas de dos comentarios de hoteles. En ambos casos, las oraciones presentan términos positivos y negativos, pero la primera de ellas es claramente negativa mientras que la segunda puede ser considerada incluso como positiva por algún potencial usuario no demasiado preocupado por el estado de la habitación y sí por su localización. Este problema es aún más evidente cuando se desea clasificar el texto de acuerdo a su intensidad emocional. Por todos estos motivos, nuestra hipótesis es que una representación del texto mediante un conjunto de emociones, en lugar de expresiones polares, resolvería el problema planteado y mejoraría los resultados en la clasificación de la polaridad y la intensidad. De esta manera, el método propuesto en esta tesis utiliza el conjunto de categorías emocionales existentes en un léxico afectivo que relaciona conceptos de WordNet con

emociones para representar el texto de entrada. El objetivo de esta representación es minimizar la perdida de información emocional importante a la hora de clasificar el texto en base a su polaridad e intensidad, mejorando así, a su vez, los resultados de la clasificación.

Por último, son pocas las aproximaciones que han estudiado el efecto de determinadas estructuras lingüísticas complejas en el análisis sentimental, como la negación, los cuantificadores o los modales. Como ya se ha dicho, estas construcciones lingüísticas pueden modificar el significado emocional del texto afectado, por lo que deben ser tomadas en consideración. La negación es tratada habitualmente en la literatura como un simple modificador que invierte la polaridad de las expresiones polares a las que afecta, mientras que los cuantificadores, a su vez, son considerados como modificadores que amplifican o disminuyen el grado o intensidad de dicha polaridad. Por otro lado, los pocos trabajos que han estudiado el problema de la modalidad en el análisis sentimental entienden el efecto de ésta como un modificador que bloquea el significado emocional de las expresiones polares a las que afecta. Sin embargo, nuestra hipótesis es que, mediante una representación del texto basada en emociones, es posible un modelado y un tratamiento más preciso de la negación, los cuantificadores y la modalidad, que a su vez permitirá mejorar sustancialmente los resultados de la clasificación. De esta forma, otra aportación importante de esta tesis es analizar los usos más frecuentes de estos modificadores desde una perspectiva emocional, su efecto separado en la clasificación de textos en base a su polaridad e intensidad, y las interacciones entre ellos en el análisis sentimental. Las conclusiones de este análisis han sido utilizadas para la construcción de un modelo basado en emociones para la negación, los cuantificadores y los modales. A día de hoy, este es el primer trabajo en analizar el efecto individual y combinado de dichos modificadores emocionales desde el punto de vista del análisis sentimental.

1.4. Estructura del documento

En esta sección, se presentan los capítulos que componen esta tesis, describiendo brevemente en cada caso su contenido.

Capítulo 1. Introducción. Este capítulo introduce los principales motivos que han llevado a la realización de esta tesis, así como la prob-

lemática y el estado actual de la disciplina. Por último, se presentan las diferentes contribuciones del trabajo realizado.

Capítulo 2. Estado del arte. Este capítulo describe en mayor detalle la disciplina que nos ocupa, presentando su origen y su historia hasta el presente. Se muestran las técnicas actuales más utilizadas para resolver las tareas más relevantes del análisis sentimental, así como para el tratamiento de los fenómenos lingüísticos de negación, cuantificación y modalidad.

Capítulo 3. Herramientas y recursos. Este capítulo describe las diferentes herramientas evaluadas y utilizadas en esta tesis, y que no han sido desarrolladas expresamente como parte de la misma.

Capítulo 4. SentiSense: un léxico afectivo basado en conceptos. Este capítulo presenta el léxico afectivo especialmente desarrollado para abordar, desde una perspectiva conceptual, distintas tareas de análisis sentimental.

Capítulo 5. Modelo lingüístico-semántico basado en emociones para la clasificación de la polaridad e intensidad. Este capítulo presenta el método propuesto para la clasificación de textos en base a su polaridad e intensidad, describiendo, para cada etapa, su misión y su funcionamiento.

Capítulo 6. Evaluación. Este capítulo describe la metodología utilizada para evaluar el sistema, a la vez que presenta los resultados obtenidos al evaluar el método propuesto en diferentes tareas y sobre colecciones de evaluación de distintos dominios.

Capítulo 7. Discusión. Este capítulo analiza y discute en profundidad los resultados obtenidos en la evaluación presentada en el capítulo anterior.

Capítulo 8. Conclusiones y trabajo futuro. Este capítulo recopila las diferentes conclusiones extraídas del trabajo realizado, y propone algunas líneas de trabajo futuro.

Capítulo 2

Estado del arte

A lo largo de este capítulo se presentarán al lector los trabajos de mayor relevancia en el campo del análisis sentimental, así como las aproximaciones más aceptadas para abordar el tratamiento de la negación, los cuantificadores y los modales en el área que nos ocupa. El capítulo se divide en cinco secciones, empezando con una breve descripción de la importancia de las emociones y los sentimientos, así como de las principales teorías psicológicas que tratan de definirlos y explicarlos. A continuación, se presenta la disciplina del análisis sentimental, describiendo su historia y las diferentes tareas que la componen, profundizando en mayor detalle en la tres tareas de mayor relevancia, y describiendo para cada una de ellas los métodos y técnicas más difundidos y aceptados. Finalmente, se presentan los diferentes enfoques utilizados en otros trabajos para abordar el problema de la detección de los modificadores emocionales (i.e., negación, cuantificadores y modales) y la resolución de su efecto sobre los términos dentro de su ámbito de acción.

2.1. La emoción y los sentimientos

Uno de los principales problemas a los que se enfrentan los sistemas de análisis sentimental es la falta de consenso a la hora de definir qué se entiende por emoción o sentimiento, y cómo se representa. A lo largo de los años, se han planteado muchas teorías psicológicas que tratan de explicar qué entiende el ser humano por emoción y cómo la interpreta, existiendo incluso a día de hoy gran diversidad de opiniones.

La identificación de los diferentes aspectos que entran en juego en los procesos emocionales, así como la clasificación de las emociones en función

de distintos criterios, han sido objeto de gran interés a lo largo de toda la historia de la humanidad, remontándose hasta tiempos de los filósofos griegos. Actualmente existen dos teorías principales, ampliamente aceptadas, para representar y clasificar las diferentes emociones humanas: la teoría de dimensiones emocionales y la teoría de categorías emocionales. La primera de ellas se basa en la idea de que las emociones se pueden representar mediante dos o tres dimensiones, variando conceptualmente según el autor, a través de las cuales se puede abarcar todo el espacio emocional humano. La teoría de las categorías emocionales defiende, por el contrario, que las distintas emociones se pueden representar mediante un conjunto de palabras que denotan dichas emociones, en las que se engloba todo el espacio emocional humano. Los principios de estas dos teorías se pueden encontrar en los trabajos realizados por los psicólogos James Russel (1980) y William James (1884), respectivamente, las cuales han ido evolucionando con los años hasta conformar las teorías actuales.

El trabajo de Russell se centra en la representación del espacio emocional humano mediante una estructura circular de dos dimensiones bipolares, el placer y la activación. Otro modelo emocional en esta misma línea deriva del estudio realizado por Watson y Tellegen (1985), en el que se presenta una estructura circular emocional similar a la de Russell, pero en este caso basada en las dimensiones de afecto positivo y afecto negativo, las cuales a su vez poseen una intensidad baja o alta. Un ejemplo de esta representación se puede ver en el gráfico de la figura 2.1.

Inspirado en esta misma idea encontramos el estudio realizado por Lang (1980), donde se presenta el estándar SAM, según el cual es posible asignar a cada palabra inglesa un valor, entre 1 y 9, para cada una de las tres dimensiones propuestas en el estándar: evaluación, activación y control. Esta forma de asignación de valores a las dimensiones está basada en la escala *Semantic Differential*, desarrollada en 1950 para el tratamiento de emociones, y cuya teoría se centra en la idea del pensamiento humano mediante polos opuestos, como el bien y el mal. En base a este estándar, Lang y Bradley (1999) crearon la *Affective Norms for English Words (ANEW)*, que pretende obtener un conjunto amplio de palabras medidas en términos de las tres dimensiones emocionales.

Por el contrario, y en relación a la teoría de las categorías emocionales, James interpreta las emociones como entidades o unidades discretas con

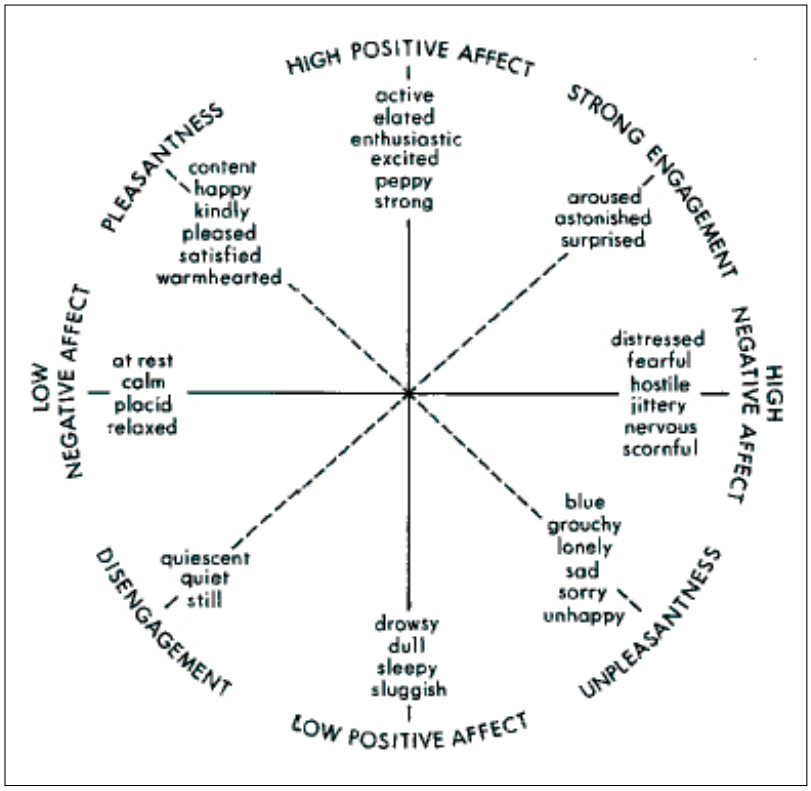


Figura 2.1: Mapa bidimensional del espacio emocional de Watson y Tellegen (1985)

límites entre ellas claramente diferenciados, y que engloban todo el espacio emocional humano. Esta teoría ha ido evolucionando a lo largo de los años hacia las categorías emocionales (Cowie y Cornelius, 2003), cuya idea se basa en la lógica interpretación de las categorías conceptuales de las emociones mediante las palabras que las denotan en cada idioma.

Uno de los puntos clave de esta teoría es la gran diversidad de opiniones que existe sobre cuáles deben ser dichas categorías, pudiendo variar desde unas pocas hasta infinitas. Realmente, los orígenes de la idea de un conjunto de emociones primitivas se podrían remontar a René Descartes (Anscombe y Geach, 1970), quien expuso la teoría de que solo existen un conjunto de emociones puras y primitivas, a partir de las cuales se derivan el resto.

Más recientemente, Izard (1977) determina que el conjunto básico de emociones está compuesto por las siguientes: *anger, contempt, disgust, distress, fear, guilt, interest, joy, shame* y *surprise*. Poco después, Plutchik (1980) propone un conjunto muy reducido de emociones básicas o primiti-

vas (*anger, anticipation, disgust, joy, fear, sadness* y *surprise*), afirmando que el resto de emociones son mezcla de éstas o derivadas. Ekman (1992) reduce todavía más el conjunto de emociones básicas, asumiendo que solo existen seis emociones claramente diferenciadas para las expresiones faciales (*anger, disgust, fear, joy, sadness* y *surprise*). Ortony et al. (1988) presentan el modelo OCC, el cual se ha ido estableciendo como estándar en síntesis de voz, y que cuenta con 22 categorías emocionales: *pride - shame, admiration - reproach, happy - resentment, gloating - pity, hope - fear, joy - distress, satisfaction - fear-confirmed, relief - disappointment, gratification - remorse, gratitude - anger* y *love - hate*. Parrot (2001), por el contrario, presenta una lista mucho más extensa de emociones, pero jerarquizada en una estructura de árbol, en cuya raíz se encuentran las emociones básicas (*love, joy, surprise, anger, sadness* y *fear*), y, más abajo en la jerarquía, las emociones secundarias y las emociones terciarias.

2.2. Análisis sentimental

A lo largo de está sección se describe en profundidad la historia de la disciplina conocida como *análisis sentimental*, presentando los trabajos que han impulsado la misma, y terminando con una breve descripción de las diferentes tareas que la conforman. Por último, se exponen y detallan las diferentes técnicas y aproximaciones utilizadas para las tres tareas más difundidas y estudiadas por la comunidad científica: la clasificación de la subjetividad, de la polaridad y de la intensidad emocional.

2.2.1. Breve historia del análisis sentimental

El término *análisis sentimental* en procesamiento del lenguaje natural hace referencia, en su más amplia acepción, a la disciplina que comprende la tarea de identificar/clasificar fragmentos de texto que contengan una carga emotiva, opinión o subjetividad. Aunque algunos autores presentan como precursores de la disciplina los trabajos de Carbonell (1979) o Wilks y Bien (1984) sobre la interpretación de creencias por parte de sistemas de inteligencia artificial, lo cierto es que fue en la última década del siglo XX cuando realmente empezaron a aparecer estudios más centrados en el análisis de textos y su interpretación subjetiva, como, por ejemplo, la interpretación de metáforas en textos (Hearst, 1992), la identificación de puntos de in-

terés (Sack, 1994; Wiebe, 1994; Wiebe y Bruce, 1995), la identificación de personajes subjetivos en narrativa (Wiebe, 1990a) o el análisis afectivo del texto (Huettner y Subasic, 2000). Sin embargo, no ha sido hasta comienzos del siglo XXI cuando la disciplina ha alcanzado su verdadera popularidad. Prueba de ello es que, durante los últimos años, el volumen de artículos publicados relacionados con el tema ha crecido exponencialmente. Este impulso definitivo ha sido motivado principalmente por dos causas: primero, el avance de los sistemas de aprendizaje máquina; y segundo, la facilidad para obtener conjuntos de entrenamiento y evaluación para dichos sistemas gracias al desarrollo de Internet.

Este carácter novel de la disciplina hace que los propios términos que la definen varíen de un autor a otro, tanto para el propio área en sí como para las múltiples sub-tareas satélites que han aparecido en torno a ella. Así, por ejemplo, es común en la literatura referirse a este campo como *análisis sentimental* (*sentiment analysis*), *minería de opiniones* (*opinion mining*) o *análisis de subjetividad* (*subjectivity analysis*) indistintamente. Por otro lado, y en cuanto a las tareas que se pueden englobar dentro de la disciplina, cabe destacar la detección de polaridad, donde los textos de entrada son analizados y clasificados en positivo o negativo según la carga emotiva que presenten; la detección de subjetividad, donde el texto es analizado y diferenciado atendiendo a la carga subjetiva expresada por parte del autor; o la detección de fragmentos del texto que contienen opiniones, donde el texto es analizado para extraer aquellos fragmentos que expresan opiniones por parte del autor. Más recientemente, han aparecido tareas más complejas como la detección de la intensidad emocional, donde el texto de entrada es clasificado atendiendo a la intensidad de su polaridad en diferentes clases (e.g., *fuertemente negativo*, *negativo*, *neutro*, *positivo* y *fuertemente positivo*), o la clasificación sentimental mediante tópicos o características del texto, donde los parámetros que ponderan la clasificación están basados en los tópicos o características de los temas tratados en los textos.

El uso del término sentimiento en relación al análisis de textos apareció por primera vez en la literatura en los trabajos de Dans y Chen (2001) y Tong (2001) sobre predicción de juicios para analizar el comportamiento de los mercados. Solamente un año más tarde, el término análisis sentimental fue acuñado por Turney (2002) y Pang et al. (2002) en referencia a sus aproximaciones para la clasificación de textos en función de su polaridad.

Este hecho supuso una revolución en el área que produjo gran cantidad de artículos relacionados con esta tarea en los sucesivos años, lo que ha llevado incluso a algunos autores a equiparar el termino análisis sentimental con la tarea de clasificación de la polaridad.

La historia del término *minería de opiniones* comenzó con Dave et al. (2003), quien definió el sistema perfecto de minería de opiniones como aquel capaz de procesar un conjunto de resultados de búsqueda generando una lista de atributos sobre cada búsqueda y asociando a cada atributo una opinión. Este artículo fue presentado en la conferencia WWW, por lo que el término, y la propia tarea en sí, ha estado estrechamente ligada a la investigación en recuperación de información y en sistemas de búsqueda en Internet.

Janyce M. Wiebe fue la gran impulsora del término *análisis de subjetividad*. En su artículo (Wiebe, 1994) se define el concepto de subjetividad sobre la idea de los *estados privados* propuestos por Quirk et al. (1985), que engloban los conceptos de emoción, opinión, evaluación y especulación. Sin embargo, el término análisis de subjetividad ha sido más asociado en la literatura al proceso de reconocimiento de fragmentos de texto con carga subjetiva frente a aquellos con carga objetiva o sin opinión inherente.

Pang y Lee (2008) defienden, en su amplia revisión sobre la disciplina, que los términos análisis sentimental y minería de opiniones hacen referencia a una misma tarea, y que a su vez podría entenderse como un subconjunto de una tarea mayor que sería el análisis de subjetividad. Esta afirmación está soportada en la idea de que, en los últimos años, ambos términos han sido utilizados indistintamente para definir tareas de más alto nivel en el análisis de sentimientos en texto. De la misma manera, pero sin equiparar minería de opiniones y análisis sentimental, Wilson et al. (2009) entienden la tarea de análisis sentimental como un subconjunto del análisis de subjetividad. Más en profundidad, Pang y Lee (2008) presentan como tareas principales las siguientes: la clasificación de la polaridad (*polarity classification*), una de las de mayor auge en los últimos años, y que muchas veces en la literatura se conoce como clasificación sentimental (*sentiment classification*); la detección de subjetividad (*subjectivity detection*), que pretende determinar cuando un fragmento de texto es subjetivo u objetivo; el análisis sentimental por tópicos (*topic-sentiment analysis*), donde las opiniones sobre ciertas características o tópicos del texto son ponderadas; y otras tareas de menor relevancia como la detección de puntos de vista (*viewpoints and perspectives*) o la interpretación

de emociones (*emotion identification*).

Por orto lado, Liu (2010) presenta en su estudio sobre el área una clasificación no solo basada en la tarea realizada, sino también en el tipo de entrada utilizada. De esta manera, Liu define la tarea de clasificación binaria de documentos en positivo o negativo según su carga emotiva como clasificación sentimental a nivel de documento (*document-level sentiment classification*), mientras que la clasificación en tres clases de intensidad (positivo, neutral y negativo), o incluso más clases, es definida como una tarea a nivel de oración en la que, en un primer paso, se realiza una clasificación de la subjetividad (*subjectivity classification*), y en un segundo paso, se realiza una clasificación de la polaridad sobre las oraciones subjetivas (*sentence-level sentiment classification*). Esta tarea es denominada por Liu como clasificación subjetiva y sentimental a nivel de oración (*sentence-level subjectivity and sentiment classification*). Liu también presenta como tarea del área el análisis sentimental basado en características (*featured-based sentiment analysis*), o la minería de opiniones definida como búsqueda y recuperación de opiniones (*opinion search and retrieval*). El análisis sentimental de oraciones comparativas (*sentiment analysis of comparative sentences*) es introducido por Liu haciendo referencia a aquellas aproximaciones que abordan la tarea de detectar este tipo de oraciones comparativas, extrayendo a su vez sus opiniones. Por último, el uso del análisis sentimental por parte de sistemas de spam es también considerado por Liu dentro de la tarea denominada spam de opiniones y la utilidad de las opiniones (*opinion Spam and utility of opinion*).

El tipo de entrada al sistema es otro aspecto importante a tener en cuenta en el análisis sentimental (Liu, 2010). En los comienzos de la disciplina, el documento era el principal elemento de entrada para la clasificación de polaridad (Pang, Lee, y Vaithyanathan, 2002; Turney, 2002), principalmente motivado por la facilidad de encontrar recursos en Internet (foros, páginas de opiniones, etc.) que asignan a un documento una puntuación positiva o negativa. Por su parte, la oración era el elemento principal para la detección de subjetividad (Wiebe, 1990b; Hatzivassiloglou y Wiebe, 2000; Riloff y Wiebe, 2003), debido a que la oración representa mejor la unidad de subjetividad. Sin embargo, en la actualidad esto no es así, y ello se debe principalmente a la opinión de muchos autores de que un documento contiene muchas oraciones tanto positivas, neutrales o negativas (Liu, 2010), independientemente de su interpretación sentimental global, y que, por tanto, la unidad natural

para la clasificación de polaridad es la oración (Kim y Hovy, 2004; Riloff, Patwardhan, y Wiebe, 2006; Meena y Prabhakar, 2007). Otros han ido más allá defendiendo que incluso una misma oración puede contener varías interpretaciones emocionales, por lo que lo correcto sería considerar como unidad primitiva el sintagma (Wilson, Wiebe, y Hoffmann, 2009). Actualmente, no existe una opinión clara y unificada sobre cuál debe ser la unidad de trabajo de los sistemas de análisis sentimental. Lo cierto es que tanto la oración como el documento son las entradas más utilizadas, y en ambos casos tienen sus beneficios y sus desventajas. El nivel de oración, por ejemplo, plantea el problema de que muchas oraciones no poseen carga emocional por sí mismas, sino como parte de un contexto mayor, por lo que en los sistemas de clasificación de la polaridad deben ser forzosamente etiquetadas como positivas o negativas pese a tener una carga emocional neutral. Por otro lado, las oraciones suelen incluir menos ruido, entendido éste como fragmentos de texto en los que se cambia de polaridad u opinión, hecho que es mucho más frecuente a nivel de documento. Finalmente, el documento en sí contiene más información, por lo que en principio su clasificación debería ser más sencilla y precisa.

Según la opinión del autor de esta tesis, y atendiendo al desarrollo observado de la disciplina durante el último lustro, el término *análisis sentimental* hace referencia a la tarea de análisis, identificación y clasificación de todo tipo de contenido emocional, subjetivo u opinionado. De esta manera, y recopilando las tareas más relevantes que se podrían englobar dentro de esta definición, se propone la siguiente clasificación de las tareas del análisis sentimental:

- *Clasificación de la subjetividad:* tarea cuyo objetivo es la identificación de fragmentos de texto que poseen un significado o una carga subjetiva, expresada por parte de la persona que ha escrito el texto, ya sea una opinión, la expresión de un sentimiento, etc.

- *Clasificación de la polaridad:* tarea que pretende, como última finalidad, clasificar fragmentos de texto, que pueden ser desde documentos hasta sintagmas, en positivo o negativo dependiendo de su significado emocional.

- *Clasificación de la intensidad:* esta tarea pretende clasificar los textos de entrada de acuerdo a la intensidad emocional expresada. De

esta manera, la mayoría de la aproximaciones que abordan este problema trabajan con lo que se denomina clasificación en tres clases de intensidad: *positivo*, *neutro* y *negativo*, mientras que los estudios más recientes están empezando a incluir la problemática de cinco clases de intensidad: *fuertemente positivo*, *positivo*, *neutral*, *negativo*, *fuertemente negativo*. Como es razonable, a mayor número de clases a las que se enfrenta un sistema, menor es su porcentaje de acierto.

- *Análisis sentimental basado en tópicos/características:* a diferencia de las tareas anteriores, en este caso la detección y el análisis emotivo del texto está basado en ciertas características o tópicos del propio texto, ponderando el análisis global. Generalmente, este tipo de sistemas suelen evaluar documentos que recogen opiniones sobre productos o servicios donde ciertos aspectos de esos productos o servicios condicionan más que otros la carga afectiva global de la opinión.

- *Minería de opiniones:* tarea más especializada en recuperación de información, en la que, dada una consulta, los sistemas deben ser capaces de determinar los documentos relevantes de dicha consulta acerca de un tema en concreto, así como clasificarlos dentro de ciertas categorías.

Aunque esta clasificación define claramente los límites entre unas tareas y otras, lo cierto es que muchas de éstas no son excluyentes entre sí, y en general muchos trabajos suelen abarcar más de una a la vez (Pang y Lee, 2004). Por ejemplo, no es extraño que la clasificación de la subjetividad se utilice como paso previo a la clasificación de la polaridad, de tal manera que solo se clasifican en positivo o negativo aquellos textos que ya de por sí presentan cierta subjetividad. Lo mismo sucede con la clasificación de la subjetividad y la clasificación de la intensidad, generalmente en la tarea de tres clases, en la que el paso previo permite determinar los fragmentos de texto neutrales (Kim y Hovy, 2004; Wilson, Wiebe, y Hoffmann, 2005; Wilson, Wiebe, y Hoffmann, 2009).

A continuación, en las siguiente secciones, se analizarán y describirán en profundidad las tres tareas de mayor relevancia del análisis sentimental, que son la clasificación de la subjetividad, la clasificación de la polaridad y la clasificación de la intensidad, por ser éstas, además, las que mayor relación guardan con el trabajo realizado en esta tesis.

2.2.2. Clasificación de la subjetividad

La detección de la carga subjetiva expresada en el texto, en contraste con los fragmentos objetivos, es una de las tareas del análisis sentimental que ha suscitado mayor interés en el mundo académico en los últimos años. La clasificación de la subjetividad aborda el problema de determinar qué partes del texto de entrada expresan la subjetividad del emisor y cuáles no. Esta tarea se ha centrado fundamentalmente en la identificación de la subjetividad a nivel de oración, de forma que es posible distinguir dos tipos de oraciones: las *oraciones objetivas*, que describen hechos, y las *oraciones subjetivas*, que describen opiniones, sentimientos o creencias.

La tarea de clasificación de la subjetividad ha sido utilizada en los últimos años como una tarea previa a otras tareas como la clasificación de la polaridad o la clasificación de la intensidad, con el objetivo de identificar las oraciones irrelevantes que no conllevan ninguna carga emotiva, evitando así posibles errores en las siguientes clasificaciones. Sin embargo, esta tarea puede ser utilizada como un fin en sí misma. Un sistema capaz de separar las oraciones que expresen hechos, como el tamaño o precio de una cámara de fotos, de otras que expresen las opiniones de los usuarios, como si el diseño es atractivo o el peso es adecuado, puede ser de gran utilidad para las empresas.

Como ya se ha comentado, la principal impulsora de esta tarea ha sido Janyce M. Wiebe, cuyos primeros trabajos se centraron en la detección de puntos de vista (Wiebe, 1994; Wiebe y Bruce, 1995) o en la identificación de personajes subjetivos en textos narrativos (Wiebe, 1990a). Decir que ha sido la principal impulsora, de hecho, es decir poco, puesto que hasta hace pocos años sus trabajos han sido la única referencia de la literatura de cierta relevancia. El enfoque planteado por Wiebe para la detección de la subjetividad se centra en el uso de técnicas supervisadas de aprendizaje máquina, abordando la tarea como una clasificación binaria donde el texto de entrada, generalmente la oración, debe ser categorizada en *subjetiva* u *objetiva*. Siguiendo esta idea, Wiebe et al. (1999) presentan un sistema basado en un clasificador Naive Bayes utilizando como variables o atributos de las instancias del clasificador la presencia en las oraciones de pronombres, adjetivos, modales excepto *will*, u otros como si la oración inicia o no un párrafo. Así mismo, en este artículo se presentan los requisitos para el correcto etiquetado de una colección de evaluación para la tarea de clasificación de la subjetivi-

dad (u otras tareas en las que el desacuerdo entre jueces pueda ser alto). Las conclusiones de esta aproximación supervisada aclaran que, aunque los resultados son satisfactorios (alrededor del 70 % de exactitud), es preciso estudiar otras variables que pueden influir más en la tarea de clasificación de la subjetividad, como son la presencia de ciertos adjetivos, la orientación de dichos adjetivos, o su graduabilidad.

Estas conclusiones fueron estudiadas y corroboradas por Bruce y Wiebe (1999), quienes determinan que, de todas las oraciones que componían la colección de evaluación desarrollada, simplemente la presencia de un adjetivo determinaba la subjetividad de la oración en un 56 % de los casos. Además, Bruce y Wiebe también identifican un tipo de adjetivos determinado, los denominados *dinámicos* por Quirk et al. (1985), que según su análisis son indicativos de la presencia de subjetividad, generando así una lista de adjetivos que incluye algunos como *careful* o *kind*. Para evaluar si estos factores realmente influyen en la clasificación automática de la subjetividad, Hatzivassiloglou y Wiebe (2000) desarrollan un sistema que determina si una oración es subjetiva en función de si contiene alguno de estos adjetivos. Los adjetivos utilizados se clasifican en: 1) adjetivos polares (i.e., negativos *vs* positivos), 2) adjetivos dinámicos, 3) adjetivos graduales y 4) todos los adjetivos identificados automáticamente. Las conclusiones extraídas de este estudio demuestran que los resultados obtenidos para los adjetivos polares, los dinámicos y los adjetivos graduales obtienen mejores resultados que cuando se evalúa con todos los adjetivos. Los resultados también demuestran la buena efectividad de los adjetivos como elementos representativos de la subjetividad, resaltando que la inclusión de este tipo de variables en sistemas de aprendizaje máquina debería mejorar los resultados obtenidos hasta el momento.

Siguiendo esta línea de identificar los elementos principales que atribuyen a una oración la subjetividad u objetividad, se enmarcan los trabajos presentados en (Riloff, Wiebe, y Wilson, 2003; Riloff y Wiebe, 2003), donde se pretende extraer de una forma automática patrones para etiquetar sustantivos subjetivos y expresiones subjetivas, respectivamente. Para ello, utilizan dos clasificadores de alta precisión para determinar oraciones fuertemente subjetivas y fuertemente objetivas, basándose para ello en la presencia de términos fuertemente subjetivos u objetivos. Una vez identificadas las oraciones de cada tipo, se extraen automáticamente los patrones de expresiones

y sustantivos subjetivos, que son añadidos de nuevo a los clasificadores de alta precisión, repitiendo así el proceso de extracción de patrones. Por último, cabe destacar el análisis más en profundidad del significado lingüístico de la subjetividad realizado por Wiebe et al. (2004), donde se presentan las construcciones lingüísticas más comunes que afectan y determinan la presencia de subjetividad en los textos.

Por otro lado, uno de los problemas del lenguaje natural, y por ende de las expresiones determinadas como subjetivas (ya sean n-gramas, sustantivos, verbos, adjetivos, etc.), es la ambigüedad léxica, fenómeno que puede hacer que una misma expresión actúe como subjetiva u objetiva dependiendo del contexto en el que se encuentre. Para poder resolver estos casos de ambigüedad, Wiebe y Wilson (2002) presentan un método basado en la *densidad subjetiva* mediante el que resuelven este tipo de conflictos de ambigüedad. El método se basa en la idea de que una expresión será más subjetiva cuantos más términos subjetivos haya en el contexto de la palabra; es decir, cuanto mayor sea la densidad subjetiva a su alrededor. Siguiendo esta idea, Wiebe y Mihalcea (2006) defienden que la subjetividad es un factor estrechamente ligado a la desambiguación semántica y que, de hecho, la detección de la subjetividad ayuda a mejorar el proceso de desambiguación semántica en palabras con significados tanto objetivos como subjetivos.

Toda esta información relevante para la tarea de clasificación de la subjetividad identificada en la literatura (i.e., listas de adjetivos, sustantivos, expresiones, etc.), fue recopilada por Wilson et al. (2004) en un sistema supervisado para clasificar sintagmas subjetivos, así como su grado de subjetividad. Además, en este trabajo Wilson et al. incluyen un conjunto de reglas lingüísticas basadas en los árboles de dependencias de las oraciones, así como una aproximación basada en la frecuencia de los términos. El trabajo recoge experimentos realizados con tres clasificadores diferentes, concluyendo que la opción de incluir toda la información (i.e., listas de adjetivos, etc., las reglas lingüísticas y las frecuencias de los términos) en las variables de las instancias del algoritmo obtiene los mejores resultados para los clasificadores en la tarea de clasificar sintagmas en base a su subjetividad y su fuerza subjetiva. Este sistema fue extendido en futuros trabajos (Wilson, Wiebe, y Hoffmann, 2005; Wilson, Wiebe, y Hoffmann, 2009), para incluir otros modificadores lingüísticos contextuales que afectan a la detección de la subjetividad, como la negación o los modificadores contrarios (como *but* o *however*). Para ello,

desarrollaron un clasificador en dos etapas, cuya primera etapa aborda el problema de la clasificación de la subjetividad a nivel de sintagma, mientras que la segunda realiza la clasificación de la polaridad solo sobre aquellos textos identificados como subjetivos.

Aunque la gran mayoría de los trabajos presentados abordan la tarea desde el punto de vista del análisis de la oración, otros como Yu y Hatzivassiloglou (2003) presentan una aproximación basada en un clasificador Naive Bayes utilizando simplemente frecuencias de palabras y obteniendo una exactitud del 97 % en la tarea de clasificar artículos de noticias y negocios (*hechos*) frente a editoriales y cartas al editor (*opiniones*). Así mismo, Yu y Hatzivassiloglou abordan la tarea de clasificación de la subjetividad a nivel de oración en este mismo artículo, mediante tres aproximaciones: similitud semántica entre oraciones, un clasificador Naive Bayes y múltiples clasificadores Naive Bayes, concluyendo, al igual que otros trabajos, que la aproximación con mayor información en las variables es la opción más adecuada, y que las mejoras introducidas por múltiples clasificadores son prácticamente inapreciables. En relación a la tarea de clasificación de la subjetividad a nivel de documento, Ng et al. (2006) experimentan con un clasificador SVM basado simplemente en la frecuencia de las palabras, obteniendo un 99 % de exactitud. Este alto rendimiento obtenido en la tarea a nivel de documento se debe, principalmente, a la gran diferencia de vocabulario entre los documentos clasificados como subjetivos y los clasificados como objetivos.

Por otro lado, durante los últimos años se han presentado distintas aproximaciones que han abordado la tarea de clasificación de la subjetividad mediante técnicas muy diversas, como la teoría de grafos (e.g., asignando pesos mayores a los nodos que contienen palabras con mayor carga subjetiva (Pang y Lee, 2004)), o el uso de expresiones polares para determinar un valor final indicativo de la carga subjetiva u objetiva del texto en función de ciertos umbrales (Kim y Hovy, 2005). Sin embargo, tal y como muestra la literatura, la gran mayoría de las aproximaciones que abordan la tarea de clasificación de la subjetividad utilizan técnicas supervisadas de aprendizaje máquina para clasificar los textos en subjetivos u objetivos (Das y Bandyopadhyay, 2010; Lambov, Dias, y Graça, 2010), variando únicamente de una aproximación a otra el tipo de variables utilizadas para representar las instancias de los algoritmos de aprendizaje máquina. Estas variables incluyen las frecuencias

de los términos, la presencia de expresiones polares, las categorías gramaticales, la sintaxis, etc. Esta mayor inclinación por este tipo de sistemas es debida, principalmente, a la sencillez que supone su desarrollo, puesto que simplemente es necesario modelar el texto de acuerdo a las variables deseadas en cada instancia del algoritmo de aprendizaje máquina, y al alto rendimiento obtenido por este tipo de sistemas en dicha tarea.

Por el contrario, el principal problema de estos sistemas es la escasa importancia dada al aspecto semántico y lingüístico. Por un lado, la mayoría de las aproximaciones no tienen en cuenta el significado emocional de las palabras, simplemente se basan en la aparición y la frecuencia de los términos; y los pocos enfoques que utilizan expresiones polares suelen trabajar con términos en lugar de conceptos, sin tener en cuenta los múltiples significados que una misma palabra puede tener y que puede afectar considerablemente a la correcta identificación de la subjetividad. Por otro lado, este tipo de sistemas no suelen tener en cuenta construcciones lingüísticas que pueden afectar a la detección de la subjetividad, como son la negación, los cuantificadores o los modales; y los pocos trabajos que han abordado esta problemática suelen limitarse a identificar su presencia, sin estudiar ni tratar su efecto. Finalmente, estos trabajos adolecen de ser excesivamente dependientes del dominio; es decir, una vez entrenado el sistema con documentos de un determinado dominio, su aplicación sobre textos de un dominio diferente, especialmente si el vocabulario utilizado es muy dispar, arroja resultados muy deficientes.

2.2.3. Clasificación de la polaridad

La clasificación de la polaridad es la tarea, dentro del análisis sentimental, que pretende determinar la polaridad del texto; es decir, clasificarlo en positivo o negativo de acuerdo a su carga emocional. La clasificación de la polaridad se ha abordado a lo largo de los años principalmente desde dos enfoques diferentes: mediante métodos supervisados y mediante métodos no supervisados. Las aproximaciones basadas en métodos supervisados tratan la clasificación de la polaridad como un problema de clasificación binaria (i.e., positivo *vs.* negativo), donde un conjunto de datos previamente etiquetado es utilizado como conjunto de entrenamiento para el algoritmo de aprendizaje máquina. Por el contrario, los métodos no supervisados suelen apoyarse en la idea de las denominadas *expresiones polares* para determinar

de una forma u otra la polaridad total de un texto. Las expresiones polares, como ya se ha expuesto, son simplemente palabras, o conjuntos de palabras, que poseen cierta polaridad en sí mismas, como, por ejemplo, las palabras *excellent* o *joy*, que son representativas de expresiones polares positivas, o las palabras *fear* o *sadness*, que son palabras representativas de expresiones polares negativas. Para realizar el cálculo global de la polaridad del texto a partir de las expresiones polares, se han utilizado multitud de técnicas, entre las que destacan fórmulas matemáticas, como la métrica PMI (del inglés *pointwise mutual information*), basada en la frecuencia de co-ocurrencia de los términos y calculada según la ecuación 2.1, o conjuntos de reglas.

$$PMI(term_1, term_2) = log_2(\frac{Pr(term_1 \wedge term_2)}{Pr(term_1) \times Pr(term_2)}) \tag{2.1}$$

Pang et al. (2002) fueron los pioneros en utilizar métodos supervisados para la clasificación de textos en positivo o negativo atendiendo a su polaridad. En su aproximación, plantean la hipótesis de si los métodos utilizados para la tarea de clasificación de tópicos son aplicables a la detección de polaridad en críticas de películas, o por el contrario, es preciso desarrollar aproximaciones específicas. Para probar su hipótesis, realizaron varios experimentos con tres algoritmos típicos de aprendizaje máquina: Naive Bayes, Maximum Entropy y Support Vector Machines, utilizando como atributos de los vectores de aprendizaje las frecuencias de los términos, la presencia o no de estos, el uso de adjetivos, n-gramas y las categorías gramaticales de los términos. Las conclusiones extraídas de los experimentos demuestran que el uso de unigramas, junto con la presencia de términos y el algoritmo *Support Vector Machines*, permite obtener los mejores resultados a la hora de clasificar críticas de películas.

Son muchas las aproximaciones que han utilizado métodos supervisados. Como en todo sistema de aprendizaje máquina, la dificultad principal de los métodos supervisados se encuentra en cómo representar los datos de entrada para generar el vector de cada instancia. Es por esto que, a lo largo de estos años, se han propuesto muchas alternativas como variables de estos vectores. A continuación se muestra una clasificación de las más utilizadas:

- *Términos*: el uso de palabras como variables en los vectores de aprendizaje ha sido, y es, una de las aproximaciones más utilizadas en la tarea de clasificación de la polaridad. Relacionado con el uso de térmi-

nos, se han evaluado variables como la presencia de estos, su frecuencia, el uso de n-gramas, de lemas o incluso su posición dentro del texto (Pang, Lee, y Vaithyanathan, 2002; Dave, Lawrence, y Pennock, 2003; Pang y Lee, 2004; Mullen y Collier, 2004; Aue y Gamon, 2005). Las funciones *tf-idf* para ponderar estas frecuencias han sido también evaluadas, obteniendo resultados muy prometedores (Martineau y Finin, 2009; Paltoglou y Thelwall, 2010). El uso de estas variables ha demostrado, al igual que en la categorización de textos, que son bastante efectivas para la tarea.

- *Categorías gramaticales*: un poco más sofisticadas que las aproximaciones basadas en términos, son aquellas que utilizan las categorías gramaticales como variable. La motivación se halla en la idea de que el rol gramatical de las palabras influye a la hora de evaluar la polaridad. Así, por ejemplo, los adjetivos han demostrado ser buenos indicadores de la afectividad del texto (Hatzivassiloglou y McKeown, 1997; Pang, Lee, y Vaithyanathan, 2002; Na et al., 2004).

- *N-gramas emocionales*: los n-gramas emocionales son un conjunto de palabras, en muchos casos combinación de dos o tres, que, por regla general, presentan una elevada carga emocional, como por ejemplo *beautiful*, *very good* o *really amazing*. En este tipo de aproximaciones se utiliza la frecuencia de estos conjuntos de palabras en los textos de entrada (Na et al., 2004; Whitelaw, Garg, y Argamon, 2005; Zaidan, Eisner, y Piatko, 2007).

- *Dependencias sintácticas*: el uso de análisis lingüísticos más en profundidad para introducir información del árbol sintáctico o de dependencias en los vectores de clasificación ha sido también ampliamente estudiado en los últimos años (Wilson, Wiebe, y Hoffmann, 2005; Ng, Dasgupta, y Arifin, 2006; Abbasi, Chen, y Salem, 2008).

- *Negación*: la negación es un factor altamente relevante a la hora de clasificar la polaridad de un texto, como se verá más adelante. Es por esto que la presencia de negación en los textos de entrada ha sido otra de las variables consideradas en los últimos años a la hora de construir los vectores de clasificación (Das y Chen, 2001; Pang, Lee, y Vaithyanathan, 2002; Na et al., 2004).

Aunque se ha realizado una división de las distintas variables utilizadas en la literatura para la creación de los vectores de aprendizaje en la clasificación de la polaridad, lo cierto es que la gran mayoría de las aproximaciones, sobre todo en los últimos años, utilizan todas o un conjunto de ellas para representar los textos de entrada, obteniendo de esta manera mejores resultados (Wilson, Wiebe, y Hoffmann, 2005; Ng, Dasgupta, y Arifin, 2006; Abbasi, Chen, y Salem, 2008; Wilson, Wiebe, y Hoffmann, 2009; Arora et al., 2010).

Por el contrario, los métodos no supervisados se basan en fórmulas matemáticas o conjuntos de reglas para inferir la polaridad de los textos de entrada a partir de las expresiones polares que los componen. Resulta difícil realizar una clasificación precisa de los diferentes trabajos presentados en la literatura a lo largo de los años por la diversidad de técnicas utilizadas. Sin embargo, se pueden diferenciar tres grandes grupos de técnicas: las fórmulas basadas en las frecuencias de los términos o las distancias entre ellos en relación a expresiones polares; las fórmulas basadas en asignación de pesos a las expresiones polares; y las aproximaciones basadas en reglas sintácticas a partir de las expresiones polares. Los métodos basados en frecuencias calculan la polaridad en función de las probabilidades de co-ocurrencia entre términos, diferenciando entre positivos y negativos. Por el contrario, las técnicas basadas en la asignación de pesos a las expresiones polares formulan modelos que hacen uso de estas polaridades y de las intensidades asociadas a las palabras para calcular la polaridad total del texto. Finalmente, las técnicas basadas en reglas utilizan peculiaridades de la sintaxis y la gramática para determinar la polaridad total del texto.

El primer trabajo en el que se propuso el uso de frecuencias de términos para resolver el problema de la detección de la polaridad fue en (Turney, 2002), donde se presenta una fórmula matemática para el cálculo de lo que Turney denomina la *orientación semántica*. La hipótesis formulada se basa en la idea de que las palabras por sí solas tienen cierta orientación emocional, principalmente los adjetivos y los adverbios, por lo que una fórmula capaz de determinar esa orientación del texto de entrada en base a su probabilidad de co-ocurrencia con términos positivos y negativos determinaría la orientación del texto completo. Para ello, Turney divide su algoritmo en tres etapas. En la primera de ellas, el texto es analizado para extraer patrones mediante las categorías gramaticales, tal y como muestra la tabla 2.1.

First word	**Second word**	**Third word** (Not extracted)
JJ	NN or NNS	anything
RB, RBR or RBS	JJ	not NN nor NNS
JJ	JJ	not NN nor NNS
NN or NNS	JJ	not NN nor NNS
RB, RBR or RBS	VBN, VBD, VBN or VBG	anything

Tabla 2.1: Patrones para la extracción de sintagmas de dos términos (Turney, 2002)

En una segunda etapa, el algoritmo estima la orientación emocional utilizando la métrica PMI (ecuación 2.1) y la ecuación 2.2, donde se calcula la relación del sintagma extraído con la expresión polar *excellent* menos la relación del sintagma con la expresión polar *poor*, positivo y negativo respectivamente. En este caso, la probabilidad de los términos de la fórmula PMI es calculada mediante los resultados obtenidos por las búsquedas del sintagma y la palabra de referencia en el motor de búsqueda AltaVista[1], utilizando para ello la relación *NEAR* del motor que devuelve los documentos donde se encuentran las palabras buscadas con una diferencia máxima de diez palabras. Por último, cuando el sistema analiza un nuevo texto determina la orientación emocional de cada sintagma extraído clasificando el texto entero en positivo si la puntuación final es positiva, o en negativo si la puntuación final es negativa.

$$SO(sintag) = PMI(sintag, \text{“}excellent\text{”}) - PMI(sintag, \text{“}poor\text{”}) \quad (2.2)$$

La asignación de pesos a las expresiones polares para la creación de fórmulas o modelos capaces de determinar la polaridad total del texto de entrada en función de sus partes, ha sido otra de las aproximaciones típicas de los trabajos no supervisados. Un ejemplo claro de estas técnicas se puede ver en el trabajo realizado por Kim y Hovy (2004), en el que formulan tres modelos para la clasificación de la polaridad a nivel de oración en función de las polaridades e intensidades de las expresiones polares de cada oración. El primer modelo propuesto se basa en la multiplicación de las polaridades de las expresiones polares asumiendo que la negación afecta más a la polaridad, y que una doble negación neutraliza el efecto de la negación. El segundo

[1] AltaVista. http://www.altavista.com/

modelo se expresa mediante la suma de las intensidades de los términos y su frecuencia; es decir, cuantos más términos positivos y más intensos haya en la oración, más positiva será ésta, y viceversa. Por último, el tercer modelo se basa en la media geométrica de las intensidades y sus polaridades. Para determinar el ámbito o ventana de los términos importantes de cada oración, el sistema utiliza varias estrategias, como, por ejemplo, asumir todas las palabras hasta el siguiente signo de puntuación desde el tópico detectado o el sujeto de la oración.

Polanyi y Zaenen (2006) presentan una propuesta similar a la de Kim y Hovy, donde la polaridad del texto global es calculada como la suma de las polaridades e intensidades de las expresiones polares. La principal diferencia se encuentra en la apreciación de que ciertos fenómenos lingüísticos afectan fuertemente al cálculo de la polaridad, como la negación, los cuantificadores o los modales. De esta manera, el cálculo de la polaridad total del texto se basa en la suma de los diferentes valores de las expresiones polares junto con la ponderación de cada uno de ellos según los modificadores que les afecten. Por ejemplo, la negación produce el efecto de cambio de polaridad en las expresiones polares dentro de su ámbito, mientras que el término *rather* disminuye en un valor fijo la intensidad de la polaridad de las expresiones polares dentro de su ámbito, ya sean positivas o negativas.

Son numerosos los trabajos que han adoptado estas técnicas, combinándolas incluso con los principios de la teoría de grafos o con técnicas supervisadas, como los trabajos presentados en (Kennedy y Inkpen, 2006; Andreevskaia y Bergler, 2007; Devitt y Ahmad, 2007; Nicolov, Salvetti, y Ivanova, 2008; Agrawal y Siddiqui, 2009; Brooke, 2009).

Por otro lado, los sistemas que hacen uso de reglas para la la detección de la polaridad tratan de modelar fenómenos lingüísticos que afectan al cálculo global de la misma. Por ejemplo, Meena y Prabhakar (2007) hacen un estudio exhaustivo de la relación entre las conjunciones y disyunciones en el cálculo de la polaridad, en el que exponen que la polaridad global de la oración puede variar cuando se presentan conjunciones y disyunciones frente a su cálculo como la simple suma de las polaridades de las expresiones polares. Para ello, determinan la polaridad de la oración mediante una propuesta similar a la de Polanyi y Zaenen, pero utilizan un conjunto de reglas para detectar conjunciones y disyunciones y calcular, a partir de ellas, la polaridad global.

Moilanen y Pulman (2007) centran su aproximación en la misma idea de

que la polaridad total se obtiene a través de la suma de las polaridades de las expresiones polares, pero basando el cálculo global en reglas lógicas sobre la sintaxis y la gramática, en lugar de en fórmulas matemáticas. Para ello, desarrollan un analizador sintáctico de dependencias para post-procesar los árboles sintácticos y gramaticales utilizando métodos de análisis sentimental de composición. Para ello, definen reglas de propagación sentimental, resolución de conflictos en polaridad, y reversión de polaridad, según los constituyentes de entrada de cada regla. De esta manera, mediante el uso de un diccionario léxico afectivo, el sistema va propagando y ajustando la polaridad de cada elemento mediante las reglas, obteniendo finalmente la polaridad global del texto de entrada.

Otros trabajos que han desarrollado conjuntos de reglas basadas en la sintaxis y la gramática para obtener la polaridad final del conjunto del texto se pueden consultar en (Shaikh, Prendinger, y Mitsuru, 2007; Choi y Cardie, 2008).

Aunque la tarea de clasificación de la polaridad ha ganado mucha relevancia en los últimos años, lo cierto es que la gran mayoría de las aproximaciones de la literatura, tanto supervisadas como no supervisadas, no tienen en cuenta un factor que desde el punto de vista del autor de esta tesis es imprescindible a la hora de analizar emocionalmente un texto, como es el aspecto semántico. Por regla general, las aproximaciones y recursos generados para esta tarea están basados en términos o palabras, incluso aquellos que utilizan la orientación de los términos, sin tener en cuenta que una misma palabra puede tener múltiples significados, y que, en muchos casos, esos significados presentan polaridades opuestas. Las pocas aproximaciones que han trabajado a nivel de conceptos en lugar de términos no utilizan ninguna estrategia de desambiguación semántica para obtener el significado apropiado de cada término de acuerdo a su contexto, por lo que la ambigüedad, y la posibilidad de trabajar con el significado erróneo de una palabra, siguen latentes.

Por otro lado, aunque se podría considerar que las aproximaciones que utilizan las expresiones polares para el cálculo de la polaridad del texto hacen uso, de una forma superficial, de conocimiento semántico, puesto que parten de la orientación polar de los términos, según la opinión del autor de esta tesis, ese conocimiento es insuficiente para la tarea, y más aun si se desea determinar no solo la polaridad sino también su intensidad, tal y como

se explicara en la sección 1.2 mediante el ejemplo de los conceptos *cancer* y *resfriado*. Recordando, ambos conceptos tienen una orientación negativa, pero mientras que el primero suele expresar o producir emociones de *miedo* o *tristeza*, el segundo denota emociones de *disgusto* o *malestar*.

Atendiendo al rendimiento de los sistemas aquí presentados, y aunque hacer una comparación exhaustiva es imposible debido a la diferencia de métodos de evaluación, lo cierto es que, en general, se observa que los sistemas supervisados obtienen mejores resultados que los no supervisados. Sin embargo, estos sistema suelen presentar otra deficiencia común, como es no tener en cuenta fenómenos lingüísticos tan importantes para la detección de la polaridad como la negación, los cuantificadores o la modalidad. Además, los pocos trabajos que hacen un estudio de los mismos, suelen utilizarlos como variables dicotómicas de los vectores de entrenamiento (e.g., la presencia o no de negación, o la presencia o no de modalidad), sin tener en cuenta, de nuevo, la semántica del texto.

Por último, como ya se ha comentado para la tarea de la clasificación de la subjetividad, las aproximaciones simples basadas en las frecuencias de los términos suelen fallar a la hora de trabajar con documentos de diferentes dominios, debido a la diferencia de vocabulario entre unos dominios y otros, lo que produce vectores de instancias muy diferentes y evaluaciones con rendimiento altamente insatisfactorio.

2.2.4. Clasificación de la intensidad

A la hora de determinar la carga afectiva de un texto no solo es importante identificar la polaridad de dicho texto, sino también la intensidad de dicha polaridad. Así, la clasificación de la intensidad es la tarea del análisis sentimental que aborda el problema de clasificar un texto en diferentes clases de intensidad de acuerdo a su carga emocional. Dentro de esta tarea, se puede diferenciar entre las aproximaciones que abordan el problema de clasificar los textos en tres clases de intensidad (i.e., positivo, neutral y negativo), y las aproximaciones que van un paso más allá y abordan el problema de clasificar los textos en más de tres clases, como por ejemplo en cinco clases de intensidad (i.e., fuertemente positivo, positivo, neutral, negativo y fuertemente negativo). En el primer caso, suele ser común que los enfoques partan de una primera etapa en la que se afronta un problema de clasificación de la subjetividad, seguida de una segunda etapa que trata el problema de la

clasificación de la polaridad, esta última realizada únicamente sobre los textos previamente clasificados como subjetivos. Por el contrario, no ocurre así con las aproximaciones que abordan el problema de la clasificación en más de tres clases.

La tarea de la clasificación de la intensidad en tres clases pretende determinar cuándo un texto es positivo, neutral o negativo, de acuerdo a su carga emocional. Es por esto, como ya se ha dicho, que la gran mayoría de los trabajos plantean el problema como una doble tarea en la que primero se clasifican los textos de entrada en subjetivos y objetivos, y posteriormente se clasifican los textos subjetivos en positivos y negativos. No han sido muchos los trabajos de la literatura que han abordado esta tarea, debido, principalmente, a su mayor complejidad. Sin embargo, haciendo una clasificación de los escasos trabajos relevantes, se puede diferenciar, al igual que en el caso de la clasificación de la polaridad, entre las aproximaciones basadas en técnicas supervisadas y las aproximaciones basadas en expresiones polares. Un ejemplo claro, y de gran relevancia, de los métodos supervisados para la tarea de clasificación de la intensidad en tres clases es el doble clasificador presentado por Wilson et al. (2005; 2009). En estos trabajos, Wilson et al. presentan un doble clasificador que, en un primer lugar, determina los textos que son susceptibles de ser clasificados en base a su polaridad, atendiendo a la carga subjetiva detectada en este primer paso. Para ello, Wilson et al. utilizan multitud de variables para las instancias del clasificador, como las categorías gramaticales, los árboles de dependencias, la presencia o no de modales, la presencia o no de expresiones polares, etc. De la misma manera, y con variables muy similares para las instancias del segundo clasificador, la aproximación realiza en un segundo paso una clasificación en positivo, negativo y ambos, de los textos clasificados como subjetivos. Kim y Hovy (2004) abordan el problema mediante expresiones polares con un método típicamente no supervisado. Para determinar que textos conllevan una carga emotiva, y por tanto son subjetivos y deben ser clasificados posteriormente en positivos o negativos, Kim y Hovy determinan el número de expresiones polares de adjetivos, verbos y nombres existentes en cada oración, clasificando como subjetivas solo aquellas que poseen al menos una expresión polar. Para determinar la polaridad de estos textos clasificados como subjetivos, los autores proponen tres modelos basados en la presencia de negación sobre las expresiones polares, la intensidad de dichas expresiones polares y la media

geométrica de las polaridades y sus intensidades. Por último, cabe resaltar el trabajo presentado por Pang y Lee (2004), donde para determinar de una manera más eficiente la polaridad de los documentos realizan un primer paso para clasificar sus oraciones en subjetivas y objetivas, eliminando estas últimas antes de realizar la clasificación de polaridad. Para identificar las oraciones subjetivas en cada documento, Pang y Lee proponen un sistema basado en la teoría de grafos donde cada oración es representada como un nodo y el objetivo es maximizar el peso de cada arista mediante diferentes métricas en relación a un conjunto de nodos con oraciones manualmente etiquetadas como subjetivas. Por último, todas aquellas oraciones que no superan un umbral determinado de relación con las oraciones subjetivas etiquetadas manualmente, son eliminadas del documento. Una vez realizado este proceso, los documentos resultantes son traducidos a vectores de frecuencias de términos y clasificados mediante dos clasificadores, Naive Bayes y SVM, según su polaridad.

Pese a esta tendencia de dividir la tarea en dos pasos, algunas aproximaciones presentadas en la literatura han resuelto la clasificación de la intensidad en tres clases en un solo paso. Yu y Hatzivassiloglou (2003), por ejemplo, presentan un sistema basado en expresiones polares donde, para cada oración, se calcula un valor a partir de las frecuencias de sus expresiones polares, clasificando una oración como positiva si el valor final excede un umbral superior, como negativa si el valor final es inferior a un umbral inferior, y como neutral si el valor calculado se encuentra entre ambos umbrales. En esta línea de desarrollar sistemas específicos para la tarea de tres clases, se enmarcó precisamente la competición SemEval 2007 Affective Text Task (Strapparava y Mihalcea, 2007). En esta competición, los sistemas presentados debían clasificar, en base a su intensidad emocional, diferentes titulares de noticias en tres categorías: positiva, neutral y negativa. Los diferentes enfoques presentados (Andreevskaia y Bergler, 2007; Chaumartin, 2007; Katz, Singleton, y Wicentowski, 2007; Sahlgren, Karlgren, y Eriksson, 2007) se basan en las mismas técnicas que los trabajos aquí expuestos, mediante técnicas supervisadas y no supervisadas, pero en este caso todos los sistemas hacen uso de un léxico afectivo basado en conceptos en lugar de términos proporcionado en la competición.

Si son escasos los trabajos que se pueden encontrar en la literatura que resuelven la clasificación de la intensidad en tres clases, son casi inexistentes

los que intentan clasificar dicha intensidad en más de tres clases. Uno de los primeros trabajos en hacerlo fue el presentado por Pang y Lee (2005), donde se realiza una serie de experimentos con distintas configuraciones de clasificadores SVM (un solo clasificador, múltiples clasificadores, clasificadores con regresión y métricas de etiquetado) para clasificar opiniones sobre películas en tres y cuatro clases de intensidad. Por el contrario, Devit y Ahmand (2007) utilizan las expresiones polares, extraídas de SentiWordNet, para clasificar la intensidad en siete clases, que varían desde muy positivo a muy negativo. Para ello, la aproximación representa cada texto de entrada como un grafo, donde cada nodo es un concepto de la base de datos léxica WordNet, extraído sin un proceso de desambiguación, y las aristas son relaciones entre conceptos de WordNet. Para determinar la intensidad de cada grafo, utilizan diferentes métricas basadas en las expresiones polares y en la estructura y topología del grafo, como la suma de todas las expresiones polares del grafo o el cálculo de la importancia o prestigio de cada nodo para ponderar dicho nodo de acuerdo al valor de la expresión polar. Los autores concluyen que, aunque el léxico afectivo es útil para la tarea de clasificación de la polaridad, sería preciso un léxico afectivo con mayor información para poder mejorar los resultados obtenidos.

También basada en el uso de expresiones polares, Brooke (2009) presenta una aproximación denominada *calculador de orientación semántica*, que calcula una orientación emocional del texto a partir de las diferentes expresiones polares que lo componen. Además, Brooke contempla el problema del efecto de los modificadores emocionales, como la negación, los cuantificadores y los modales, sobre las expresiones polares dentro de su ámbito de acción. Brooke evalúa su aproximación en la tarea de clasificación de la intensidad en tres, cuatro y cinco clases, concluyendo que el rendimiento decrece a razón de casi el 15 % de media en exactitud al incluir cada nueva clase.

Por último, Saggion et al. (2010) plantean la hipótesis de que la generación de resúmenes guiados, mediante consultas y expresiones polares, de documentos con opiniones, puede ayudar a mejorar los resultados de la tarea de clasificación de la intensidad en cinco clases. Para ello, desarrollan un sistema supervisado mediante clasificadores SVM que evalúan sobre los textos completos de las opiniones y sobre los resúmenes generados con diferentes tasas de compresión (entre el 10 % y el 50 % del texto original). Para en-

trenar el clasificador utilizan, al igual que muchos otros trabajos previos, diferentes métricas para asociar intensidades y frecuencias a las expresiones polares de cada texto.

Todos estos trabajos plantean debilidades similares a las expuestas para la tarea de la clasificación de la polaridad. Por un lado, las diferentes aproximaciones que abordan la tarea de clasificación de la intensidad mediante técnicas supervisadas no suelen tener en cuenta fenómenos lingüísticos como la negación, los cuantificadores y los modales, así como su efecto a la hora de determinar la intensidad emocional del texto. Además, los pocos trabajos que han tratado estos fenómenos los enfocan de una manera muy simplista, teniendo en cuenta únicamente su presencia o frecuencia en los textos, y representando estos como una variable más de la instancias del algoritmo de aprendizaje máquina. Por otro lado, tanto los métodos supervisados como los no supervisados suelen utilizar el término o palabra como unidad principal para determinar las expresiones polares, lo que supone posibles errores al considerar significados erróneos, puesto que una misma palabra, como ya se ha dicho, puede tener múltiples significados de los cuales algunos pueden connotar emociones positivas, y otros, negativas. Por otro lado, las pocas aproximaciones que hacen uso de conceptos en lugar de términos no suelen realizar un proceso previo de desambiguación semántica.

Es también importante tener en cuenta el componente semántico de la tarea, que al igual que en la clasificación de la polaridad, puede influir en gran medida en los resultados obtenidos. Es por esto que en este tipo de sistemas, con mayor importancia que en los planteados para la tarea de clasificación de la polaridad, es imprescindible discernir y matizar los sentimientos o emociones expresados en el texto, ya que en muchos casos la información de si una palabra es positiva o negativa no es suficiente, y es preciso un modelado más complejo y rico en significado para determinar correctamente la intensidad emocional.

Por último, los enfoques que plantean la tarea de clasificación de la intensidad como un doble proceso, no permiten una extrapolación de los sistemas a tareas más complejas con diferentes clases de intensidades. Además, el segundo proceso suele ser muy especializado para la tarea de clasificación de la polaridad, lo que impide aún más el abordar otro tipo de tareas que, aun siendo similares, requieren mayor análisis y conocimiento al aumentar considerablemente la dificultad de la tarea con cada nueva clase incluida.

2.3. La negación en el análisis sentimental

La importancia de la negación en las tareas del análisis sentimental ha sido ampliamente estudiada en los últimos años (Na et al., 2004; Whitelaw, Garg, y Argamon, 2005; Andreevskaia y Bergler, 2007; Jia, Yu, y Meng, 2009). Los trabajos tradicionales del análisis sentimental, como ya se ha comentado, utilizan métodos supervisados que modelan el texto como un conjunto de frecuencias de palabras (Pang, Lee, y Vaithyanathan, 2002), construyendo un vector donde las entradas corresponden a palabras individuales, y donde a cada palabra se le asigna su frecuencia en el texto o un valor binario indicando su presencia. Otros han evaluado diferentes estrategias basadas en las frecuencias como el uso del n-gramas en lugar de palabras individuales, el uso de categorías gramaticales, etc. (Pang, Lee, y Vaithyanathan, 2002; Dave, Lawrence, y Pennock, 2003; Na et al., 2004; Moilanen y Pulman, 2007; Shaikh, Prendinger, y Mitsuru, 2007).

Estas aproximaciones simples no tienen en cuenta ningún conocimiento explícito de las expresiones emocionales, por lo que la negación precisa ser modelada de una manera muy superficial. Das y Chen (2001), por ejemplo, presentan un sistema para la detección y tratamiento de la negación que simplemente se basa en añadir el sufijo *NOT* a las palabras que se encuentran cerca de un término de negación para diferenciar en el vector de frecuencias de términos entre las palabras negadas y las no negadas. De esta manera, en la oración *I did<u>n't</u> like the movie*, el término *like* es transformado en el término *like-NOT*. Respecto al ámbito de acción de la negación, en este tipo de aproximaciones es habitual asumir como ámbito todos aquellos términos entre la señal de negación y el primer signo de puntuación (Pang, Lee, y Vaithyanathan, 2002), o considerar como ámbito de acción un número fijo de términos después del término de negación (Hu y Liu, 2004). Sin embargo, puesto que no todas las palabras próximas a un término de negación están necesariamente afectadas por él, este tipo de aproximaciones superficiales fallan a la hora de delimitar y resolver el efecto, por ejemplo, del término *best* en la oración *<u>No</u> doubt this hotel is the best*, puesto que este término no está dentro del ámbito de acción del término de negación *No*. Como resultado, y tal y como defienden Wiegand et al. (2010), este tipo de representaciones son lingüísticamente ineficientes, por lo que se traducen, por regla general, en mejoras casi despreciables en comparación con las aproximaciones que no consideran dicho efecto.

Las aproximaciones más recientes en el área de análisis sentimental representan el texto mediante las denominadas *expresiones polares*, que, como ya se ha dicho, son términos o conjuntos de términos que poseen una polaridad en sí mismos. Las expresiones polares se obtienen generalmente de léxicos afectivos basados en la polaridad de los términos, que anotan palabras o conceptos como negativos, neutrales o positivos, como por ejemplo el léxico afectivo *General Inquirer* (Stone et al., 1966) o SentiWordNet (Esuli y Sebastiani, 2006b). En este contexto, las negaciones son consideradas como un modificador que invierte la polaridad de las expresiones polares, de manera que las expresiones polares negadas son transformadas en expresiones polares sin negación con la polaridad opuesta.

Una de las aproximaciones más populares en este tipo de sistemas es la propuesta por Polanyi y Zaenen (2006) en su estudio sobre los modificadores polares. En este trabajo, los autores modelan los textos mediante expresiones polares, asignando valores positivos a las expresiones polares positivas y valores negativos a la expresiones polares negativas. De esta manera, si una expresión polar se encuentra dentro del ámbito de acción de una negación, el efecto de esta última es simplemente invertir la polaridad del valor asignado a la expresión polar. Por ejemplo, si la palabra *successful* es etiquetada con un valor positivo de valencia de *+2*, el mismo término precedido de un término de negación, como en *not successful*, es asignado el valor *-2*. Una aproximación muy similar se puede encontrar en el trabajo de Kennedy y Inkpen (2006). En este caso, los autores desarrollan un sistema con el que comprueban que el procesamiento de la negación contribuye a mejorar el rendimiento del mismo al evaluarlo en la tarea de clasificación de la polaridad a nivel de documento. Sin embargo, nada se dice en el artículo acerca de la mejora exacta conseguida, por lo que no hay datos reales sobre la efectividad de estas aproximaciones cuando se procesa la negación. Así mismo, tampoco se describe el proceso seguido para detectar el ámbito de acción de la negación.

También inspirados en esta idea de tratar la negación como un modificador de la polaridad, Wilson et al. (2005; 2009) modelan la negación como tres tipos de variables o atributos de las instancias del clasificador: *variables de negación*, que representan, por ejemplo, la presencia de una negación en una ventana fija de cuatro palabras antes de la expresión polar; *modificadores de variable*, que hacen referencia a la presencia de diferentes modificadores de

la polaridad, como por ejemplo *little*; y *variables de modificación de la polaridad*, que son palabras que no son explícitamente negaciones pero modifican las expresiones polares, como por ejemplo *disappointed* o *hope*. Estas tres variables de negación son añadidas como variables a un vector de aprendizaje compuesto de frecuencias de términos, dependencias sintácticas, número de expresiones polares, etc., y utilizado como entrada de un algoritmo supervisado para la tarea de clasificación de la polaridad a nivel de sintagma. La principal limitación de esta aproximación es el no detectar el ámbito de acción de la negación adecuadamente, así como no tratar directamente el efecto de dicha negación sobre las expresiones polares, sino como variables binarias en las instancias del algoritmo.

Un trabajo más avanzado se presenta en (Choi y Cardie, 2008). Su hipótesis es asumir que las palabras o constituyentes dentro de una expresión interactúan unos con otros para determinar la polaridad total de la expresión. La polaridad de las palabras y constituyentes es obtenida mediante un léxico afectivo, y la polaridad total entre palabras y constituyentes es obtenida mediante un conjunto de reglas de inferencia para derivar dicha polaridad total de la polaridad individual de sus constituyentes. Mediante estas reglas, además, se restringe el ámbito de acción de la negación a constituyentes específicos, en lugar de utilizar el ámbito de acción de las expresiones completas.

Todas estas aproximaciones presentan ciertas limitaciones a la hora de abordar el efecto de la negación desde el punto de vista del análisis sentimental. Por un lado, casi todas consideran la negación como un modificador que siempre invierte la polaridad de la opinión expresada, pero en muchos casos la negación también disminuye la fuerza de dicha polaridad. Consideremos, por ejemplo, el término de negación *not* en la oración *Football is <u>not</u> my favorite sport*. En este caso, la negación no significa necesariamente que el emisor odie el fútbol, sino más bien que no siente entusiasmo por él. Es también importante resaltar que, en muchos casos, el efecto de la negación es tratado como una simple variable más dentro del conjunto de atributos de los algoritmos de aprendizaje máquina. Esto supone un tratamiento muy superficial de la negación, que en mucho casos conlleva una perdida de información importante que de otra manera permitiría determinar con mayor precisión la carga emocional del texto. Por otro lado, pocos son los trabajos que intentan delimitar el ámbito de acción de la negación. Por último, en

todos los casos se asume que la tarea de clasificación de la polaridad depende exclusivamente de la orientación polar de las palabras o conceptos inmersos en el texto, sin tener en cuenta las emociones concretas expresadas, y que en muchos casos pueden ayudar a capturar de una manera mucho más eficiente el significado afectivo del texto, así como su polaridad.

2.4. Los cuantificadores en el análisis sentimental

Los cuantificadores son términos especiales que modifican el grado o intensidad de las emociones dentro de su ámbito de acción. El efecto de los cuantificadores puede ser amplificar la fuerza de las emociones o, por el contrario, disminuir dicha fuerza. Por ejemplo, en la oración *The hotel is terribly good*, la expresión *terribly good* es más positiva que simplemente *good*. Por el contrario, el cuantificador *less* en la oración *It is less funny than I was expecting* disminuye la intensidad de la palabra *funny*. Los cuantificadores son también comúnmente denominados *intensificadores/amplificadores* o *disminuidores* (Whitelaw, Garg, y Argamon, 2005; Andreevskaia y Bergler, 2007; Meena y Prabhakar, 2007).

El método más utilizado en la literatura para la detección de cuantificadores consiste en el uso de listas de palabras que actúan como cuantificadores cuando juegan un rol gramatical determinado en la oración (modificadores adverbiales y adjetivales). Una de las principales referencias a la hora de desarrollar una lista de cuantificadores es el libro de Quirk et al. (1985), en el que los autores hacen un exhaustivo análisis de las estructuras lingüísticas del inglés. Sin embargo, la gran mayoría de las aproximaciones que abordan el problema de los cuantificadores no suelen realizar un estudio en profundidad para incluir todos los cuantificadores posibles, sino que, en general, se limitan a contemplar los más frecuentemente utilizados en el inglés, y que por ende serán los que mayor impacto tendrán en la tarea a abordar.

La mayoría de los trabajos que utilizan estas listas de cuantificadores suelen asociar un valor fijo para todos los cuantificadores que amplifican la fuerza de las emociones, y otro valor fijo para todos los cuantificadores que la disminuyen. Polanyi y Zaenen (2006), por ejemplo, utilizan una lista de términos para detectar la presencia de los cuantificadores y determinan su efecto añadiendo o restando un “punto” a la valencia asignada a las

expresiones polares dentro del ámbito de acción de los cuantificadores. De esta manera, el cuantificador *very* en la expresión *very suspicious*, donde la expresión polar *suspicious* tiene asignada una valencia de *-2*, amplifica dicha valencia, asignando a *suspicious* el valor final *-3*; mientras que, por el contrario, el cuantificador *little* en la expresión *little efficient* disminuye la intensidad de la expresión polar *efficient* de *+2* a *+1*.

En esta misma línea, Kennedy y Inkpen (2006) asignan a cada expresión polar positiva un valor de *+2*. Atendiendo a esto, si una expresión polar positiva es precedida por un cuantificador que amplifique dicha expresión el valor asignado final es *+3*, mientras que si la expresión polar positiva es precedida por un cuantificador que disminuya dicha expresión, el valor final asignado es *+1*. En el caso de las expresiones polares negativas, Kennedy y Inkpen asignan un valor a cada una de *-2*, que será *-1* o *-3* si van precedidas por cuantificadores que amplifiquen o disminuyan, respectivamente, dicha expresión polar. La versión mejorada de este sistema utiliza una aproximación supervisada en la que las variables de las instancias del algoritmo de aprendizaje son bi-gramas que contienen información combinada de la presencia de cuantificadores y la expresión polar. De esta manera, por ejemplo, el bi-grama *very good* es representado mediante la variable *int_good*, donde *int* indica la presencia de cualquier cuantificador que amplifique su intensidad; mientras que el bi-grama *barely good* es representado por la variable *dim_good*, donde *dim* indica la presencia de cualquier cuantificador que disminuya dicha intensidad.

Wilson et al. (2009) siguen un enfoque similar a la aproximación mejorada de Kennedy y Inkpen, modelando los cuantificadores como dos variables binarias de las instancias del algoritmo de aprendizaje máquina. La variable *precedido por intensificador* es activada si la palabra o expresión polar es precedida por un cuantificador que amplifique. La variable *intensificador en sí mismo* es activada si el propio término es un cuantificador que amplifique. A diferencia de otros trabajos, Wilson et al. no distinguen entre cuantificadores que amplifican y cuantificadores que disminuyen, puesto que estos últimos son tratados en su aproximación como términos especiales de negación, centrándose simplemente en los cuantificadores que amplifican, a los que denominan *intensificadores*. De esta manera, una palabra es considerada un *intensificador* si dicha palabra aparece en una lista de intensificadores y es precedida por una palabra con el apropiado rol gramatical.

Sin embargo, todas estas aproximaciones presentan importantes limitaciones. Por un lado, y al igual que en la negación, son muchos los trabajos que no abordan su efecto real sobre las emociones o expresiones polares, sino que simplemente utilizan la presencia o no de los cuantificadores como variables en las instancias de los algoritmos de aprendizaje máquina. De esta manera, aunque se determina su presencia y parte de su impacto, se pierde mucha información real sobre cómo el cuantificador afecta a los términos dentro de su ámbito de acción. Por otro lado, las aproximaciones que utilizan valores fijos para determinar el efecto de los cuantificadores sobre los términos no son capaces de distinguir la propia intensidad del efecto de los cuantificadores, puesto que no todos los cuantificadores intensifican o disminuyen de la misma manera. Por ejemplo, la oración *The movie is terribly funny* es indudablemente más positiva que la oración *The movie is quite funny*, debido principalmente a la diferencia de intensidades de los cuantificadores *terribly* y *quite*. De la misma manera, los cuantificadores que disminuyen poseen diferentes intensidades, como por ejemplo, los cuantificadores *hardly* y *almost* en las oraciones *I hardly love you* y *I almost love you*. Es por esto que parece necesario un sistema que asigne a cada cuantificador un efecto acorde a su intensidad, así como que dicho efecto se traduzca en el impacto real de dichos cuantificadores sobre los términos dentro de su ámbito de acción.

2.5. La modalidad en el análisis sentimental

Los operadores modales permiten a los humanos distinguir entre eventos o situaciones que han ocurrido, que están ocurriendo o que van a ocurrir (eventos *reales*), y aquellos que podrían o probablemente deberían suceder (eventos *irreales*) (Polanyi y Zaenen, 2006). A lo largo de los años, se han propuesto múltiples definiciones y clasificaciones de los modales en la literatura, cada una de ellas basada en diferentes aspectos como el aspecto semántico, lingüístico o filosófico. Por ejemplo, von Fintel (2006) define la modalidad en relación a la necesidad y posibilidad, clasificándola de esta manera en función de su significado en *modalidad alética*, *modalidad epistémica*, *modalidad deóntica*, *modalidad bouletic*, *modalidad circunstancial* y *modalidad teleológica*. Palmer (2001) propone, por el contrario, una clasificación de la modalidad basada en el punto de vista del receptor, distinguiendo entre *modalidad*

proposicional, dividida en *epistémica* y *evidencial*, y *modalidad de eventos*, dividida a su vez en *deóntica* y *dinámica*. Por último, Baker et al. (2010) clasifican la modalidad en base a las indicaciones percibidas por el emisor o el receptor, clasificándola en *factual*, *evidencial* y *sentimental*.

El área de investigación de la modalidad en procesamiento del lenguaje natural ha sido tradicionalmente asociado, e incluso comparado, con el estudio de la presencia de especulación o incertidumbre en el texto a analizar. Esta equiparación ha sido principalmente motivada por los estudios realizados sobre la presencia de certeza o incertidumbre en textos biomédicos (Light, Qiu, y Srinivasan, 2004; Velupillai, 2010; Goldstein y Uzuner, 2010), donde el objetivo principal es separar los fragmentos de texto que expresan veracidad sobre una afirmación o proposición, de los que no. De hecho, la modalidad ha sido principalmente interpretada desde esta perspectiva por parte de la comunidad científica y comúnmente denominada como *hedging* ("palabras cuyo trabajo es hacer las cosas más o menos difusas" (Lakoff, 1973)) o *factuality* ("información que determina cuándo un evento mencionado en el texto se corresponde con una situación real o, por el contrario, con una situación con cierto grado de incertidumbre" (Sauri y Pustejovsky, 2009)).

En este sentido, Light et al. (2004) presentan un estudio preliminar de cómo la incertidumbre es expresada en artículos de biomedicina, concluyendo con la estimación de que alrededor del 11 % de las oraciones en los resúmenes de los artículos pueden ser clasificadas como especulativas. Light et al. también presentan un sistema automático para la clasificación de textos en especulativos o no, obteniendo unos resultados altamente satisfactorios para todos los experimentos realizados (≈90 % de exactitud). Medlock y Briscoe (2007) presentan un sistema automático para la clasificación de oraciones en las categorías de especulativo y no especulativo, donde la principal contribución es el exhaustivo análisis realizado por los autores para la descripción de la tarea, así como la definición de qué debe y qué no debe ser una especulación. La detección de la especulación en artículos de biomedicina se ha convertido en un importante área de investigación, dada su utilidad contrastada en tareas como la extracción automática de relaciones biológicas a partir de artículos científicos, o la tarea de determinar automáticamente síntomas de enfermedades a partir de historiales medicos (Rokach, Romano, y Maimon, 2008; Szarvas, 2008; Van Asch, Morante, y Daelemans, 2010;

Dalianis y Skeppstedt, 2010). De hecho, este auge de la investigación en la clasificación automática de textos atendiendo a la especulación ha motivado recientes competiciones como la realizada en la conferencia CONLL (*Conference on Natural Language Learning*) en el año 2010, donde la tarea principal consistía en detectar fragmentos de incertidumbre en artículos de biomedicina y artículos de Wikipedia.

Desde un punto de vista lingüístico, la modalidad puede ser definida como la actitud o punto de vista del emisor hacia el estado actual del mundo. Carter y McCarthy (2006) destacan dos significados principales para la modalidad desde el punto de vista lingüístico. El primero de ellos tiene que ver con la probabilidad, posibilidad y certeza, con las evaluaciones de posibles hechos o eventos, y con las deducciones hechas en base a esas evaluaciones. El segundo significado hace referencia a los eventos o situaciones realizadas, o a la toma de control del curso de dichos eventos o situaciones, generalmente expresando cuándo algo es obligado, necesario, deseable, o debe estar prohibido o permitido. Carter y McCarthy defienden que las expresiones lingüísticas modales más importantes en inglés están representadas por los verbos modales, que a su vez se dividen en *verbos modales principales* (*can, could, may, might, will, shall, would, should* y *must*), y *verbos semi-modales* (*dare, need, ought to* y *used to*). Los autores también incluyen como formas modales otras expresiones típicas del inglés, como *had better, be meant to, be obliged to* o *be supposed to*, y otros verbos que, aun no siendo modales, actúan como tales en algunas situaciones, como los verbos *hope, manage, suppose, seem, wish* o *want*

La modalidad es considerada como un factor importante en el análisis sentimental (Polanyi y Zaenen, 2006), pero ha sido en los últimos años cuando realmente ha atraído el interés de la comunidad científica. Tal y como los autores exponen, la modalidad puede neutralizar la polaridad de las expresiones polares dentro de su ámbito de acción, como sucede en la oración *Breakfast <u>could</u> be much better considering the room rate*, donde el verbo modal *could* neutraliza la positividad de la oración *much better*. La incertidumbre expresada mediante las formas modales puede también afectar al significado emocional de los textos. Sin embargo, como afirman Baker et al. (2010), la incertidumbre generalmente no modifica el significado emocional, puesto que las connotaciones afectivas de los textos son positivas o negativas independientemente de si el evento o situación es un hecho o no. La

oración *I would like to kill you* claramente expresa que la acción de matar no ha sido realizada, pero incluso en esta situación la polaridad negativa de la oración es incuestionable. De hecho, desde la perspectiva del análisis afectivo, la modalidad generalmente modifica la polaridad e intensidad de los textos dentro de su ámbito de acción cuando las formas modales utilizadas expresan deseos, necesidades o quejas, que es prácticamente el segundo significado propuesto por Carter y McCarthy (2006) en su descripción sobre las formas modales.

Son muy pocas, a día de hoy, las aproximaciones que han estudiado y tratado de resolver el efecto de las formas modales en sistemas de análisis sentimental. Un trabajo pionero en el estudio de la modalidad en el contexto de la tarea de clasificación de la subjetividad es el realizado por Wiebe et al. (1999), donde se estudia si la presencia o no de formas modales afecta a la detección de la subjetividad. Las conclusiones de este trabajo, junto con otras, se presentan en (Wilson, Wiebe, y Hoffmann, 2009). En este trabajo, la modalidad es utilizada en un método supervisado como una variable binaria que indica simplemente la presencia o no de formas modales en el texto, y que es añadida a la instancia que representa el texto para el algoritmo de aprendizaje máquina. Brooke (2009) también analiza el uso de la modalidad en los textos afectivos, pero en este caso el estudio se centra en determinar los fragmentos de texto que contienen términos *irrealistas* para neutralizar las expresiones polares que se encuentran dentro de su ámbito de acción. Sin embargo, aunque se han realizado algunos estudios sobre la modalidad, ningún trabajo previo en la literatura ha estudiado el impacto real de la modalidad en las distintas tareas del análisis sentimental, así como tampoco el efecto de las diferentes formas modales en las emociones y sentimientos dentro del ámbito de acción de dichas construcciones lingüísticas.

Capítulo 3

Herramientas y recursos

La finalidad de este capítulo es describir las diferentes herramientas y recursos que han sido analizados y utilizados para la realización del sistema expuesto en esta tesis, y que no han sido desarrolladas por el autor. Así mismo, para cada herramienta se argumentan los motivos que han llevado a esa elección, describiendo las ventajas y limitaciones de los sistemas elegidos frente a otras alternativas disponibles.

3.1. Herramientas lingüísticas

Las herramientas lingüísticas se pueden definir como programas o aplicaciones que implementan técnicas y métodos para realizar análisis lingüísticos a distintos niveles dado un texto de entrada. Ejemplos claros de estas herramientas son los analizadores sintácticos, los etiquetadores gramaticales o los sistemas para la división del texto en oraciones.

Como en la mayoría de los sistemas de procesamiento del lenguaje natural que pretenden abordar tareas complejas, en el sistema desarrollado en esta tesis es preciso realizar un pre-análisis del texto de entrada a partir del cual se extrae la información lingüística necesaria para la correcta realización de las sucesivas etapas, como la identificación de conceptos del texto de entrada, la detección de la negación y su ámbito, etc. Puesto que el desarrollo de estás técnicas y métodos se escapa del ámbito de esta tesis, se ha optado por evaluar un conjunto de herramientas disponibles para investigación, determinando, para cada una de ellas, su adecuación al sistema desarrollado, y seleccionando finalmente las más convenientes. Por ello, en las siguientes secciones se exponen las alternativas elegidas, describiendo su metodología

y su modo de operación, así como su salida e integración con el sistema. En primer lugar, se explica el analizador sintáctico de dependencias Stanford Parser, y a continuación, se presenta la arquitectura para ingeniería textual Gate, exponiendo en cada caso los motivos de su elección y su función en el sistema.

3.1.1. Stanford Parser

El analizador sintáctico de dependencias Stanford Parser[1] (Klein y Manning, 2003), es un analizador estadístico de código libre desarrollado por la Universidad de Stanford que combina el análisis sintáctico de un texto con el análisis de dependencias. Un analizador sintáctico es un programa que pretende determinar las estructuras gramaticales de las oraciones, como por ejemplo, qué conjunto de palabras forman una entidad y qué tipo de entidad es (e.g., sujeto, predicado u objeto directo). Los analizadores sintácticos probabilísticos utilizan conocimientos del lenguaje mediante conjuntos de entrenamiento realizados a mano, tratando de reproducir mediante reglas estadísticas un análisis similar sobre oraciones nuevas. Aunque aún a día de hoy este tipo de herramientas no son precisas al 100 %, lo cierto es que los analizadores estadísticos supusieron un gran avance en el procesamiento del lenguaje en los años 90, siendo actualmente la mejor aproximación para este tipo de tareas. En concreto, el Stanford Parser utiliza una gramática probabilística libre de contexto y no-lexicalizada para determinar la estructura sintáctica de las nuevas oraciones y sus dependencias, donde cada sintagma anotado está asociado a una probabilidad a partir de la cual se determina el conjunto de las categorías de la oración.

El funcionamiento del Stanford Parser es muy similar al de otros analizadores léxicos de dependencias. Presenta tres posibles salidas para una misma oración de entrada, pudiendo elegir entre solo una o todas. La primera de ellas muestra la salida del analizador para la tarea de etiquetado gramatical de los diferentes términos de la oración de entrada (lo que en inglés se conoce como *Part of Speech*). Es decir, dada una oración el analizador devuelve una salida donde cada palabra está etiquetada con su categoría gramatical (nombre, verbo, adjetivo, adverbio, etc.). Por ejemplo, la figura 3.1 muestra la salida del Stanford Parser para la oración de entrada *This hotel is new or recently renovated with big nice rooms*. El etiquetado utilizado

[1] Stanford Parser. http://nlp.stanford.edu/software/lex-parser.shtml

en el analizador sigue el esquema propuesto en el *Penn Treebank Project*[2], donde, por ejemplo, el tag *NN* identifica a los sustantivos comunes, el *VBZ* a los verbos en presente y tercera persona del singular, el *JJ* a los adjetivos o el *RB* a los adverbios, y que se puede considerar un estándar, ya que es el utilizado por otras muchas herramientas de procesamiento del lenguaje natural, como por ejemplo Gate.

```
This/DT hotel/NN is/VBZ new/JJ or/CC recently/RB renovated/VBN
with/IN big/JJ nice/JJ rooms/NNS
```

Figura 3.1: Etiquetado gramatical del Stanford Parser

La segunda opción de salida del Stanford Parser devuelve el árbol sintáctico generado para la oración de entrada. En esta representación se muestran tanto el etiquetado gramatical de cada palabra como las relaciones entre cada grupo de palabras desde el punto de vista sintáctico, representando estas relaciones mediante una estructura de árbol en el que las hojas son las palabras de la oración. Al igual que en el etiquetado gramatical, el etiquetado en esta salida sigue el esquema propuesto en el *Penn Treebank Project*. En la figura 3.2 se puede ver la salida del analizador para la oración de ejemplo, donde cada grupo de palabras es etiquetado con su respectiva función sintáctica dentro de la oración y encapsulado entre paréntesis. Por ejemplo, el sintagma *This hotel* es identificado y etiquetado como *NP*; es decir, como sintagma nominal, mientras que, por ejemplo, el sintagma *is new or recently renovated with big nice rooms* es etiquetado como *VP*; es decir, como sintagma verbal.

Por último, la tercera salida del Stanford Parser devuelve las dependencias generadas para la oración de entrada. Estas dependencias están basadas en las relaciones gramaticales existentes entre los diferentes términos relacionados mediante la sintaxis para formar estructuras como objeto directo, modificadores adverbiales, etc. Es por esto que, en esta representación, cada dependencia esta formada por una función de dos parámetros; es decir, por los dos términos que están relacionados, donde el nombre de la función expresa la relación gramatical correspondiente. Cada término lleva a su vez un número asociado indicando la posición de dicho término en la oración de entrada. Las etiquetas utilizadas para determinar las dependencias entre

[2]Penn Treebank Project. http://www.cis.upenn.edu/~treebank/

```
(ROOT
  (S
    (NP (DT This) (NN hotel))
    (VP (VBZ is)
      (UCP
        (ADJP (JJ new))
        (CC or)
        (VP
          (ADVP (RB recently))
          (VBN renovated)
          (PP (IN with)
            (NP (JJ big) (JJ nice) (NNS rooms))))))))
```

Figura 3.2: Árbol sintáctico del Stanford Parser

los términos están basadas en las relaciones gramaticales de las palabras y estructuradas jerárquicamente siguiendo en cierto modo las aproximaciones lingüísticas expuestas en (Carroll, Minnen, y Briscoe, 1999), y cuya explicación, composición y estructura se puede consultar en el trabajo (Marneffe, Maccartney, y Manning, 2006). En la figura 3.3 se presenta un ejemplo de este formato de salida, donde, por ejemplo, la función *amod* representa un modificador adjetival de la palabra *big* sobre la palabra *rooms*, y la función *pobj* indica que el objeto de la preposición *with* recae sobre la palabra *rooms*.

```
det(hotel-2, This-1)
nsubjpass(new-4, hotel-2)
auxpass(new-4, is-3)
cc(new-4, or-5)
advmod(renovated-7, recently-6)
conj(new-4, renovated-7)
prep(renovated-7, with-8)
amod(rooms-11, big-9)
amod(rooms-11, nice-10)
pobj(with-8, rooms-11)
```

Figura 3.3: Dependencias del Stanford Parser

Otra característica importante de este analizador sintáctico es su capacidad para trabajar en distintos idiomas, simplemente entrenando al analizador convenientemente para ello. Para poder trabajar en otros idiomas, el analizador necesita de un conjunto de oraciones etiquetado sintácticamente.

El formato de este conjunto de entrenamiento puede seguir el estándar del *Penn Treebank Project* o incluso uno diferente. Actualmente el analizador está disponible para el inglés, alemán, chino y árabe.

Por último, Stanford Parser proporciona dos modos de ejecución, mediante consola y mediante una interfaz de desarrollo. La ejecución mediante consola simplemente requiere los parámetros apropiados, así como determinar el tipo de salida y especificar si se desea obtenerla en un archivo o en consola. La interfaz de desarrollo permite ejecutar, entrenar y realizar modificaciones sobre el analizador para obtener mejoras o nuevas capacidades. El grupo de investigación responsable proporciona una API para poder llevar a cabo esta tarea.

Uno de los principales motivos que ha llevado a la selección del analizador sintáctico Stanford Parser para el desarrollo del sistema presentado en esta tesis es su elevado porcentaje de acierto, en torno al 90 %, así como sus múltiples formatos de salida que permiten realizar todas las tareas necesarias para el desarrollo del sistema. Además, este analizador dispone de un *plugin* para la herramienta Gate, que se describirá en la siguiente sección, haciendo su uso más sencillo y transparente. Por último, la posibilidad de portar el analizador del idioma inglés al español facilita la adaptación del sistema al español en un trabajo futuro.

3.1.2. Gate

La herramienta Gate[3], del inglés *General Architecture for Text Engineering* (Cunningham et al., 2002), es una arquitectura genérica para la realización, de una forma fácil, visual y transparente para el usuario, de análisis lingüísticos complejos que, por regla general, conllevan la utilización de más de un recurso. Gate ha sido desarrollada por la Universidad de Sheffield (Reino Unido) en respuesta a la necesidad de realizar un proceso de estandarización para el desarrollo de herramientas de análisis lingüísticos. El proyecto surgió en 1995 y actualmente sigue activo con múltiples actualizaciones y mejoras, tanto en relación a su rendimiento como a la incorporación de nuevos recursos y funcionalidades.

Por regla general, y tal y como exponen los autores de Gate, un sistema de ingeniería lingüística hace uso de tres componentes básicos: algoritmos, datos de entrada y una interfaz gráfica para representar los datos. Aunque

[3]GATE. http://gate.ac.uk/

en ciertas ocasiones un sistema puede prescindir de alguno de estos tres componentes, principalmente de la visualización, lo cierto es que la gran mayoría de los sistemas actuales incluyen los tres. Es por esto que en Gate los recursos están clasificados en las siguientes tres categorías:

- *Recursos lingüísticos*: en esta categoría se encuadran todos aquellos recursos como córpora, ontologías o léxicos.
- *Recursos de procesamiento*: que incluyen todos los algoritmos de procesamiento del texto de entrada, como analizadores sintácticos, generadores de lenguaje natural, etc.
- *Recursos de visualización*: los métodos para visualización y edición de los datos analizados se enmarcan en esta categoría.

Gate incluye un núcleo que implementa la arquitectura para la interpretación y ejecución de los diferentes recursos, y un amplio conjunto de recursos o *plugins* para la realización de la tareas. Estos *plugins* son recursos lingüísticos, de procesamiento o de visualización, y pueden ser desarrollados por cualquier persona, ente público o privado, siempre y cuando se adapten al estándar de Gate CREOLE (del inglés *Collection of REusable Objects for Language Engineering*). Los recursos en CREOLE necesitan de un XML específico que Gate sabe interpretar, y donde es preciso incluir información como cuál es la clase que implementa el recurso, los parámetros necesarios, los iconos, las dependencias con otros recursos, etc. Gate utiliza todos estos archivos para determinar cuántos *plugins* hay en el sistema, así como su funcionamiento y forma de ejecución.

Existen dos formas de ejecutar una aplicación en Gate, mediante la interfaz gráfica y mediante su interfaz de programación. Un ejemplo de la interfaz gráfica se puede ver en la figura 3.4. La interfaz gráfica muestra con claridad cómo se crean y ejecutan aplicaciones lingüísticas complejas mediante Gate. Como se puede ver en la imagen, a la izquierda de la ventana se encuentran los recursos creados y utilizados; en este caso, se ha creado la aplicación *Analizador de texto*, que dispone de los recursos lingüísticos y de los recursos de procesamiento listados abajo. En el centro de la ventana se observan los diferentes recursos, a la izquierda los disponibles y a la derecha los utilizados en su orden adecuado. Como se puede ver, al seleccionar los diferentes recursos de procesamiento se muestran los parámetros de config-

uración, donde, por ejemplo, se indica el recurso lingüístico a analizar, en este caso, el documento *GATE Document_0002F* con el texto de entrada.

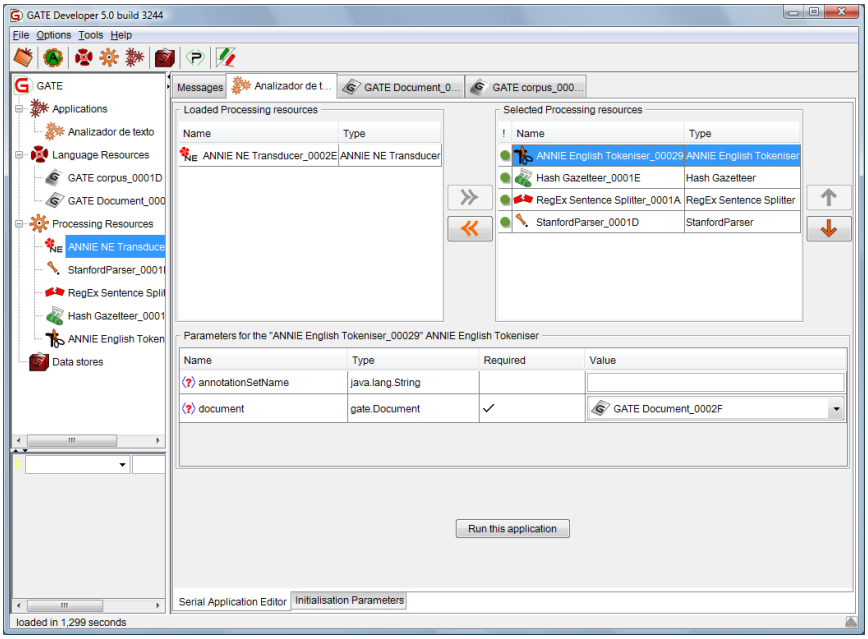


Figura 3.4: Interfaz gráfica de Gate

La ejecución de una aplicación lingüística compleja sigue un mecanismo de pila por el que se ejecutan secuencialmente los diferentes recursos que se desean utilizar, y que, en muchos casos, no se puede alterar ya que los resultados producidos por un recurso son requeridos por el siguiente. Como se puede ver en la imagen, el analizador sintáctico de dependencias Stanford Parser posee un *plugin* para acceder a él a través de Gate, por lo que su integración en una aplicación lingüística compleja es sumamente sencilla. La interfaz para programación de Gate es prácticamente igual de fácil y transparente para el usuario que la interfaz gráfica, lo que permite realizar tareas complejas de análisis de texto de forma eficiente. Es más, Gate dispone de un mecanismo para crear y guardar una aplicación mediante la interfaz gráfica que puede ser utilizada mediante la interfaz de programación con solo cargar el archivo.

Los datos de salida de los análisis de Gate son encapsulados en un archivo XML en el que se representan pares atributo/valor. Los atributos son cadenas de caracteres, mientras que los valores son objetos java representados en XML. De esta manera, el tratamiento de los resultados obtenidos por una

aplicación lingüística desarrollada en Gate se realiza de forma uniforme y sencilla, independientemente del recurso utilizado, mediante cualquier aplicación en Java. La interfaz gráfica de Gate permite, además, realizar una visualización más intuitiva sobre el texto analizado según se van seleccionando las diferentes etapas de análisis realizadas.

Por todos los motivos expuestos, se ha decidido utilizar Gate como parte del sistema desarrollado en esta tesis. En concreto, la facilidad para crear e importar desde un sistema externo aplicaciones lingüísticas complejas de una forma sencilla y transparente, la representación de la salida mediante un archivo XML estructurado, fácilmente interpretable mediante aplicaciones desarrolladas en Java, así como su amplio uso dentro de la comunidad científica para el procesamiento del lenguaje natural han sido las principales motivaciones para seleccionar esta herramienta para el pre-procesamiento del texto de entrada. Además, la facilidad para incluir herramientas externas, como es el caso del analizador sintáctico Stanford Parser, mediante el sistema de *plugins* que Gate proporciona, sin necesidad de realizar interfaces para cada herramienta utilizada, ha sido otro factor muy decisivo.

3.2. Weka

Weka[4] (*Waikato Environment for Knowledge Analysis*) es un programa de código abierto desarrollado por la University of Waikato, Nueva Zelanda, cuyo objetivo es hacer disponible, de una forma sencilla y transparente para el usuario, los principales algoritmos de aprendizaje máquina para su uso en tareas de minería de datos, clasificación automática, etc. Weka está implementado completamente en el lenguaje de programación Java, y dispone de una interfaz gráfica y de una interfaz de código para su ejecución directamente desde un entorno de desarrollo.

Weka presenta herramientas para pre-procesamiento de datos, clasificación, regresión, agrupamiento, asociación de reglas y visualización. En la figura 3.5 se observa la pantalla principal de Weka, donde se pueden ver cuatro acciones principales:

- **Explorer**: la herramienta principal de Weka, permite pre-procesar los datos utilizando distintos filtros (discretizar, normalizar, etc.), utilizar los clasificadores implementados para realizar modelos de predic-

[4] http://www.cs.waikato.ac.nz/ml/weka/

ción (árboles de decisión, funciones de regresión, etc.) o generar agrupamientos de los datos (K-Means, Cobweb, etc.), entre otras funciones.

- **Experimenter**: esta herramienta permite realizar evaluaciones sistemáticas, como la comparación entre diferentes esquemas de aprendizaje o algoritmos de clasificación, utilizar diferentes metodologías para evaluar los resultados obtenidos, o la aplicación de tests de significancia estadística sobre dichos resultados.

- **KnowledgeFlow**: esta herramienta permite hacer prácticamente las mismas operaciones que el *Explorer* mediante una interfaz gráfica diferente que introduce, por ejemplo, una interfaz entre Java-Beans y algoritmos de aprendizaje, clasificadores, etc. Además, permite definir flujos de operaciones para, por ejemplo, realizar un filtro sobre los datos, clasificarlos y evaluarlos.

- **Simple CLI**: esta herramienta proporciona una línea de comandos para realizar las mismas tareas que el *Explorer* o el *KnowledgeFlow* mediante secuencias de comandos.

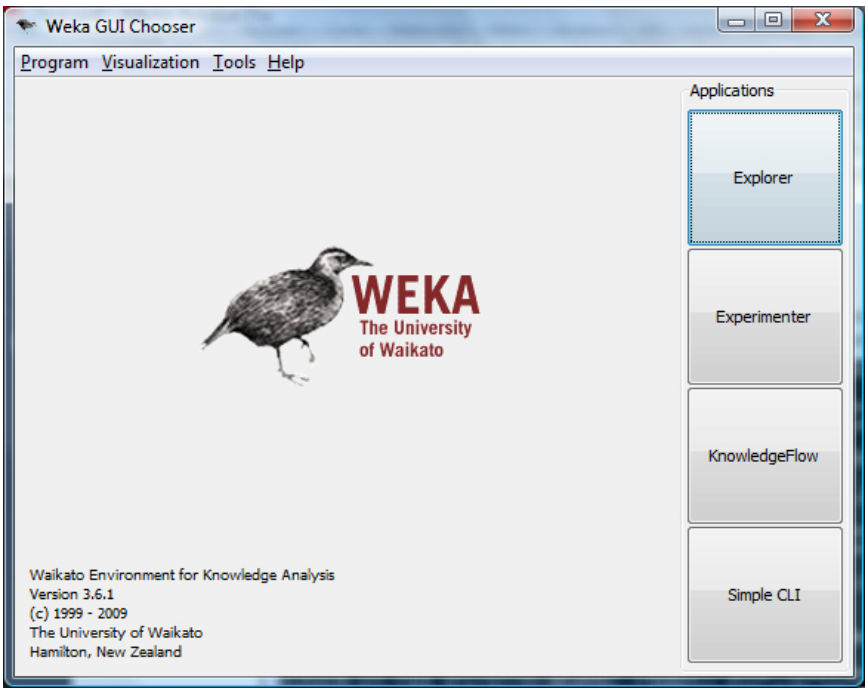


Figura 3.5: Pantalla principal de Weka

La representación de los datos utilizada por Weka responde al acrónimo de *Attribute-Relation File Format (arff)*. El formato *arff* está compuesto por una estructura claramente diferenciada en tres partes:

- **Cabecera**: sección del formato en la que se define el nombre de la relación.
- **Declaraciones de atributos**: en esta sección se declaran los atributos que compondrán el vector o instancia de cada ejemplo. Cada atributo está formado por la tripla *etiqueta/nombre_relación/tipo*, y donde los tipos aceptados por Weka son: *NUMERIC*, *INTEGER*, *REAL*, *DATE*, *STRING* y *ENUMERATED*. El último atributo de esta sección es la clase en la que se clasifica cada instancia.
- **Sección de datos**: en esta sección se declara cada instancia, es decir, cada ejemplo compuesto por todos sus atributos separados por comas, y donde el último de ellos es la clase de la instancia.

En la sección de datos se permite cierta incertidumbre en los atributos de cada instancia, que será representada por el símbolo "?", y que representa el dato desconocido de ese atributo para esa instancia. La figura 3.6 muestra un fichero en este formato que modela instancias ejemplo sobre cuándo jugar al aire libre dependiendo de parámetros atmosféricos. En la última versión de Weka, no obstante, se han introducido métodos para importar datos de diferentes formatos como CVS o mediante acceso SQL.

Centrándonos en la herramienta *Explorer*, que es la utilizada en este trabajo, se puede ver en la figura 3.7 que las diferentes tareas que se pueden llevar a cabo sobre los datos se presentan como pestañas, que a su vez engloban las diferentes herramientas disponibles para el tratamiento de los datos según la tarea. De todas las pestañas, cabe destacar las pestañas *Preprocess* y *Classify*.

La pestaña *Preprocess* permite, entre otras tareas, cargar los conjuntos de datos, explorar las características de los atributos y las instancias, o visualizar datos estadísticos de los mismos. Mediante esta pestaña también se pueden aplicar diferentes filtros sobre los datos de entrada. Estos filtros se clasifican en supervisados, que tienen en cuenta la clase a la que pertenecen los datos, y no supervisados, que no tienen en cuenta esta información. Entre los más destacados se encuentran:

```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute class {play, not_play}

@data
sunny,85,85,FALSE,not_play
sunny,80,90,TRUE,not_play
overcast,83,86,FALSE,play
rainy,70,96,FALSE,play
rainy,68,80,FALSE,play
rainy,65,70,TRUE,not_play
overcast,64,65,TRUE,play
sunny,72,95,FALSE,not_play
sunny,69,70,FALSE,play
rainy,75,80,FALSE,play
sunny,75,70,TRUE,play
overcast,72,90,TRUE,play
overcast,81,75,FALSE,play
rainy,71,91,TRUE,not_play
```

Figura 3.6: Ejemplo de conjunto de datos en formato *arff*

- ***Discretize***: convierte un rango de atributos numéricos en atributos nominales.
- ***Resample***: obtiene del conjunto de datos un subconjunto aleatorio de tamaño especificado. También puede ser utilizado para mantener la distribución de la clase en la muestra resultante.
- ***Attribute Selection***: permite el uso de varios tipos de clasificadores y búsquedas para realizar una selección de los atributos relevantes de la base de datos.
- ***Normalize***: normaliza todos los valores numéricos del conjunto de datos.
- ***Randomize***: reordena el conjunto de datos de manera aleatoria.

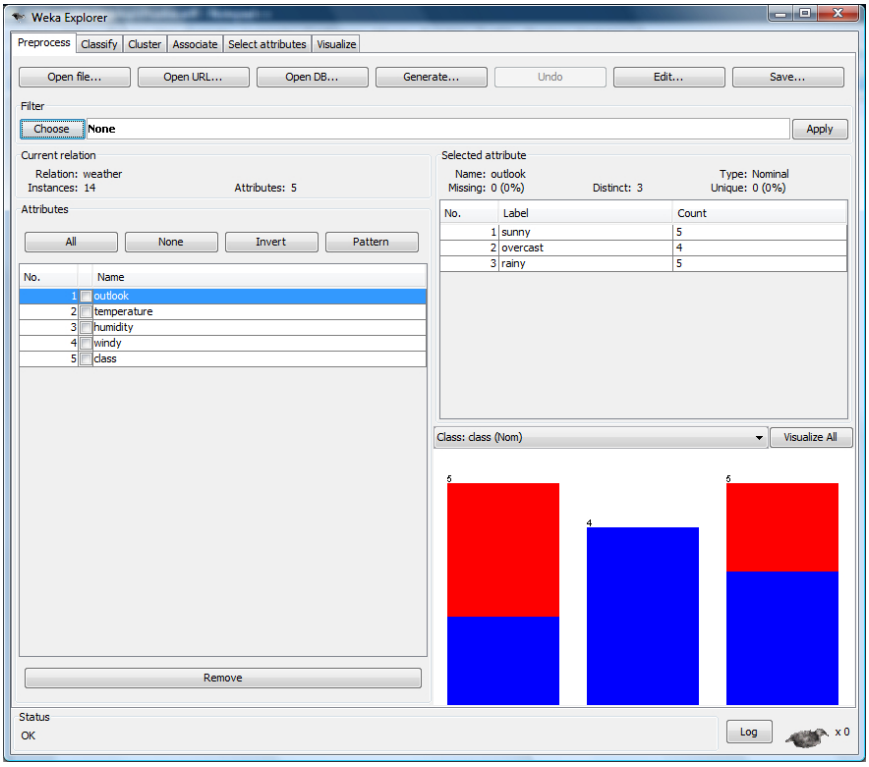


Figura 3.7: Pestaña *Preprocess* del *Explorer* de Weka

Por último, la pestaña *Classify* permite definir modelos de predicción con los datos de entrada a partir de los algoritmos incluidos en Weka. Como se puede ver en la figura 3.8, Weka permite elegir entre múltiples algoritmos de clasificación, distribuidos entre las categorías *bayes*, *functions*, *lazy*, *meta*, *mi*, *misc*, *rules* y *trees*. Una vez seleccionado el algoritmo deseado, los parámetros de éste se pueden modificar para ajustarlos a los datos de entrada.

Esta interfaz también permite ejecutar el clasificador sobre el conjunto de datos cargados, así como seleccionar diferentes tipos de evaluación mediante el panel *Test options*, y cuyos resultados serán mostrados en el *Classifier output*. En concreto, Weka proporciona cuatro métodos principales de evaluación:

- ***Use training set***: este método utiliza como conjunto de test el propio conjunto de entrenamiento.

- ***Supplied test set***: utiliza para la evaluación un conjunto de test independiente proporcionado por el usuario.

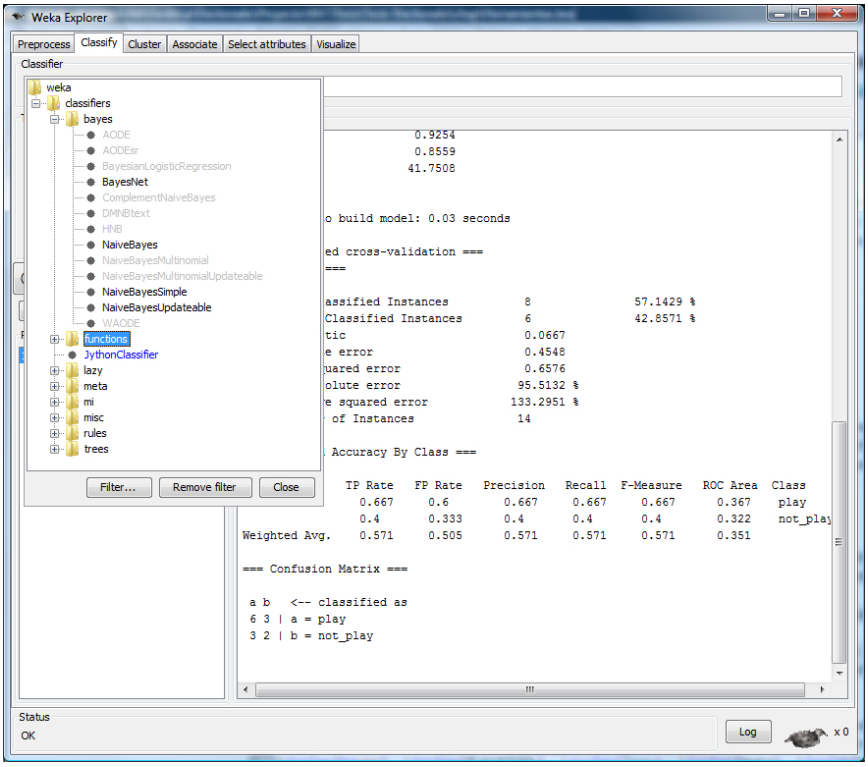


Figura 3.8: Pestaña *Classify* del *Explorer* de Weka

- ***Cross Validation***: quizá el mas utilizado, este método realiza el test utilizando el algoritmo de validación cruzada. Dicho algoritmo consta de dos etapas: en la primera de ellas se procede a la división del conjunto de datos en k subconjuntos del mismo tamaño. En la segunda, cada subconjunto se utiliza para el test, mientras que el resto de datos se utilizan para el entrenamiento. El error se estima como la media de los distintos errores obtenidos.

- ***Percentage Split***: este método divide el conjunto de datos suministrado en dos subconjuntos, uno destinado al entrenamiento y otro al test, según un porcentaje especificado por el usuario.

Weka calcula automáticamente las métricas más utilizadas para este tipo de tareas como son el *accuracy* (exactitud), *precision* (precisión), *recall* (cobertura) y *F-measure* (medida-F) de cada clase y la media global. Muestra, además, la matriz de confusión de todas las instancias clasificadas.

3.3. WordNet

WordNet[5] es una extensa base de datos léxica para el inglés desarrollada bajo la dirección de George A. Miller e iniciada en el año 1985, en cuyo proyecto han participado diferentes organizaciones públicas y privadas como la Universidad de Princeton, el Departamento de Investigación Naval o el Departamento de Defensa de los Estados Unidos (Miller, 1995). Actualmente el proyecto sigue activo y es mantenido por el grupo de investigación WordNet de la Universidad de Princeton.

WordNet nació con dos propósitos. Por un lado, la creación de una combinación entre léxico y tesauro fácil de usar y de comprender; por otro lado y más importante, proporcionar una herramienta básica para el desarrollo de aplicaciones complejas de análisis textual y procesamiento del lenguaje natural. Es por ello que WordNet se organizó en torno a cinco categorías léxicas básicas: nombres, verbos, adjetivos, adverbios y elementos funcionales, de las cuales actualmente solo existen las cuatro primeras. El principal problema de esta organización es la excesiva información redundante almacenada.

A diferencia de otros recursos similares, WordNet esta basado en conceptos, en lugar de términos o palabras, cuya representación dentro de la base de datos se realiza mediante los denominados *synonym sets* o *synsets*, que constituyen grupos de elementos con un mismo significado. Los *synsets* están representados mediante un identificador unívoco, una definición denominada *gloss*, que en muchos casos sirve para matizar el significado del concepto, y el conjunto de términos o palabras que representan dicho concepto en inglés. En la figura 3.9 se muestra un ejemplo de la estructura de los *synsets* a los que pertenece la palabra *love*, donde entre paréntesis aparece el *gloss* del concepto precedido por los términos representativos del *synset*.

Además de esta relación de sinonimia, WordNet ofrece otro conjunto de relaciones entre *synsets* como *hiperonimia*, *homonimia*, *antonimia*, *relacionado con*, *derivado de*, etc. Estas relaciones, también denominadas punteros, dependen del tipo de categoría gramatical a la que pertenezca el *synset* (nombres, verbos, adverbios y adjetivos), aunque todas las categorías tienen las relaciones de sinonimia y antonimia. A continuación se profundizará en las relaciones más importantes en las diferentes categorías utilizadas en este trabajo: nombres, verbos, adjetivos y adverbios.

Sustantivos. WordNet presenta alrededor de 117.798 sustantivos rep-

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

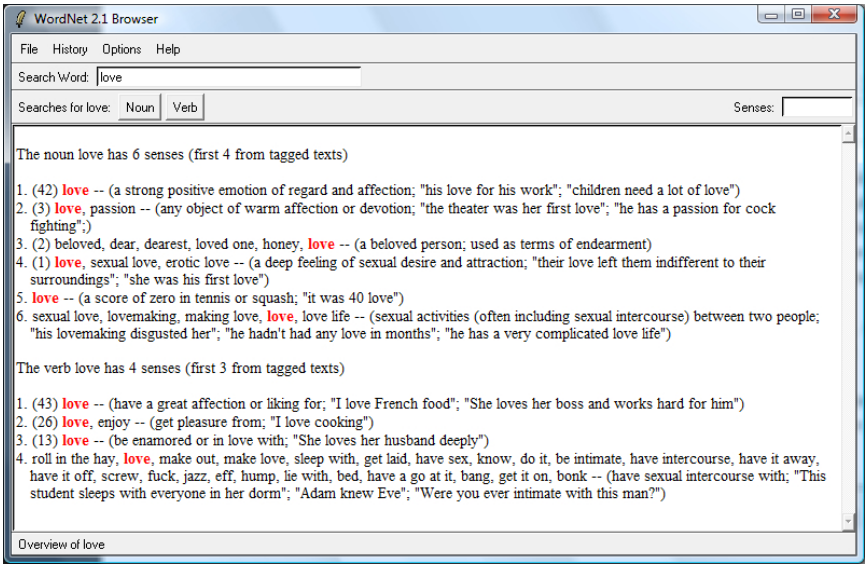


Figura 3.9: Resultados de la búsqueda de la palabra *love* en WordNet

resentados por unos 82.115 *synsets* estructurados jerárquicamente mediante la relación de hiperonimia, donde el concepto padre de uno dado generaliza a dicho concepto, creando un sistema de herencia léxica donde un concepto hereda los rasgos distintivos de su ancestro. Así, por ejemplo, el primer significado de la palabra *love* presenta la jerarquía de *synsets* expuesta en la figura 3.10.

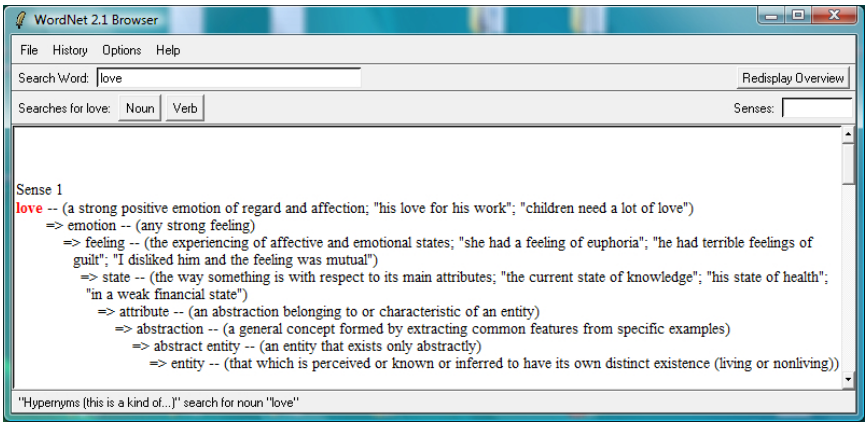


Figura 3.10: Jerarquía del primer significado de *love* en WordNet

Además de esta relación de hiperonimia, y de las ya expuestas relaciones de sinonimia y antonimia, los sustantivos en WordNet también presentan las siguientes relaciones:

- **Hiponimia (*hyponym*)**: Y es un hipónimo de X si todo Y es un X. Por ejemplo, *atún* es un hipónimo de *pescado* puesto que un atún es un tipo de pescado.

- **Holonimia (*holonym*)**: Y es un holónimo de X si X es parte de Y. Por ejemplo, *procesador* es un holónimo de *computadora* ya que el procesador es parte de la computadora.

- **Meronimia (*meronym*)**: Y es un merónimo de X si Y es parte de X. Por lo tanto, *ordenador* es un merónimo de *procesador*.

- **Términos coordinados (*coordinate terms*)**: dos términos se consideran términos coordinados si ambos comparten un mismo hiperónimo. Por ejemplo, *edificio* y *puente* son términos coordinados ya que comparten el mismo hiperónimo *construcción*.

Verbos. WordNet contempla unos 11.529 términos verbales representados mediante 13.767 *synsets*. Las formas verbales se organizan en WordNet mediante implicación léxica (*lexical entailment*), donde un verbo V_1 está relacionado con otro V_2 cuando la oración *alguien* V_1 implica lógicamente a la oración *alguien* V_2. Un ejemplo de esta relación se puede ver en los verbos *roncar* y *dormir*, puesto que la oración *Está roncando* implica a la oración *Está dormido*. Es importante aclarar que esta relación no es simétrica; es decir, que el verbo V_1 implique al verbo V_2 no significa necesariamente que el verbo V_2 implique al verbo V_1. Además de esta relación, y de las de antonimia y sinonimia, los verbos en WordNet poseen las siguientes relaciones:

- **Hiperonimia (*hyperonym*)**: el verbo V_1 es un hiperónimo del verbo V_2 si la acción de V_2 está representada por V_1. Por ejemplo, el verbo *moverse* es un hiperónimo del verbo *viajar*.

- **Hiponimia (*hyponym*)**: el verbo V_1 es un hipónimo del verbo V_2 si la acción de V_1 representa la de V_2. *Viajar* es, pues, un hipónimo de *moverse*.

- **Troponimia (*troponym*)**: el verbo V_1 es tropónimo del verbo V_2 si la actividad de V_1 implica de algún modo realizar la actividad de V_2. Por ejemplo, el verbo *roncar* es tropónimo del verbo *dormir*, puesto que roncar implica dormir.

- **Términos coordinados (*coordinate terms*)**: dos verbos V_1 y V_2 son términos coordinados si tienen en común un mismo hiperónimo, al igual que en los sustantivos.

- **Causalidad (*cause*)**: la relación de causa se establece entre dos verbos que actúan, respectivamente, como causativo y resultado. Un ejemplo de esta relación lo encontraríamos entre la pareja de verbos *enseñar* y *aprender*.

Adjetivos. WordNet consta de unos 21.479 adjetivos organizados en 18.156 *synsets* y clasificados en dos categorías, descriptivos y relacionales. Los adjetivos descriptivos son aquellos que acompañan a los sustantivos y les otorgan valores de atributos bipolares, por lo que se organizan mediante las relaciones de sinonimia y antonimia. Los adjetivos *divertido* y *horrible* son ejemplos claros de este tipo de adjetivos. Por otro lado, los adjetivos relacionales representan una relación de "relativo a" o "asociado al" sustantivo, como por ejemplo el adjetivo *higiene* en relación al sustantivo *dental*. Además de estas relaciones, es interesante resaltar la existencia de las siguientes:

- **Participio de un verbo (*Participle of verb*)**: esta relación identifica aquellos conceptos donde el adjetivo tiene la forma del participio de un verbo pero actúa como adjetivo. Un ejemplo lo encontramos en el adjetivo *horrorizado*.

- **Perteneciente (*Pertainym*)**: esta relación pretende asociar la forma adjetival con el sustantivo del que se deriva, como por ejemplo, el adjetivo *presidencial* que deriva del sustantivo *presidente*.

Adverbios. WordNet presenta alrededor de 4.481 adverbios distribuidos en 3.621 *synsets*, donde no hay ninguna estructura de representación determinada. Los adverbios en WordNet, a diferencia de los sustantivos y los verbos, no poseen estructura de árbol, ni están divididos en categorías, como los adjetivos, ni poseen ningún tipo de clasificación como el tipo de modificador (tiempo, lugar o grado). Además de la relación de sinonimia entre términos, y en algunos casos, la relación de antonimia, cabe destacar la siguiente relación:

- **Derivado de un adjetivo (*Derived from adjective*)**: esta relación une los *synsets* de aquellos adverbios derivados de adjetivos, con los

synsets de dichos adjetivos. Por ejemplo, el adverbio *terriblemente* presenta una relación de este tipo con el adjetivo *terrible*.

Actualmente, WordNet está formado por alrededor de 155.287 términos estructurados en 117.659 *synsets* distribuidos entre las diferentes categorías gramaticales. WordNet está disponible para su uso gratuito, lo que ha promovido el desarrollo de un gran número de proyectos satélites que proporcionan herramientas para abordar diversas tareas lingüísticas, como por ejemplo la desambiguación semántica o el etiquetado emocional. WordNet dispone de una interfaz *online* así como de una interfaz gráfica, cuya última version disponible para el sistema operativo Windows es la 2.1. Para el acceso a WordNet mediante entornos de desarrollo se han implementado múltiples interfaces, entre las que cabe destacar la *Java WordNet Interface*[6], que ha sido utilizada en esta tesis por su facilidad de uso e idoneidad para el desarrollo de aplicaciones en Java.

En cuanto a la utilización de WordNet como base de datos léxica se refiere, su elección estuvo motivada por la posibilidad que ofrece de trabajar a nivel de conceptos en lugar de términos, a la vez que por su estructura de relaciones, principalmente la sinonimia y la antonimia. Además, su gran difusión científica en el campo del procesamiento del lenguaje natural posibilita el uso de otras herramientas y recursos asociados que permiten obtener mayores funcionalidades, como sistemas de desambiguación semántica. Por último, debido al auge que WordNet está teniendo en el mundo científico, distintos países de la Unión Europea han realizado herramientas similares en otros idiomas, entre ellos el español, dentro del proyecto EuroWordNet[7], lo que proporciona la posibilidad de portar fácilmente el sistema, en un futuro, a otros idiomas.

3.4. WordNet::SenseRelate

WordNet::SenseRelate[8] es un paquete Perl desarrollado por Ted Pedersen (2005) para abordar el problema de la desambiguación semántica (*Word Sense Disambiguation*) en WordNet, permitiendo, por tanto, identificar un término con su concepto adecuado en relación al contexto en el que se utilice.

[6] Java WordNet Interface. http://projects.csail.mit.edu/jwi/
[7] EuroWordNet. http://www.illc.uva.nl/EuroWordNet/
[8] WordNet::SenseRelate. http://www.d.umn.edu/~tpederse/senserelate.html

WordNet::SenseRelate ha sido desarrollado para su uso en cualquier disciplina de procesamiento del lenguaje natural, como la generación automática de resúmenes, la simplificación de texto, la clasificación y extracción de información, etc. (Plaza, Díaz, y Gervás, 2011), tratando de solventar un problema muy importante como es la identificación del significado correcto para cada palabra cuando se analiza un texto.

WordNet::SenseRelate hace uso de las medidas de relación y similitud semántica implementadas en el paquete WordNet::Similarity[9] del mismo autor. El paquete de desambiguación semántica WordNet::SenseRelate presenta tres opciones para realizar la desambiguación de términos:

- **WordNet::SenseRelate::AllWords**: esta versión del paquete permite, dado un texto de entrada, obtener el concepto adecuado para todos y cada uno de los términos presentes en el texto.
- **WordNet::SenseRelate::WordToSet**: en esta versión del paquete el método asigna el concepto más adecuado al término de entrada de acuerdo a las relaciones de este con un conjunto de palabras dado.
- **WordNet::SenseRelate::TargetWord**: esta versión realiza la desambiguación de una única palabra obteniendo el concepto adecuado y de acuerdo al contexto en el que se encuentra.

En este trabajo se ha optado por utilizar la versión AllWords, ya que permite, dado un texto de entrada, obtener los conceptos asociados a todos sus términos. Para ello, el programa necesita como parámetros el fichero donde se encuentra el texto a analizar, el tipo de formato del texto, el tamaño de la ventana de contexto para cada término (por defecto cuatro), la métrica de similitud a utilizar y, opcionalmente, una lista de parada. Para cada oración del texto de entrada, WordNet::SenseRelate::AllWords ejecuta el algoritmo expuesto en la tabla 3.1. La medida de similitud se calcula, como se ha expuesto, mediante el paquete WordNet::Similarity, mediante la llamada al programa *wdls.pl* en el parámetro de entrada que representa la métrica de similitud de WordNet::SenseRelate.

Una vez ejecutado este algoritmo de desambiguación, el programa devuelve como salida el texto de entrada en el que cada término ha sido etiquetado con su correspondiente categoría gramatical y su *sense* o significado

[9] WordNet::Similarity. http://wn-similarity.sourceforge.net/

```
for each word w in sentence
   disambiguate-single-word (w)
disambiguate-single-word (w)
   for each sense s_{t_i} of target word t, where i = 0..N
      let score_i = 0
      for each word w_j in context-window
         next if j = t
         for each sense s_{j_k} of w_j
            temp − score_k = relatedness (s_{t_i}, s_{j_k})
         best-score = max temp-score
         if best-score > pairScore
            score_i = score_i + best-score
   return s_{t_i} such that
   score_i > score_j for all j in s_{t_0}, ...,s_{t_N}
      and score_i > contextScore
```

Tabla 3.1: Algoritmo de desambiguación del paquete WordNet::SenseRelate::AllWords

en WordNet. A modo de ejemplo, la figura 3.11 muestra el resultado de ejecutar WordNet::SenseRelate::AllWords sobre el texto *Staff was very helpful and welcome in reception, breakfast is good.*

```
Staff#n#1 be#v#1 very#r#1 helpful#a#1
and welcome#a#1 in reception#n#2 ,
breakfast#n#1 be#v#1 good#a#1
```

Figura 3.11: Salida del paquete WordNet::SenseRelate::AllWords para el texto de entrada *Staff was very helpful and welcome in reception, breakfast is good*

WordNet::SenseRelate ha sido seleccionado para el desarrollo del sistema de esta tesis debido a su fácil y transparente adaptación con la base de datos léxica WordNet, permitiendo de esta manera disponer de los conceptos adecuados para cada término de un texto de entrada, y pudiendo así trabajar en el sistema a nivel de conceptos en lugar de términos.

3.5. Léxicos afectivos

Los léxicos afectivos se pueden entender como bases de datos donde sus entradas, ya sean términos o conceptos, están etiquetadas con algún tipo de información emocional o afectiva. A la hora de hablar de léxicos afectivos se pueden distinguir dos grandes clasificaciones: por un lado, los léxicos

basados en términos o palabras; y por otro, los léxicos basados en conceptos o significados. Los léxicos afectivos basados en términos tienen como unidad primitiva el término o palabra, el cual es etiquetado con cierta información emocional, mientras que los léxicos afectivos basados en conceptos utilizan el significado de las palabras como unidad primitiva, siendo el concepto el que es etiquetado emocionalmente.

Así mismo, tanto en los léxicos basados en términos como en los basados en conceptos se puede diferenciar entre los que se inspiran en la teoría psicológica de las categorías emocionales para etiquetar sus entradas, y los que se basan en la teoría psicológica de las dimensiones emocionales (ver sección 2.1). Los léxicos que se sustentan en la primera de ellas etiquetan las entradas con categorías emocionales que representan las emociones básicas de las teorías psicológicas, como *sadness* o *fear*, mientras que los segundos etiquetan las entradas de acuerdo a diferentes dimensiones emocionales, como por ejemplo la *polaridad* o la *activación*.

Atendiendo a los léxicos basados en términos y que se apoyan en la teoría de las categorías emocionales, cabe destacar el LIWC Dictionary (*Linguistic Inquiry and Word Count Dictionary*) (Pennebaker, Francis, y Booth, 2001) y la lista LEW (*List of Emotional Words*)(Francisco, Gervás, y Peinado, 2010). El LIWC Dictionary está compuesto por unos 2.300 términos y lemas etiquetados con una o varias categorías, entre las que se encuentran, por ejemplo, *sadness*, *negative emotion* o *overall affect*. La lista LEW dispone de alrededor de 1.500 lemas asociados con su categoría gramatical y etiquetados con una o varias categorías emocionales, las cuales están, además, organizadas jerárquicamente.

De entre los léxicos basados en términos y en la teoría de las dimensiones emocionales, cabe destacar el diccionario *General Inquirer* (GI) (Stone et al., 1966) como el más reconocido y utilizado. El diccionario GI presenta diversas categorías afectivas como, por ejemplo, *active*, *passive*, *strong* o *weak*, entre las que cabe destacar la lista de palabras positivas, compuesta por 1.915 términos, y la lista de palabras negativas, compuesta por 2.291 términos. Otro recurso que ha sido muy utilizado por la comunidad científica para el análisis sentimental en los últimos años ha sido la lista ANEW (*Affective Norms for English Words*) (Bradley y Lang, 1999), donde los términos están etiquetados con valores en el intervalo 1 a 9 para las dimensiones emocionales *pleasure*, *arousal* y *dominance*, de acuerdo con el estándar SAM

(*Self-Assessment Manikin*) (Lang, 1980). Por último, cabe mencionar el diccionario DAL (*Whissell´s Dictionary of Affect Language*) (Whissell, 1989), compuesto por 8.742 palabras etiquetadas con las dimensiones emocionales *activation*, *evaluation* e *imagery*.

El principal problema de estas aproximaciones radica en el uso de términos o palabras como unidad principal de los léxicos, sin tener en cuenta los múltiples significados que un mismo término puede tener cuando se utiliza en distintos contextos, y que en muchos casos tendrán connotaciones emocionales muy diferentes. Un ejemplo de esto se encuentra en la palabra *cancer*, cuyo término posee 5 entradas en WordNet de las cuales solo una hace referencia a "crecimiento maligno o tumor", mientras que las restantes están relacionadas con "astrología" o con la "constelación cáncer". Por este motivo, en los últimos años han surgido diversas aproximaciones que utilizan el concepto o significado como unidad primitiva para el etiquetado emocional, y que, por regla general, están basadas en WordNet (ver sección 3.3).

Puesto que uno de los principales objetivos de este trabajo es precisamente el uso de conceptos en lugar de términos a la hora de analizar sentimentalmente el texto, en las siguientes secciones se profundizará en los léxicos afectivos más relevantes que han seguido la línea de utilizar conceptos en lugar de términos, y cuyos máximos exponentes son WordNet Affect y SentiWordNet.

3.5.1. WordNet Affect

WordNet Affect[10] (Strapparava y Valitutti, 2004) es un subproyecto del proyecto WordNet Domains (Bentivogli et al., 2004), en el que cada *synset* de WordNet ha sido etiquetado con al menos una categoría semántica de dominio. Las categorías semánticas en WordNet Domains están estructuradas jerárquicamente formando un conjunto final de 200 categorías, en cuyo nivel superior se encuentran las categorías *doctrines*, *free_time*, *applied_science*, *pure_science*, *social_science*, *factotum* y *metrology*.

Tal y como defienden los autores de WordNet Affect, el propósito de este léxico afectivo es etiquetar con una jerarquía emocional los conceptos de WordNet que portan un significado afectivo. La elección de WordNet como herramienta básica para el desarrollo del proyecto se debió a su amplia

[10] WordNet Affect. http://wndomains.fbk.eu/wnaffect.html

cobertura del inglés, su estructura de conceptos en torno a *synsets* y las interfaces disponibles para poder recorrer la base léxica completa a través de sus relaciones. WordNet Affect ha sido realizado mediante un proceso semi-automático, donde en un primer paso se etiquetó manualmente con categorías emocionales un conjunto inicial o *core* de *synsets*, expandiéndolo automáticamente mediante las relaciones de estos *synsets* en un segundo paso.

En la primera etapa, los autores utilizaron un conjunto de 1.903 términos, denominado AFFECT, etiquetados con diferentes atributos emocionales que tradujeron a los apropiados *synsets* de WordNet. Para la segunda etapa, los autores analizaron las relaciones léxicas y semánticas de WordNet para determinar cuales preservaban las connotaciones emocionales, concluyendo que las relaciones *antonymy*, *similarity*, *derived-from*, *pertains-to*, *attribute*, *also-see* preservaban dicha información y se podían añadir automáticamente los *synsets* relacionados con los *synsets* del *core* mediante ellas. Por otro lado, se observó que las relaciones *hyperonymy*, *entailment*, *causes* o *verb-group* solo preservaban la información afectiva en un porcentaje, por lo que los *synsets* expandidos mediante estas relaciones fueron revisados a mano.

WordNet Affect está basado en la versión 1.6 de WordNet y dispone de dos versiones, WordNet Affect 1.0 y WordNet Affect 1.1. La primera de ellas consta de 2.874 *synsets* etiquetados, que cubren unas 4.800 palabras. En esta versión los conceptos están etiquetados con su origen (*core* o la relación a partir de la que se han obtenido), y con categorías que indican si es una emoción, si es un estado físico, si es un respuesta emocional, etc. La segunda versión de WordNet Affect consta de 911 *synsets* etiquetados, que cubren alrededor de 2.300 términos. A diferencia de la versión anterior, los *synsets* de esta versión están etiquetados con categorías emocionales como *joy* o *sadness*. En esta versión, WordNet Affect presenta una jerarquía de categorías emocionales compuesta por más de 300 categorías, y cuya estructura se puede ver en la figura 3.12, donde los nodos (ya sean hojas o no) con texto representan las categorías emocionales. Los nodos con el formato *n#número* representan los *synsets* etiquetados con la categoría padre, donde "n" identifica a los sustantivos, "a" a los adjetivos, "r" a los adverbios y "v" a los verbos, y "número" el identificador del *synset*. Por último, las hojas cuyo formato es *@término* representan las palabras asociadas al *synset* padre.

Las principales deficiencias de este recurso radican en la extensa jerarquía

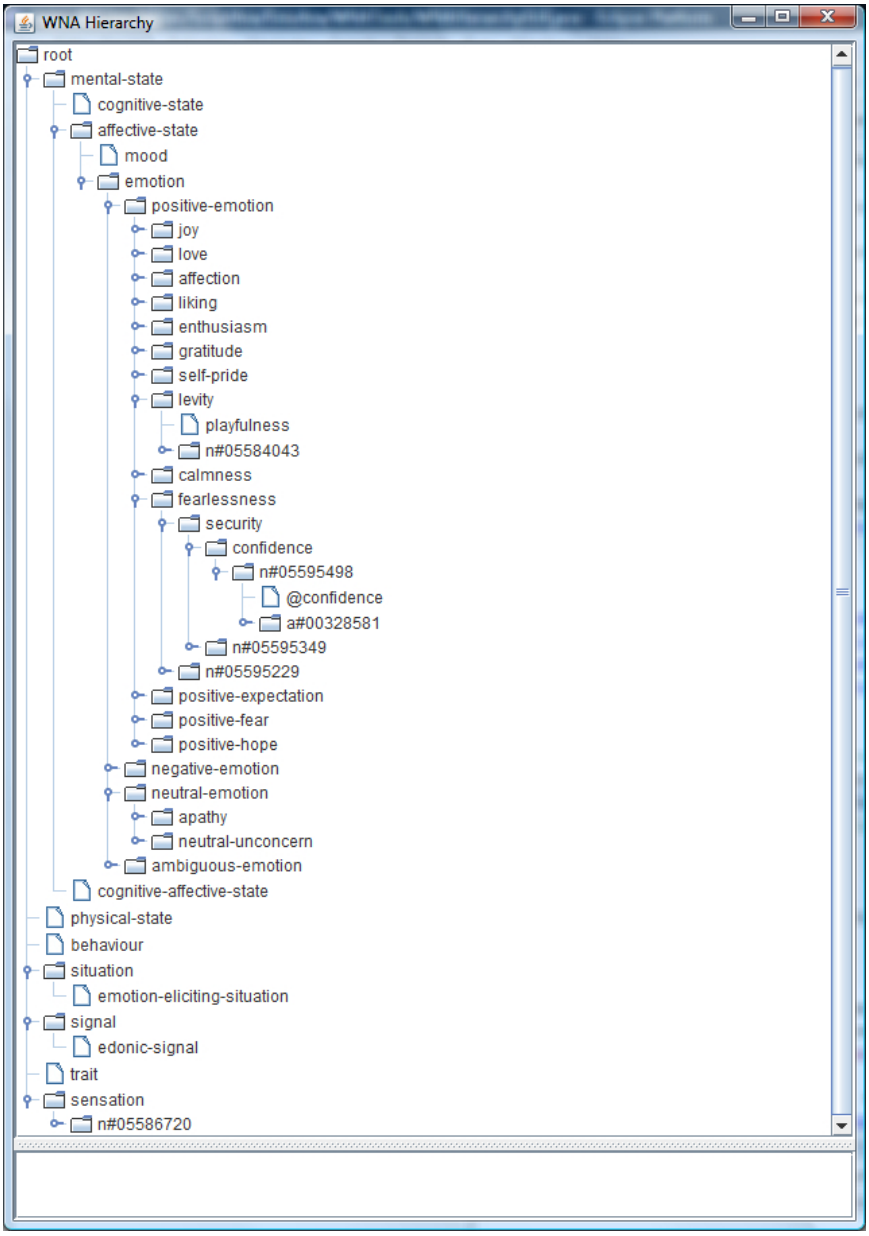


Figura 3.12: Jerarquía de las categorías emocionales de WordNet Affect

de categorías emocionales propuesta para el etiquetado de conceptos, donde para realmente obtener categorías emocionales es preciso llegar al segundo nivel dentro del grupo *emotion*. Además, el número de categorías de este nivel es 32, lo que supone una representación demasiado amplia, y no muy acorde con las teorías psicológicas de las categorías emocionales (Plutchik, 1980; Ortony y Turner, 1990; Ekman, 1992; Parrot, 2001). Por otro lado, se

podría plantear el uso del tercer nivel de la jerarquía, en el que los conceptos están clasificados en positivos, neutrales, negativos y ambiguos. El principal problema es que WordNet Affect está diseñado para su uso con categorías emocionales, existiendo otros recursos mejor preparados para la clasificación de conceptos en positivos y negativos como SentiWordNet, que será descrito en la siguiente sección.

Cabe también subrayar que WordNet Affect contiene conceptos etiquetados con más de una categoría emocional, lo que puede llevar a situaciones en las que no se pueda determinar la emoción correcta sin un proceso de desambiguación específico, o que la categoría emocional recuperada no sea la más apropiada. Por último, es importante también tener en cuenta la versión de WordNet sobre la que se sustenta WordNet Affect, puesto que la última version de WordNet para el sistema operativo Windows es la 2.1, mientras que WordNet Affect esta basado en la versión 1.6, lo que supone la necesidad de realizar una traducción entre versiones de WordNet o perder las mejoras introducidas en las últimas versiones.

3.5.2. SentiWordNet

SentiWordNet[11] (Esuli y Sebastiani, 2006b; Baccianella, Esuli, y Sebastiani, 2010) es un léxico afectivo disponible para investigación, también construido sobre la base de datos léxica WordNet, donde cada *synset* ha sido etiquetado con tres valores numéricos en las categorías *Pos(s)*, *Neg(s)* y *Obj(s)*, representando, respectivamente, la polaridad positiva, la polaridad negativa y la neutralidad del concepto. Los valores numéricos de cada categoría varían en el rango 0 a 1, y el valor suma de todas las categorías debe ser 1. La principal motivación de este léxico es la elaboración de un recurso que de soporte a los sistemas que abordan la tarea de clasificación de la polaridad mediante el tratamiento de conceptos en lugar de términos. En la figura 3.13 se puede ver un ejemplo de la última versión online de SentiWordNet, basada en WordNet 3.0, en la que se muestra la información disponible para la palabra *good*.

El léxico afectivo SentiWordNet ha sido desarrollado mediante la clasificación automática de los *synsets* de WordNet, para lo que los autores adaptaron sus métodos para el cálculo de *PN-polarity* (Esuli y Sebastiani, 2005) y *SO-polarity* (Esuli y Sebastiani, 2006a) para trabajar con *synsets* en lu-

[11] SentiWordNet. http://sentiwordnet.isti.cnr.it/

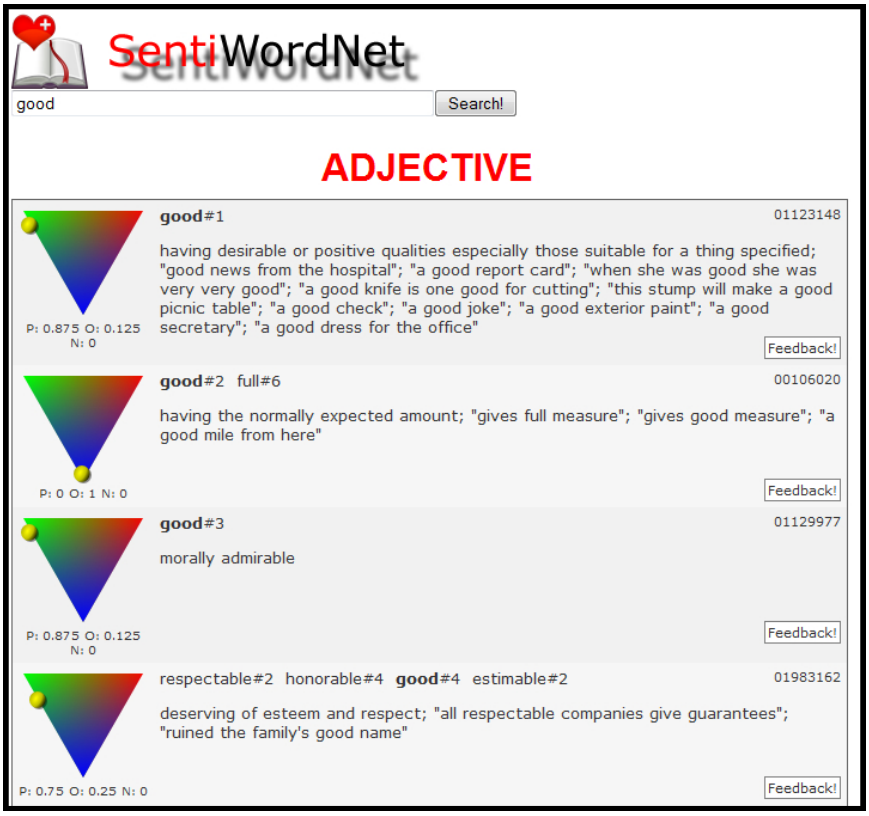


Figura 3.13: Entrada en SentiWordNet para el término *good*

gar de términos. El método *PN-polarity* calcula la orientación de términos expandiendo un conjunto inicial de términos etiquetados como positivos y negativos a través de las relaciones de WordNet, como la sinonimia o antonimia. Una vez que el conjunto inicial ha sido expandido con todos los posibles términos, el método crea una representación textual de cada término a partir de los *glosses* de todos los *synsets* al que pertenece dicho término. Esta representación textual es transformada a vectores de términos *tf-idf*, que conforman el conjunto de entrenamiento de un algoritmo binario de clasificación. El cálculo de *SO-polarity* es prácticamente similar al propuesto para *PN-polarity*, con la única salvedad de que en este caso el algoritmo determina si el término es positivo, negativo o neutral.

Para realizar la clasificación de los *synsets* de WordNet, los autores desarrollaron un conjunto de clasificadores ternarios basados en el análisis cuantitativo de los *glosses* de los *synsets*, donde los distintos clasificadores se diferencian en la forma de entrenamiento. Cada uno de estos clasificadores

es capaz de determinar si un *synset* es positivo, negativo o neutral, obteniendo una puntuación final para cada categoría. Los valores finales para cada *synset* se obtienen mediante la normalización de los resultados de todos los clasificadores, ponderando en caso de discrepancia entre los distintos clasificadores de acuerdo al número de clasificadores que han asignado los valores. Por último, en la última versión los autores incluyen un paso final para ajustar los valores de cada *synset* basado en el cálculo de caminos aleatorios (*random-walk*) sobre la jerarquía de WordNet.

SentiWordNet cubre todos los *synsets* de WordNet. Aunque ésta es quizá una de sus mayores aportaciones, es también una de sus mayores debilidades, puesto que, en muchos casos, conceptos que no estén bien relacionados con el conjunto inicial de *synsets* etiquetados manualmente no obtendrán puntuaciones muy acordes. Un ejemplo de esto se puede observar en la figura 3.14, donde el concepto representativo de *tumor* o *cancer* obtiene una puntuación muy baja (0,125) en la categoría *Neg*.

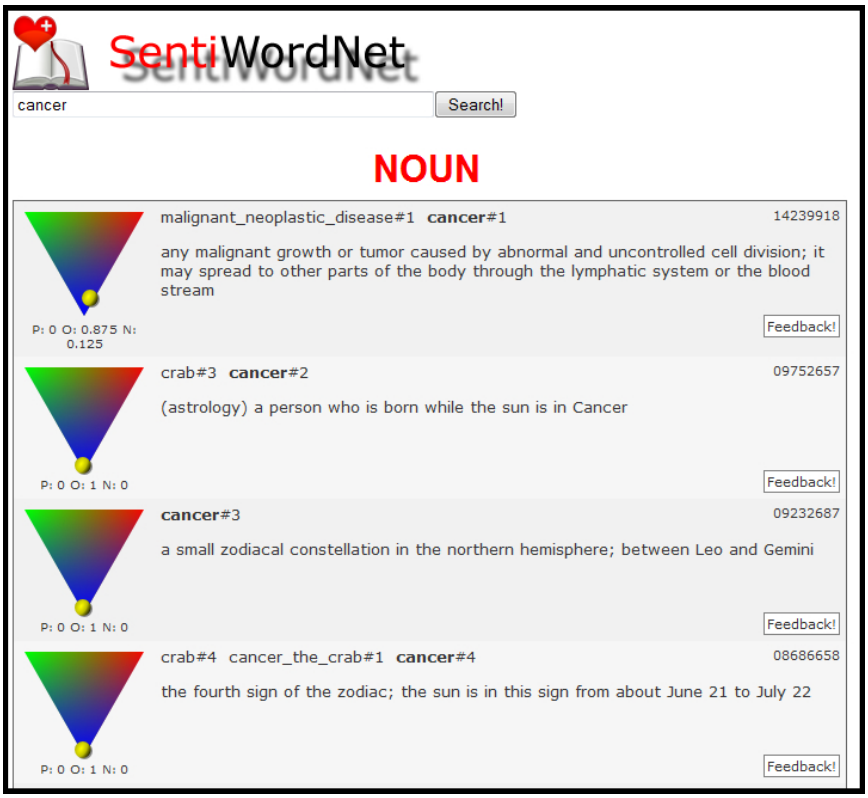


Figura 3.14: Entrada en SentiWordNet para el término *cancer*

Por otro lado, aunque un léxico afectivo basado en la polaridad de los

conceptos puede ser una opción muy adecuada para la tarea de clasificación de la polaridad, puede no aportar información suficiente al abordar la tarea de clasificación de la intensidad.

Capítulo 4

SentiSense: un léxico afectivo basado en conceptos

El propósito de este capítulo es presentar al lector el léxico afectivo basado en conceptos *SentiSense*, desarrollado como parte de esta tesis doctoral con el objetivo de paliar las deficiencias encontradas en los léxicos disponibles en la literatura. Para ello, en primer lugar se analizan los motivos que nos han impulsado a elaborar SentiSense, y a continuación, se describe su diseño y formato, a la vez que se presenta una serie de herramientas desarrolladas para facilitar su uso y mantenimiento.

4.1. Motivación

Tal y como se expusiera en la sección 3.5, existen en la literatura diferentes léxicos emocionales, cada uno de ellos apoyado por una teoría psicológica y con características y formatos que los diferencian. Sin embargo, la práctica totalidad de ellos presentan un problema fundamental, que consiste en utilizar el término o palabra como unidad primitiva del etiquetado. Ello es debido a que la gran mayoría de los sistemas tradicionales de análisis sentimental utilizan como unidad principal de trabajo el término (Pang, Lee, y Vaithyanathan, 2002; Dave, Lawrence, y Pennock, 2003; Pang y Lee, 2004; Turney, 2002), ya sea mediante aproximaciones basadas en frecuencias o en el uso de expresiones polares. Es por esto que la literatura recoge multitud de léxicos de este tipo (Whissell, 1989; Bradley y Lang, 1999; Pennebaker, Francis, y Booth, 2001). La principal limitación de este tipo de léxicos es

que no tienen en cuenta la polisemia de los términos o palabras; es decir, el hecho de que un término puede tener más de un significado dependiendo del contexto en el que se utiliza, y que, en muchos casos, esos significados tienen connotaciones emocionales diferentes. Así, por ejemplo, la palabra *cold* puede referirse tanto a un desorden respiratorio como a la ausencia de calor. Si bien la primera acepción connota, indudablemente, un significado emocional negativo, la segunda puede ser considerada neutral desde el punto de vista emocional.

Por todo ello, en los últimos años se ha popularizado la creación de léxicos afectivos basados en conceptos. Dichos léxicos se apoyan en bases de datos como WordNet (ver sección 3.3), donde la unidad principal es el significado y todos los términos asociados a cada concepto están agrupados en conjuntos de sinónimos. De esta manera, cualquier sistema puede determinar el concepto adecuado para cada término mediante el uso de una estrategia de desambiguación semántica, recuperando así mismo la etiqueta emocional apropiada para cada concepto desambiguado. Entre los más relevantes de este nuevo grupo de léxicos afectivos, y disponibles para su uso en investigación, se encuentran WordNet Affect (Strapparava y Valitutti, 2004) y SentiWordNet (Esuli y Sebastiani, 2006b), ambos construidos sobre WordNet. WordNet Affect (ver sección 3.5.1) es un léxico emocional estructurado jerárquicamente sobre un conjunto de categorías emocionales, donde su tercer nivel contiene las categorías emocionales *positive-emotion*, *negative-emotion*, *neutral-emotion* y *ambiguous-emotion*, y que engloban a su vez en el siguiente nivel un conjunto más extenso de categorías (un total de 32), entre las que se encuentran las típicas de las teorías psicológicas, tales como *joy*, *fear* o *calmness*. De cada categoría emocional hoja penden, finalmente, los conceptos etiquetados. Por el contrario, SentiWordNet (ver sección 3.5.2) es un léxico afectivo donde cada concepto está etiquetado con tres puntuaciones, según su polaridad, en las categorías *Obj* (cómo de objetivo es el concepto), *Pos* (cómo de positivo es) y *Neg* (cómo de negativo es). Los valores de cada categoría varían en el rango 0 a 1, siendo 1 el valor suma de todas las categorías.

En ambos casos, dichos recursos plantean problemas que hacen que su uso en esta tesis sea inapropiado para los objetivos marcados. Por un lado, WordNet Affect presenta un conjunto emocional demasiado amplio, ya que 32 categorías emocionales de primer nivel útil incluye excesiva generalidad

para que un sistema de análisis sentimental pueda, a través de dicho conjunto, extraer automáticamente las peculiaridades de las diferentes categorías emocionales. Además, WordNet Affect presenta una estructura jerárquica, con muchos más niveles y categorías emocionales, y muy poca concordancia con las principales teorías psicológicas (Plutchik, 1980; Ortony y Turner, 1990; Ekman, 1992; Parrot, 2001). Otro inconveniente importante de este léxico afectivo lo hemos encontrado al evaluarlo, ya que hemos identificado varias entradas que poseen múltiples categorías emocionales, siendo imposible para un sistema automático discernir la más adecuada (por ejemplo, el *synset* "a#00117872 *angered, enraged, furious, infuriated, maddened*" está etiquetado como *anger*, *fury* y *infuriation*). Por último, WordNet Affect está basado en la versión 1.6 de WordNet, mientras que la última versión disponible para el sistema operativo Windows es la 2.1, lo que supone realizar un mapeo entre ambas versiones si se desea utilizar las mejoras introducidas en la última versión de WordNet, pudiendo producirse errores en dicho mapeo.

Por otro lado, el principal propósito del léxico afectivo SentiWordNet es dar soporte a tareas de clasificación de textos en base a su polaridad, puesto que solo proporciona información sobre la orientación polar de los conceptos. Sin embargo, esta excesiva generalización de la carga afectiva de cada concepto, basada simplemente en la polaridad, puede no ser suficiente para abordar otras tareas más complejas como la clasificación de textos en base a su intensidad, ya que estas tareas requieren una información de grano más fino sobre el significado emocional de cada concepto. Un ejemplo de esto se puede ver en los conceptos *cáncer* y *resfriado*, donde, aunque ambos expresan emociones negativas, el primero claramente conlleva un sentimiento de *miedo* y *tristeza*, mientras que el segundo expresa *malestar* o *disgusto*.

Por todos estos motivos, se ha decidido implementar, como parte de esta tesis doctoral, un léxico emocional basado en conceptos, denominado SentiSense, y apoyado en la teoría psicológica de las categorías emocionales, lo que nos permite capturar no solo la polaridad sino también las connotaciones de intensidad asociadas a cada categoría a la hora de evaluar emocionalmente un texto. Sin embargo, y para soslayar los problemas de WordNet Affect, el conjunto emocional es mucho más reducido, y en concordancia con las diferentes teorías psicológicas de las categorías emocionales. Además, en el léxico afectivo un concepto solamente puede estar etiquetado con una categoría

emocional, eliminando así toda posible ambigüedad. Aunque este léxico ha sido desarrollado para atender los requisitos del sistema planteado en este trabajo, puede ser utilizado en cualquier aproximación que trabaje a nivel de conceptos. La metodología adoptada en su desarrollo, así como su estructura, serán descritas en detalle en las siguientes secciones.

4.2. Diseño de SentiSense

SentiSense es un léxico en el que cada entrada ha sido etiquetada con una emoción dentro de un conjunto de categorías emocionales. A diferencia de otros léxicos afectivos, el elemento principal de SentiSense no es el término, sino el significado, obteniendo de esta forma un mismo etiquetado para todas las palabras con un mismo significado, o conjunto de sinónimos. Para poder llevar a cabo esta tarea es preciso disponer de un tesauro basado en conceptos y no en palabras. Es por esto que para su desarrollo se ha utilizado la base de datos léxica WordNet. WordNet (ver sección 3.3) es un tesauro para el inglés donde cada entrada está basada en la idea de los *synonym sets* o *synsets*, que engloban grupos de elementos con un mismo significado. Los *synsets* están representados mediante un significado o definición denominado *gloss* y el conjunto de términos o palabras que denotan dicho concepto. Además de esta relación de sinonimia, WordNet ofrece otro conjunto de relaciones entre *synsets* como *hiperonimia*, *homonimia*, *antonimia*, *relacionado con*, *derivado de*, etc. Dichas relaciones dependen del tipo de categoría gramatical a la que pertenezca el *synset* (nombre, verbo, adverbio y adjetivo). WordNet ha sido desarrollado para el idioma inglés, aunque la idea ha sido portada a otros idiomas a través del proyecto EuroWordNet [1], por lo que resulta sencillo extender SentiSense a otros idiomas.

El conjunto de categorías emocionales disponibles en SentiSense está compuesto por 14 categorías y ha sido elegido teniendo en cuenta las diferentes aproximaciones psicológicas presentadas a lo largo del último siglo (Plutchik, 1980; Ortony y Turner, 1990; Ekman, 1992; Parrot, 2001). En concreto, las categorías emocionales disponibles en SentiSense son: *like*, *compassion*, *joy*, *anger*, *hate*, *expectancy*, *love*, *sadness*, *calmness*, *ambiguous*, *fear*, *hope*, *surprise* y *disgust*. Además, SentiSense incluye una relación de antonimia entre aquellas categorías emocionales fuertemente opuestas como

[1] EuroWordNet. http://www.illc.uva.nl/EuroWordNet/

joy-sadness. Se ha decidido introducir esta relación por dos motivos principales: primero, la relación de antonimia hace posible un mejor tratamiento de la negación desde un punto semántico, y segundo, hace posible que el léxico sea extendido automáticamente a través de la relación de antonimia presente en WordNet. La tabla 4.1 muestra la relación de antonimia para cada categoría emocional en SentiSense.

Categoría	**Antónimo**	**Categoría**	**Antónimo**
joy	sadness	hope	-
fear	calmness	compassion	-
surprise	-	ambiguous	-
anger	calmness	sadness	joy
disgust	like	calmness	-
love	hate	like	disgust
expectancy	-	hate	love

Tabla 4.1: Categorías emocionales y sus respectivos antónimos en SentiSense

SentiSense ha sido desarrollado semi-automáticamente en dos etapas siguiendo la idea presentada en (Strapparava y Valitutti, 2004): en una primera fase, se realiza el etiquetado manual de un conjunto de conceptos para crear una semilla o conjunto inicial; en una segunda fase, y partiendo de esta semilla, se expande el léxico afectivo mediante las diferentes relaciones de WordNet. En la primera fase, se etiquetaron manualmente 1200 conceptos con las categorías emocionales de SentiSense, creando de esta manera el conjunto inicial o semilla. En la segunda etapa, este conjunto semilla se extendió automáticamente mediante las relaciones de *sinonimia*, *antonimia*, *derivado de* y *relacionado con* de WordNet. Aunque existen más relaciones entre conceptos que podrían ser utilizadas en el proceso de expansión, por regla general esas relaciones introducen excesivo ruido; es decir, se etiquetan muchos conceptos que no deberían ser etiquetados con dichas categorías, por lo que se ha preferido obtener un léxico con menor número de conceptos etiquetados pero con mayor acierto en la asociación concepto/categoría emocional. De esta manera, el léxico afectivo SentiSense consta de 2.214 *synsets* que cubren un total de 5.551 términos. La tabla 4.2 muestra la distribución de los *synsets* etiquetados entre las diferentes categorías emocionales y gramaticales.

Categoría	**Nombres**	**Verbos**	**Adjetivos**	**Adverbios**
like	146	70	184	25
compassion	11	3	8	2
joy	63	30	52	19
anger	23	11	24	6
hate	2	4	11	1
expectancy	74	35	46	9
love	30	4	28	6
sadness	62	30	73	25
calmness	18	6	35	10
ambiguous	24	15	20	3
fear	104	43	38	14
hope	32	14	14	6
surprise	13	9	15	3
disgust	222	134	254	51
Total	824	408	802	180

Tabla 4.2: Distribución de los *synsets* en SentiSense por categorías emocionales y gramaticales

4.3. Herramienta de etiquetado manual de SentiSense

Para realizar la primera etapa de una forma más sencilla, puesto que el proceso de etiquetado recae sobre conceptos y es preciso realizar un proceso de desambiguación semántica de cada palabra, se ha desarrollado una herramienta que facilita el etiquetado manual de conceptos. Como se muestra en la figura 4.1, la pantalla de la herramienta consta de tres partes diferenciadas. La primera, en la parte superior, permite elegir entre cuatro córpora o colecciones de textos ampliamente difundidos en el dominio del análisis sentimental: SemEval 2007 (Strapparava y Mihalcea, 2007), Sentence Polarity Dataset (Pang y Lee, 2005), MovieReview Dataset (Pang, Lee, y Vaithyanathan, 2002) y HotelReview Corpus (Carrillo de Albornoz et al., 2011). El botón *Start tagging* permite iniciar el proceso de etiquetado sobre el corpus seleccionado y con un rango de documentos dado (campos *Starts in* y *Ends in*).

La segunda parte de la aplicación, en la parte izquierda de la pantalla, muestra el texto del documento a etiquetar así como la intensidad emocional con la que ha sido clasificado de acuerdo al etiquetado manual de cada corpus. La tabla *Terms in document* muestra los términos que componen el texto del documento. Cada vez que se selecciona una palabra en la tabla se

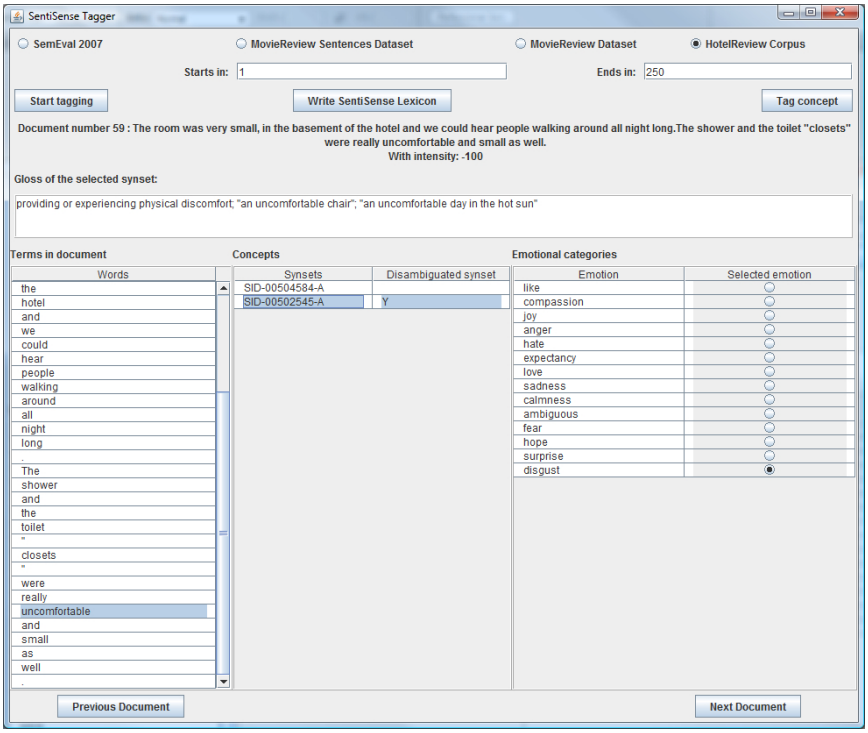


Figura 4.1: Pantalla de la aplicación de etiquetado de SentiSense

muestran automáticamente en la tabla *Concepts* los posibles conceptos en WordNet asociados al término seleccionado (columna *synsets*), así como el concepto desambiguado por el algoritmo de desambiguación (la herramienta usa el algoritmo *lesk*, al igual que el sistema) marcándolo con "Y" (columna *Disambiguated Synset*). Además, cada vez que un *synset* es seleccionado en la tabla *Concepts*, su definición o *gloss* se muestra en el área *Gloss of the selected synset*.

La última parte de la herramienta, en la parte derecha de la ventana, muestra la tabla con las diferentes categorías emocionales del léxico SentiSense. Cada vez que un *synset* es seleccionado, la aplicación muestra automáticamente la emoción que posee en el léxico, si el concepto ha sido etiquetado previamente. Si el concepto no hubiera sido etiquetado, simplemente basta con elegir la emoción adecuada y pulsar el botón *Tag concept* y el concepto será etiquetado y añadido al léxico. Por último, los botones *Next document* y *Previous document* permiten avanzar o retroceder por el conjunto de documentos del rango seleccionado, mientras que el botón *Write SentiSense Lexicon* permite escribir en un archivo los cambios realizados.

4.4. Herramienta de expansión de SentiSense

Aunque para el desarrollo de la segunda fase no es necesaria una aplicación gráfica, se ha decido crear una que permita visualizar, de una forma sencilla y amigable, el léxico afectivo, mostrando tanto los *synsets* como las palabras de cada uno y la emoción asignada. La figura 4.2 muestra una captura de pantalla del visor del léxico SentiSense. Como se puede ver, cada vez que un *synset* es seleccionado la aplicación automáticamente muestra la emoción asignada a dicho concepto, así como el *gloss* o definición del mismo en el campo *Gloss of the selected synset*.

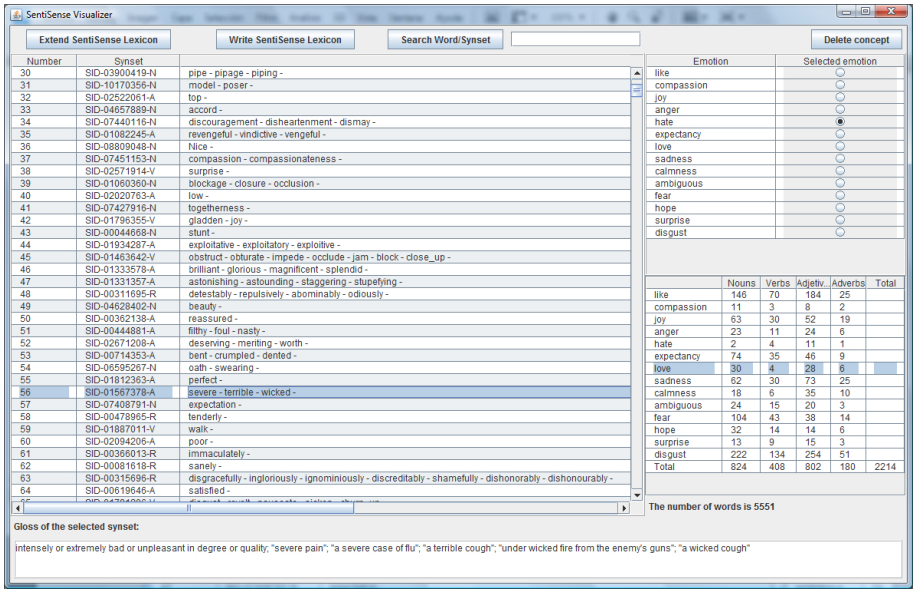


Figura 4.2: Pantalla de la aplicación de visualización de SentiSense

El botón *Extend SentiSense Lexicon* permite expandir automáticamente el léxico con las relaciones entre conceptos descritas anteriormente. De esta manera, para cada concepto etiquetado en el léxico, la herramienta identifica todos los conceptos enlazados a él mediante las relaciones de *sinonimia*, *antonimia*, *derivado de* y *relacionado con*, y etiqueta dichos conceptos en caso de que no tuvieran ya una categoría emocional asignada. En el caso de la sinonimia, los conceptos nuevos son etiquetados con la misma categoría emocional. Lo mismo sucede con las relaciones derivado de y relacionado con. Para etiquetar los conceptos antónimos, se utiliza la relación de antonimia entre categorías emocionales de SentiSense. De este modo, cuando un concepto esta relacionado con otro mediante una relación de antonimia, se

obtiene la categoría emocional del primer concepto y se etiqueta el segundo con la categoría emocional antónima. Cuando no existe relación de antonimia entre las categorías emocionales, el nuevo concepto no es etiquetado.

Puesto que la extensión automática del léxico puede incluir conceptos mal etiquetados, la aplicación dispone de un botón para eliminar conceptos, *Delete concept*, así como una búsqueda por palabras y *synsets*, *Search Words/Synsets*. Al igual que en la herramienta de etiquetado manual, el botón *Write SentiSense Lexicon* permite escribir el léxico en un archivo. Por último, como se puede ver en la figura 4.2, la herramienta muestra en la parte inferior derecha unas estadísticas del número total de conceptos etiquetados, así como su distribución entre categorías emocionales y categorías gramaticales.

4.5. Formato de SentiSense

Puesto que la posibilidad, y el deseo, de que SentiSense sea utilizado en otros trabajos ha estado presente desde un principio en su desarrollo, el formato utilizado para su difusión es XML. Se ha optado por este formato por su sencillez y por la multitud de herramientas disponibles para su procesamiento en la gran mayoría de lenguajes de programación.

El conocimiento creado en este léxico afectivo se ha configurado en dos archivos de texto. Por un lado, el archivo con las categorías emocionales y la relación de antonimia. Por otro lado, el archivo que relaciona los conceptos y su correspondiente categoría emocional. Como se puede ver en la figura 4.3, el archivo de categorías emocionales simplemente recoge, en cada línea, una categoría emocional y su correspondiente categoría antónima.

```
<SentiSenseEmotionalCategories>
    <EmotionalCategory name="joy" antonym="sadness"/>
    <EmotionalCategory name="fear" antonym="calmness"/>
    <EmotionalCategory name="surprise" antonym=""/>
    <EmotionalCategory name="anger" antonym="calmness"/>
    ...
</SentiSenseEmotionalCategories>
```

Figura 4.3: Extracto del archivo de categorías emocionales de SentiSense

En la figura 4.4, se muestra un extracto del archivo de SentiSense que recoge la información de cada concepto etiquetado y su categoría emocional

asociada. En este archivo, cada entrada representa un concepto de WordNet, cuyo identificador es el identificador del *synset* tal y como WordNet lo proporciona a través de la función *toString()*. Así mismo, para su mejor identificación y brindar la posibilidad en el futuro del estudio de posibles mejoras de expansión mediante técnicas de similitud semántica, se ha decidido almacenar también el *gloss* o definición de cada concepto, así como su categoría gramatical. Finalmente, el último atributo de cada entrada es la categoría emocional de SentiSense a la que se encuentra asociado el concepto.

```
<SentiSenseCorpus>
    <Concept synset="SID-00152712-A" pos="adjective"
    gloss="lacking cordiality; unfriendly..." emotion="disgust"/>
    <Concept synset="SID-00050667-R" pos="adverb"
    gloss="in a joyous manner;..." emotion="joy"/>
    <Concept synset="SID-03430539-N" pos="noun"
    gloss="a weapon that discharges a missile ..." emotion="fear"/>
    <Concept synset="SID-01799986-V" pos="verb"
    gloss="give displeasure to" emotion="disgust"/>
</SentiSenseCorpus>
```

Figura 4.4: Extracto del archivo de *synsets* etiquetados en SentiSense

Capítulo 5

Modelo lingüístico-semántico basado en emociones para la clasificación de la polaridad e intensidad

A lo largo de este capítulo se describirá el sistema automático basado en emociones para la clasificación de texto en función de su polaridad e intensidad emocional. El sistema está basado en aprendizaje supervisado y hace uso de análisis semántico y de un conjunto de reglas lingüísticas para abordar los diferentes retos que el análisis sentimental plantea. Además, cada término del texto es desambiguado mediante un algoritmo de desambiguación semántica permitiendo así al sistema trabajar a nivel de conceptos en lugar de a nivel de términos, y obviando los problemas que la ambigüedad puede introducir. Por último, el sistema utiliza el conocimiento encapsulado en un léxico afectivo para etiquetar los conceptos identificados con categorías emocionales. De esta manera, el texto de entrada es traducido a un conjunto de categorías emocionales que representan el significado afectivo del texto.

Como ya se ha expuesto, la presencia de negaciones, cuantificadores y modales puede afectar a la polaridad e intensidad del texto. La negación puede invertir o disminuir la polaridad del texto, mientras que los cuantificadores pueden amplificar o disminuir su intensidad. La modalidad puede invertir, disminuir o incluso no afectar en nada a la polaridad e intensidad del texto. De esta manera, una vez que las categorías emocionales de los

conceptos han sido identificadas, el método detecta en el texto dichos modificadores emocionales y trata convenientemente su efecto sobre las categorías emocionales dentro de su ámbito de acción.

El sistema ha sido preparado para trabajar con el idioma inglés. La elección de éste en detrimento del español se basa en la mayor disponibilidad y precisión de herramientas para realizar análisis de textos en inglés. Aún así, el algoritmo diseñado es independiente del idioma, por lo que puede ser adaptado fácilmente al español.

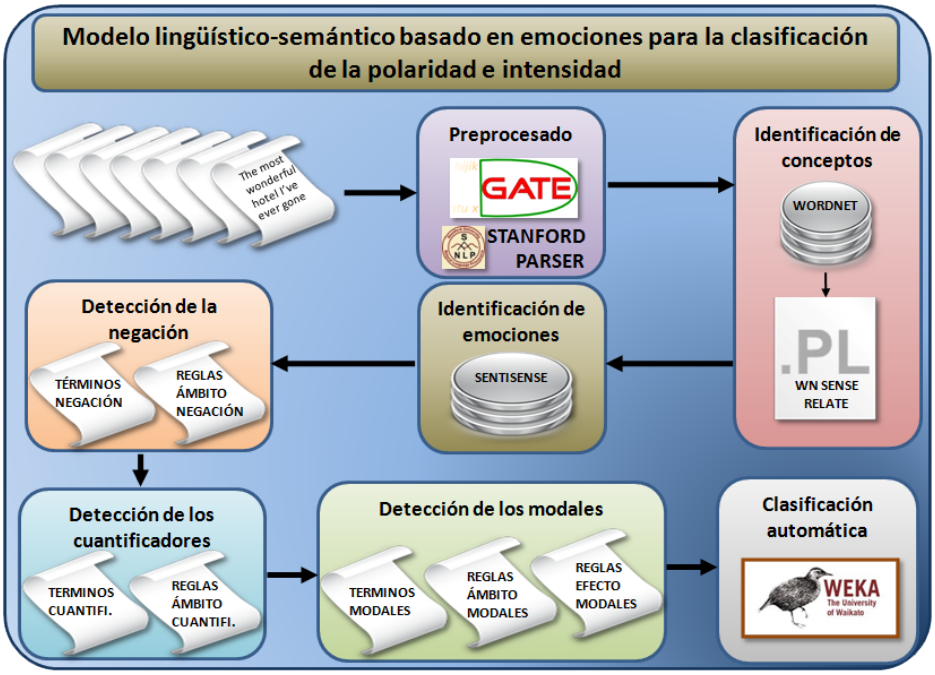


Figura 5.1: Diagrama de ejecución del sistema

El algoritmo propuesto consta de siete fases secuenciales, tal y como se puede apreciar en la figura 5.1. En la primera etapa, el sistema analiza el texto extrayendo la información lingüística necesaria para llevar a cabo el resto de etapas, realizando tareas como la división del texto en oraciones, el etiquetado de cada término con su categoría gramatical o la generación del árbol sintáctico de cada oración. La segunda fase del sistema resuelve la ambigüedad del texto analizado obteniendo, para cada término, su correspondiente concepto en la base de datos léxica WordNet. Tras esto, se identifican aquellos conceptos que poseen carga emocional mediante su comparación con los presentes en un léxico afectivo, asignando a cada concepto la cate-

goría emocional adecuada. Una vez identificadas las categorías emocionales del texto de entrada, la siguiente fase identifica las negaciones y aplica su efecto a los conceptos afectados por éstas. La etapa cinco consiste en identificar los cuantificadores presentes en el texto, así como la resolución de su efecto. Por último, se analiza el texto para detectar la presencia de modales y resolver su efecto sobre los conceptos dentro de su ámbito de acción. Finalmente, toda la información obtenida en los pasos previos es transformada en un *Vector de Intensidades Emocionales* (VEI, por las siglas en inglés de *Vector of Emotional Intensities*), que será la entrada de un algoritmo de clasificación. Las siguientes secciones explican en detalle el funcionamiento de cada una de las siete etapas que conforman el algoritmo de clasificación emocional.

5.1. Preprocesado: análisis sintáctico

La primera etapa del sistema realiza las acciones de división del texto de entrada en oraciones, identificación de los términos que las componen, etiquetado de cada término con su categoría gramatical (*POS*, del inglés *Part of Speech*), generación del árbol sintáctico de cada oración e identificación de las dependencias entre sus términos. Esta fase del método es necesaria para recopilar la información lingüística que permitirá realizar las sucesivas etapas del sistema. Por ejemplo, la categoría gramatical de cada palabra es necesaria para la correcta desambiguación de cada término de acuerdo a su contexto, mientras que el árbol sintáctico es utilizado para detectar correctamente el ámbito de acción de los modificadores emocionales.

Para llevar a término este análisis, son necesarias un conjunto de herramientas de procesamiento del lenguaje natural y su ejecución en un orden determinado. Para realizar esta tarea de una forma transparente para el usuario, se ha optado por utilizar la arquitectura general de ingeniería textual GATE (por sus siglas en inglés, *General Architecture for Text Engineering* (Cunningham et al., 2002)), que proporciona un sistema de pila con múltiples herramientas para el procesamiento del lenguaje natural (ver sección 3.1.2). Además, se ha decidido utilizar el analizador sintáctico *Stanford Parser* (Klein y Manning, 2003) para generar el árbol sintáctico de cada oración e identificar las dependencias entre los términos (ver sección 3.1.1). Stanford Parser es un analizador sintáctico estadístico con una gran

eficacia y que permite, además, obtener las dependencias entre los términos de cada oración. Por último, GATE dispone también de un mecanismo fácil y transparente para integrar el analizador sintáctico Stanford Parser en la pila de ejecución.

A continuación se detallan los diferentes módulos de GATE utilizados, así como la tarea desempeñada por cada uno de ellos:

- *Annie English Tokeniser:* módulo que permite determinar, a partir de un texto de entrada, los diferentes términos que lo componen. El módulo divide el texto en términos simples como marcas de puntuación, números o palabras. Además, en este proceso el módulo identifica cierto tipo de reglas especiales para el inglés que permiten realizar la división de una forma más efectiva, como son, por ejemplo, las formas contraídas de negación (e.g., *don't*) o el genitivo sajón.

- *Hash Gazetter:* módulo que permite identificar algunas entidades, siglas o palabras compuestas, como si de un solo término se tratase (e.g., Estados Unidos o J.R.R. Tolkien).

- *RegEx Sentence Splitter:* módulo para dividir el texto de entrada en oraciones según reglas gramaticales. Estas reglas están implementadas mediante expresiones regulares en Java.

- *Stanford Parser:* módulo de acceso al analizador sintáctico y de dependencias Stanford Parser, que permite etiquetar los términos detectados con su categoría gramatical, así como generar el árbol sintáctico y generar las dependencias entre los diferentes términos.

Los módulos se han descrito exactamente en el mismo orden en que han de ejecutarse dentro de la pila de GATE, ya que la salida de uno de ellos se requiere como entrada para el siguiente. El resultado final obtenido se escribe en un archivo en formato XML, donde cada oración aparece delimitada junto con sus términos, así como con el resto de información necesaria para la realización de las sucesivas etapas del método. Una vez terminado el análisis, el sistema interpreta el archivo XML generado por GATE creando una estructura interna del texto que permitirá llevar a cabo el resto de fases. Aunque GATE permite evitar la escritura del archivo e interpretar los resultados directamente desde código, se ha optado por esta opción para poder ejecutar el método en varias etapas cuando los conjuntos de evaluación son

demasiado grandes, evitando así sobrecargar el sistema y tener que analizar un texto repetidas veces con GATE.

5.2. Identificación de conceptos

Finalizada la etapa de análisis sintáctico y gramatical del texto, y habiendo identificado las oraciones y los términos que las componen, el sistema realiza la conversión de dichos términos a conceptos. De esta forma, se consigue que el método trabaje a nivel semántico; es decir, con conceptos en vez de términos, eliminado toda posible ambigüedad. Además, este análisis puede entenderse también como una primera clasificación entre los términos cuyos significados pueden contener connotaciones subjetivas u objetivas dependiendo del contexto en que se encuentran, como muestra el estudio de Wiebe y Mihalcea (2006), que relaciona la desambiguación con la subjetividad.

Para llevar a cabo está tarea, el sistema hace uso de uno de los múltiples algoritmos de desambiguación semántica disponibles para WordNet (ver sección 3.3). Estos algoritmos se basan en las relaciones de similitud existentes entre los distintos significados posibles para un término y el conjunto de palabras de su contexto. Existen diversas aproximaciones, entre las que destacan las que se basan en grafos, sistemas estadísticos, etc. (ver (Agirre y Edmonds, 2006) para un estudio exhaustivo de técnicas de desambiguación). En concreto, en este sistema se ha optado por utilizar el algoritmo *Lesk* (1986) implementado en el paquete *WordNet::SenseRelate* (ver sección 3.4).

La información extraída en la primera etapa es utilizada para realizar adecuadamente la desambiguación de cada palabra; en concreto, se utilizan la raíz del término y su categoría gramatical. Con esta información, y con un contexto o ventana de cuatro palabras, el algoritmo *Lesk* resuelve la ambigüedad de cada palabra devolviendo el lema y el significado asociado a la misma en la base de datos léxica WordNet. La información devuelta por WordNet::SenseRelate es utilizada para recuperar el *synset* apropiado mediante la interfaz de Java para WordNet JWI[1] (del inglés *Java Wordnet Interface*). El proceso de desambiguación es realizado solamente sobre los términos que pertenecen a las categorías gramaticales *nombre*, *verbo*, *adjetivo* y *adverbio* en WordNet, ya que son las únicas que pueden expresar una emoción.

[1] Java WordNet Interface. http://projects.csail.mit.edu/jwi/

Además de recuperar todos los *synsets* de las palabras del texto, el sistema recupera también, para cada uno de ellos, todos sus hiperónimos. Este proceso se realiza para enriquecer la etapa de identificación de categorías emocionales del método. La recuperación de todos los hiperónimos de cada *synset* desambiguado se realiza mediante las relaciones disponibles entre los *synsets* de WordNet.

Los beneficios de trabajar a nivel de concepto en vez de a nivel de término pueden verse en la oración *Test to predict breast cancer relapse is approved.* El concepto *cancer* presenta cinco significados en WordNet, y solo uno de ellos hace relación a un "crecimiento maligno o tumor", mientras que los otros hacen referencia a la "astrología" y a la "constelación". Sin un algoritmo de desambiguación semántica, el término *cancer* en esta oración puede ser erróneamente interpretado, lo que sin duda conllevaría la incorrecta identificación de la emoción subyacente.

5.3. Identificación de categorías emocionales

Una vez el sistema ha determinado todos los conceptos de cada oración, el siguiente paso consiste en traducir cada concepto identificado a su correspondiente categoría emocional de un léxico afectivo, si el concepto tiene una categoría emocional asociada. Para realizar este paso se precisa, por tanto, de un léxico afectivo que asocie conceptos de WordNet con categorías emocionales, atendiendo al significado o carga afectiva del concepto. Es por ello que, para realizar esta etapa, se ha desarrollado el léxico afectivo SentiSense, descrito en detalle en el capítulo 4, y que se ajusta a todos los requisitos de la idea propuesta para representar el texto mediante categorías emocionales, sorteando los problemas existentes en los léxicos afectivos disponibles. Cabe mencionar, no obstante, que esta etapa puede realizarse con cualquier otro léxico afectivo que asigne categorías emocionales a conceptos de WordNet. De esta manera, para cada concepto de cada oración del texto de entrada, el método comprueba si el concepto está etiquetado en el léxico afectivo, asignando la categoría emocional recuperada a dicho concepto en caso afirmativo.

Puesto que, por regla general, los léxicos emocionales solamente representan un pequeño porcentaje de las palabras existentes en los diccionarios o tesauros como WordNet, en este sistema se ha utilizado la relación de hiper-

onimia disponible en WordNet para mejorar este proceso de identificación de categorías emocionales. La definición de la Real Academia de la Lengua Española de hiperónimo es "Palabra cuyo significado incluye al de otra u otras"; es decir, un término cuyo significado es una generalización de otro u otros. Así, por ejemplo, la palabra *pájaro* es un hiperónimo de *jilguero* o *gorrión*. Nuestra hipótesis se basa en considerar que un hiperónimo conlleva la misma carga emotiva que los conceptos a los que incluye, pero con una intensidad o "fuerza" menor. De esta forma, el sistema intenta recuperar la categoría emocional asociada a cada concepto del texto de entrada en SentiSense. En caso de que el concepto no se encuentre etiquetado en el léxico afectivo, el sistema trata de recuperar la categoría emocional de su hiperónimo más cercano hasta alcanzar un cierto nivel, puesto que una excesiva generalización conlleva una perdida total de la carga emotiva. Este umbral ha sido establecido empíricamente a 3.

La oración *Siesta cuts risk of heart disease death study finds* ilustra claramente el proceso descrito. En esta oración, los conceptos *risk*, *death* and *disease* se etiquetan con las categorías emocionales *fear*, *fear* y *disgust*[2], respectivamente. Sin embargo, mientras que para los dos primeros conceptos dichas categorías se obtienen directamente de SentiSense, el tercero no se encuentra etiquetado en el léxico, por lo que es etiquetado con la emoción asignada a su hiperónimo de primer nivel, cuyo *synset* contiene términos tales como *illness* y *sickness*.

5.4. Detección de negación y su ámbito

Una vez etiquetados todos los conceptos de cada oración con sus correspondientes categorías emocionales, el sistema trata el efecto de la negación sobre las emociones en su ámbito. La negación es un aspecto importante en los sistemas que abordan el análisis sentimental, tanto desde el punto de vista de la detección de la polaridad como de la intensidad de dicha polaridad. La negación puede cambiar la polaridad de una oración, como en *I don't like this hotel*, o simplemente disminuir su intensidad, como en *The hotel is not bad*.

La detección de la negación es un proceso complejo que no solo abarca la identificación de las palabras o términos que expresan la negación, sino

[2]Utilizamos las categorías emocionales del léxico afectivo SentiSense para asignar emociones a los conceptos.

también su ámbito de acción; es decir, las palabras afectadas por dicha negación. Es por ello que esta etapa se divide, a su vez, en dos sub-etapas: identificación de la negación e identificación de su ámbito de acción. En la primera etapa, el sistema analiza cada oración comprobando la existencia de partículas o términos de negación, mientras que en la segunda identifica todos aquellos conceptos de la oración que se encuentran afectados por la negación identificada.

En este trabajo la detección de negación se centra en las llamadas *negaciones explícitas* (del inglés *explicit negations*) (Choi y Cardie, 2008); es decir, aquellas introducidas por términos como *not* o *never*. Por contra, este trabajo considera las llamadas *negaciones suaves o implícitas* (del inglés *soft negators*), como el adverbio *rarely* o el adjetivo *less*, como cuantificadores. La identificación de la negación en el sistema se realiza mediante un mecanismo sencillo que compara cada término de la oración con un conjunto de términos de negación presentes en una lista. Esta lista se ha realizado tras el estudio de diferentes aproximaciones previas (Councill, McDonald, y Velikovich, 2010; Morante, 2010). Este mecanismo permite, además, detectar errores frecuentes en la escritura como la omisión de apóstrofe en las negaciones en inglés. La lista de términos incluidos en el sistema se puede ver en la tabla 5.1. Esta lista ha sido desarrollada en formato XML para facilitar de esta manera su adaptación a futuros cambios.

No	None	Non	Nor	Nothing	Neither	Nobody
Never	Nowhere	Not	n't	Don't	Dont	Doesn't
Doesnt	Didn't	Didnt	Haven't	Havent	Can't	Cant
Cannot	Couldn't	Couldnt	Needn't	Neednt	Wasn't	Wasnt
Weren't	Werent	Daren't	Darent	Hadn't	Hadnt	Isn't
Isnt	Aren't	Arent	Oughtn't	Oughtnt	Wouldn't	Wouldnt
Hasn't	Hasnt	Mightn't	Mightnt	Mustn't	Mustnt	Shouldn't
Shouldnt	Won't	Wont	Shan't	Shant		

Tabla 5.1: Términos incluidos en la lista de negación

Sin embargo, en este mecanismo de detección de la negación se deben tener en cuenta algunas expresiones especiales en las que los términos de negación no constituyen una negación de la opinión o emoción, como por ejemplo ocurre en la oración *Not only the room was dirty but the staff was very rude*, donde el término *Not*, como parte de la expresión *Not only* no se comporta como una partícula de negación. De esta manera, cuando el término de negación pertenece a alguna de estas expresiones, se considera

que no existe tal negación. Ejemplos de este tipo de expresiones son: *not only*, *not just*, *not merely*, *not even*, *not to mention*, *nothing if not* y *no wonder*.

Una vez identificados los términos de negación de la oración, el ámbito de acción de dicha negación se determina utilizando el árbol sintáctico generado por el Stanford Parser en la primera etapa del método. Para cada término de negación, el sistema busca el primer ancestro común entre el término de negación y la palabra inmediatamente después de él, asumiendo todos los nodos hoja descendientes del ancestro, y que se encuentran a la derecha del término de negación, como dentro del ámbito de la negación. Este proceso se puede ver en la figura 5.2, donde se muestra el árbol sintáctico de la oración *Reception staff had limited English for an international hotel and were* <u>*not*</u> *very helpful*. En esta oración el sistema identifica el término de negación *not* y recupera el primer ancestro, ADJP, común entre éste y la palabra inmediatamente después de él, *very*, asumiendo por tanto como ámbito de la negación los conceptos *very helpful*.

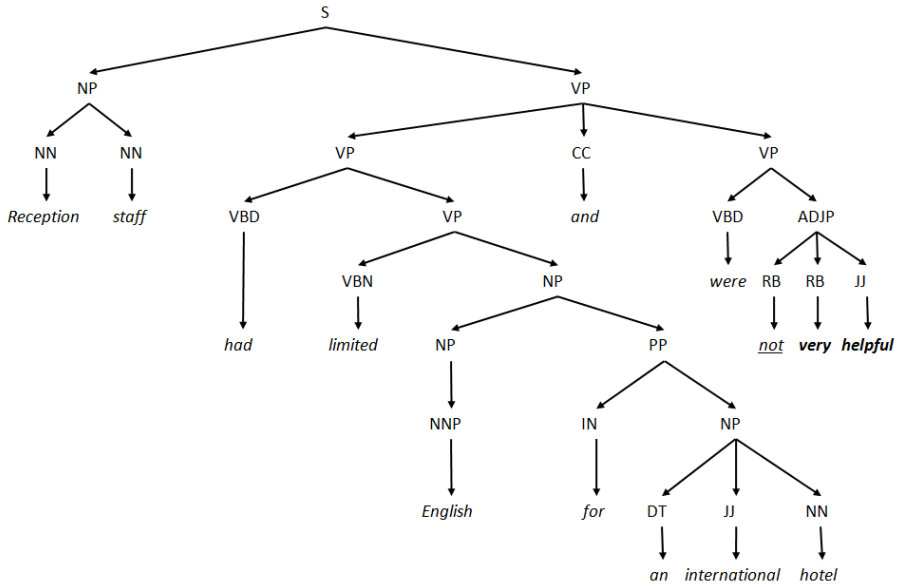


Figura 5.2: Árbol sintáctico de la oración *Reception staff had limited English for an international hotel and were not very helpful*. El término de negación *not* está subrayado. Los términos incluidos en su ámbito de acción se muestran en negrita

Sin embargo, asumir todos los descendientes del primer ancestro común como ámbito de la negación puede introducir ruido en ciertos casos como oraciones subordinadas o casos especiales de puntuación. Para evitar esto, el

módulo de negación incluye un conjunto de reglas que permiten delimitar el ámbito de la negación en estos casos especiales. Estas reglas están basadas en términos específicos (i.e., *because*, *until*, *why*, etc.) que introducen oraciones subordinadas o pueden delimitar el ámbito de la negación cuando actúan con determinados roles gramaticales. De esta manera, cuando se detecta una negación en una oración, se analiza cada término de su ámbito determinando si alguno coincide con los presentes en las reglas y actúa con el rol gramatical apropiado. En dichos casos, la regla correspondiente es lanzada delimitando así el ámbito de la negación. Los delimitadores del ámbito de la negación, así como sus roles gramaticales apropiados, se pueden ver en la tabla 5.2.

Término	**POS**
so, as, because, that, if until, since, unless, before	IN
while, about, than, despite	IN
who, what, whose	WP
Why, when, where	WRB
which	WDT
however	RB
that	RB
“,”, - , :, (,), !, ?	-

Tabla 5.2: Lista de delimitadores del ámbito de la negación

Un ejemplo de este proceso se puede ver en la oración *I did<u>n't</u> need any transportation <u>because</u> the hotel place is good.* El sistema detecta la negación y determina como posible ámbito el texto *need any transportation because the hotel place is good.* Sin embargo, como en ese ámbito se encuentra el término *because* que determina el inicio de una oración subordinada que debería estar fuera del ámbito de la negación, la regla apropiada es ejecutada y el ámbito de la negación para esta oración se reduce a *need any transportation.* El árbol sintáctico de esta oración se muestra en la figura 5.3.

Este método de detección de la negación y su ámbito de acción está basado en el propuesto por Jia et al. (2009), que ha demostrado mejorar los métodos tradicionales basados en determinar el ámbito de la negación mediante una ventana fija de palabras, o hasta el siguiente signo de puntuación o la siguiente expresión polar.

Incluso más importante que identificar el término de la negación y su ámbito de acción, es determinar su impacto o efecto sobre los conceptos y las emociones dentro de su ámbito de acción. Como ya se ha expuesto, la negación en análisis sentimental ha sido generalmente considerada como un

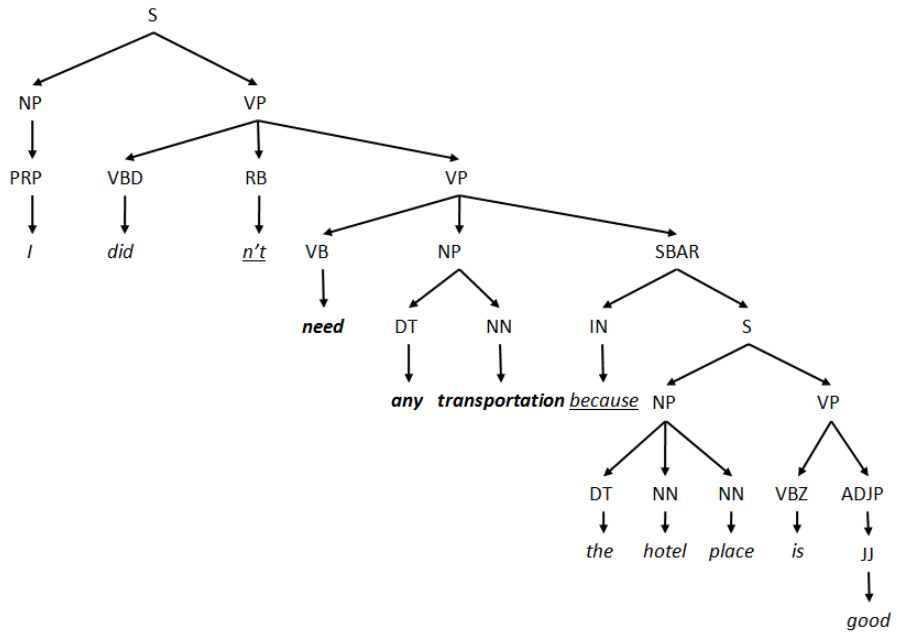


Figura 5.3: Árbol sintáctico para la oración *I didn't need any transportation because the hotel place is good*. El término de negación *n't* y el delimitador del ámbito de la negación *because* están subrayados. Los términos incluidos en su ámbito de acción aparecen en negrita

modificador que invierte la polaridad de los términos a los que afecta, cambiando las expresiones polares de positivas a negativas y viceversa (Agrawal y Siddiqui, 2009). Por el contrario, en este trabajo se propone un sofisticado modelo de la negación basado en la idea de que ésta cambia el significado emocional de los conceptos afectados dentro de su ámbito, y no solo su polaridad. Considérense, por ejemplo, las siguientes oraciones:

No deaths in the accident.

I've never been afraid of the future.

Aunque el concepto *deaths* de la primera oración expresa tristeza y miedo (*sadness* y *fear*), el hecho de que no haya habido ninguna muerte suscita felicidad y calma (*happiness* y *calmness*). En el segundo ejemplo, el concepto *afraid* denota miedo (*fear*), pero como está dentro del ámbito de acción de la negación, la oración en sí expresa una emoción de calma (*calmness*). Siguiendo esta idea, para cada concepto etiquetado con un categoría emocional que se encuentra en el ámbito de acción de una negación, el método cambia dicha

categoría emocional por su antónima. De esta manera, el léxico afectivo utilizado debe tener una relación de antonimia entre las categorías. Si no se encuentra la categoría emocional antónima en el léxico (porque la categoría en sí no tenga un antónimo claro), el concepto es etiquetado sin categoría emocional. La fuerza de la emoción antónima es, a su vez, disminuida en un valor α que debe ser determinado empíricamente. La razón para disminuir la fuerza de la emoción antónima es considerar a la negación como un modificador que invierte la emoción, pero que generalmente también disminuye su intensidad en un cierto grado. Así, por ejemplo, no tener miedo no implica necesariamente sentir calma, al igual que no querer a alguien no significa odiarla.

Por último, en la tabla 5.3 se pueden observar algunos ejemplos de oraciones con negación extraídas de diversas opiniones de productos.

There was no way to open the window or balcony door.
None of the window switches are lit at night, same goes to power locks.
Neither the bed, nor the pillows are very comfortable.
I have nothing else good to say about this camera.
We picked a non smoking room and the floor and the room smelt badly.
If you're looking to buy a really good 5.8GHz system, then seriously consider this one if you don't mind the price.
We have never experienced any interference on the phone by the other devices.

Tabla 5.3: Ejemplos de oraciones con expresiones negadas

5.5. Detección de cuantificadores y su ámbito

Los cuantificadores son el segundo grupo de modificadores emocionales a tener en cuenta cuando se plantea desarrollar un sistema de análisis sentimental. Los cuantificadores se definen como términos capaces de amplificar o disminuir la intensidad emocional del texto al que afectan, como, por ejemplo, los términos *very* y *less* en las oraciones *The room is very confortable* y *The party was less funny that I was expected.* Como se puede ver en estos ejemplos, el primero de ellos amplifica lo confortable que fue la estancia, *very confortable*, mientras que el segundo disminuye lo divertido que fue la fiesta, *less funny.* Cabe notar que no todos los cuantificadores tienen la misma intensidad cuando amplifican o disminuyen la carga emocional de un texto, puesto que no es lo mismo decir *The room was absolutely good* que *The room was quite good.*

La detección de cuantificadores y su ámbito sigue un proceso similar al realizado para la negación. Al igual que en la negación, los términos que pueden actuar como cuantificadores son identificados mediante una lista de cuantificadores. Existen multitud de listas y referencias en la literatura sobre términos que actúan como cuantificadores en los textos (Quirk et al., 1985; Brooke, 2009). En concreto, para este sistema se ha recopilado de la literatura una lista con los cuantificadores más comunes en inglés, que se muestra en la tabla 5.4. Es importante mencionar que estos términos solo actúan como cuantificadores cuando desempeñan un determinado rol gramatical (modificadores adverbiales y de adjetivos), por lo que cada elemento de la lista tiene asociados los roles gramaticales que puede adoptar cuando actúa como cuantificador. Un ejemplo de esto se puede ver en las oraciones *The rooms where so beautiful* y *The rooms where small so we did not like them*, donde el término *so* solo actúa como cuantificador en la primera oración. Al igual que en la negación, la lista de cuantificadores ha sido desarrollada en formato XML para facilitar de esta manera su adaptación a cambios futuros.

Términos	**%**	**Términos**	**%**	**Términos**	**%**
Very	75	Small	-50	So	75
Great	60	Really	80	More	35
Much	35	Little	-75	Bit	-35
Some	40	Less	-55	Barely	-75
Hardly	-85	Almost	-20	Slightly	-75
Relatively	-50	Somewhat	-60	Fairly	65
Pretty	75	Rather	75	Quite	35
Perfectly	75	Obviously	75	Certainly	75
Completely	85	Definitely	95	Absolutely	95
Highly	75	Truly	75	Especially	70
Particularly	45	Significantly	45	Awfully	75
Totally	85	Fully	85	Entirely	75
Strongly	55	Extremely	95	Incredibly	85
Terribly	75	Immensely	75	Such	55
Exceptionally	75	Exceedingly	85	Vastly	75
Super	75	Thoroughly	65	Tremendously	85
Ridiculously	-65	Extraordinarily	75	Total	75
Huge	75	Tremendous	85	Complete	75
Absolute	75	Incredible	75	Utter	75
Big	65	Bigger	35	Biggest	75
Extra	20	Major	35	High	55
Higher	45	Highest	65	Real	20
Minor	-75	Low	-75	Lower	-65
Lowest	-85	Few	-75	Fewer	-65
Fewest	-85	Lot	75	Difficult	-65
Only	-25				

Tabla 5.4: Términos incluidos en la lista de cuantificadores

Los cuantificadores son generalmente tratados en la literatura como modificadores que siempre afectan con una misma intensidad a las palabras o conceptos en su ámbito de acción. Es decir, se considera que todos amplifican o disminuyen con una misma intensidad (Polanyi y Zaenen, 2006; Kennedy y Inkpen, 2006). Sin embargo, como ya se ha comentado, no es lo mismo decir *Lunch was really good* que decir *Lunch was quite good*. Es por ello que en este sistema se ha seguido la idea presentada por Brooke (2009), donde cada cuantificador lleva asociado un porcentaje que varía según la intensidad con la que potencie o disminuya el texto al que afecta. De esta manera, la tabla 5.4 muestra, para cada cuantificador, su porcentaje asociado, donde los valores positivos representan a los cuantificadores que intensifican la intensidad emocional, mientras que los valores negativos representan los cuantificadores que la disminuyen.

El mecanismo seguido en el sistema para la detección de cuantificadores en los textos consiste en la comparación de los términos de cada oración con todas las entradas de la lista. Cuando el sistema detecta una coincidencia, comprueba a su vez que el rol gramatical del término identificado es el adecuado, y si es así, éste es marcado como cuantificador. El ámbito de acción de cada cuantificador es determinado de la misma manera que en la negación, mediante el árbol sintáctico de la oración generado por el Stanford Parser en la primera etapa del algoritmo. Así mismo, la lista de delimitadores de ámbito de acción expuesta en la tabla 5.2 para la negación, es la misma que se utiliza para delimitar el ámbito de acción de los cuantificadores para oraciones subordinadas.

Una vez que los cuantificadores han sido detectados, y que su ámbito de acción ha sido determinado, los conceptos afectados son marcados con los porcentajes de los cuantificadores, cuyo valor se usará posteriormente en la etapa de clasificación automática para aumentar o disminuir la intensidad emocional de esos conceptos. Para ilustrar este proceso, consideremos la oración:

> *We also got the rate that included breakfast, which was very convenient for us and less expensive than going somewhere else.*

En esta oración, el método identifica los cuantificadores *very* y *less*. El primero de ellos amplifica la intensidad emocional de los conceptos inmersos en el texto *convenient for us*, de forma que, como el concepto *convenient*

está etiquetado con la categoría emocional *like*, es marcado con el porcentaje del cuantificador *very* (75,0 %). En el caso del cuantificador *less*, su ámbito incluye los conceptos *expensive than going somewhere else*. Como el concepto *expensive* está etiquetado con la categoría emocional *disgust*, el concepto es marcado con el porcentaje de *less* (-65,0 %).

Por último, la tabla 5.5 muestra algunos ejemplos de oraciones extraídas de opiniones de diferentes productos.

The characters are not very well described, but only vaguely outlined.
Breakfast is expensive - much better to go buy some croissants in the morning.
The steering is perfectly weighted, quick, and sharp.
All the staff were extremely helpful.
This memory is cheapest, lowest quality on the market.
Our room and bathroom were cramped and we could barely fit our suitcases.
Braking tends to feel a bit mushy at highway speeds.
It was a noisy place, we could hardly sleep.

Tabla 5.5: Ejemplos de oraciones con cuantificadores

5.6. Detección de modales y su ámbito

La modalidad es otro de los factores a tener en cuenta a la hora de analizar un texto desde el punto de vista emocional, puesto que puede modificar su significado. Sin embargo, y a diferencia de los modificadores ya estudiados, el efecto concreto de la modalidad no es tan claro, y depende en gran medida de la forma en que se exprese dicha modalidad.

Para analizar la presencia o no de modalidad en el texto, en este trabajo se utiliza la clasificación lingüística de formas modales del inglés propuesta por Carter y McCarthy (2006). Esta clasificación distingue entre tres grandes categorías de formas modales: *verbos modales*, *expresiones modales* y *otros verbos*. Los verbos modales son, a su vez, clasificados en *verbos modales principales* y *verbos semi-modales* (ver tabla 5.6). Partiendo de esta clasificación, realizamos un análisis inicial para determinar cómo los términos modales son utilizados en los textos afectivos. Para ello, analizamos la presencia de las formas modales en diferentes córpora de opiniones de productos, tras lo que determinamos centrar nuestra atención en los verbos modales principales, puesto que los verbos semi-modales, las expresiones modales y los otros verbos con significado modal raramente aparecen en dichos córpora, y cuando lo hacen no modifican el significado emocional de los conceptos dentro de

su ámbito de acción. Un ejemplo de esto se puede ver en la oración *I got worse and needed to go back to hospital*, donde el verbo semi-modal *need* no modifica el significado emocional de la oración.

Verbos modales principales		**Verbos semi-modales**
Can	Would	Dare
Will	Should	Need
Could	Must	Ought to
May	Might	Used to
Shall		

Tabla 5.6: Lista de verbos modales

Centrándonos en los verbos modales principales, en nuestro análisis identificamos tres tipos de patrones de uso que determinan un efecto diferente en las emociones a las que afectaban: *verbos modales principales*, *verbos modales principales + to be* y *verbos modales principales + pretérito perfecto (have -ed)*. Por ello, centramos nuestro estudio y aplicación del modelo a estos tres patrones de verbos modales principales.

La identificación de los verbos modales es realizada en el método de la misma forma que con los modificadores anteriores, mediante una lista de formas modales donde cada término tiene asociado su correspondiente categoría gramatical para evitar errores de ambigüedad (por ejemplo el uso del término *may* como mes). Cada forma modal de la lista, a su vez, tiene asociada una o más reglas de acuerdo a los tres patrones identificados en el análisis. Análogamente, el ámbito de acción de los verbos modales es delimitado de la misma manera que en la negación y los cuantificadores (ver tabla 5.2).

Como ya se ha dicho, el efecto de la modalidad desde un punto de vista emocional no es tan homogéneo como en el caso de la negación y los cuantificadores. Es importante destacar que, para desarrollar un sistema capaz de capturar el significado emocional exacto de las diferentes formas modales y su efecto, éste debería incluir un alto grado de conocimiento pragmático que está fuera del alcance de la investigación actual en procesamiento del lenguaje natural. Esta es, quizá, la principal razón por la que muy pocas aproximaciones en la literatura han abordado el problema de la modalidad en el análisis sentimental, y las pocas que lo han hecho consideran las formas modales como meros modificadores que bloquean la polaridad de los términos a los que afecta su ámbito (Polanyi y Zaenen, 2006; Brooke, 2009).

La modalidad tiene muchas y muy diversas interpretaciones, pero la

mayoría de ellas se centran en la idea de incertidumbre y probabilidad frente a eventos o situaciones. Esto parece sugerir que la modalidad pudiera actuar como un modificador que invierte la polaridad, puesto que no hay certeza de que el evento o situación haya sucedido realmente. Sin embargo, desde un punto de vista emocional, la incertidumbre de un evento o situación no afecta por regla general al significado emocional del conjunto del texto, como se puede ver en la oración *We would obviously stay there again when we return.*

Otra interpretación importante de la modalidad, como exponen Carter and McCarthy (2006), hace referencia a su uso para expresar deseos, necesidades u obligaciones acerca de eventos que han pasado, que están pasando o que pasarán. Esta interpretación de la modalidad es, quizás, la más frecuentemente utilizada en los textos afectivos y la que verdaderamente afecta a la polaridad e intensidad emocional del texto, como se puede ver en las oraciones *It would have been nice to have another window*, *The staff at the reception could be more pleasant* o *The hotel should have been more organized.*

De esta manera, los diferentes usos de la modalidad determinan los diferentes efectos sobre el significado emocional del texto:

1. **Efecto I:** ningún efecto, cuando la modalidad es utilizada para expresar un grado bajo de incertidumbre.

2. **Efecto II:** decremento de la intensidad de las emociones dentro de su ámbito de acción, cuando la modalidad es utilizada para expresar deseos o un alto grado de incertidumbre.

3. **Efecto III:** inversión de la polaridad de las emociones, decrementando además la intensidad de las mismas, cuando la modalidad es utilizada para expresa necesidades, obligaciones y quejas o desacuerdos.

Tras un análisis cuidadoso de los diferentes usos de los verbos modales principales en los textos con carga emocional, hemos asociado cada forma modal con uno de los posibles efectos identificados. Las principales conclusiones de este análisis, que están resumidas en la tabla 5.7, son:

- El análisis ha revelado que los verbos modales principales *could* y *should* son, por regla general, los utilizados por las personas para expresar necesidades, obligaciones y quejas o desacuerdos. Estos verbos

sirven, generalmente, para comunicar la perspectiva de las personas sobre cómo deberían ser los eventos o situaciones, cambiando así el significado emocional de los términos dentro de su ámbito de acción. Algunos ejemplos de uso de estos verbos son:

Internet could be free.

They should offer a better quality toilet paper.

Sin embargo, es importante destacar que el verbo *could* por sí solo es más utilizado como el pasado del verbo *can*, y en raras ocasiones expresa modalidad, como ocurre en la oración:

I couldn't go to your party because I was sick.

- Los deseos, por el contrario, son generalmente expresados mediante el uso del verbo modal *would*, cuyo significado denota una sugerencia sobre cómo los eventos o situaciones deberían realizarse para mejorarlos, decrementando así la intensidad emocional de los términos afectivos en su ámbito. Un ejemplo de este uso del verbo *would* se puede ver en la oración:

  It would be nice if there is a kettle in the room.

  Por el contrario, el patrón *would + have -ed* normalmente expresa desacuerdo, por lo que cambia el significado emocional del texto disminuyendo, a la vez, su intensidad, como se puede ver en la oración:

  Tea and coffee making facilities in the rooms would have been good.

- Los verbos modales principales *must* y *might* raramente cambian el significado emocional del texto, expresando principalmente un grado pequeño de incertidumbre, como se puede ver en las oraciones:

  Pictures online must be a different hotel.

  There is a shopping mall next to the hotel where you can get anything you might need.

Sin embargo, cuando el verbo *might* es utilizado con el patrón *might + have -ed* generalmente expresa desacuerdo, por lo que, al igual que *would*, invierte el invierte el significado emocional y disminuye la intensidad de los términos afectados. Un ejemplo de esto se puede ver en la oración:

A touch of romance <u>might</u> have been nice for the story.

- El verbo *may* es utilizado principalmente para expresar un alto grado de incertidumbre. De esta manera, su efecto disminuye la intensidad emocional del texto afectado por su ámbito, como muestran las oraciones:

  Too little sleep <u>may</u> mean too fat kids.

  The surroundings <u>may</u> be good for some people but not me.

- El análisis de los verbos modales principales *can*, *will* and *shall*[3] ha revelado que estos verbos son utilizados generalmente para expresar situaciones con un grado muy bajo de incertidumbre o probabilidad, por lo que raramente modifican el significado emocional del texto. Un ejemplo de esto se puede ver en la oraciones:

  You <u>can</u> hear the toilet flushing.

  If you run into problems this hotel <u>will</u> do everything possible to help.

  I <u>shall</u> prevent the slightest shadow of partiality.

- En el análisis hemos querido también determinar el efecto de los verbos semi-modales *dare*, *need*, *ought to* and *used to*. Las conclusiones extraídas han sido que estos verbos juegan un papel similar a los verbos *can*, *will* and *shall*, por lo que no modifican el significado emocional del texto al que afectan.

Por otro lado, hemos estudiado también si el efecto de la negación puede cambiar el efecto de los verbos modales desde el punto de vista emocional. Las conclusiones extraídas muestran que la negación no modifica el efecto de

[3]Nos referimos al uso de *shall* para describir eventos futuros.

Forma Modal	**Efecto**
Could	Ningún efecto
Could + be	Emoción antónima + disminución
Could + have -ed	Emoción antónima + disminución
Should	Emoción antónima + disminución
Should + be	Emoción antónima + disminución
Should + have -ed	Emoción antónima + disminución
Would	Ningún efecto
Would + be	Disminución
Would + have -ed	Emoción antónima + disminución
Must	Ningún efecto
Must + be	Ningún efecto
Must+ have + ed	Ningún efecto
May	Disminución
May + be	Disminución
May + have -ed	Disminución
Might	Ningún efecto
Might + be	Ningún efecto
Might + have -ed	Emoción antónima + disminución

Tabla 5.7: Verbos modales y su efecto

la modalidad. Para ilustrar esto, consideremos la oración *Breakfast couldn't have been better (**like**) for the price paid.* El primer paso sería considerar el efecto de la modalidad sin tener en cuenta la negación. La forma modal *could* cambia el significado emocional de los términos afectivos, dentro del conjunto *better (**like**) for the price*, por su significado antónimo (***disgust***), disminuyendo además sus intensidades. Sin embargo, como la forma *could* se encuentra dentro del ámbito de una negación, una vez que su efecto ha sido tratado se debe tener en cuenta el efecto de esta negación. En concreto, la negación invierte de nuevo el significado emocional de *better (**disgust**) for the price* por su antónimo (***like***).

Para finalizar, la tabla 5.8 muestra algunos ejemplos de oraciones con presencia de formas modales obtenidas de diferentes opiniones de diversos productos.

We could get new towels at the desk.
The bathroom could have been cleaner.
The staff should communicate more with each other.
I would recommend my friends to stay in this hotel.
A traditional American breakfast with eggs would have been nice to have.
The surroundings may be good for some people but not me.
If you pay in euros you might receive the change in Romanian currency.
This must have been the worst room in the building.

Tabla 5.8: Ejemplos de oraciones con formas modales

5.7. Clasificación automática

Un vez analizadas todas las oraciones, identificadas las categorías emocionales de sus conceptos y resueltos los modificadores emocionales de negación, cuantificadores y modalidad, el texto es traducido a un *Vector de Intensidades Emocionales* (VEI, por las siglas en inglés de *Vector of Emotional Intensities*), que será la entrada de un algoritmo de aprendizaje máquina. El VEI es un vector de 14 posiciones, cada una de ellas representando a una categoría emocional del léxico afectivo SentiSense. El texto es transformado en un VEI de la siguiente manera:

- Por cada concepto, C_i, etiquetado con una categoría emocional, E_j, el valor asignado por ese concepto a esa categoría emocional es 1,0.

- Si no se ha etiquetado el concepto con una categoría emocional, entonces la categoría de su primer hiperónimo etiquetado, E_{hiper_j}, es utilizada. Como la hiperonimia generaliza el significado del concepto, el valor asignado por el concepto a la categoría emocional del hiperónimo es calculado mediante la ecuación 5.1.

$$valor(C_i, E_{hiper_j}) = \frac{1}{prof(hiper_j) + 1} \tag{5.1}$$

- Si el concepto está dentro del ámbito de una negación, el sistema utiliza la categoría antónima, E_{anton_j}, si hubiera. El valor del concepto para la categoría emocional antónima es multiplicado por $0,6$. Para determinar este valor se han realizado diferentes experimentos, obteniendo los mejores resultados con este valor. La razón de utilizar un valor inferior a $1,0$ para las categorías emocionales dentro del ámbito de una negación se basa en la creencia del autor de que una negación invierte la polaridad del significado emocional, aunque generalmente también disminuye su intensidad.

- Si el concepto está dentro del ámbito de un cuantificador, el valor del concepto para su correspondiente categoría emocional es incrementado/decrementado con el porcentaje del cuantificador asignado al concepto en la etapa cinco, según la ecuación 5.2.

$$valor(C_i, E_j) = valor(C_i, E_j) \times (1 + cuantificador\ porcentaje) \tag{5.2}$$

- Si el concepto se encuentra en el ámbito de acción de una forma modal, entonces el valor del concepto para su categoría emocional es calculado como sigue (ecuación 5.3), donde los valores $0,5$ y $0,6$ han sido empíricamente determinados.

$$valor(C_i, E_j) = \begin{cases} valor(C_i, E_j) & \text{si efecto modal de tipo I} \\ valor(C_i, E_j) * 0,5 & \text{si efecto modal de tipo II} \\ valor(C_i, E_{anton_j}) * 0,6 & \text{si efecto modal de tipo III} \end{cases} \tag{5.3}$$

- Por último, la posición en VEI de la categoría emocional asignada al concepto es incrementada con el valor del concepto previamente calculado.

Por ejemplo, la oración *Cleaning starts very early even on Sunday, furnitures are moved, very noisy, not possible to relax and sleep* produce el VEI *[1,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 2,35]* (*like, compassion, joy, anger, hate, expectancy, love, sadness, calmness, ambiguous, fear, hope, surprise, disgust*). En esta oración, los conceptos *cleaning* y *relax* son etiquetados por el sistema con la categoría emocional *like*, mientras que el concepto *noisy* es etiquetado con la categoría emocional *disgust*. Sin embargo, como el concepto *relax* se encuentra dentro del ámbito de una negación, el sistema recupera su categoría emocional antónima, *disgust*, y añade 0,6 a la posición en VEI. Además, el sistema identifica el término *very* que actúa como cuantificador sobre el concepto *noisy* amplificando su intensidad en un 75,0 %. De esta manera, la posición de la categoría emocional *like* en el VEI obtiene un valor final de 1,0, mientras que la posición de la categoría emocional *disgust* consigue un valor de 2,35 (0,6 de *relax* + 1,75 de *very noisy*).

Una vez finalizado el análisis de todo el texto y generado su VEI, éste es utilizado como entrada de un algoritmo de aprendizaje máquina donde será clasificado de acuerdo a su polaridad e intensidad emocional.

Capítulo 6

Evaluación

A lo largo de este capítulo se presenta la metodología de evaluación, así como los resultados de los diferentes experimentos realizados para evaluar el rendimiento del sistema propuesto. Los resultados son, además, comparados con los obtenidos por dos aproximaciones basadas, respectivamente, en frecuencias de términos y expresiones polares. Con estos experimentos, se pretende evaluar el rendimiento del método propuesto para las tareas de clasificación de textos en base a su polaridad y a su intensidad, en ambos casos a nivel de oración y a nivel de documento. Además, se pretende cuantificar el efecto de los modificadores emocionales y de la ambigüedad sobre la calidad de los resultados obtenidos.

6.1. Metodología de evaluación

6.1.1. Métricas de evaluación

Las métricas más utilizadas en tareas de clasificación emocional de textos son la *exactitud* (*accuracy*), la *precisión*(*precision*), la *cobertura* (*recall*) y la *medida-F* (*F-measure* o *F1*). Para un correcto entendimiento de estas métricas, es necesario explicar lo que se conoce con el nombre de *tabla de verdad*, que representa todas las posibles combinaciones de acierto/error del clasificador. Considérese la tabla 6.1, donde se presentan las posibles combinaciones de clasificación en dos clases de un sistema, por ejemplo, de detección de polaridad. Se pueden ver en la tabla tanto las predicciones realizadas por el sistema automático, como las clasificaciones correctas según un etiquetado manual. Si un documento es clasificado automáticamente en

la misma categoría que la clasificación manual, estaremos ante un *verdadero positivo (TP, true positive)* o un *verdadero negativo (TN, true negative)*, mientras que si el documento es clasificado automáticamente en una categoría diferente a la clasificación manual estaremos ante un *falso negativo (FN, false negative)* o un *falso positivo (FP, false positive)*.

	Automático	
Manual	Positivo	Negativo
Positivo	TP	FN
Negativo	FP	TN

Tabla 6.1: Posibles combinaciones de resultados en la clasificación en dos clases

Atendiendo a esta tabla, las diferentes métricas expuestas se definen como sigue:

- La ***exactitud*** *(accuracy)* mide el porcentaje de aciertos en relación al total de documentos para el clasificador en ambas clases. La exactitud es una medida global de acierto del clasificador en la tarea, y se calcula según la ecuación 6.1.

$$exactitud = \frac{TP + TN}{TP + FN + FP + TN} \tag{6.1}$$

- La ***precisión*** *(precision)* mide el número de aciertos del clasificador para una clase entre el total de clasificados en esa clase. Es decir, la precisión mide la exactitud del clasificador para esa clase, puesto que una mayor precisión indica un menor numero de *falsos positivos*, y por el contrario, una baja precisión indica muchos *falsos positivos*. La precisión se calcula según la ecuación 6.2.

$$precisión_P = \frac{TP}{TP + FP} \tag{6.2}$$

- La ***cobertura*** *(recall)* mide el número de aciertos para una clase en relación al número de documentos que deberían haber sido clasificados en esa clase. Es decir, la cobertura se puede ver como la sensibilidad del clasificador para esa clase, ya que a mayor cobertura menor número de *falsos negativos*, mientras que una cobertura baja indica mayor número de *falsos negativos*. La cobertura se calcula según la ecuación 6.3.

$$cobertura_P = \frac{TP}{TP + FN} \tag{6.3}$$

- La ***medida-F*** *(F-Score)* es una combinación de las medidas anteriores, precisión y cobertura, que representa la media armónica de la precisión y la cobertura. La medida-F se calcula según la ecuación 6.4.

$$medida - F = \frac{2 \times precisión \times cobertura}{precisión + cobertura} \tag{6.4}$$

Estas métricas, aunque presentadas aquí en relación a una *tabla de verdad* de 2x2, pueden ser extrapoladas a cualquier número de clases. Las ecuaciones 6.5 y 6.6 muestran la precisión y cobertura, respectivamente, para un clasificador de tres clases, y cuyos parámetros se obtienen de la tabla de verdad de dimensión 3x3 mostrada en la figura 6.2. Por regla general, cuando se realizan evaluaciones con más de dos categorías, los resultados que se presentan son las medias de la precisión, cobertura y medida-F obtenidas para las diferentes clases.

$$precisión_A = \frac{tp_A}{tp_A + e_{BA} + e_{CA}} \tag{6.5}$$

$$cobertura_A = \frac{tp_A}{tp_A + e_{AB} + e_{AC}} \tag{6.6}$$

	Automático		
Manual	A	B	C
A	tp_A	e_{AB}	e_{AC}
B	e_{BA}	tp_B	e_{BC}
C	e_{CA}	e_{CB}	tp_C

Tabla 6.2: Posibles combinaciones de resultados en la clasificación en tres clases

Los principales motivos por los que se han popularizado estas métricas en la evaluación de tareas de análisis sentimental son, en primer lugar, que la mayoría de los problemas son planteados como tareas de clasificación de documentos en diferentes categorías; en segundo lugar, que el análisis sentimental surge, en parte, como una tarea dentro de la categorización de documentos donde la clasificación depende del contenido emocional del texto; y por último, su facilidad de cálculo y cómputo, así como la no necesidad de intervención humana.

A la hora de utilizar estas métricas es importante tener en cuenta la distribución de los documentos entre las diferentes clases, puesto que si una clase obtiene una distribución inferior al 20 % o 30 % los resultados para la exactitud pueden ser ligeramente superiores a lo que deberían ser, debido al mayor porcentaje de documentos de la clase predominante. Esta problemática es aún mayor en los sistemas supervisados, donde el método de aprendizaje puede sobre-aprender la clase mayoritaria, en perjuicio de las restantes clases. Debido a estas posibles desviaciones, es frecuente presentar los resultados de las métricas *exactitud*, *precisión* y *cobertura*, o *exactitud* y *medida-F*, puesto que de esta manera se pueden ver las desviaciones de una u otra clase, respecto de la media armónica.

Otro aspecto a tener en cuenta a la hora de evaluar sistemas de clasificación supervisados es la distinción que hay que hacer entre el conjunto de entrenamiento (grupo de documentos a partir de los cuales el clasificador aprenderá) y el conjunto de evaluación (grupo de documentos que serán clasificados automáticamente por el sistema y sobre los que se calculan las métricas de evaluación). Es necesario separar ambos conjuntos para obtener una evaluación correcta, ya que si no se realiza esta división los resultados del sistema podrían ser ligeramente mejores al estar el sistema sobre-adaptado al conjunto utilizado para el aprendizaje. Es por esto que, normalmente, se realizan k divisiones equitativas sobre el conjunto de documentos totales, donde *k-1* partes son utilizadas para aprender y la restante para evaluar. Este método, considerado como el más ecuánime para realizar la evaluación de este tipo de sistemas, se denomina *cross fold validation*. Por regla general, se utilizan diez divisiones (*10 fold-cross validation*); es decir, se divide el conjunto total de documentos en 10 conjuntos. Cada subconjunto es utilizado como conjunto de evaluación, mientras que el resto es utilizado como conjunto de entrenamiento, evaluando el sistema diez veces y obteniendo como resultado final la media de las k evaluaciones parciales.

6.1.2. Colecciones de evaluación

Las colecciones de evaluación son conjuntos de datos etiquetados con información relevante para la tarea para la cual han sido desarrollados. En concreto, las colecciones de evaluación en análisis sentimental están compuestas por fragmentos de texto, ya sean oraciones, párrafos o documentos completos, que pueden ser de diversos temas, como noticias, revisiones de

productos o servicios, o blogs, y que están etiquetados con categorías como *positivo*, *negativo*, *neutral*, o incluso con una categoría emocional representativa de la emoción global, como *joy*, *anger* o *sadness*.

Las colecciones de evaluación permiten a los sistemas determinar el rendimiento de éstos, así como compararlo con el de otros sistemas mediante métricas estandarizadas y bajo los mismos supuestos de evaluación. Además, en los sistemas supervisados, estas colecciones permiten entrenarlos utilizando un subconjunto de datos etiquetados de la colección, y evaluarlos sobre otro subconjunto.

En concreto, para este trabajo se han seleccionado tres colecciones de evaluación que se ajustan al propósito del sistema desarrollado, y que permiten evaluar el método para las tareas de clasificación de polaridad e intensidad, ambas a nivel de oración y a nivel de documento. Las dos primeras colecciones (el conjunto SemEval 2007 y el SFU corpus) son colecciones ampliamente utilizadas en la literatura. El conjunto SemEval 2007 es utilizado en esta evaluación para realizar experimentos de clasificación de polaridad e intensidad a nivel de oración, mientras que el SFU corpus es utilizado para evaluar la clasificación de polaridad a nivel de documento. La tercera colección de evaluación, el HotelReview corpus, ha sido desarrollado como parte de esta tesis y permite evaluar el método para las tareas de clasificación de polaridad e intensidad, ambas a nivel de oración y documento. A continuación se describen en detalle cada una de estas colecciones.

SemEval 2007: Affective Text Task

La colección de evaluación *SemEval 2007 Affective Text Task*[1] está compuesta por titulares de noticias extraídas de páginas web como Google o CNN, así como de distintos periódicos. El propósito de la elección de noticias para el desarrollo de la colección de evaluación, según apuntan sus autores (Strapparava y Mihalcea, 2007), es doble: por un lado, se persigue evaluar sistemas de detección de polaridad y la detección de emociones sobre este género de documentos; y por otro, dado que los titulares de noticias se presentan, por regla general, mediante una sola oración, la colección permite realizar una evaluación de la detección de polaridad a nivel de oración. Esta colección se presenta en dos conjuntos, uno de 1000 oraciones y otro de 250, donde cada oración está etiquetada con dos propósitos: la detección de la polaridad

[1] SemEval 2007 Affective Text Task. http://www.cse.unt.edu/~rada/affectivetext/

y de la intensidad emocional de cada titular, y la detección de la emoción predominante de entre un conjunto de 6 emociones básicas.

La colección de evaluación presenta para las tareas de detección de polaridad e intensidad, que son las utilizadas en este trabajo, un valor numérico para cada oración en el intervalo [-100, 100], donde -100 indica la máxima carga negativa del titular, 0 representa un titular neutral y 100 un titular altamente positivo. El etiquetado de las oraciones dentro de este intervalo es quizá la mayor diferencia de esta colección con respecto a otras, y posibilita la realización de evaluaciones tanto para la detección de polaridad como para la detección del grado o intensidad de dicha polaridad. Precisamente, éste ha sido el principal motivo para su utilización en esta tesis doctoral, ya que permite evaluar el sistema a nivel de oración tanto para la tarea de detección de polaridad como para la detección de su intensidad. Por otro lado, esta colección nos permite evaluar el sistema en un dominio no tan frecuente en este área, como son las noticias periodísticas, permitiendo así determinar cómo se comporta el sistema frente a textos, en principio objetivos, pero con una carga de polaridad por su semántica.

En concreto, para acometer la evaluación de la clasificación de polaridad, hemos dividido las 1000 oraciones del conjunto de entrenamiento en dos clases (negativa = [-100, 0), positiva = [0,100]), obteniendo 526 titulares para la clase negativa y 474 titulares para la positiva. Para la clasificación de intensidad, hemos dividido los 1000 titulares en tres clases de acuerdo con la división propuesta por la competición SemEval del año 2007 (negativo = [-100, -50), neutral = [-50, 50) y positivo = [50, 100]); y en cuatro clases (fuertemente negativo = [-100, -50), negativo = [-50, 0), positivo = [0, 50) y fuertemente positivo = [50, 100]). Como resultado, en la primera distribución (3 clases) se dispone de 251 instancias para la clase negativa, 598 instancias para la clase neutral y 151 instancias para la clase positiva. En la segunda distribución (4 clases) disponemos de 251 instancias para la clase fuertemente negativa, 279 instancias para la clase negativa, 319 instancias para la clase positiva y 151 instancias para la clase fuertemente positiva.

SFU review corpus

La colección *Simon Fraser University review corpus*[2] (SFU) (Taboada y Grieve, 2004) está formada por 400 comentarios de productos y servicios ex-

[2]SFU corpus. http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html

traídos de la página web *epinions.com*[3]. Estos 400 comentarios se agrupan en conjuntos de 50 comentarios, 25 etiquetados como positivos y 25 etiquetados como negativos, para un total de 8 productos o servicios: libros, coches, ordenadores, utensilios de cocina, hoteles, películas, música y teléfonos. La polaridad de estos comentarios ha sido extraída mediante la etiqueta de "recomendado" o "no recomendado" de la página web *epinions.com*.

El uso de comentarios o *revisiones* de productos o servicios para la creación de colecciones de evaluación se ha convertido, en los últimos años, en una práctica muy habitual en el área de análisis sentimental. Los dos principales motivos para ello son, por un lado, el interés tanto de empresas como gobiernos por determinar automáticamente la polaridad de los comentarios realizados por los usuarios sobre sus productos, servicios y decisiones; y por otro, la facilidad de realizar colecciones de evaluación automáticamente sin la necesidad de etiquetar manualmente esos conjuntos gracias a la existencia de un gran número de páginas web, foros, etc., donde los usuarios realizan comentarios que ellos mismos etiquetan en base a su polaridad.

Por último, la principal razón que nos ha llevado a elegir este corpus para la evaluación del sistema es la posibilidad de realizar experimentos del método a nivel de documento para la tarea de detección de polaridad. Además, al ser una colección formada por grupos de 50 documentos sobre diferentes productos, permite evaluar el sistema en lo que se denomina *multi-dominio*; es decir, entrenar el algoritmo de aprendizaje máquina sobre 7 dominios y evaluarlo sobre otro diferente mediante *cross fold validation*, puesto que los documentos en la colección están ordenados por productos.

HotelReview Corpus

Puesto que las dos colecciones anteriores solo permiten evaluar el sistema en las tarea de detección de polaridad e intensidad a nivel de oración en el dominio periodístico (SemEval 2007), y en la tarea de detección de polaridad a nivel de documento para comentarios de productos (SFU review corpus), decidimos realizar una nueva colección de evaluación que permitiera evaluar el sistema en la tarea de detección de polaridad tanto a nivel de oración como de documento, y en la tarea de predicción del grado de polaridad a nivel de documento, ambos en el dominio de comentarios de productos o servicios. No se planteó la posibilidad de realizar una colección de evaluación

[3]Epinions.com. http://www10.epinions.com/

sobre noticias periodísticas ya que, aunque los titulares pueden contener cierta orientación afectiva, en general las noticias suelen ser en su mayoría objetivas.

La colección de evaluación se ha realizado a partir de comentarios de hoteles extraídos de la página web *booking.com*[4]. Los comentarios en esta página se dividen entre *lo que te ha gustado* y *lo que te ha disgustado* del hotel, pudiendo de esta manera asumir que las oraciones en el primer grupo son positivas y las oraciones en el segundo negativas. Además, el comentario global sobre el hotel lleva asociada una etiqueta con una puntuación en el rango 1 a 10, donde 1 corresponde a una muy mala experiencia del usuario, y 10 corresponde a una experiencia excelente.

Para construir el corpus, se extrajeron automáticamente 25 comentarios por hotel de un total de 60 hoteles (25 x 60 = 1500 comentarios), donde cada comentario contiene la siguiente información:

- La ciudad donde se ubica el hotel.
- La nacionalidad del usuario que escribió el comentario.
- La fecha en la que se escribió el comentario.
- El tipo de usuario, de entre las siguientes siete categorías: *solo traveler*, *young couple*, *mature couple*, *people with friends*, *group*, *family with young children* y *family with older children*.
- Una puntuación, entre 1 y 10, evaluando la opinión general del usuario. Esta puntuación no es asignada directamente por el usuario, sino que es calculada automáticamente por *booking.com* mediante unas puntuaciones que el usuario asigna a 6 aspectos del hotel: *personal del hotel*, *servicios/facilidades*, *limpieza de la habitación*, *confort*, *calidad por precio* y *localización*. Desafortunadamente, estas puntuaciones desagregadas no están accesibles.
- El texto del comentario, donde se distinguen dos secciones separadas en las que el usuario expone lo que le ha gustado y lo que le ha disgustado del hotel.

Una vez extraídos los 1500 comentarios, se analizaron los textos en relación a la puntuación que *booking.com* genera automáticamente, y se llegó

[4]Booking.com. http://www.booking.com

a dos conclusiones principales. Por un lado, por regla general, ninguna puntuación es inferior a 4, lo que puede deberse a que hay muchos factores de evaluación que se compensan unos con otros. Por otro lado, y más importante, generalmente esa puntuación no representa la información recogida en el texto del comentario, llegando incluso a darse casos en los que dos comentarios con puntuaciones similares representan opiniones muy diferentes. Por ejemplo, las dos opiniones siguientes obtienen una puntuación de 6,5, pero la segunda es claramente más negativa que la primera:

- *Good location. Nice roof restaurant - (I have stayed in the baglioni more than 5 times before). Maybe reshaping/redecorating the lobby.*
- *Noisy due to road traffic. The room was extremely small. Parking awkward. Shower screen was broken and there was no bulb in the bedside light.*

Como solución a este problema, se solicitó a dos personas que etiquetaran cada comentario de la colección de evaluación con dos categorías, la primera de entre las siguientes cinco opciones: *Excellent*, *Good*, *Fair*, *Poor* y *Very Poor*; y la segunda de entre las tres siguientes opciones: *Good*, *Fair* y *Poor*. Para resolver discrepancias de juicio entre los dos anotadores, todos aquellos comentarios con conflicto fueron eliminados de la colección. Se seleccionaron, finalmente, 1000 comentarios para conseguir 200 comentarios para cada categoría en la distribución de cinco clases; y 349 comentarios negativos, 292 comentarios neutrales y 359 comentarios positivos en la distribución de tres clases. La colección está disponible para su uso en investigación[5] en formato XML. La tabla 6.3 muestra, a modo de ejemplo, un comentario extraído del corpus.

De esta manera, la colección permite evaluar sistemas en la tarea de detección de polaridad a nivel de documento, mediante los textos etiquetados con las categorías *NegativeOpinion*, documento negativo, y *PositiveOpinion*, documento positivo, que forman un total de 1585 documentos distribuidos en 733 documentos negativos y 852 documentos positivos. Además, utilizando un sistema que divida el texto de cada categoría en oraciones, se puede utilizar el corpus para evaluar la tarea de detección de polaridad a nivel de oración, resultando en una colección de 4340 oraciones distribuidas en 2187

[5]HotelReview corpus. http://nil.fdi.ucm.es/index.php?q=node/456

<HotelReview *idDoc*="D_8" *hotelID*="H_2" *hotelLocation*="Paris" *reviewerCategory*="Young couple" *reviewerNationality*="Belgium" *date*="February 10, 2010" *score*="9.3" *5_classes_intensity*="Good"> *3_classes_intensity*="Good">

<PositiveOpinion> I liked the location, breakfast was nice as well as Tea Time Buffet, that was really nice. Parking on weekends is free (on the street, and it's safe). We got to the room and it was very smelly (cigarettes) so we asked and changed and got a nice room. I'd recommend this hotel definitely. But hey... it's not a 4 star hotel...</PositiveOpinion>

<NegativeOpinion>Staff is nice except for one receptionist (a man) at night, he was not helpful at all, I asked him for directions and he said there's nothing I can do if you don't know Paris. Anyways, everybody else was nice.</NegativeOpinion>

</HotelReview>

Tabla 6.3: Comentario extraído de la colección HotelReview corpus

oraciones negativas y 2153 oraciones positivas. Por otro lado, utilizando las etiquetas *5_classes_intensity* y *3_classes_intensity*, se pueden evaluar sistemas en la tarea de clasificación de la intensidad a nivel de documento.

6.1.3. Algoritmos de aprendizaje máquina

Para determinar los algoritmos de aprendizaje máquina que mejor se comportan para el sistema desarrollado en las tareas de clasificación de la polaridad e intensidad, se han probado los distintos clasificadores presentes en Weka (ver sección 3.2), entrenados y evaluados mediante *10-cross fold validation*. Finalmente, y para no mostrar los resultados de todos los clasificadores, se han elegido los tres que mejor comportamiento promedio han mostrado: un modelo lógico de regresión (*Logistic*), un árbol funcional (*Functional Tree (FT)*) y máquinas de vectores soporte (*Support Vector Machine (LibSVM)*). Conviene mencionar que los mejores resultados de los experimentos fueron obtenidos con los valores por defecto en Weka para los tres clasificadores.

6.1.4. Líneas base

En el campo del análisis sentimental, la comparación de unas aproximaciones con otras es prácticamente inexistente. El principal motivo de esta situación radica en la imposibilidad de replicar exactamente las condiciones de evaluación, bien por no disponer de la colección utilizada, bien por falta de detalles en la especificación de la metodología. Es por esto que en este trabajo se han desarrollado dos líneas base con las que comparar los resultados obtenidos por el sistema diseñado. La primera de ellas se basa en la

investigación de Pang et al. (2002), que plantea un sistema de SVM (*Support Vector Machines*) sobre bolsas o vectores de unigramas para la clasificación de textos en función de su carga emocional. El método se define como sigue: sea $u_1, ..., u_m$ un conjunto de m unigramas que pueden aparecer en el documento a evaluar. Sea $f_i(d)$ el número de veces que u_i ocurre en el documento d. El documento d es representado por el vector $R = f_1(d), ..., f_m(d)$ y utilizado como entrada para el clasificador LibSVM de Weka. El conjunto m de unigramas ha sido limitado a los unigramas que aparecen al menos 4 veces en la colección de evaluación. A pesar de su simplicidad, este método es considerado mucho más efectivo que otras aproximaciones más complejas que utilizan bi-gramas, la posición de las palabras o la categoría gramatical de los términos, y, por tanto, se trata de una línea base difícil de batir.

La segunda línea base es una modificación no supervisada del sistema presentado en esta tesis doctoral, aunque basado en expresiones polares en lugar de emociones, en el que cada documento es transformado en un número que indica la polaridad e intensidad del mismo. Para ello, se ha asignado a cada categoría emocional del léxico afectivo SentiSense un valor positivo, neutral o negativo, de acuerdo con la categoría emocional, tal y como se muestra en la tabla 6.4.

Categoría	**Valor**	**Categoría**	**Valor**
joy	+1,0	hope	+1,0
fear	-1,0	compassion	0,0
surprise	0,0	ambiguous	0,0
anger	-1,0	sadness	-1,0
disgust	-1,0	calmness	+1,0
love	+1,0	like	+1,0
expectancy	0,0	hate	-1,0

Tabla 6.4: Categorías emocionales y sus respectivos valores de polaridad

De esta manera, el resultado final del documento a analizar se obtiene de la suma de todos los valores de los conceptos etiquetados con categorías emocionales. La negación y la modalidad se tratan como es habitual en los sistemas basados en expresiones polares; es decir, son tratadas como fenómenos que cambian la polaridad, por lo que cuando un concepto se encuentra en el ámbito de una negación o de un modal su valor obtenido de la categoría emocional es multiplicado por -1. Los cuantificadores y los hiperónimos son tratados de la misma manera que en la última etapa del método propuesto, con la única diferencia de que el valor ponderado es el

recuperado de la categoría.

Por lo tanto, en relación a la adaptación del sistema para trabajar con expresiones polares en lugar de emociones, los tres primeros pasos del método no sufren ningún cambio respecto a la idea original (pre-procesado, identificación de conceptos e identificación de categorías emocionales). El tratamiento de la negación, como ya se ha expuesto, sí sufre cambios, y la emoción antónima no es recuperada. Los pasos para resolver los cuantificadores y las formas modales no se modifican. Por último, en el paso de clasificación automática, cada concepto etiquetado con una categoría emocional es traducido a su correspondiente valor según la tabla 6.4. Además, si el concepto fue etiquetado mediante un hiperónimo, entonces el valor de la polaridad asignado es calculado según la ecuación 5.1. Si el concepto se encuentra dentro del ámbito de una negación, el valor de polaridad calculado simplemente se invierte. Si el concepto se encuentra dentro del ámbito de un cuantificador, el valor de su polaridad es calculado según la ecuación 5.2. De la misma manera, el efecto de las formas modales es resuelto utilizando la ecuación 5.3, excepto que la emoción antónima no es recuperada y el valor es invertido al igual que en la negación. Por último, se suman todos los valores polares calculados para cada concepto. Para la tarea de clasificación de polaridad, si el valor obtenido es positivo, entonces el texto es clasificado como positivo. En caso contrario, si el valor calculado es negativo, el texto es clasificado como negativo. Para la tarea de clasificación de intensidad, el valor del texto global calculado es traducido al intervalo [-100,100], clasificando el texto de acuerdo a los subconjuntos creados en el intervalo en clases de intensidad.

Para ilustrar este proceso, considérese de nuevo la oración *Cleaning starts very early even on Sunday, furnitures are moved, very noisy, not possible to relax and sleep*, cuyos conceptos están etiquetados con las categorías emocionales {*like* (cleaning), *disgust* (noisy) y *like* (relax)}. Siguiendo el esquema presentado, dichas categorías son traducidas a su correspondiente valor polar de acuerdo con la tabla 6.4: {1,0 (cleaning), -1,0 (noisy) y 1,0 (relax)}. Además, como el concepto *relax* se encuentra en el ámbito de una negación, su valor polar es invertido y multiplicado por 0,6 (-0,6). Por otro lado, el concepto *noisy* es intensificado por el cuantificador *very* cuyo porcentaje es 75 %, multiplicando así el valor polar por 1,75 (-1,75). De esta manera, el valor polar final para la oración es: 1,0 (cleaning) + (-1,75) (noisy) + (-0,6)

(relax) = -1,35, por lo que la oración es clasificada como negativa.

6.1.5. Experimentos

Los experimentos realizados pretenden responder a las siguientes preguntas planteadas durante la realización del sistema:

- ¿Cómo se comporta el sistema en la tarea de clasificación de la polaridad tanto a nivel de documento como de oración?
- ¿Cómo se comporta el sistema en la tarea de clasificación de la intensidad emocional tanto a nivel de documento como de oración?
- ¿Cuál es el efecto del tratamiento de la negación, los cuantificadores y los modales en el rendimiento del sistema?. ¿Existen sinergias entre los modificadores emocionales?
- ¿Presenta la ambigüedad léxica un problema para el rendimiento del sistema?. En caso afirmativo, ¿puede ser resuelto mediante técnicas de desambiguación semántica?

En las siguientes secciones, se presentan los experimentos realizados para dar respuesta a cada una de estas preguntas, así como los resultados alcanzados.

6.2. Clasificación de la polaridad

Este primer grupo de experimentos trata de evaluar el método propuesto en la tarea de clasificación de textos en base a su polaridad, tanto a nivel de oración como a nivel de documento. Para ello, dado un conjunto de textos de entrada, el sistema los clasifica como positivos o negativos en función de la carga emotiva de cada uno.

6.2.1. Clasificación de la polaridad a nivel de oración

El primer experimento realizado trata de evaluar el método en la tarea de clasificación de la polaridad (i.e., *positivo* vs. *negativo*) a nivel de oración. Para este cometido, se han utilizado dos colecciones de evaluación: la colección de titulares de noticias SemEval 2007 y el HotelReview corpus. Para ello, cada oración de la colección de titulares ha sido categorizada en una de

las clases -100/100 (-100 = [-100, 0), 100 = [0,100]), donde -100 representa la clase negativa y 100 la positiva. Respecto a la colección HotelReview corpus, todas las oraciones dentro de la categoría "like" han sido consideradas como positivas, mientras que todas las oraciones dentro de la categoría "dislike" han sido consideradas como negativas.

La tabla 6.5 muestra los resultados promedio de exactitud (*Acc.*) y medida-F (*F-1*) de estos experimentos. Los mejores resultados de exactitud se obtienen con el clasificador LibSVM para la colección de evaluación SemEval 2007 (75,1 %), y con el clasificador Logistic para el HotelReview corpus (85,0 %). Sin embargo, es importante destacar que el comportamientos de los tres clasificadores es similar en ambas colecciones de evaluación.

Método	**SemEval**		**HotelReview**	
	Acc.	F-1	Acc.	F-1
LibSVM	**75,1**	**75,0**	84,8	84,7
Logistic	74,3	74,3	**85,0**	**84,9**
FT	74,3	74,3	84,9	**84,9**
Expresiones polares	73,2	74,7	74,7	78,0
Unigramas	52,2	35,9	77,6	77,6

Tabla 6.5: Resultados promedio para diferentes clasificadores en la tarea de clasificación de la polaridad a nivel de oración

Comparando los resultados obtenidos por el método con las dos líneas bases, cabe destacar que al evaluar la aproximación basada en unigramas sobre la colección SemEval 2007, su exactitud decrece bruscamente (52,2 % de exactitud), en comparación con el resto de aproximaciones y con los resultados obtenidos en la colección de evaluación HotelReview. El motivo de estos resultados es que el algoritmo clasifica todas las instancias como negativas, motivo por el cual además la medida-F disminuye aun más. Por el contrario, el rendimiento de esta misma aproximación es mucho mejor para la misma tarea en la colección de evaluación HotelReview (77,6 % de exactitud y medida-F). Con respecto a los resultados obtenidos por la aproximación basada en expresiones polares, las diferencias en exactitud para la colección de evaluación SemEval 2007 con respecto al mejor clasificador de nuestro método no son tan marcadas (73,2 % de exactitud), siendo la medida-F incluso mejor que la obtenida por los clasificadores Logistic y FT. Por el contrario, las diferencias con respecto a nuestro método en los resultados obtenidos en la colección de evaluación HotelReview aumentan considerablemente (74,7 % de exactitud), siendo además similares a los obtenidos

por la aproximación basada en unigramas.

6.2.2. Clasificación de la polaridad a nivel de documento

El segundo experimento persigue evaluar el método en la tarea de clasificación de la polaridad a nivel de documento. Para ello, se han utilizado dos colecciones de evaluación: la colección HotelReview, que agrupa un conjunto de opiniones sobre hoteles, y el SFU corpus, que recoge diversas opiniones sobre productos y servicios de distintos tipos. En el caso de la colección de hoteles, los textos dentro de la categoría "like" de cada opinión han sido considerados como positivos, mientras que los textos de cada opinión dentro de la categoría "dislike" han sido considerados como negativos. En cuanto al SFU corpus se refiere, recordamos que los 400 documentos ya disponen de una etiqueta indicando si son positivos o negativos.

Los resultados obtenidos para este experimento se muestran en la tabla 6.6. Como se puede observar, el mejor clasificador es Logistic, obteniendo un 93,6 % de exactitud en la colección HotelReview, y un 73,2 % de exactitud en el SFU corpus. Al igual que a nivel de oración, los resultados de los tres clasificadores en la tarea de clasificación de la polaridad a nivel de documento para el HotelReview corpus son similares. Por el contrario, los resultados obtenidos por LibSVM en el corpus SFU difieren ligeramente de los obtenidos por Logistic y FT, siendo estos últimos comparables entre sí.

Método	**HotelReview**		**SFU**	
	Acc.	F-1	Acc.	F-1
LibSVM	93,1	93,1	62,5	60,7
Logistic	**93,6**	**93,6**	**73,2**	**73,3**
FT	93,3	93,3	71,2	71,2
Expresiones polares	88,4	89,2	62,5	68,3
Unigramas	85,6	85,5	66,7	66,7

Tabla 6.6: Resultados promedio para diferentes clasificadores en la tarea de clasificación de la polaridad a nivel de documento

En relación a la comparación de los resultados obtenidos por nuestro método con las dos líneas bases, en este caso las diferencias con respecto a la aproximación basada en unigramas decrecen en comparación con las obtenidas a nivel de oración en ambas colecciones de evaluación (ver tabla 6.5). Por otro lado, las diferencias del método respecto a la aproximación basada en expresiones polares para el corpus de evaluación HotelReview no

son tan marcadas ($\approx 5\,\%$), como en la clasificación a nivel de oración ($\approx 10\,\%$), aunque sí son importantes en la colección SFU ($\approx 11\,\%$).

6.3. Clasificación de la intensidad

El segundo grupo de experimentos pretende responder a la pregunta de cómo se comporta el sistema en la tarea de clasificación de la intensidad emocional, tanto a nivel de documento como a nivel de oración. Para ello, dado un texto de entrada, el método lo clasificará dependiendo de su carga emotiva en distintas clases de intensidad emocional.

6.3.1. Clasificación de la intensidad a nivel de oración

Además de evaluar el sistema para la detección de la polaridad, queremos determinar su rendimiento a la hora de clasificar oraciones en base a su intensidad emocional. Para ello, hemos realizado dos distribuciones de intensidad en la colección de evaluación SemEval 2007. La primera distribución consta de 3 clases de intensidad (negativo [-100, -50), neutral [-50, 50) y positivo [50, 100]), a las que han sido asignadas cada una de las oraciones de la colección. Esta distribución coincide con la utilizada en la competición SemEval del 2007. La segunda distribución consta de 4 clases de intensidad: fuertemente negativo [-100, -50), negativo [-50, 0), positivo [0, 50) y fuertemente positivo [50, 100].

Los resultados obtenidos por el sistema para estos experimentos se muestran en la tabla 6.7. Como se puede apreciar, el rendimiento del sistema decrece considerablemente al aumentar el número de clases, ya que la tarea se vuelve más compleja. Es importante destacar que, al igual que en la clasificación de polaridad, los tres clasificadores se comportan de manera similar en ambas distribuciones, siendo la diferencia entre ellos menor del $2\,\%$.

En relación a los resultados obtenidos por las dos líneas base, sorprendentemente el rendimiento de la aproximación basada en unigramas para la tarea de clasificación en 3 clases mejora con respecto a la clasificación en 2 clases, aunque continúa siendo peor que el obtenido por los tres clasificadores de nuestro método. Comparando los resultados de nuestro método para 4 clases con respecto a la aproximación de unigramas, comprobamos que se obtienen resultados muy prometedores y significativamente mejores que los alcanzados por la línea base. En esta última, además, los resulta-

Método	**2 clases**		**3 clases**		**4 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
LibSVM	**75,1**	**75,0**	63,2	55,7	45,8	41,1
Logistic	74,3	74,3	63,7	59,3	**46,2**	**43,6**
FT	74,3	74,3	**64,1**	**59,5**	45,0	42,5
Expresiones polares	73,2	74,7	53,7	50,4	32,1	38,2
Unigramas	52,2	35,9	59,7	44,7	32,1	15,6

Tabla 6.7: Resultados promedios para diferentes clasificadores en la tarea de clasificación de la intensidad a nivel de oración (colección de evaluación de la competición SemEval 2007)

dos decrecen considerablemente con respecto a la tarea de clasificación de 3 clases ($\approx 30\,\%$).

El rendimiento de la aproximación basada en expresiones polares es bastante similar al del método propuesto en esta tesis para la tarea de clasificación de la polaridad (2 clases), como ya se ha comentado. Sin embargo, su rendimiento decrece bruscamente, incluso más que en la aproximación basada en unigramas, en el caso de la tarea de clasificación en 3 clases ($\approx$ 20 % con respecto a 2 clases). Los resultados obtenidos para la tarea de clasificación en 4 clases son similares a los obtenidos por la línea base de unigramas, cuyas diferencias con respecto a los resultados obtenidos por el mejor de los clasificadores de nuestro método son considerables (14,1 % de exactitud). Este comportamiento sugiere que la línea base de expresiones polares captura adecuadamente la polaridad del texto, pero no es suficiente si se pretende predecir su intensidad emocional.

6.3.2. Clasificación de la intensidad a nivel de documento

En esta ocasión, los experimentos están dirigidos a evaluar el sistema en la tarea de clasificación de la intensidad a nivel de documento, utilizando para ello la colección de evaluación HotelReview. Como se mostró en la sección 6.1.2, cada documento de esta colección está etiquetado con una categoría dentro de una escala de 5 clases (i.e., *Excellent*, *Good*, *Fair*, *Poor*, *Very Poor*), y una categoría dentro de una escala de 3 clases (i.e., *Good*, *Fair*, *Poor*).

La tabla 6.8 resume los resultados promedios obtenidos por los tres clasificadores del método y las dos líneas base. Al igual que ocurriera en la evaluación de la tarea a nivel de oración, el rendimiento del sistema decrece con el número de clases, puesto que la tarea aumenta su complejidad con cada

nueva clase introducida. Sin embargo, los resultados obtenidos para 5 clases son incluso mejores que los obtenidos para 4 clases a nivel de oración.

Método	**2 clases**		**3 clases**		**5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
LibSVM	93,1	93,1	68,4	68,7	45,9	44,7
Logistic	**93,6**	**93,6**	**70,8**	**70,8**	**50,0**	**48,3**
FT	93,3	93,3	68,6	68,5	46,2	45,0
Expresiones polares	88,4	89,2	47,2	55,7	29,8	35,6
Unigramas	85,6	85,5	54,4	46,5	30,9	25,4

Tabla 6.8: Resultados promedio para diferentes clasificadores en la tarea de clasificación de la intensidad a nivel de documentos (colección de evaluación HotelReview)

No obstante, el deterioro que se produce al incluir cada nueva clase es mayor en este experimento ($\approx$ 20 % al pasar de 2 a 3 clases y $\approx$ 40 % al pasar de 2 a 5 clases) que al tratar con oraciones (tabla 6.7). Esta diferencia se debe al mayor porcentaje de acierto del método en la tarea de 2 clases a nivel de documento, en detrimento de la misma tarea a nivel de oración. De hecho, los resultados obtenidos por el método para 3 clases, y para 4 y 5 clases respectivamente, son bastante similares tanto a nivel de oración como a nivel de documento.

Los resultados obtenidos por la aproximación basada en unigramas muestran una disminución mucho mayor con respecto a la tarea de 2 clases ($\approx$ 30 % para 3 clases y $\approx$ 50 % para 5 clases) que la experimentada por cualquiera de las versiones de nuestro método, e incluso por la obtenida por la misma línea base a nivel de oración (ver tabla 6.7).

Respecto a la línea base basada en expresiones polares, el rendimiento de esta aproximación es bastante similar al del método propuesto para la tarea de dos clases, pero decrece bruscamente cuando se aumenta el número de clases ($\approx$ 40 % para 3 clases y $\approx$ 60 % para 5 clases), incluso más que en la línea base basada en unigramas. Estos resultados confirman nuestra hipótesis de que las aproximaciones basadas en expresiones polares no son capaces de capturar adecuadamente la intensidad emocional de un texto.

6.4. El efecto de los modificadores emocionales

El tercer grupo de experimentos persigue dos objetivos principales. Por un lado, se pretende evaluar la importancia de procesar los modificadores emo-

cionales estudiados (i.e., negación, cuantificadores y modales), en las tareas de clasificación de polaridad e intensidad, ambas a nivel de oración y documento; y por otro, comparar nuestro modelo basado en emociones con la aproximación basada en expresiones polares, descrita en la sección 6.1.4, y que es capaz de abordar el problema de los modificadores emocionales en el análisis sentimental. Para este grupo de experimentos solo se muestran los resultados obtenidos por el clasificador con mejor rendimiento, *Logistic*.

Para evaluar si el tratamiento de los modificadores emocionales mejora el rendimiento en la tarea de clasificación de la polaridad e intensidad, investigamos el comportamiento del sistema con y sin resolver dichos modificadores. Además, queremos analizar la influencia de los modificadores tanto a nivel de oración (i.e., cuando se clasifican oraciones de acuerdo a su polaridad e intensidad emocional), como a nivel de documento (i.e., cuando se clasifican documentos de acuerdo a los mismo criterios). Los experimentos se realizan sobre las colecciones de evaluación SemEval 2007 para las tareas de clasificación de la polaridad e intensidad a nivel de oración, y HotelReview corpus para las mismas tareas a nivel de documento.

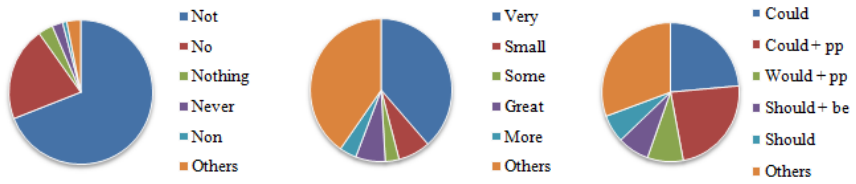
La tabla 6.9 muestra el porcentaje de negaciones, cuantificadores y modales existentes en las colecciones de evaluación SemEval 2007 y HotelReview por oración y documento respectivamente. Atendiendo a la tabla, podemos observar que la presencia de modificadores emocionales es significativamente mayor en la colección de evaluación HotelReview que en la colección SemEval 2007. Una posible explicación para este suceso se encuentra en el tipo de documentos de cada colección, puesto que los modificadores suelen ser mucho más frecuentes en las opiniones que en los titulares de noticias. Dado el bajo porcentaje de oraciones que contienen un modificador emocional en la colección de SemEval 2007 (alrededor del 4 %), se espera que el efecto del tratamiento de los modificadores emocionales en dicha colección sea insignificante. Sin embargo, sí se espera que el efecto de procesar los modificadores emocionales en la colección HotelReview se traduzca en mejoras considerables en el rendimiento del método. Como se puede ver en la tabla 6.9, los cuantificadores son los modificadores más comunes, seguidos de la negación, y, finalmente, las formas modales.

Por otro lado, los gráficos de las figuras 6.1(a), 6.1(b) y 6.1(c) muestran los 5 términos más frecuentes de negación, cuantificadores y modales, respectivamente, así como su frecuencia relativa. Como era de esperar, el

		Negaciones	Cuantificadores	Modales
SemEval	Por oración	3,64 %	4,23 %	3,24 %
HotelReview	Por oración	20,85 %	33,36 %	6,9 %
	Por documento	36,09 %	60,32 %	16,15 %

Tabla 6.9: Distribución de la negación, los cuantificadores y los modales en oraciones y documentos respectivamente, para las colecciones de evaluación SemEval 2007 y HotelReview

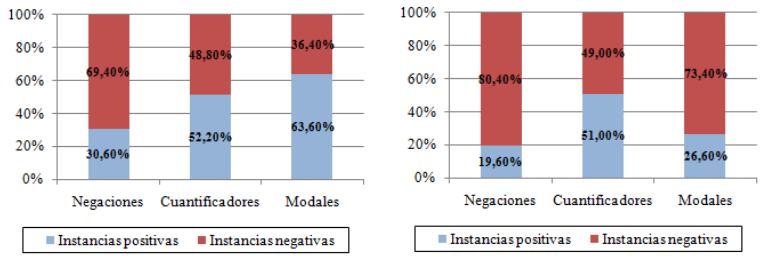
término de negación en inglés más utilizado es *not* (69,1 %), seguido por *no* (21,1 %), consiguiendo de esta forma ambos términos una representación del 90 % de la negaciones de ambos córpora. El cuantificador más utilizado es *very*, que representa cerca del 40 % de las ocurrencias de todos los cuantificadores. Referente a las formas modales, el verbo inglés *could* es el más frecuente en ambos córpora, ya sea solo o acompañado del presente perfecto, representando aproximadamente el 47 % de todas las formas modales.



(a) Negaciones (b) Cuantificadores (c) Modales

Figura 6.1: Cinco términos más frecuentes de negación, cuantificación y modalidad, respectivamente, y su frecuencia relativa en las colecciones de evaluación

Finalmente, las figuras 6.2(a) y 6.2(b) muestran la distribución de cada modificador en clases de polaridad (instancias positivas y negativas) para cada colección de evaluación. Llama la atención el mayor porcentaje de uso de formas negadas en las instancias negativas, suceso que ocurre en ambas colecciones de evaluación, lo que nos lleva a la conclusión de que es más común negar emociones y opiniones positivas que negativas. Los cuantificadores, por el contrario, están distribuidos equitativamente en ambas clases de polaridad, mientras que las formas modales dependen de la colección evaluada.



(a) Colección SemEval 2007 (b) Colección HotelReview

Figura 6.2: Distribución de los modificadores en clases de polaridad en las colecciones de evaluación

6.4.1. Efecto de la negación

El primer grupo de experimentos pretende cuantificar el impacto de la negación y la adecuación del modelo emocional propuesto en dos tareas de análisis sentimental, clasificación de la polaridad e intensidad, a nivel de oración. Para ello, evaluamos nuestra aproximación basada en emociones y la aproximación basada en expresiones polares en ambas tareas, con y sin realizar procesamiento de la negación. Para ello, se utiliza la colección de evaluación SemEval 2007 y la misma distribución de clases que en los experimentos anteriores. Los resultados obtenidos se muestran en la tabla 6.10.

Método	**Polaridad**		**3 clases**		**4 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	72,8	72,8	63,0	58,4	45,3	42,9
Emociones + Negación	**73,2**	**73,1**	**63,3**	**58,7**	**46,0**	**43,3**
Expresiones polares + Sin modificadores	72,9	74,4	**54,4**	**51,0**	**32,1**	**38,2**
Expresiones polares + Negación	**73,2**	**74,7**	53,6	50,5	31,2	37,4

Tabla 6.10: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de la negación

Respecto de la tarea de clasificación de la polaridad, se puede ver en la tabla que ambas aproximaciones obtienen unos resultados muy similares, consiguiendo incluso una mejora prácticamente igual cuando realizan tratamiento de la negación ($\approx 0{,}4\,\%$). Por el contrario, cuando el número de clases aumenta la aproximación propuesta basada en emociones obtiene resultados mucho mejores respecto de la línea base de expresiones polares

($\approx$ 15 % de mejora en exactitud para la tarea de 4 clases). El efecto de la negación también varía en ambas aproximaciones en la clasificación de la intensidad: en nuestro método, el tratamiento del efecto de la negación se traduce en una mejora promedio del 0,5 %, mientras que en la aproximación basada en expresiones polares el tratamiento de la negación disminuye los resultados de exactitud ($\approx$ 0,7 % para la tarea de 3 clases y $\approx$ 0,4 % en la tarea de 4 clases).

Estos resultados indican que, como esperábamos, aunque el uso de expresiones polares para representar el texto puede ser suficiente a la hora de abordar la tarea de clasificación de la polaridad, su uso para tareas de grano más fino donde el texto debe ser clasificado en diferentes clases de intensidad se plantea inadecuado. En estos casos, además, el tratamiento de la negación como un simple modificador que invierte la polaridad es insuficiente, por lo que la representación del texto mediante categorías emocionales puede capturar mejor su significado emocional.

No obstante, es importante mencionar que el efecto de la negación en estos experimentos es casi inapreciable en ambas aproximaciones, debido, como ya se ha dicho, a la baja frecuencia de términos de negación en la colección de evaluación SemEval 2007.

El segundo grupo de experimentos en relación al tratamiento de la negación está dirigido a evaluar su efecto en las dos mismas tareas que los experimentos anteriores, esta vez a nivel de documento. Para ello, en estos experimentos se ha utilizado la colección de evaluación HotelReview, con la misma distribución que en los experimentos anteriores. Los resultados se muestran en la tabla 6.11.

Método	**Polaridad**		**3 clases**		**5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	88,5	88,6	63,6	63,6	43,4	41,1
Emociones + Negación	**92,0**	**92,1**	**68,1**	**66,6**	**46,4**	**44,7**
Expresiones polares + Sin modificadores	78,2	80,5	42,8	51,6	27,3	32,0
Expresiones polares + Negación	**85,4**	**86,5**	**45,5**	**52,5**	**29,3**	**34,7**

Tabla 6.11: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de la negación

A diferencia de los experimentos realizados a nivel de oración, en este caso los resultados del método basado en emociones son sustancialmente mejores en exactitud que los de la aproximación basada en expresiones po-

lares para la tarea de clasificación de la polaridad. Por el contrario, y sorprendentemente, la mejora obtenida al procesar las formas de negación es mejor en esta última aproximación que en la basada en emociones (7,2 % de exactitud *vs.* 3,5 %). Sin embargo, al aumentar el número de clases el comportamiento de ambos sistemas es similar al mostrado en la tabla 6.10 para la tarea de clasificación de intensidad a nivel de oración: por un lado, el modelo basado en emociones obtiene mejores resultados que la línea base de expresiones polares ($>$ 20 % de mejora en exactitud para la tarea de 3 clases y $>$ 15 % para la tarea de 5 clases); por otro lado, la exactitud mejora cuando se evalúa el sistema de emociones con tratamiento de la negación. Finalmente, a diferencia de los resultados obtenidos a nivel de oración en la tarea de clasificación de la intensidad, los resultados de la aproximación basada en expresiones polares mejoran al tratar el efecto de la negación, aunque en un porcentaje muy bajo y en menor medida que los del método propuesto.

Por último, cabe resaltar que, comparando las tablas 6.10 y 6.11, se confirma nuestra hipótesis de que el efecto de la negación es mucho más acusado en la colección HotelReview que en la colección SemEval 2007. Este hecho es debido a que más del 36 % de los documentos de la colección HotelReview presentan formas negadas, mientras que en la colección SemEval 2007 solo un 3,64 % de las oraciones contienen negaciones.

6.4.2. Efecto de los cuantificadores

Al igual que en la negación, este primer grupo de experimentos está dirigido a determinar si el tratamiento de los cuantificadores junto con el modelo emocional propuesto mejoran los resultados en la tarea de clasificación de la polaridad e intensidad a nivel de oración. De la misma manera, la colección de evaluación SemEval 2007 es la utilizada para llevar a cabo los experimentos.

Los resultados obtenidos se muestran en la tabla 6.12. Como se puede observar, la mejora conseguida con el procesamiento de los cuantificadores en ambas aproximaciones es prácticamente despreciable (0,2 % de media para la línea base de expresiones polares y 0,13 % de media para nuestra aproximación basada en emociones). Este resultado era esperado, ya que la presencia de cuantificadores en la colección SemEval 2007 es muy escasa. Sin embargo, aunque el porcentaje de oraciones con al menos un cuantificador es

mayor que el porcentaje de oraciones con negaciones (4,23 % *vs.* 3,64 %), el efecto de la negación es mucho más significativo que el de los cuantificadores. Tampoco este hallazgo era inesperado, si se tiene en cuenta que el efecto de la negación es invertir la emoción de los conceptos afectados, mientras que los cuantificadores solo ponderan la intensidad de la misma. Es interesante mencionar que Kennedy y Inkpen (2006) obtienen unos resultados similares a los expuestos en esta tesis, concluyendo de la misma manera que la consideración del efecto de la negación en el análisis sentimental es mucho más importante que el de los cuantificadores que disminuyen la intensidad.

Método	**Polaridad**		**3 clases**		**4 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	72,8	72,8	63,0	58,4	**45,3**	**42,9**
Emociones + Cuantificadores	**73,1**	**73,1**	**63,2**	**58,8**	**45,3**	**42,9**
Expresiones polares + Sin modificadores	**72,9**	**74,4**	54,4	51,0	32,1	38,2
Expresiones polares + Cuantificadores	**72,9**	**74,4**	**54,5**	**51,2**	**32,6**	**38,8**

Tabla 6.12: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de los cuantificadores

El segundo grupo de experimentos está dirigido a valorar el impacto de los cuantificadores en la clasificación de la polaridad e intensidad, esta vez a nivel de documento, utilizando como colección de evaluación el HotelReview corpus.

Como se puede ver en la tabla 6.13, los resultados de estos experimentos muestran una mejora consistente en ambas aproximaciones y para todas las distribuciones de polaridad e intensidad. Al igual que en la negación, la mejora es mayor cuando se clasifican documentos en lugar de oraciones ($\approx$ 1,6 % de mejora en promedio para ambas aproximaciones, que es más de ocho veces la mejora en promedio obtenida en las tareas de clasificación a nivel de oración). De nuevo, estos resultados eran esperados, debido al mayor porcentaje de documentos afectados por los cuantificadores respecto del porcentaje de oraciones, así como a la diferencias entre ambas colecciones de evaluación.

Al igual que en los experimentos a nivel de oración, la mejora observada al procesar los cuantificadores en ambas aproximaciones es considerablemente menor que la obtenida a nivel de documento para la negación ($\approx$ 2 % de mejora promedio en exactitud en ambas aproximaciones para las distintas tareas). Así mismo, las mejoras de la aproximación basada en emociones

Método	**Polaridad**		**3 clases**		**5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	88,5	88,6	63,6	63,6	43,4	41,1
Emociones + Cuantificadores	**89,2**	**89,3**	**65,8**	**63,7**	**45,3**	**43,4**
Expresiones polares + Sin modificadores	78,2	80,5	42,8	51,6	27,3	32,0
Expresiones polares + Cuantificadores	**79,5**	**81,6**	**44,5**	**53,1**	**29,0**	**34,0**

Tabla 6.13: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de los cuantificadores

respecto de la línea base de expresiones polares siguen la misma tónica que en los experimentos anteriores, siendo menor en la tarea de clasificación de la polaridad ($\approx$ 10 % de exactitud) que en las tareas de clasificación de la intensidad ($\approx$ 20 % de exactitud para 3 clases y $\approx$ 15 % para la de 5 clases).

6.4.3. Efecto de la modalidad

El siguiente grupo de experimentos pretende responder a la pregunta de si capturar y tratar convenientemente el efecto de la modalidad puede mejorar los resultados obtenidos en las tareas de clasificación de polaridad e intensidad emocional a nivel de oración. Al igual que en los casos anteriores, evaluamos ambas aproximaciones con y sin procesamiento de la modalidad.

La tabla 6.14 muestra los resultados obtenidos en estos experimentos. Como se puede ver, los resultados no mejoran al procesar la modalidad en la tarea de clasificación de polaridad para la aproximación basada en expresiones polares. Por otro lado, al aumentar el número de clases la misma aproximación obtiene resultados incluso peores que al no procesar la modalidad (una disminución del 0,3 % para la tarea de 3 clases y del 0,2 % para la tarea de 4 clases). Por el contrario, los resultados en todas las tareas evaluadas mejoran en nuestra aproximación basada en emociones cuando se trata el efecto de la modalidad. Aun así, esta mejora es casi despreciable ($\approx$ 0,16 % de media en exactitud) debido al bajo porcentaje de formas modales que aparecen en las oraciones de la colección de evaluación SemEval 2007.

En segundo lugar, y al igual que con los anteriores modificadores, se ha evaluado el sistema para las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de la modalidad, en las mismas circunstancias que en los experimentos anteriores.

Como muestran los resultados de la tabla 6.15, el procesamiento de la

Método	**Polaridad**		**3 clases**		**4 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	72,8	72,8	63,0	58,4	45,3	42,9
Emociones + Modales	**73,1**	**73,0**	**63,1**	**58,5**	**45,4**	**43,4**
Expresiones polares + Sin modificadores	**72,9**	**74,4**	**54,4**	**51,0**	**32,1**	**38,2**
Expresiones polares + Modales	**72,9**	**74,4**	54,1	50,6	31,9	37,9

Tabla 6.14: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de la modalidad

modalidad mejora los resultados en prácticamente todas las tareas y ambas aproximaciones. En concreto, los resultados mejoran para todas las tareas en la aproximación basada en emociones en aproximadamente un 0,7 % de exactitud de media. En el caso de la línea base de expresiones polares, los resultados mejoran en la tarea de polaridad y en la tarea de intensidad para 3 clases en 1,5 % de exactitud y 0,1 % de exactitud, respectivamente. Sin embargo, los resultados empeoran ligeramente en la tarea de clasificación en 5 clases. Por último, al igual que ocurriera con los modificadores anteriores, la mejora de los resultados al procesar la modalidad es mayor al clasificar documentos que oraciones, debido, como ya se ha dicho, al mayor porcentaje de documentos con presencia de formas modales (ver tabla 6.9)

Método	**Polaridad**		**3 clases**		**5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	88,5	88,6	63,6	63,6	43,4	41,1
Emociones + Modales	**89,8**	**89,8**	**64,1**	**63,9**	**43,7**	**41,3**
Expresiones polares + Sin modificadores	78,2	80,5	42,8	51,6	**27,3**	**32,0**
Expresiones polares + Modales	**79,7**	**81,7**	**42,9**	**51,8**	26,8	31,5

Tabla 6.15: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de la modalidad

6.4.4. Efecto de todos los modificadores emocionales

El último grupo de experimentos aborda la pregunta ¿cómo se traduce en los resultados el tratar todos los modificadores anteriores simultáneamente?. Es decir, queremos comprobar si existen sinergias entre los diferentes modificadores que hagan posible mejorar los resultados cuando se procesan todos juntos respecto a cuando se procesan por separado.

Siguiendo la misma metodología que en los experimentos anteriores, se

evalúa el sistema en las tarea de clasificación de la polaridad e intensidad, tanto a nivel de oración como a nivel de documento, con y sin procesar los modificadores, en la mismas condiciones y con las mismas colecciones de evaluación.

Los resultados obtenidos al procesar todos los modificadores en los experimentos a nivel de oración se muestran en la tabla 6.16. Lo más interesante que revela esta tabla es que, en lo que concierne a la aproximación basada en expresiones polares, el uso de los tres modificadores juntos no redunda en mejoras de exactitud en comparación con no procesar ningún modificador o procesar solamente algunos de ellos; y esto ocurre para todas las tareas que han sido evaluadas. Por el contrario, nuestro método basado en emociones obtiene mejores resultados para todas las tareas propuestas cuando se procesa el efecto de todos modificadores en comparación con no procesar ninguno o con procesar únicamente algunos de ellos. Sin embargo, esta mejora es menor de lo esperado, debido al bajo porcentaje de oraciones afectadas por algún modificador en la colección de evaluación SemEval 2007.

Método	**Polaridad**		**3 clases**		**4 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	72,8	72,8	63,0	58,4	45,3	42,9
Emociones + Modificadores	**74,3**	**74,3**	**63,7**	**59,3**	**46,2**	**43,6**
Expresiones polares + Sin modificadores	72,9	74,4	**54,4**	**51,0**	**32,1**	**38,2**
Expresiones polares + Modificadores	**73,2**	**74,7**	53,7	50,4	**32,1**	**38,2**

Tabla 6.16: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de oración, con y sin procesamiento de los modificadores emocionales

Respecto a las tareas de clasificación de la polaridad e intensidad a nivel de documento, la tabla 6.17 muestra que el procesamiento de todos los modificadores mejora la línea base *sin modificadores* para ambas aproximaciones en términos de exactitud y medida-F. En concreto, el tratamiento de los tres modificadores a la vez se traduce en una mejora del 10,2 % para la tarea de clasificación de la polaridad en la aproximación basada en expresiones polares, reduciéndose drásticamente esta mejora al aumentar el número de clases (4,4 % para la tarea de 3 clases y 2,5 % para la tarea de 4 clases). Por el contrario, la mejora obtenida con el tratamiento del efecto de los tres modificadores en el método basado en emociones es similar para las tres tareas de clasificación ($\approx$ 6,25 % de exactitud promedio). La consistencia de estas mejoras en ambas aproximaciones hace resaltar la importancia de

tratar adecuadamente el efecto de estos modificadores emocionales en las tareas de clasificación de la polaridad e intensidad.

Método	**Polaridad**		**3 clases**		**5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emociones + Sin modificadores	88,5	88,6	63,6	63,6	43,4	41,1
Emociones + Modificadores	**93,6**	**93,6**	**70,8**	**70,8**	**50,0**	**48,3**
Expresiones polares + Sin modificadores	78,2	80,5	42,8	51,6	27,3	32,0
Expresiones polares + Modificadores	**88,4**	**89,2**	**47,2**	**55,7**	**29,8**	**35,6**

Tabla 6.17: Resultados promedio de exactitud y medida-F en las tareas de clasificación de la polaridad e intensidad a nivel de documento, con y sin procesamiento de los modificadores emocionales

Por último, la observación más importante es, especialmente en nuestra aproximación basada en emociones, que la mejora obtenida por el tratamiento del efecto combinado de todos los modificadores supera en algunos casos la suma de las mejoras obtenidas al tratarlos por separado. De esta manera, estos resultados vienen a confirmar nuestra hipótesis de la existencia de sinergias entre los distintos modificadores emocionales y de que su tratamiento conjunto es necesario para mejorar los sistemas de clasificación de la polaridad e intensidad.

6.5. El efecto de la ambigüedad

Por último, nos interesa evaluar el efecto de la ambigüedad léxica en las tareas de clasificación de la polaridad y la intensidad, ambas a nivel de oración y a nivel de documento, y comprobar si el hecho de utilizar algoritmos de desambiguación semántica puede mejorar el comportamiento de los distintos clasificadores. Con este propósito, repetimos los experimentos realizados para dichas tareas con las colecciones de evaluación SemEval 2007 y HotelReview, utilizando en este caso diferentes estrategias de desambiguación semántica. Para ello, ejecutamos el método propuesto basado en emociones utilizando el algoritmo de desambiguación semántica Lesk (1986) para obtener el significado de cada término en WordNet de acuerdo al contexto en que se utiliza, tal y como se explica en el capítulo 5. En las mismas condiciones, repetimos los experimentos pero asignando a cada término su primer significado en WordNet. Por último, repetimos los experimentos seleccionando aleatoriamente el significado de cada término de entre las distintas alternativas disponibles en WordNet.

Los resultados de estos experimentos se pueden ver en las tablas 6.18 y 6.19, a nivel de oración y a nivel de documento respectivamente. Se puede observar que, tanto a nivel de oración como a nivel de documento, el uso del algoritmo Lesk para desambiguar mejora los resultados para todos los clasificadores respecto de las otras dos estrategias. Sin embargo, las diferencias con respecto a la estrategia de elegir el primer significado posible no son tan marcadas como cabría esperar, obteniendo la primera estrategia una mejora promedio menor del 2 % de exactitud sobre la segunda. Como era de esperar, las diferencias con respecto a la estrategia aleatoria son mucho mayores.

WSD	**Método**	**2 clases**		**3 clases**		**4 clases**	
		Acc.	F-1	Acc.	F-1	Acc.	F-1
Lesk	*LibSVM*	**75,1**	**75,0**	63,2	55,7	45,8	41,1
	Logistic	74,3	74,3	63,7	59,3	**46,2**	**43,6**
	FT	74,3	74,3	**64,1**	**59,5**	45,0	42,5
1er Significado	*LibSVM*	71,8	71,7	62,9	54,9	44,6	40,1
	Logistic	72,8	72,7	**63,1**	**58,4**	**45,4**	**43,6**
	FT	**73,0**	**73,0**	63,0	58,3	43,6	41,7
Aleatorio	*LibSVM*	63,2	61,9	60,1	49,6	**42,9**	39,5
	Logistic	**63,5**	62,7	**61,7**	**55,0**	42,2	**40,0**
	FT	63,3	**62,8**	61,2	54,4	40,8	38,1

Tabla 6.18: Resultados promedio para diferentes clasificadores en las tareas de clasificación de la polaridad e intensidad a nivel de oración utilizando diferentes estrategias de desambiguación semántica

WSD	**Método**	**2 clases**		**3 clases**		**5 clases**	
		Acc.	F-1	Acc.	F-1	Acc.	F-1
Lesk	*LibSVM*	93,1	93,1	68,4	68,7	45,9	44,7
	Logistic	**93,6**	**93,6**	**70,8**	**70,8**	**50,0**	**48,3**
	FT	93,3	93,3	68,6	68,5	46,2	45,0
1er significado	*LibSVM*	91,7	91,8	67,0	66,2	45,3	44,2
	Logistic	**91,9**	**92,0**	**67,9**	66,1	**46,8**	**44,7**
	FT	91,3	91,4	67,4	**67,4**	44,6	43,7
Aleatorio	*LibSVM*	80,8	80,8	56,4	51,8	34,9	33,9
	Logistic	81,8	81,7	**58,9**	55,3	**37,7**	35,6
	FT	**82,5**	**82,3**	56,1	**55,6**	35,9	**35,7**

Tabla 6.19: Resultados promedio para diferentes clasificadores en las tareas de clasificación de la polaridad e intensidad a nivel de documento utilizando diferentes estrategias de desambiguación semántica

Por último, se observa que la mejora que se produce en la tarea de clasificación de la intensidad emocional al realizar desambiguación semántica

es mayor cuando se trabaja con documentos que cuando se trabaja con oraciones.

Capítulo 7

Discusión

El objetivo del capítulo que nos ocupa es analizar los resultados obtenidos en los diferentes experimentos realizados hasta este punto, y que han sido presentados en el capítulo anterior. Con ello se pretende comprender el comportamiento del sistema propuesto, así como descubrir sus fortalezas y debilidades e identificar posibles alternativas de solución. El capítulo se organiza como sigue. En primer lugar, se estudian los resultados de la tarea de clasificación de la polaridad, tanto a nivel de oración como a nivel de documento. En segundo lugar, se discuten estos mismos resultados para la tarea de predicción de la intensidad emocional. La tercera sección investiga el efecto de los denominados modificadores emocionales en el rendimiento obtenido en las tareas de clasificación de la polaridad e intensidad. Por último, se analiza el impacto de la ambigüedad en las dos tareas anteriores.

7.1. Clasificación de la polaridad

En primer lugar, analizamos los resultados obtenidos en relación a la tarea de clasificación de la polaridad a nivel de oración. Los resultados presentados en la tabla 6.5 muestran que nuestro método se comporta mejor que otras aproximaciones del estado del arte, con independencia del dominio de los textos y del conjunto de evaluación utilizado. Estos buenos resultados parecen confirmar nuestra hipótesis de que un modelo basado en emociones, junto con un análisis lingüístico en profundidad, resulta más apropiado para abordar la tarea de clasificación de la polaridad que las aproximaciones superficiales basadas en frecuencias de términos, e incluso que otras aproximaciones más complejas basadas en expresiones polares.

En concreto, la diferencia entre el mejor de los clasificadores del método propuesto y la aproximación basada en unigramas es de 22,9 % de exactitud para la colección de evaluación SemEval 2007, y de 7,4 % de exactitud para la colección HotelReview. Esta notable diferencia es debida al pobre rendimiento de la aproximación basada en unigramas cuando se evalúa sobre la colección SemEval 2007; y ello, a su vez, se explica por el vocabulario utilizado en dicha colección. Por un lado, hay que tener en cuenta que la colección SemEval 2007 comprende un conjunto de titulares de noticias de temas muy dispares, mientras que todas las oraciones de la colección HotelReview expresan opiniones sobre hoteles, por lo que poseen un vocabulario muy homogéneo en comparación con el vocabulario de la colección SemEval 2007. Por otro lado, la aproximación basada en unigramas utiliza frecuencias de términos para representar el texto a clasificar, por lo que parece obvio que capturará mejor la polaridad de las oraciones en una colección cuyo vocabulario se repita a través de las distintas instancias. En conclusión, los resultados indican que, aunque los sistemas basados en frecuencias de términos pueden ser una alternativa sencilla para clasificar textos homogéneos en base a su polaridad, se muestran insuficientes a la hora de clasificar textos con un vocabulario amplio y poco homogéneo. Por el contrario, los buenos resultados obtenidos por nuestro sistema en ambas colecciones de evaluación parecen indicar que la aproximación propuesta captura con mayor precisión el significado emocional de las oraciones, independientemente del dominio o la temática de las mismas.

En cuanto a la aproximación basada en expresiones polares, las diferencias con respecto a nuestro método no son tan señaladas en la colección SemEval 2007 (1,9 % de exactitud), aunque sí considerables respecto a la colección HotelReview (10,3 % de exactitud). Sin embargo, en esta ocasión las discrepancias no son debidas al comportamiento inadecuado de la línea base en la primera colección de evaluación, sino al mejor comportamiento de nuestro sistema en la colección HotelReview. Estos resultados sugieren que aunque la aproximación basada en expresiones polares es una alternativa viable a tener en cuenta a la hora de abordar la tarea de clasificación de la polaridad, la representación del texto mediante emociones es capaz de mejorar sustancialmente los resultados incluso en esta tarea.

Respecto a nuestro método, la tabla 6.5 muestra también diferencias sustanciales entre los resultados obtenidos en ambas colecciones (9,9 % de

exactitud entre los mejores clasificadores). Tras analizar en profundidad las clasificaciones producidas, hemos determinado que una de las causas de esta disparidad es la diferencia en el número de oraciones no etiquetadas con ninguna emoción entre ambas colecciones. Aunque el léxico afectivo SentiSense tiene una cobertura de vocabulario prácticamente igual para ambas colecciones (20,5 % para HotelReview y 22 % para SemEval 2007), el número de oraciones no etiquetadas con ninguna emoción en la colección SemEval 2007 es del 27 %, mientras que en la colección HotelReview es del 16 %. Este fenómeno se debe, principalmente, a que un gran número de las oraciones que componen la colección SemEval 2007 no poseen ningún significado emocional *per se*. Por ejemplo, la oración *Looking beyond the iPhone* no expresa ninguna emoción por si misma, excepto que el lector sea un fan o detractor de Apple. Por el contrario, las oraciones de la colección HotelReview poseen una alta carga emotiva, puesto que provienen de comentarios donde los usuarios exponen lo que les ha gustado y disgustado de su experiencia en los diferentes hoteles.

Otra de las posibles causas de esta variación entre los resultados de nuestro método para ambas colecciones podría ser la diferencia en el número de oraciones que componen ambas colecciones (1000 oraciones en la colección SemEval 2007 *vs.* 4340 oraciones en el HotelReview corpus). Al presentar la segunda colección muchas más instancias, podría ocurrir que el modelo de predicción aprendido sea más exacto.

Para comprobar empíricamente si esto es así, hemos repetido los experimentos sobre la colección HotelReview seleccionando aleatoriamente 1000 oraciones, y entrenando y evaluando el sistema solo sobre este subconjunto. Sin embargo, los resultados son muy similares a los obtenidos sobre el total de las oraciones de la colección (83,5 % de exactitud para *libSVM*, 85,2 % de exactitud para *Logistic* y 84,5 % de exactitud para *FT*). Este hecho parece indicar que el modelo propuesto es capaz de aprender a partir de un número de instancias relativamente pequeño, y que, a partir de dicho número, el beneficio que se consigue es prácticamente despreciable.

Por último, cabe mencionar que un buen porcentaje de los errores cometidos por el método propuesto a la hora de clasificar las oraciones, cuando ha sido evaluado sobre la colección HotelReview, se debe a oraciones mal clasificadas por los propios usuarios, y que son utilizadas en el conjunto de entrenamiento (por ejemplo, la oración *Anyway, everybody else was nice* de una

opinión está incorrectamente localizada en la sección <NegativeOpinion>). Otra fuente importante de errores son las faltas de ortografía cometidas por los usuarios al escribir. Estas faltas hacen que el sistema no pueda traducir correctamente los términos a los conceptos adecuados, y por tanto, no puedan ser etiquetados con ninguna emoción.Por el contrario, este tipo de errores no han sido observados en la colección SemEval 2007, puesto que ha sido desarrollada expresamente para una tarea competitiva, y a partir de titulares de noticias, que raramente contienen errores ortográficos.

Los resultados de la evaluación de los distintos sistemas en la tarea de clasificación de la polaridad a nivel de documento, mostrados en la tabla 6.6, revelan que nuestro sistema se comporta incluso mejor que a nivel de oración para la colección HotelReview. De hecho, únicamente se clasifican mal 6 de cada 100 opiniones, resultado que incluso se podría comparar con el de la clasificación realizada por un humano, si se tiene en cuenta la subjetividad inherente a la tarea. En nuestra opinión, la principal causa de este mejor resultado en la clasificación de documentos en lugar de oraciones del corpus HotelReview se debe a que en las opiniones generalmente se introducen algunas oraciones neutrales que no expresan ningún sentimiento por sí mismas (por lo que no son etiquetadas con ninguna emoción), pero que tienen que ser clasificadas necesariamente como positivas o negativas cuando se evalúa el método a nivel de oración. Por el contrario, el número de documentos no etiquetados con ninguna emoción es casi despreciable (3 %). Otra posible razón de esta disparidad de resultados se puede deber a la longitud de los textos, puesto que la representación de un documento mediante el *VEI* incluirá más información que la representación de una única una oración, lo que supondrá mayor información para los clasificadores, que por ende aprenderán mejor el modelo de predicción.

Por otro lado, los resultados de la tabla 6.6 también muestran que el rendimiento de nuestro método es menor cuando se evalúa sobre la colección SFU. Una de las razones para esta diferencia entre ambas colecciones se encuentra en la cobertura de vocabulario en el léxico afectivo SentiSense. Como ya se ha dicho, SentiSense cubre el 20,5 % del vocabulario existente en la colección de evaluación HotelReview; es decir, el 20,5 % de los conceptos de la colección están etiquetadas con una emoción. Por el contrario, solo el 9,5 % de los conceptos presentes en la colección SFU están presentes en el léxico afectivo. Esta diferencia de cobertura hace que los *VEI* contengan

menos información emocional, hecho que se traduce en un peor aprendizaje por parte de los clasificadores. Otra razón importante de esta divergencia es la diferencia de extensión y contenido entre los documentos de ambas colecciones. Mientras que los documentos de la colección HotelReview constan en media de unos 150 términos, los documentos de la colección SFU presentan un promedio de 1200 términos. Al ser los comentarios del corpus SFU tan extensos, frecuentemente contienen oraciones neutrales, e incluso fragmentos con opiniones contradictorias, empezando, por ejemplo, con opiniones positivas y terminando con opiniones negativas, lo que supone que su clasificación automática sea más compleja. Sin embargo, a pesar de estas limitaciones, el sistema obtiene unos resultados altamente satisfactorios clasificando bien más de 7 de cada 10 documentos.

Es importante también destacar la diferencia entre la colección HotelReview y la colección SFU a la hora de evaluar a nivel de documento el método, y que explica en cierta medida esta diferencia en los resultados. Como ya se ha expuesto, la colección HotelReview está compuesta por un conjunto de opiniones de hoteles, por lo que todos los documentos versan sobre el mismo dominio. Por el contrario, la colección SFU contiene 50 documentos (25 documentos positivos y 25 documentos negativos) de 8 productos diferentes. Los 400 documentos se encuentran ordenados por productos en la colección de evaluación, lo que hace que al realizar una evaluación cruzada de 10 pliegues, se tomen grupos de 40 documentos para evaluar y 360 documentos para el entrenamiento en cada iteración de la evaluación. De esta manera, los documentos en el conjunto de evaluación pertenecen en su mayoría a productos distintos de los utilizados para aprender en cada iteración. Por lo tanto, la evaluación sobre la colección SFU se puede ver como un tipo de evaluación multi-dominio para la tarea de clasificación de la polaridad, donde los algoritmos de aprendizaje máquina aprenden de instancias de un dominio y el modelo aprendido se utiliza para clasificar instancias de otro dominio diferente.

Por otro lado, merece la pena mencionar que nuestro método se comporta adecuadamente con independencia del algoritmo de aprendizaje máquina utilizado, tal y como se puede apreciar en las tablas 6.5 y 6.6, lo que confirma la adecuación de la representación utilizada para modelar el texto a la tarea abordada.

Finalmente, la tabla 6.6 muestra que el rendimiento de nuestro sistema

es considerablemente mejor que el de las líneas base para ambas colecciones de evaluación, obteniendo una mejora promedio del 6,6 % de exactitud para la colección HotelReview, y una mejora promedio del 8,6 % para la colección SFU. Comparando las tablas 6.5 y 6.6, se puede observar que las diferencias entre nuestro sistema y la aproximación basada en unigramas se han reducido considerablemente en la evaluación a nivel de documento (8 % de exactitud para la colección HotelReview y 6,5 % de exactitud para la colección SFU), siendo incluso mejores los resultados que los de la aproximación basada en expresiones polares en la colección SFU. Sin embargo, este comportamiento era esperado, puesto que la línea base de unigramas utiliza las frecuencias de los términos de los textos para aprender el modelo de predicción. De esta manera, a mayor número de términos por instancia mayor probabilidad de acierto; es decir, al tener los documentos mayor información que las oraciones cabe esperar que el rendimiento del sistema sea mejor.

La tabla 6.6 muestra, así mismo, que el rendimiento de la aproximación basada en expresiones polares es solo un poco inferior al obtenido por nuestro sistema en la colección HotelReview (5,2 % de exactitud), aunque considerablemente menor en la colección SFU (10,7 %). Esta diferencia importante entre ambas colecciones de evaluación es motivada, principalmente, por la menor cobertura del léxico afectivo SentiSense, al igual que sucede para el método propuesto. De la misma manera, comparando los resultados obtenidos en las tablas 6.5 y 6.6, se puede extraer la conclusión de que aunque la representación mediante expresiones polares obtiene peor rendimiento que nuestro método, puede ser una opción interesante a la hora de abordar la tarea de la clasificación de la polaridad, ya que obtiene unos resultados interesantes y estables en todas las evaluaciones planteadas.

7.2. Clasificación de la intensidad

El análisis de los resultados en la tarea de clasificación de la intensidad a nivel de oración, mostrados en la tabla 6.7, confirman nuestra hipótesis de que modelar los textos mediante emociones permite capturar de una forma más eficiente el significado afectivo de los mismos. Como era de esperar, el rendimiento del método decrece al aumentar el número de clases de intensidad, puesto que la dificultad de la tarea se incrementa con cada clase. De hecho, esta dificultad es evidente incluso si la tarea la abordan personas,

debido a la subjetividad a la hora de decidir en qué categoría clasificar cada instancia. Además, resulta obvio que, tanto para para los algoritmos automáticos como para las personas, cuanto mayor es el número de clases más difusos se convierten los límites entre ellas. En particular, el rendimiento del sistema disminuye en un 11 % de exactitud cuando se pasa de 2 a 3 clases, y en un 17,9 % de exactitud cuando se pasa de 3 clases a 4 clases.

Sin embargo, pese a esta disminución de rendimiento al incrementar el número de clases, los tres clasificadores del método obtienen mejores resultados para todas las tareas que las dos líneas base, clasificando bien aproximadamente 2 de cada 3 oraciones en la tarea de 3 clases, y aproximadamente 1 de cada 2 en la tarea de 4 clases. Es importante recordar, además, que esta colección de evaluación contiene gran cantidad de oraciones que por sí mismas no poseen ningún significado emocional, pero que forzosamente han de ser clasificadas en una u otra clase, hecho que dificulta aun más la tarea de clasificación de la intensidad.

Respecto a la comparación del método con la aproximación basada en unigramas, nuestro método mejora en un 4,7 % de exactitud para la tarea de 3 clases, lo que no es una diferencia importante. Es interesante resaltar que la línea base de unigramas obtiene mejores resultados para la tarea de 3 clases que para la tarea de 2 clases. La razón es que esta línea base clasifica todas las instancias como neutrales en la tarea de 3 clases, y ello se debe a la distribución no uniforme de las instancias en clases de la colección SemEval 2007 (251 instancias para la clase negativa, 598 para la clase neutral y 151 para la positiva). De esta manera, al clasificar todas las instancias como neutrales la exactitud aumenta, pero la medida-F se ve penalizada. Por el contrario, la diferencia entre exactitud y medida-F en nuestro sistema no es tan marcada, lo que significa que el clasificador aprende de manera similar todas las clases. En cuanto a la tarea de 4 clases, la diferencia de nuestro método respecto a esta línea base es de 14,1 % de exactitud, lo que demuestra que la aproximación presentada en esta tesis es capaz de capturar mucho mejor la intensidad emocional que los métodos tradicionales basados en frecuencias.

Al igual que en nuestro método, al aumentar el número de clases el rendimiento de la aproximación basada en expresiones polares disminuye. Sin embargo, esta disminución es mucho más marcada en esta línea base, alcanzando el 19,5 % en exactitud al pasar de 2 a 3 clases, y el 21,6 % al

pasar de 3 a 4 clases. Este resultado corrobora nuestra hipótesis de que las expresiones polares pueden ser una buena representación para abordar la tarea de detección de la polaridad, pero son insuficientes cuando se pretenden resolver tareas más complejas como la clasificación en distintos grados de intensidad. Por el contrario, nuestra aproximación es capaz de determinar de una manera más eficiente las sutilezas emocionales que permiten clasificar textos en base a su intensidad, obteniendo una mejora promedio en exactitud del 12 % respecto a la aproximación basada en expresiones polares.

Respecto a la evaluación de la clasificación de la intensidad a nivel de documento, los resultados obtenidos son de nuevo mejores que los alcanzados a nivel de oración (ver tablas 6.7 y 6.8). Al igual que en la tarea de clasificación de la polaridad, esto es debido a que los documentos contienen más información que las oraciones, lo que hace que el número de documentos no etiquetados con ninguna emoción sea prácticamente nulo. Se obtiene así una mejora del 6,7 % de exactitud en la tarea de 3 clases respecto al nivel de oración, y una mejora del 3,8 % de exactitud en la tarea de 5 clases a nivel de documento, comparada con la de 4 clases a nivel de oración, obteniendo la primera incluso mejores resultados pese a tener una clase más.

Como era de esperar, la exactitud decrece al aumentar el número de clases en comparación con la tarea de detección de la polaridad. Como ya se ha dicho, aumentar el número de clases hace que los límites entre las clases sean más difusos, estando incluso sujetos a apreciaciones subjetivas. Prueba de ello es que se han encontrado en la colección HotelReview algunas opiniones etiquetadas con diferentes clases, pero que han sido determinadas por diferentes jueces como similares en relación a su intensidad emocional. Por ejemplo, las dos opiniones siguientes son etiquetadas en la colección como *Very Poor* y *Poor* respectivamente, pero son, en opinión de dos jueces diferentes, representativas de un mismo grado de negatividad:

Only the location was a positive aspect. The room was very small, in the basement of the hotel and we could hear people walking around all night long. The shower and the toilet closets were really uncomfortable and small as well.

Very near transport to London central. On arrival the apartment was very smelly of old spicy food and standard of cleanliness was poor. Entrance to apartment (hallway) needs a paint job.

Sin embargo, incluso con esta limitación, el comportamiento del sistema en la tarea de clasificación de la intensidad es sorprendentemente bueno, clasificando correctamente 3 de cada 4 documentos en la tarea de 3 clases y 1 de cada 2 en la tarea de 5 clases.

La comparación de estos resultados con los obtenidos por las aproximaciones basadas en unigramas y en expresiones polares ha revelado que, aunque el rendimiento de ambas aproximaciones es bueno cuando se aborda la tarea de clasificación de la polaridad, su eficacia disminuyen drásticamente en la tarea de clasificación de 3 clases (31,2 % para la línea base *Unigramas* y 41,2 % para la línea base *Expresiones polares*), lo que supone aproximadamente un 20 % menos de exactitud en comparación con nuestro método. Es interesante resaltar que, aunque esta colección de evaluación es, desde cierto punto de vista, más propensa a mejores resultados por los sistemas automáticos debido a su temática y a la estructura de sus documentos, ambas líneas base han obtenido resultados menores que a nivel de oración. Esta disminución se debe a la distribución equitativa de las instancias en clases de la colección HotelReview (349 en *Poor*, 292 en *Fair* y 359 en *Good*), lo que hace que los clasificadores no sobre-aprendan la clase neutral, como ocurriera a nivel de oración, y obtengan mejores resultados al ser ésta la clase predominante. Por el contrario, nuestro método obtiene mejores resultados que a nivel de oración, incluso pese a la distribución equitativa, lo que nos hace pensar que el modelado mediante emociones realmente captura correctamente el significado emocional y es capaz de obtener buenos resultados con independencia del dominio de la colección o su distribución, aprendiendo adecuadamente todas las clases. De la misma manera, el rendimiento de ambas aproximaciones para la tarea de clasificación de 5 clases es considerablemente peor que el obtenido por nuestro método (≈20 % de exactitud), clasificando correctamente menos de 1 de cada 3 documentos.

Atendiendo a los resultados obtenidos en la tarea de clasificación de la polaridad, tanto a nivel de oración como a nivel de documento, y a los obtenidos en la tarea de clasificación de la intensidad, tanto a nivel de oración como a nivel de documento, se puede concluir el buen comportamiento de nuestro método, y corroborar así nuestra hipótesis de que una representación del texto basada en emociones captura mucho mejor los matices afectivos, lo que se traduce en una mayor exactitud en la clasificación automática no solo de la intensidad sino también de la polaridad. Análogamente, y aunque

la aproximación basada en unigramas puede ser una opción de bajo coste viable, sus resultados son considerablemente inferiores en la clasificación de polaridad, y altamente insatisfactorios en la clasificación de intensidad, por lo que su aplicación efectiva es cuestionable. Por otro lado, los resultados de la aproximación basada en expresiones polares, aunque ligeramente menores que los obtenidos por nuestro método, demuestran que el uso de expresiones polares permite capturar eficientemente la polaridad, y por tanto, pueden ser utilizadas para esta tarea. Por el contrario, no resultan apropiadas para abordar la tarea de clasificación de intensidad.

7.3. El efecto de los modificadores emocionales

Los resultados presentados hasta este punto han demostrado que las aproximaciones tradicionales basadas en frecuencias de términos y en expresiones polares no son suficientes a la hora de abordar tareas complejas del análisis sentimental, como la clasificación de oraciones y documentos en diferentes grados de positividad y negatividad, lo que hace necesario disponer de representaciones emocionales más precisas del texto a clasificar. De acuerdo con esto, al representar el texto mediante emociones se hace también necesario un modelo acorde para tratar con los denominados modificadores emocionales (i.e., negación, cuantificación y modalidad), capaz de resolver el efecto de dichos modificadores sobre las emociones dentro de su ámbito de acción.

Los experimentos realizados y presentados en la sección 6.4 han demostrados que la presencia de negaciones, cuantificadores y modales puede modificar el significado emocional de un texto, y que, por lo tanto, es posible mejorar el rendimiento de los sistemas que resuelven las tareas de clasificación de la polaridad e intensidad mediante el correcto modelado y tratamiento de dichas construcciones lingüísticas. Sin embargo, el autor de esta tesis no conoce ningún trabajo previo que haya cuantificado el impacto real de los modificadores emocionales en las tareas de análisis sentimental, así como su interacción entre ellos. De acuerdo a esto, las principales conclusiones del estudio realizado en este trabajo se resumen en los siguientes puntos:

- Una conclusión interesante extraída de los resultados obtenidos es que, cuando el texto es modelado mediante expresiones polares, el efecto de los modificadores emocionales no es uniforme, sino que depende de la

tarea en cuestión. De hecho, mientras los resultados mejoran cuando se procesa el efecto de la negación, los cuantificadores y los modales en la tarea de clasificación de la polaridad respecto de no procesarlos, el tratamiento de dichos modificadores disminuye el rendimiento de la aproximación en la tarea de clasificación de la intensidad. Como ya se ha expuesto, la razón de esta disminución de rendimiento parece ser que las expresiones polares no son una representación suficiente para discriminar entre diferentes clases de intensidad.

- Centrándonos en el método propuesto en esta tesis, y a la vista de los resultados obtenidos, se puede concluir que no todos los modificadores emocionales tienen el mismo impacto sobre las emociones a las que afectan. Este hecho se puede observar claramente en la tabla 7.1, donde los incrementos en exactitud y medida-F son mostrados respecto de los obtenidos en la línea base *Sin modificadores* del método propuesto. Como se puede ver, mientras que el efecto de las negaciones es bastante marcado (3,7 % de mejora en exactitud de media para los experimentos realizados en la colección de evaluación HotelReview), el efecto de los cuantificadores y los modales es mucho menor (1,6 % y 0,7 % de mejora media en exactitud, respectivamente, sobre la misma colección). Atendiendo a esto, se puede concluir que la negación es, de todos los modificadores emocionales, el que más influye en tareas de análisis sentimental, y por tanto, aquel cuyo tratamiento produce un beneficio mayor.

Clasificador	**Polaridad**		**3 clases**		**4 y 5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
SemEval 2007						
Emociones + Negación	0,4	0,3	0,3	0,3	0,7	0,4
Emociones + Cuantificadores	0,3	0,3	0,2	0,4	0,0	0,0
Emociones + Modales	0,3	0,2	0,1	0,1	0,1	0,5
Emociones + Todos los modificadores	1,5	1,5	0,7	0,9	0,9	0,7
HotelReview Corpus						
Emociones + Negación	3,5	3,5	4,5	3,0	3,0	3,6
Emociones + Cuantificadores	0,7	0,7	2,2	0,1	1,9	2,3
Emociones + Modales	1,3	1,2	0,5	0,3	0,3	0,2
Emociones + Todos los modificadores	5,1	5,0	7,2	7,2	6,6	7,2

Tabla 7.1: Porcentajes de incrementos en exactitud y medida-F respecto de los obtenidos en la línea base *Sin modificadores* del método propuesto, para diferentes combinaciones de los modificadores emocionales evaluados

- Como se puede ver en la tabla 7.1, al aumentar el número de clases de intensidad, los resultados decrecen progresivamente, debido, como ya se ha expuesto, al aumento de la dificultad de la tarea. Sin embargo, como era de esperar, la mejora obtenida con el procesamiento de los modificadores emocionales se incrementa ligeramente con el número de clases. Como ya se ha expuesto, al aumentar el número de clases de intensidad los límites entre estas clases se hacen más difusos. Es por esto que un correcto tratamiento del efecto de estos modificadores emocionales sobre las emociones inmersas en el texto es crucial para determinar de una forma más precisa la intensidad emocional.

- El efecto de los distintos modificadores emocionales no es independiente entre sí, sino que interactúan unos con otros. De esta forma, en muchas ocasiones es posible obtener mayores mejoras en rendimiento cuando su efecto es tratado de manera conjunta, que la suma de las mejoras que se obtienen al tratarlos por separado. Una explicación a este hecho es que, por ejemplo, cuando no se procesa la negación o la modalidad, si las emociones asignadas a algunos conceptos en el texto son incorrectas (porque esos conceptos se encuentran en el ámbito de acción de una negación o un modal), el tratamiento de los cuantificadores sobre estos conceptos puede estar ponderando erróneamente las emociones incorrectas, decrementando así el rendimiento global del sistema. Para ilustrar esto, consideremos las siguientes oraciones:

  Neither the bed nor the pillows are very comfortable.

  Staff could be more friendly.

  La primera oración presenta una negación (*neither...nor*) y un cuantificador (*very*) que afecta al concepto *comfortable*. Este concepto es etiquetado con la categoría emocional *like* en el léxico afectivo SentiSense. Nótese que, si la negación no es procesada, el cuantificador *very* intensificará erróneamente esta emoción positiva en un 75 %, por lo que el sistema clasificaría esta oración como positiva atendiendo al *VEI*. Por el contrario, el procesamiento del efecto de la negación hace que el concepto *comfortable* sea etiquetado con su categoría emocional antónima, *disgust*, intensificando de esta manera el cuantificador *very* el peso de dicha categoría emocional negativa en un 75 %. De la misma

manera, el no procesar los modales en la segunda oración haría que el cuantificador *more* intensificará erróneamente el concepto *friendly*, que induciría a una clasificación de la oración como positiva en lugar de negativa.

- Esta interacción entre los cuantificadores y los otros modificadores emocionales explica la baja mejora obtenida al procesar los cuantificadores por separado (ver tabla 7.1), a pesar de que su presencia en las colecciones de evaluación es mucho mayor que la del resto de modificadores emocionales. Además, es importante aclarar de nuevo que, al contrario que las negaciones, los cuantificadores no cambian el significado emocional de los textos a los que afectan, sino que simplemente aumentan o disminuyen su intensidad, por lo que su efecto se diluye en gran medida en un sistema que capture apropiadamente el significado emocional, ya que simplemente afectará a los casos límites. Esta conclusión explica, así mismo, por qué los beneficios de procesar los cuantificadores son mayores en las tareas de clasificación de la intensidad que en la tarea de clasificación de la polaridad, donde los casos límites son mucho más difusos al haber un mayor número de clases. Por último, el poco efecto en los resultados derivado del tratamiento de los modales se debe, principalmente, a la baja frecuencia de estas construcciones lingüísticas en ambas colecciones de evaluación, como ya se expuso en la sección 6.4, así como al hecho de que algunas de las construcciones de modalidad no afectan al significado emocional de los conceptos dentro de su ámbito.

- Los resultados presentados en la tabla 6.16 muestran que el efecto de los modificadores emocionales en la clasificación de titulares de noticias de la colección SemEval 2007 es casi inapreciable, debido, como ya se dicho, a que estos modificadores emocionales son muy poco frecuentes en este tipo de textos. Sin embargo, a pesar de la alta frecuencia con la que los modificadores emocionales aparecen en la colección de evaluación HotelReview, la mejora obtenida es menor de la esperada (ver tabla 6.17). Para poder evaluar el efecto real de los modificadores emocionales en las tareas de clasificación de la polaridad e intensidad, hemos repetido los experimentos realizados a nivel de documento con el método propuesto sobre la colección de evaluación HotelReview, pero

eliminando todos aquellos documentos que no contienen negaciones, cuantificadores y formas modales, respectivamente. De esta manera, se han repetido los experimentos sobre tres colecciones diferentes: *HotelReview-Negaciones*, *HotelReview-Cuantificadores* y *HotelReview-Modales*, midiendo de esta manera el efecto de la negación, los cuantificadores y los modales, respectivamente, en cada una de las colecciones creadas.

Clasificador	**Polaridad**		**3 clases**		**5 clases**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
HotelReview-Negaciones						
Emociones	89,2	88,6	61,6	60,1	40,0	39,0
Emociones + Negación	92,8	92,6	66,4	66,1	43,3	43,0
HotelReview-Cuantificadores						
Emociones	90,4	90,4	64,3	62,3	43,5	41,8
Emociones + Cuantificadores	91,5	91,5	65,9	65,3	43,6	42,1
HotelReview-Modales						
Emociones	87,5	87,3	55,2	53,2	41,5	40,4
Emociones + Modales	91,8	91,7	58,1	56,3	44,0	42,6

Tabla 7.2: Exactitud y medida-F media en la tarea de clasificación de la polaridad e intensidad a nivel de documento, cuando solo son evaluados los documentos que contienen negaciones, cuantificadores y modales

Los resultados de estos experimentos se muestran en la tabla 7.2. Como se puede ver en la tabla, los resultados obtenidos son similares a los conseguidos por el método cuando se evalúa sobre la colección completa, excepto para el caso de los modales. Estos últimos obtienen una mejora media de rendimiento del 3,2 % en exactitud. A partir de estos resultados, se puede concluir que la importancia de los modificadores depende del tipo y dominio de documentos analizados, así como de la tarea a evaluar. Aunque su tratamiento en tareas de análisis sentimental permite alcanzar mejoras de rendimiento importantes, entorno al 6,3 % de exactitud de media para la colección HotelReview, su procesamiento no es tan crítico como en otras tareas y dominios, por ejemplo, en la clasificación de historiales médicos.

- Por último, queremos evaluar si el impacto y la presencia de los modificadores emocionales dependen de la polaridad y la intensidad de los documentos; es decir, si son más frecuentes o su impacto varía dependiendo de cómo de positivo o negativo es un documento. Para ello, repetimos los experimentos anteriores aunque únicamente en la

tarea de clasificación de la polaridad e intensidad para 3 clases. La tabla 7.3 muestra los resultados de precisión (*Pr.*) y cobertura (*Re.*) para cada clase en las colecciones anteriores, *HotelReview-Negaciones*, *HotelReview-Cuantificadores* y *HotelReview-Modales*, respectivamente. Curiosamente, se observa que las instancias negativas (i.e., *Poor*) son mejor clasificadas que las positivas (i.e., *Good*). Un análisis más detallado de los documentos evaluados pone de manifiesto que los usuarios suelen utilizar más conceptos con carga emocional para describir circunstancias de desagrado o disconformidad que para describir situaciones de satisfacción o agrado. Por el contrario, cuando se considera una clase intermedia (i.e., *Fair*), ésta se convierte en la clase peor aprendida, principalmente debido a que aunque los documentos son considerados como neutrales, conllevan cierta carga emotiva que en muchos casos se encuentra en los límites entre lo positivo y lo negativo. Cabe también mencionar que, según los resultados obtenidos, parece que los modificadores emocionales tienen un mayor impacto en las instancias positivas que en el resto. Este hecho se debe, en parte, a que los conceptos negativos son más representativos de expresiones con carga emocional, por lo que en mucho casos, se utiliza la negación de juicios negativos para presentar juicios positivos.

Clasificador	**2 clases**				**3 clases**					
	Good		**Poor**		**Good**		**Fair**		**Poor**	
	Pr.	Re.	Pr.	Re.	Pr.	Re.	Pr.	Re.	Pr.	Re.
HotelReview-Neg										
Emociones	78,4	61,6	91,1	95,9	50,0	36,4	48,6	52,3	72,0	80,6
Emociones + Neg	87,4	74,1	93,9	97,4	62,2	50,5	52,8	54,1	76,6	80,6
HotelReview-Cuan										
Emociones	88,8	92,8	92,2	87,8	66,4	81,1	48,9	30,4	69,6	72,8
Emociones + Cuan	90,5	93,2	92,7	89,7	69,0	82,9	52,3	35,2	72,7	77,9
HotelReview-Mod										
Emociones	79,0	72,1	90,2	93,1	48,6	52,2	33,3	22,5	67,8	79,6
Emociones + Mod	86,2	82,4	93,7	95,2	54,8	59,7	36,7	25,4	68,9	79,6

Tabla 7.3: Precisión y cobertura media por clase en la tarea de clasificación de la polaridad e intensidad para 3 clases

Por último, la comparación de nuestro método basado en emociones, así como el modelado del efecto de los modificadores emocionales sobre dichas emociones, respecto de la línea base de expresiones polares demuestra que, a tenor de los resultados, en casi todos los experimentos la aproximación

de emociones obtiene mayores mejoras que la aproximación basada en expresiones polares. Es por tanto razonable afirmar que la representación del texto mediante emociones y el tratamiento de los modificadores sobre dichas emociones redunda en mejores resultados que el uso de expresiones polares, donde solo se tiene en cuenta si el concepto es negativo o positivo y los modificadores se tratan como simples inversores de polaridad.

7.4. El efecto de la ambigüedad

Finalmente, analizamos el impacto de la ambigüedad en las tareas de clasificación de la polaridad y la intensidad. Como se puede ver en las tablas 6.18 y 6.19, el uso de un algoritmo de desambiguación semántica para resolver la ambigüedad mejora los resultados de ambas tareas, tanto a nivel de oración como a nivel de documento. Mediante el uso de un algoritmo de desambiguación semántica se puede determinar de una manera más precisa el significado correcto de un concepto de acuerdo al contexto en que se utiliza, lo que se traduce en la creación de un *VEI* que captura más fielmente el significado emocional del texto.

La mejora obtenida es, en el caso de la tarea de clasificación de la polaridad, más evidente a nivel de oración que cuando se clasifican documentos completos. Este resultado no es sorprendente, ya que cuanta más información disponible haya para la predicción de la instancia, menor es el efecto de asignar un significado erróneo a alguno de sus términos. Por el contrario, este hecho se invierte al abordar la tarea de clasificación de la intensidad. Al conllevar esta tarea más dificultad, los fallos de asignar un significado erróneo a varios conceptos dentro de un documento pueden suponer peores clasificaciones, mientras que a nivel de oración al haber menor número de conceptos menor es el número de posibles de errores, y por tanto, menor la diferencia entre las estrategias.

Otro resultado interesantes es que, aunque el uso de un algoritmo de desambiguación semántica mejora los resultados de exactitud y medida-F, esta mejora es menor de la esperada con respecto a la estrategia del primer significado. Esto se debe a que WordNet organiza los significados de los términos de acuerdo a su frecuencia, por lo que el primer significado es siempre el más probable y frecuente para cada término. Es por esto que la estrategia del significado más frecuente en tareas de desambiguación semántica es con-

siderada una línea base difícil de batir. Por el contrario, la mejora respecto de la estrategia aleatoria es bastante acusada (aproximadamente el 10 % de mejora en exactitud en promedio).

Atendiendo a estos resultados, se puede concluir que, aunque el estado del arte de los algoritmos de desambiguación semántica está aun lejos de conseguir una calidad óptima (la efectividad del algoritmo Lesk está entorno al 30 o 40 por ciento de los términos analizados), lo cierto es que su uso mejora considerablemente los resultados frente a la estrategia aleatoria y, aunque en menor medida, frente a los resultados de la estrategia del primer significado. Esto pone de manifiesto, por un lado, la importancia de determinar el significado correcto de cada término de cara a asignarle la emoción adecuada en las tareas de análisis sentimental; y por otro, que el porcentaje de error que introducen los algoritmos de desambiguación afecta a los resultados de la clasificación, y que con la mejora de dichos algoritmos es de esperar que mejore también la exactitud de los clasificadores emocionales.

Capítulo 8

Conclusiones y trabajo futuro

8.1. Conclusiones

En esta tesis se ha presentado una nueva aproximación que pretende resolver distintas tareas dentro del área del análisis sentimental, como son la clasificación de textos según su polaridad y su intensidad emocional. El sistema realiza un análisis lingüístico en profundidad más complejo que las aproximaciones actuales del estado del arte, y hace un uso intensivo de conocimiento e información contextual para abordar problemas típicos del procesamiento del lenguaje natural, como la ambigüedad léxica y la detección de la presencia y el ámbito de acción de negaciones, cuantificadores y modales, en textos afectivos.

El método diseñado se basa en la traducción de los términos o palabras del texto a conceptos de la base de datos léxica WordNet, utilizando para ello diferentes estrategias de desambiguación semántica para obtener el significado correcto de cada término de acuerdo con su contexto. Estos conceptos son, a su vez, traducidos a categorías emocionales del léxico afectivo SentiSense, desarrollado como parte de este trabajo, y que liga conceptos con categorías emocionales. A diferencia de los enfoques anteriores, el texto de entrada es modelado como un conjunto de emociones en lugar de términos o expresiones polares, lo que permite capturar con mayor fidelidad la polaridad e intensidad del texto afectivo. Así mismo, se han desarrollado técnicas lingüísticas avanzadas para la identificación de negaciones, cuantificadores

y modales, así como su ámbito de acción y su efecto sobre las emociones a las que afectan.

Por otro lado, se ha realizado una evaluación exhaustiva para determinar la viabilidad del modelo propuesto y valorar el desempeño del sistema desarrollado en dos tareas típicas del análisis sentimental; por un lado, la clasificación de textos según su polaridad (i.e., positivo *vs.* negativo); y por otro, su clasificación en función de su intensidad emocional (i.e., diferentes grados de positividad y negatividad), en ambos casos a nivel de oración y a nivel de documento. Para ello, se han utilizado tres colecciones de evaluación, compuestas por documentos de diferentes tipos y dominios: una colección de titulares de noticias periodísticas, un conjunto de opiniones de usuarios de hoteles, y una colección de opiniones de usuarios de 8 productos y servicios diferentes y muy dispares. Los resultados obtenidos con nuestro sistema mejoran los conseguidos por los otros sistemas evaluados bajo las mismas condiciones experimentales y en las mismas tareas. Estos resultados demuestran la factibilidad de nuestra hipótesis, y cómo el modelado de los textos mediante emociones, junto con el tratamiento de la ambigüedad y el efecto de los modificadores emocionales, pueden mejorar sustancialmente los resultados en tareas de análisis sentimental.

La detección automática de modificadores lingüísticos en textos afectivos ha despertado recientemente el interés de la comunidad científica en el área del análisis sentimental. Las negaciones, los cuantificadores y los modales pueden modificar y cambiar el significado emocional de los textos a los que afectan, por lo que es importante determinar su ámbito de acción, así como sus efectos, para poder clasificar adecuadamente oraciones y documentos de acuerdo a diferentes clases de polaridad e intensidad. Sin embargo, el problema es mucho más complejo que simplemente invertir, disminuir o aumentar la polaridad de las expresiones polares que se encuentran dentro de ámbito de acción de los modificadores emocionales, y requiere modelar su efecto sobre las emociones concretas expresadas en el texto.

A razón de los resultados obtenidos, la investigación realizada en esta tesis demuestra que es posible mejorar el rendimiento de las tareas de clasificación de la polaridad y la intensidad emocional mediante la representación de los textos como un conjunto de emociones y mediante el tratamiento de la negación, los cuantificadores y los modales como modificadores de dichas emociones. De esta manera, representar el efecto de los modificadores co-

mo acciones sobre las emociones (i.e., utilización de la emoción antónima, disminución o intensificación de las emociones, etc.) permite representar de una forma más precisa el efecto de estas construcciones lingüísticas desde el punto de vista emocional. Para demostrar esto, comparamos los resultados obtenidos por nuestra aproximación con los obtenidos por una de las estrategias más comúnmente aceptada para el tratamiento de estos modificadores, y que se basa en el uso de expresiones polares. Esta comparación revela los beneficios de nuestra nueva aproximación, permitiéndonos afirmar que el tratamiento de estas construcciones lingüísticas como modificadores emocionales mejora considerablemente el rendimiento de los sistemas de análisis sentimental.

Uno de los principales objetivos de este trabajo es cuantificar el impacto real de los modificadores emocionales en la tareas del área de análisis sentimental. En conocimiento del autor de esta tesis, ningún trabajo previo ha cuantificado el efecto de estas construcciones lingüísticas en la interpretación de los textos desde un punto de vista afectivo. En este trabajo, se ha experimentado con dos tareas del análisis sentimental, la clasificación de la polaridad y la clasificación en función de la intensidad, evaluando el impacto directo de los modificadores emocionales sobre el rendimiento en el desempeño de ambas tareas, y sobre textos de diferentes dominios. Sin embargo, mediante este análisis se ha concluido que, aunque el procesamiento de los modificadores emocionales produce mejoras significativas en todas las tareas y tipos de documentos, su efecto es menor del esperado. Este hecho sugiere que el procesamiento de los modificadores emocionales, aunque ayuda a los sistemas de análisis sentimental a capturar de una forma más precisa el significado emocional, especialmente en la tarea de clasificación de la intensidad, no es un elemento tan crítico como puede serlo en otras disciplinas como el procesamiento de textos biomédicos.

Por otro lado, el estudio realizado de los modificadores emocionales ha revelado que no todos afectan a las emociones de la misma forma. Así, la negación es el modificador que más influye, mientras que el efecto de los cuantificadores y los modales es mucho más sutil. La principal explicación para este hecho es que, mientras que la negación invierte las emociones asociadas a los conceptos dentro de su ámbito de acción, los cuantificadores simplemente incrementan o decrementan la intensidad de esas emociones. En el caso de los modales, a su vez, su efecto es no afectar al significado

emocional o simplemente decrementar su intensidad emocional, y solo en ciertas ocasiones invierte el significado emocional. Otra conclusión interesante es el hecho de que el procesamiento combinado de todos los modificadores produce, en la mayoría de los casos, mejores resultados que la suma del beneficio que produce su tratamiento por separado. Estos resultados confirman nuestra hipótesis de que existen sinergias entre los diferentes efectos de los modificadores emocionales, y que, por tanto, su tratamiento conjunto permite capturar mejor el significado emocional del texto analizado.

Los resultados de la evaluación confirman así mismo nuestra intuición de que un sistema que trabaje a nivel de conceptos en lugar de términos, y que haga uso de un algoritmo de desambiguación semántica para determinar el significado adecuado de cada palabra de acuerdo a su contexto, obtendrá mejores resultados. De esta manera, se reduce el número de errores cometidos al asumir un significado erróneo de un término no desambiguado, que suelen ser muy frecuentes en las aproximaciones basadas en términos y que hacen uso de léxicos afectivos basados en términos o palabras en lugar de conceptos. Estos resultados demuestran, además, que aunque el porcentaje de acierto de los algoritmos de desambiguación semántica sea insatisfactorio, su uso mejora sustancialmente los resultados del sistema de clasificación emocional, lo que presupone un beneficio aún mayor cuando el estado del arte de los algoritmos de desambiguación mejore.

Por último, es importante mencionar también que los buenos resultados obtenidos con el sistema se han conseguido utilizando un léxico afectivo que simplemente cubre aproximadamente el 20 % del vocabulario de las colecciones de evaluación, o incluso menos en el caso de la colección SFU (9,5 % de cobertura del léxico afectivo SentiSense). Este hecho sugiere que no es necesario disponer de un léxico afectivo que cubra todo el vocabulario del texto para obtener unos resultados altamente satisfactorios, puesto que simplemente con 1 de cada 5 palabras etiquetadas en el léxico afectivo se han obtenido resultados que han superado ampliamente las dos líneas base implementadas.

8.2. Trabajo futuro

Aunque los resultados obtenidos en este trabajo son prometedores, se han detectado algunos aspectos que deben ser estudiados con mayor consideración.

Una de las principales debilidades del sistema es su comportamiento ante textos, especialmente oraciones, cuya carga emocional es muy baja. Este tipo de textos hacen que el sistema no sea capaz de determinar en muchas ocasiones la polaridad o intensidad de los mismos. Para resolver este problema, habría que tener en cuenta información adicional del dominio y del contexto de los textos. Por ejemplo, para el caso de titulares de noticias periodísticas, sería interesante tener en cuenta información relacionada con la sección del periódico en que se encuentran, el cuerpo de la noticia, etc. Para el caso de opiniones de productos, sería interesante conocer cierta información relativa al producto, como por ejemplo las características del mismo que más importan a los usuarios, con el objetivo de ponderar como afectan las opiniones de acuerdo a la característica expresada.

Otro aspecto importante, también relacionado con el tratamiento de textos con poca carga emocional, sería la adaptación del modelo para introducir un paso previo que determinara aquellas oraciones que expresan hechos y aquellas que expresan opiniones, sentimientos o creencias. Es decir, adaptar el sistema para introducir un paso previo de clasificación de la subjetividad, donde se separaran las oraciones con carga emocional (subjetivas) de las que no expresan emoción alguna (objetivas), analizando en un segundo paso solamente las oraciones subjetivas. De esta manera, se limitaría el efecto de estas oraciones con poca carga emotiva y, previsiblemente, se mejorarían los resultados tanto en la tarea de clasificación de la polaridad como en la tarea de clasificación de la intensidad, al simplemente analizar los textos con carga emocional. Para ello habría, además, que seleccionar una colección de evaluación bien creada y estructurada que nos permitiera medir adecuadamente el impacto de este nuevo paso previo.

Por otro lado, aunque el uso de un algoritmo de desambiguación semántica ha supuesto mejoras sustanciales en las tareas evaluadas, lo cierto es que los errores de desambiguación debidos a la efectividad del algoritmo utilizado han supuesto una limitación a estas mejoras. Es por esto que en un futuro nos gustaría realizar una evaluación extrínseca de diferentes algoritmos de desambiguación semántica, como por ejemplo los presentados en (Agirre y Soroa, 2009). De esta manera, podríamos evaluar de una forma más precisa el impacto sobre el sistema del uso de diferentes algoritmos de desambiguación semántica y con diferente rendimiento, lo que supondría, en principio, mejoras adicionales para el sistema.

Como ya se ha expuesto, los buenos resultados obtenidos se han logrado con una cobertura del léxico afectivo de aproximadamente el 20 % del vocabulario de las colecciones de evaluación, e incluso mucho inferior para el caso de la colección de evaluación SFU. Sin embargo, sería interesante estudiar el límite de convergencia entre la mejora posible de los resultados y la cobertura de vocabulario del léxico afectivo utilizado. Es decir, evaluar qué porcentaje de cobertura es necesario para obtener mejoras de rendimiento, y a partir del cual dicho rendimiento se estabiliza. De esta manera, se podría determinar la cobertura de vocabulario necesaria para portar el método a cualquier otro dominio.

Por otro lado, aunque en este trabajo se ha estudiado y propuesto un modelo para el tratamiento de los modificadores emocionales, se han quedado fuera de este modelo fenómenos como la ironía o el sarcasmo. Aunque es cierto que estos modificadores no son tan frecuentes como los estudiados, no cabe duda de que su uso puede afectar a la interpretación de la carga emotiva del texto. El análisis y tratamiento de la ironía o el sarcasmo es, probablemente, una de las tareas más complejas dentro del análisis de los modificadores emocionales, ya que no existen modelos conceptuales sobre cómo se expresan y el efecto que producen. Es por esto que son muy pocas las aproximaciones que los han estudiado, y las que lo han hecho, solo abordan el problema de forma muy superficial. El estudio de este tipo de formas lingüísticas requiere de gran cantidad de conocimiento y de un análisis exhaustivo del contexto del texto, puesto que estos fenómenos son difíciles de interpretar incluso para los humanos. Es por ello que este estudio se plantea como un reto interesante de trabajo futuro.

Otro aspecto interesante que habría que considerar en relación a las tareas del análisis sentimental a nivel de documentos, sería el posible efecto que tiene la estructura del discurso. Existen ciertos patrones de la estructura del discurso, especialmente utilizados en cierto tipo de documentos, que pueden afectar al análisis afectivo global del mismo. Por ejemplo, una estructura típica del discurso es aquella en la que las primeras oraciones expresan una polaridad, ya sea positiva o negativa, pero terminan con una oración expresando la polaridad contraria y cambiando por completo la polaridad global del documento. Este tipo de estructuras son muy utilizadas, por ejemplo, en debates o en opiniones que comparan varios productos de un dominio, donde en las primeras oraciones se hacen concesiones pero se concluye con la

opinión real sobre el producto. Este análisis se podría incluir como parte del método propuesto mediante una etapa de post-procesamiento que analizara cada oración del documento y resolviera la polaridad final de acuerdo a los diferentes patrones de la estructura del discurso identificados.

Aunque la tarea de determinar la emoción predominate de un texto no ha obtenido mucha popularidad en el área de procesamiento del lenguaje natural, principalmente debido a la subjetividad y dificultad de la tarea en sí, incluso para los seres humanos, sería interesante evaluar el método propuesto para dicha tarea. En realidad, el sistema ya está, de por sí, realizando un análisis de la emoción predominante del texto analizado, por lo que simplemente habría que disponer de una colección de evaluación que cumpliera los requisitos necesarios para evaluar el sistema en dicha tarea. Así, podríamos determinar el comportamiento de nuestro sistema no solo en las tareas de clasificación de la polaridad y la intensidad, sino también en la tarea de detección de la emoción predominante.

Por último, nos gustaría traducir el modelo emocional y semántico propuesto en esta tesis para abordar las mismas tareas en otros idiomas, especialmente el español. Para ello, habría que evaluar las diferentes herramientas utilizadas y su disponibilidad en el idioma elegido, así como estudiar los distintos patrones de uso de los modificadores emocionales en dicho idioma, con el objetivo de resolver su efecto. De esta manera, se podría, además, analizar la dificultad de cada idioma en las tareas estudiadas en esta tesis.

Bibliografía

Abbasi, A., H. Chen, y A. Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Trans. Inf. Syst.*, 26:12:1–12:34.

Agirre, E. y P. Edmonds. 2006. *Word Sense Disambiguation: Algorithms and Applications*, volumen 33 de *Text, Speech and Language Technology*. Springer, 1 edición.

Agirre, E. y A. Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. En *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, páginas 33–41.

Agrawal, S. y T. J. Siddiqui. 2009. Using syntactic and contextual information for sentiment polarity analysis. En *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, páginas 620–623. ACM.

Andreevskaia, A. y S. Bergler. 2007. CLaC and CLaC-NB: Knowledge-based and corpus-based approaches to sentiment tagging. En *Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 117–120. Association for Computational Linguistics.

Anscombe, E. y P. Geach. 1970. *Descartes Philosophical Writings*. Nelson: The Open University.

Arora, S., E. Mayfield, C. Penstein Rosé, y E. Nyberg. 2010. Sentiment classification using automatically extracted subgraph features. En *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches*

to Analysis and Generation of Emotion in Text, páginas 131–139. Association for Computational Linguistics.

Aue, A. y M. Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. En *Proceedings of the International Conference on Recent Advances in NLP*.

Baccianella, S., A. Esuli, y F. Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. En Nicoletta Calzolari (Conference Chair) Khalid Choukri Bente Maegaard Joseph Mariani Jan Odijk Stelios Piperidis Mike Rosner, y Daniel Tapias, editores, *Proceedings of the Seventh conference on International Language Resources and Evaluation*. European Language Resources Association.

Baker, K., M. Bloodgood, B. Dorr, N. W. Filardo, L. Levin, y C. Piatko. 2010. A modality lexicon and its use in automatic tagging. En *Proceedings of Language Resources and Evaluation Conference*, páginas 1402–1407. European Language Resources Association.

Bentivogli, L., P. Forner, B. Magnini, y E. Pianta. 2004. Revising the wordnet domains hierarchy: Semantics, coverage and balancing. En *Proceedings of the Workshop on Multilingual Linguistic Ressources*, páginas 101–108. Association for Computational Linguistics.

Bradley, M.M. y P.J. Lang. 1999. Affective norms for english words (ANEW): Instruction manual and affective ratings. Technical report c-1, The Center for Research in Psychophysiology, University of Florida.

Brooke, J. 2009. A semantic approach to automated text sentiment analysis. Master's thesis, SIMON FRASER UNIVERSITY.

Bruce, R. F. y J. M. Wiebe. 1999. Recognizing subjectivity: a case study in manual tagging. *Nat. Lang. Eng.*, 5:187–205.

Brustein, O. 2010. Nation's political pulse, taken using net chatter. Published October 31 2010. Visto el 15 de diciembre de 2010 en http://www.nytimes.com/2010/11/01/technology/01sentiment.html. The New York Times.

Carbonell, J. 1979. *Subjective understanding: Computer models of belief systems*. Ph.D. tesis, Yale.

Carrillo de Albornoz, J., L. Plaza, P. Gervás, y A. Díaz. 2011. A joint model of feature mining and sentiment analysis for product review rating. En *Proceedings of the 33rd European Conference on Information Retrieval*.

Carroll, J., G. Minnen, y T. Briscoe. 1999. Corpus annotation for parser evaluation. En *Proceedings of the European Chapter of the Association for Computational Linguistics, workshop on Linguistically Interpreted Corpora*, páginas 35–41.

Carter, R. y M. McCarthy. 2006. *Cambridge grammar of english. A compresive guide spoken and written english grammar and usage*. Cambridge University Press.

Chaumartin, F-R. 2007. Upar7: a knowledge-based system for headline sentiment tagging. En *Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 422–425. Association for Computational Linguistics.

Choi, Y. y C. Cardie. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, páginas 793–801. Association for Computational Linguistics.

ComScore. 2007. Online consumer-generated reviews have significant impact on offline purchase behavior. Visto el 15 de diciembre, 2010, en http://www.comscore.com/press/release.asp?press=1928.

Councill, I., R. McDonald, y L. Velikovich. 2010. What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. En *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, páginas 51–59. University of Antwerp.

Cowie, R. y R. R. Cornelius. 2003. Describing the emotional states that are expressed in speech. *Speech Commun.*, 40(1-2):5–32.

Cunningham, H., D. Maynard, K. Bontcheva, y V. Tablan. 2002. GATE: A framework and graphical development environment for robust NLP

tools and applications. En *Proceedings of the 40th Annual Meeting of the ACL*.

Dalianis, H. y M. Skeppstedt. 2010. Creating and evaluating a consensus for negated and speculative words in a swedish clinical corpus. En *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, páginas 5–13. Association for Computational Linguistics.

Das, A y S. Bandyopadhyay. 2010. Subjectivity detection using genetic algorithm. En *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*.

Das, S. y M. Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. En *Proceedings of the Asia Pacific Finance Association Annual Conference*.

Dave, K., S. Lawrence, y D. M. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. En *Proceedings of the 12th international conference on World Wide Web*, páginas 519–528. ACM.

Devitt, A. y K. Ahmad. 2007. Sentiment polarity identification in financial news: A cohesion-based approach. En *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, páginas 984–991. Association for Computational Linguistics.

Ekman, P. 1992. Are there basic emotions? *Psychological Review*, 99(3):550–553.

Elango, P. 2005. Coreference resolution: A survey. Informe técnico, University of Wisconsin Madison.

Esuli, A. y F. Sebastiani. 2005. Determining the semantic orientation of terms through gloss classification. En *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, páginas 617–624. ACM.

Esuli, A. y F. Sebastiani. 2006a. Determining term subjectivity and term orientation for opinion mining. En *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*.

Esuli, A. y F. Sebastiani. 2006b. SENTIWORDNET: A publicly available lexical resource for opinion mining. En *Proceedings of the 5th Conference on Language Resources and Evaluation*, páginas 417–422.

Francisco, V., P. Gervás, y F. Peinado. 2010. Ontological reasoning for improving the treatment of emotions in text. *Knowledge and Information Systems*, 24(2):23.

Goldstein, I. y O. Uzuner. 2010. Does Negation really matter? En *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, páginas 23–27. Association for Computational Linguistics.

Hatzivassiloglou, V. y K. R. McKeown. 1997. Predicting the semantic orientation of adjectives. En *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, páginas 174–181. Association for Computational Linguistics.

Hatzivassiloglou, V. y J. M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. En *Proceedings of the 18th Conference on Computational linguistics*, páginas 299–305. Association for Computational Linguistics.

Hearst, M. 1992. Direction-based text interpretation as an information access refinement. En Paul Jacobs, editor, *Text-Based Intelligent Systems*. Lawrence Erlbaum Associates, páginas 257–274.

Hitlin, P y L. Rainie. 2004. The use of online reputation and rating systems. *Pew Internet & American Life Project Memo*.

Hu, M. y B. Liu. 2004. Mining and summarizing customer reviews. En *Proceedings of the tenth ACM International Conference on Knowledge Discovery and Data Mining*, páginas 168–177. ACM.

Huettner, A. y P. Subasic. 2000. Fuzzy typing for document management. En *ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes*, páginas 26–27.

Ide, N. y J. Véronis. 1998. Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24:2–40.

Izard, C. E. 1977. Human emotions. *Plenum Press,New York*.

James, W. 1884. *What is emotion?*, volumen Mind 9.

Jia, L., C. Yu, y W. Meng. 2009. The effect of negation on sentiment analysis and retrieval effectiveness. En *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, páginas 1827–1830. ACM.

Katz, P., M. Singleton, y R. Wicentowski. 2007. Swat-mp: the semeval-2007 systems for task 5 and task 14. En *Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 308–313. Association for Computational Linguistics.

Kennedy, A. y D. Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22:2006.

Kim, S-M. y E. Hovy. 2004. Determining the sentiment of opinions. En *Proceedings of the 20th International Conference on Computational Linguistics*. Association for Computational Linguistics.

Kim, S-M. y E. Hovy. 2005. Automatic Detection of Opinion Bearing Words and Sentences. En *Companion Volume to the Proceedings of the Second International Joint Conference on Natural Language Processing*.

Klein, D. y C. D. Manning. 2003. Accurate unlexicalized parsing. En *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, páginas 423–430. Association for Computational Linguistics.

Lakoff, G. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2:458–508.

Lambov, D., G. Dias, y J. Graça. 2010. Multi-view learning for text subjectivity classification. En *Proceedings of the 1st Workshop on omputational Approaches to Subjectivity and Sentiment Analysis*.

Lang, P. J. y M. M. Bradley. 1999. Affective norms for english words (anew). En *Technical Manual and Affective Rattings*.

Lang, P.J. 1980. Behavioral treatment and bio-behavioral assessment: Computer applications. *Ablex Publishing, Norwood.*

Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. En *Proceedings*

of the 5th Annual International Conference on Systems Documentation, páginas 24–26. ACM.

Light, M., X. Y. Qiu, y P. Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. En *Proceedings of the HLT-NAACL. Workshop: BioLINK, Linking Biological Literature, Ontologies and Databases*, páginas 17–24. Association for Computational Linguistics.

Liu, B. 2010. Sentiment analysis and subjectivity. En Nitin Indurkhya y Fred J. Damerau, editores, *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL. ISBN 978-1420085921.

Marneffe, M., B. Maccartney, y C. Manning. 2006. Generating typed dependency parses from phrase structure parses. En *Proceedings of the Language Resources and Evaluation Conference*, páginas 449–454.

Martineau, J. y T. Finin. 2009. Delta TFIDF: An improved feature space for sentiment analysis. En *Proceedings of the Third International Conference on Weblogs and Social Media*.

Medlock, B. y T. Briscoe. 2007. Weakly supervised learning for hedge classfication in scientific literature. En *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, páginas 992–999. Association for Computational Linguistics.

Meena, A. y T. V. Prabhakar. 2007. Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. En *Proceedings of the 29th European Conference on Information Retrieval Research*, páginas 573–580. Springer-Verlag.

Miller, G. A. 1995. WordNet: a lexical database for English. *Commun. ACM*, 38:39–41.

Moilanen, K. y S. Pulman. 2007. Sentiment composition. En *Proceedings of Recent Advances in Natural Language Processing*, páginas 378–382.

Morante, R. 2010. Descriptive analysis of negation cues in biomedical texts. En Nicoletta Calzolari (Conference Chair) Khalid Choukri Bente Maegaard Joseph Mariani Jan Odijk Stelios Piperidis Mike Rosner, y Daniel

Tapias, editores, *Proceedings of the Seventh Conference on International Language Resources and Evaluation*. European Language Resources Association (ELRA).

Mullen, T. y N. Collier. 2004. Sentiment analysis using support vector machines with diverse information sources. En Dekang Lin y Dekai Wu, editores, *Proceedings of the Empirical Methods in Natural Language Processing*, páginas 412–418. Association for Computational Linguistics.

Na, J. C., H. Sui, C. Khoo, S. Chan, y Y. Zhou. 2004. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. En I. C. McIlwaine (Ed.), editor, *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference*, páginas 49–54.

Ng, V., S. Dasgupta, y S. M. N. Arifin. 2006. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. En *Proceedings of the COLING/ACL on Main conference poster sessions*, páginas 611–618. Association for Computational Linguistics.

Nicolov, N., F. Salvetti, y S. Ivanova. 2008. Sentiment analysis: Does coreference matter? En *Proceedings of the Symposium on Affective Language in Human and Machine*, páginas 37–40.

Ortony, A., G. Clore, y A. Collins. 1988. *The cognitive structure of emotions*. Cambridge University Press.

Ortony, A. y T. J. Turner. 1990. What's basic about basic emotions? *Psychological Review*, 97(3):315–331.

Palmer, F. R. 2001. *Mood and modality*. Cambridge University Press.

Paltoglou, G. y M. Thelwall. 2010. A study of information retrieval weighting schemes for sentiment analysis. En *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, páginas 1386–1395. Association for Computational Linguistics.

Pang, B. y L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. En *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, páginas 271–278.

Pang, B. y L. Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. En *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, páginas 115–124. Association for Computational Linguistics.

Pang, B. y L. Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.

Pang, B., L. Lee, y S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, páginas 79–86.

Parrot, W.G. 2001. *Emotions in social psychology: Essential readings.* Psychology Press.

Patwardhan, S., S. Banerjee, y T. Pedersen. 2005. SenseRelate::TargetWord: A generalized framework for word sense disambiguation. En *Proceedings of the ACL on Interactive Poster and Demonstration Sessions*, páginas 73–76. Association for Computational Linguistics.

Pennebaker, J. W., M. E. Francis, y R. J. Booth. 2001. *Linguistic Inquiry and Word Count (LIWC): LIWC2001.*

Plaza, L., A. Díaz, y P. Gervás. 2011. A semantic graph-based approach to biomedical summarisation. *Artificial Intelligence in Medicine.*

Plutchik, R., 1980. *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, capítulo A general psychoevolutionary theory of emotion, páginas 3–33. New York: Academic.

Polanyi, L. y A. Zaenen. 2006. Contextual valence shifters. En W. Bruce Croft James Shanahan Yan Qu, y Janyce Wiebe, editores, *Computing Attitude and Affect in Text: Theory and Applications*, volumen 20 de *The Information Retrieval Series.* Springer Netherlands, páginas 1–10.

Quirk, R., S. Greenbaum, G. Leech, y J. Svartvik. 1985. *A comprehensive grammar of the English language.* Longman.

Riloff, E., S. Patwardhan, y J. Wiebe. 2006. Feature subsumption for opinion analysis. En *Proceedings of the 2006 Conference on Empirical Meth-*

ods in Natural Language Processing, páginas 440–448. Association for Computational Linguistics.

Riloff, E. y J. Wiebe. 2003. Learning extraction patterns for subjective expressions. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, páginas 105–112. Association for Computational Linguistics.

Riloff, E., J. M. Wiebe, y T. Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. En *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4*, páginas 25–32. Association for Computational Linguistics.

Rokach, L., R. Romano, y O. Maimon. 2008. Negation recognition in medical narrative reports. *Inf. Retr.*, 11:499–538.

Russell, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Sack, W. 1994. On the computation of point of view. En *Proceedings of AAAI*, página 1488.

Saggion, H., E. Lloret, y M Palomar. 2010. Using text summaries for predicting rating scales. En *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, páginas 44–51.

Sahlgren, M., J. Karlgren, y G. Eriksson. 2007. Sics: valence annotation based on seeds in word space. En *Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 296–299. Association for Computational Linguistics.

Sauri, R. y J. Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43:227–268.

Shaikh, M. A., H. Prendinger, y I. Mitsuru. 2007. Assessing sentiment of text by semantic dependency and contextual valence analysis. En *Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction*, páginas 191–202. Springer-Verlag.

Stone, P. J., D. C. Dunphy, M. S. Smith, y D. M. Ogilvie. 1966. *The General Inquirer: A computer approach to content analysis.* MIT Press.

Strapparava, C. y R. Mihalcea. 2007. SemEval-2007 task 14: affective text. En *Proceedings of the 4th International Workshop on Semantic Evaluations*, páginas 70–74. Association for Computational Linguistics.

Strapparava, C. y A. Valitutti. 2004. WordNet-Affect: An affective extension of WordNet. En *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volumen 2004, páginas 1083–1086.

Szarvas, G. 2008. Hedge Classification in Biomedical Texts with a Weakly Supervised Selection of Keywords. En *Proceedings of 46th Meeting of the Association for Computational Linguistics*, páginas 281–289.

Taboada, M. y J. Grieve. 2004. Analyzing appraisal automatically. En *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, páginas 158–161.

Tong, R. M. 2001. An operational system for detecting and tracking opinions in on-line discussion. En *Proceedings of the Workshop on Operational Text Classification*.

Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. En *Proceedings of the Association for Computational Linguistics*, páginas 417–424.

Van Asch, V., R. Morante, y W. Daelemans. 2010. Towards improving the precision of a relation extraction system by processing negation and speculation. En *Proceedings of the 4th Symposium on Semantic Mining in Biomedicine*, páginas 164–165.

Velupillai, S. 2010. Towards a better understanding of uncertainties and speculations in Swedish clinical text: analysis of an initial annotation trial. En *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, páginas 14–22. Association for Computational Linguistics.

Von Fintel, K., 2006. *Encyclopedia of Philosophy*, capítulo Modality and Language. MacMillan Reference.

Watson, D. y A. Tellegen. 1985. Toward a consensual structure of mood. *Psychological bulletin*, 98(2):219–235.

Whissell, C.M, 1989. *Emotion: Theory, Research, and Experience*, capítulo The dictionary of affect in language, páginas 113–131. New York: Academic Press.

Whitelaw, C., N. Garg, y S. Argamon. 2005. Using appraisal groups for sentiment analysis. En *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, páginas 625–631. ACM.

Wiebe, J. M. 1990a. Identifying subjective characters in narrative. En *Proceedings of the International Conference on Computational Linguistics*, páginas 401–408.

Wiebe, J. M. 1990b. *Recognizing subjective sentences: a computational investigation of narrative text*. Ph.D. tesis.

Wiebe, J. M. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.

Wiebe, J. M. y R. Bruce. 1995. Probabilistic classifiers for tracking point of view. En *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, páginas 181–187.

Wiebe, J. M., R. F. Bruce, y T. P. O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. En *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, páginas 246–253. Association for Computational Linguistics.

Wiebe, J. M. y R. Mihalcea. 2006. Word sense and subjectivity. En *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, páginas 1065–1072. Association for Computational Linguistics.

Wiebe, J. M. y T. Wilson. 2002. Learning to disambiguate potentially subjective expressions. En *proceedings of the 6th conference on Natural language learning - Volume 20*, páginas 1–7. Association for Computational Linguistics.

Wiebe, J. M., T. Wilson, R. Bruce, M. Bell, y M. Martin. 2004. Learning subjective language. *Computational Linguistics*, 30:277–308.

Wiegand, M., A. Balahur, B. Roth, D. Klakow, y A. Montoyo. 2010. A survey on the role of negation in sentiment analysis. En *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, páginas 60–68. Association for Computational Linguistics.

Wilks, Y. y J. Bien. 1984. Beliefs, points of view and multiple environments. En *Proceedings of the international NATO Symposium on Artificial and Human Intelligence*, páginas 147–171. Elsevier North-Holland, Inc.

Wilson, T., J. M. Wiebe, y P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. En *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, páginas 347–354. Association for Computational Linguistics.

Wilson, T., J. M. Wiebe, y P. Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433.

Wilson, T., J. M. Wiebe, y R. Hwa. 2004. Just how mad are you? Finding strong and weak opinion clauses. En *Proceedings of the 19th National Conference on Artifical intelligence*, páginas 761–767. AAAI Press.

Yu, H. y V. Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, páginas 129–136. Association for Computational Linguistics.

Zaidan, O., J. Eisner, y C. Piatko. 2007. Using "Annotator Rationales" to improve machine learning for text categorization. En *Proceedings of the Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*, páginas 260–267. Association for Computational Linguistics.

Apéndice A

Publicaciones

A continuación se muestran las publicaciones que han resultado de la realización de esta tesis doctoral.

A.1. Clasificación de la polaridad y la intensidad

1. Carrillo de Albornoz J, Plaza L, Gervás P. 2010. Improving Emotional Intensity Classification using Word Sense Disambiguation. En Journal on Research in Computing Science, volumen 46. Special issue: Natural Language Processing and its Applications, páginas 131-142.

2. Carrillo de Albornoz J, Plaza L, Gervás P. 2010. A Hybrid Approach to Emotional Sentence Polarity and Intensity Classification. En Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL 2010), páginas 153-161. Uppsala, Suecia.

3. Carrillo de Albornoz J, Plaza L, Gervás P, Díaz A. 2011. A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating. En Proceedings of the 33rd European Conference on Information Retrieval (ECIR 2011), páginas 55-66. Dublin, Irlanda.

4. Plaza L, Carrillo de Albornoz J. 2011. Sentiment Analysis in Business Intelligence: A survey. Capítulo de libro en: Customer Relationship Management and the Social and Semantic Web: Enabling Cliens Conexus. IGI-Global. Editores: R. Colomo-Palacios, J. Varajao y P. Soto-Acosta.

A.2. Aplicación del análisis sentimental a otros dominios de la inteligencia artificial

1. Carrillo de Albornoz J, Garmendia L, Gervás P. 2008. Sistema automático de decisión del plano de cámara en videojuegos y entornos virtuales. En Proceedings del Segundo Congreso Internacional de Matemáticas en la Ingeniería y la Arquitectura, páginas 219-230.

2. León C, Carrillo de Albornoz J, Gervás P. 2008. A Framework for Building Creative Objects From Heterogeneous Generation Systems. En Proceedings of the 5th International Joint Workshop on Computational Creativity. Madrid, España.

3. Carrillo de Albornoz J, Gervás P. 2009. Automated Camera Management Based on Emotional Analysis of the Film Script. En Proceedings of the Computer Graphics, Visualization, Computer Vision and Image Processing (IADIS Intenational Conference), páginas 41-48. Algarve, Portugal.

A.3. Aplicación de procesos de análisis de textos a otros dominios

1. Plaza L, Carrillo de Albornoz J, Prados J. 2010. Sistemas de Acceso Inteligente a la Información Biomédica: una Revisión. En Revista Internacional de Ciencias Podológicas, volumen 4, número 1, páginas 7-15.

A.4. Aplicación de otras técnicas de inteligencia artificial

1. Gómez-Albarrán M, Bautista S, Carrillo de Albornoz J. 2008. Personalized Access and Students'Coauthoring in Repositories of Learning Objects: The Case of a Repository of Programming Examples. En Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies.

Combining Linguistic and Semantic Information in an Emotion-based Model for Polarity and Intensity Classification

PhD Dissertation

Submitted by
Jorge Carrillo de Albornoz Cuadrado
in partial fulfillment of the requirements for the degree of Doctor of Computer Science
Department of Software Engineering and Artificial Intelligence
Universidad Complutense de Madrid
Madrid, September 2011

Abstract

The present work describes a novel approach to text polarity and intensity classification based on semantic analysis and linguistic rules. The aim is to determine whether a sentence or document expresses a positive, negative or neutral sentiment, as well as its intensity. The method performs word sense disambiguation in order to work with concepts rather than terms, and makes use of the knowledge in an affective lexicon, *SentiSense*, to label these concepts with emotional categories.

Negation, quantification and modality are common linguistic constructions that may modify the emotional meaning of the text affected by them, and therefore need to be taken into consideration in sentiment analysis. Negation is usually considered as a simple polarity shifter, while quantifiers are regarded as amplifiers or diminishers of the strength of such polarity. Modality, in turn, has only been addressed in a very naive fashion, so that modal forms are believed to neutralize the polarity of the terms within their scope. However, processing these constructions as mere polarity modifiers may be adequate for polarity classification, but it is clearly not enough for more complex sentiment analysis tasks (e.g., intensity classification), for which a more fine-grained model based on emotions instead of polar expressions is needed.

In this work, we also study the effect of modifiers on the emotions affected by them, and propose a model of negation, quantification and modality especially conceived for sentiment analysis tasks. We compare our emotion-based strategy with the traditional approach based on polar expressions and found that representing the text as a set of emotions increases accuracy in different classification tasks, and that this representation allows a more accurate modeling of modifiers that results in further classification improvements. We also study the most common uses of modifiers in affective texts and quantify their impact in two different sentiment analysis tasks, polarity

recognition and intensity classification. Finally, we analyze the combined effect of emotional modifiers and found that there exist interesting synergies between them.

An extensive evaluation on texts from different domains (news headlines and various product reviews) is performed in order to determine the feasibility of our approach in different polarity and intensity classification tasks. The effect of word ambiguity and the presence of negations, quantifiers and modality in opinionated text is also analyzed and quantified. The results obtained compare favorably with those achieved by other systems addressing similar evaluations. They also demonstrate that using word sense disambiguation techniques and dealing with emotional modifiers may considerably improve the performance of sentiment analysis related systems.

Index

1. Introduction **1**

1.1. Motivation 1

1.2. Sentiment Analysis: Stating the Problem 3

1.3. Main Contributions and Goals 8

1.4. Thesis Road Map 9

2. Summary of Previous Work **11**

2.1. The Emotion 11

2.2. Sentiment Analysis 12

2.2.1. The Origins of Sentiment Analysis 12

2.2.2. Subjectivity Classification 13

2.2.3. Polarity and Intensity Classification 14

2.2.4. Corpora and Lexicons 16

2.3. Negation in Sentiment Analysis 18

2.4. Quantifiers in Sentiment Analysis 20

2.5. Modality in Sentiment Analysis 22

3. An Emotion-based Model for Polarity and Intensity Classification **25**

3.1. Pre-processing 26

3.2. Concept Identification 26

3.3. Emotion Identification 27

3.4. Solving Negation 29

3.5. Solving Quantifiers 33

3.6. Solving Modality 35

3.7. Classification 39

4. Evaluation **43**
4.1. Evaluation Methodology 43
4.1.1. Evaluation Metrics 43
4.1.2. Evaluation Collections 43
4.1.3. Baselines 46
4.1.4. Machine Learning Methods 48
4.1.5. Experiments 48
4.2. Evaluating Polarity Detection 49
4.2.1. Sentence Level Polarity Detection 49
4.2.2. Document Level Polarity Detection 50
4.3. Evaluating Intensity Classification 50
4.3.1. Sentence Level Intensity Classification 51
4.3.2. Document Level Intensity Classification 52
4.4. Evaluating the Effect of Emotional Modifiers 52
4.4.1. Negation 54
4.4.2. Quantifiers 56
4.4.3. Modality 57
4.4.4. All Modifiers 58
4.5. Evaluating the Effect of Word Ambiguity 60

5. Discussion **61**
5.1. Polarity Detection 61
5.2. Intensity Classification 63
5.3. The Effect of Emotional Modifiers 65
5.4. The Effect of Word Ambiguity 69

6. Conclusions and Future Work **71**

References **75**

A. Publications **85**
A.1. Polarity and Intensity Classification 85
A.2. Practical Applications of Sentiment Analysis 85
A.3. Other Natural Language Processing Applications 86
A.4. Other Artificial Intelligence Applications 86

Chapter 1

Introduction

1.1. Motivation

Consumers´ opinions provide valuable information for companies. First, they help to understand how their products and services are perceived. They yield clues about customers´ satisfaction and expectations that can be used to determine their current and future needs and preferences. Second, they may be of help in understanding what product dimensions or attributes are important to each consumer segment, and even to discover new segments based on groups of consumers concerned about the same attributes. Third, customers´ opinions on the products offered by competitors provide essential information to accomplish a successful competitor analysis.

However, seeking consumers´ opinions has traditionally been a difficult and time-consuming task, which usually involves hiring expensive consulting firms to conduct surveys on representative groups of customers. The results of these surveys are restricted to a very limited sample of customers and need to be extrapolated from the sample to the population. Moreover, these surveys are only valid for a short time period, since customers´ opinions change fast nowadays, and frequently due to things that are completely out of the companies´ control.

With the growth of the WWW and the popularization of review sites, forums, blogs and discussion groups, consumer's information is not limited to the word of mouth, since they may easily share their experiences and opinions, positive or negative, regarding any product or service, so that the cost of gathering information for the purchase decision process is low. In fact, according to two surveys (Hitlin and Rainie, 2004; ComScore, 2007),

81 % of Internet users in USA have done online research on a product at least once; and, among readers of online reviews of restaurants, hotels, and various services (e.g., travel agencies or doctors), between 73 % and 87 % report that reviews had a significant influence on their purchase (Pang and Lee, 2008). Thus, it seems clear that the influence of online reviews on the purchase decision and the companies´ reputation should not be neglected.

However, the huge amount of online reviews that are generated every day makes any kind of manual processing unfeasible. The challenge is, therefore, the development of new technologies to efficiently process this information and produce useful knowledge for the company. But, as stated by Liu (2010), monitoring opinions on the Web is a very difficult task and the understanding and knowledge of the problem and its solution are still limited.

The information in online reviews and forums (commonly referred to as *user-generated content*) is usually presented as non-structured free-text so that automatically extracting users´ opinions requires using advanced Natural Language Processing (NLP) techniques. In particular, a discipline within the NLP has emerged to deal with the computational treatment of opinions, sentiments and subjectivity in text, called *Sentiment Analysis* and sometimes referred to as *Opinion Mining*. Sentiment analysis, in turn, involves different research tasks, such as subjectivity detection, polarity recognition, emotional intensity classification, opinion reason mining, product comparison and opinion summarization. Subjectivity detection aims to discover subjective or neutral terms, phrases or sentences, and it is frequently used as a previous step in polarity and intensity classification with the aim of separating subjective information from opinionated one (Wiebe, Bruce, and O'Hara, 1999; Pang and Lee, 2004; Kim and Hovy, 2004). Polarity recognition attempts to classify texts into positive or negative (Pang, Lee, and Vaithyanathan, 2002; Turney, 2002; Esuli and Sebastiani, 2006a). The emotional intensity classification (or rating inference) task goes a step further and tries to identify different degrees of positivity and negativity, e.g. *strongly-negative*, *negative*, *fair*, *positive* and *strongly-positive* (Esuli and Sebastiani, 2006a; Brooke, 2009; Wilson, Wiebe, and Hoffmann, 2009). Opinion reason mining aims to discover the different features or topics that are evaluated in an opinionated text, as well as the text containing the user's judgment on such features (Carenini, Ng, and Pauls, 2006; Kim and Hovy, 2006). Product comparison aims to compare users´ opinions of different

products (typically, competitive products) (Liu, Hu, and Cheng, 2005; Jindal and Liu, 2006). Finally, opinion summarization focuses on generating summaries of opinionated text by compressing the main opinions expressed in it (Carenini and Cheung, 2008; Balahur et al., 2009).

1.2. Sentiment Analysis: Stating the Problem

The problem of automatically extracting and rating opinions from text is not trivial, as the research community has noticed in the last years. Focusing on product and service reviews, many challenging NLP issues need to be addressed. We use the next example extracted from a phone review[1] to illustrate some of them:

```
(1) We bought this phone to replace a cordless phone we
had for about 9 years.
(2) The phone we replaced was very cheap but worked
perfectly for us until I dropped it into a bucket of water.
(3) We like Panasonic products so decided to give
this phone a try.
(4) The sound quality wasn't good at all for us.
```

Figure 1.1: Text fragment from a phone review

In order to analyze a review, one must first identify the sentences where the opinions, beliefs and sentiments of the user are expressed. That is, one must distinguish between two main groups of textual information: *facts* and *subjective expressions*. Facts are textual fragments where the user expresses objective statements about the topics of the text, while subjective expressions denote the user feelings and sentiments about those topics. Subjective expressions can be presented in many forms, such as desires, opinions, beliefs or speculations, making the task a great challenge. For instance, in the example in figure 1.1 the sentences (1) and (2) are good representations of facts while the sentences (3) and (4) state the user opinions about different product features. To properly identify subjective expressions, it is necessary to capture the semantic of the words and their orientation (i.e., objective vs. subjective). To do this, special lexicons must be used where each word is labeled according to its orientation. However, words usually have multiple senses so that a word can act as subjective or objective within a sentence

[1] SFU review corpus. http://www.sfu.ca/m̃taboada/research/SFU_Review_Corpus.html

depending on its context. The example in figure 1.2 clearly illustrates this problem. The word *disaster* in the first sentence entails a subjective opinion or experience about the service of a hotel, while the same word in the second sentence has an objective orientation. Therefore, to correctly determine the orientation of sentences, it is necessary to disambiguate the meaning of its words using Word Sense Disambiguation (WSD) techniques (Ide and Véronis, 1998; Agirre and Edmonds, 2006).

```
(1) The hotel service is a disaster.
(2) The disaster has caused 10 victims.
```

Figure 1.2: Different orientation of the word *disaster* depending on the context

Moreover, even for humans classifying sentences into facts or subjective expressions may be difficult. Consider, for instance, the fragment of a review extracted from *booking.com*[2] shown in the figure 1.3. In this review, the sentences (1), (5), (6), (7) and (10) express user feelings, while the sentences (2), (3), (4), (8) and (9) express facts about the hotel. In spite of this, these factual sentences should not be considered as neutral, since they express an implicit opinion about the hotel, such as its good location in front of the Eiffel Tower or the lack of free breakfast.

```
(1) The view of Eiffel Tower was very good.
(2) We were able to see it every day and
night from our room. (3) Cooking facilities
are sufficient with Microwave and hobs.
(4) Cutleries are provided too. (5) Everything
is quite clean. (6) Sofa bed is comfortable.
(7) Location is quite far from shopping centers
metro stations. (8) No breakfast included.
(9)  Kettle did not work. (10) View of Eiffel
Tower can not be guaranteed so you have to try your luck.
```

Figure 1.3: Text fragment from a hotel review

Focusing on the subjective sentences, it may be observed from the example in figure 1.1 that the sentence (3) denotes a positive point of view of the reviewer about the Panasonic products, while the sentence (4) clearly expresses a negative opinion about the phone. In order to determine the

[2]HotelReview corpus. http://nil.fdi.ucm.es/index.php?q=node/456

polarity of an opinionated sentence, i.e. to classify it into positive or negative, the process followed is similar to that explained for the subjectivity problem. Determining polarity might seem an easy task, as many words have some polarity by themselves. However, words do not always express the same feeling, and in most cases the polarity of a word depends on the context in which the word is used. So, terms that clearly denote negative feelings or emotions can be neutral, or even positive, depending on their context. The degree or strength of polarity is also an interesting point to be notice. For sure, any opinionated sentence can be classified into positive or negative, but it is clear that not all sentences express the same negative or positive intensity. For example, the sentence *This phone is AWFUL* is clearly more negative than the sentence *The sound quality wasn´t good at all for u.*

On the other hand, when evaluating an opinion, it is also important to identify the target or object of the opinion, since it is not the same, for instance, a sentence expressing a negative opinion about a room of a hotel than another complaining about the color of the wall in the hotel reception. Three main questions need to be solved when evaluating the opinion target and its effect on polarity: first, how to determine which topics or features are important for users when evaluating a certain type of products; second, which sentences in the review talk about such topics or features; and third, how to adjust the polarity of those sentences according to the features to which they refer. However, determining what the focus of a sentence is may be difficult, since frequently the feature or topic the sentence is talking about is not clear. For instance, sentence (2) in the example in figure 1.4 expresses an opinion about the hotel staff. However, the staff is not explicitly referenced in the sentence, so that it is difficult to automatically determine that "everybody" refers to the hotel staff. This is usually regarded as a kind of co-reference problem that ought to be faced using Anaphora and Co-reference Resolution techniques (Elango, 2005; Nicolov, Salvetti, and Ivanova, 2008).

```
(1) Staff is nice except for one receptionist
(a man) at night, he was not helpful at all,
I asked him for directions and he said there's
nothing I can do if you don't know Paris.
(2) Anyways, everybody else was nice.
```

Figure 1.4: Text fragment from a hotel review

Although the analysis of opinions in sentences is an interesting problem, the main goal is usually to rank the entire review or document into different grades or strengths of polarity (e.g., strongly-positive, positive, fair, negative and strongly-negative). Reviews are usually a combination of positive and negative opinions, so that the problem becomes more difficult as it is necessary to take into account the weight of each sentence in the whole review, the topics or features to which each sentence refers and their influence on the overall user opinion. The discourse structure is also important when working at the document level, since it can influence the final rating of the text. For instance, the example in figure 1.5 starts with two positive sentences about the hotel and ends with a negative one that changes the overall polarity of the review. This is a typical discourse structure in debates and reviews where the introduction makes some concessions and the conclusion ends with a statement expressing the opposite point of view.

```
(1) The hotel is in a nice location. (2) Also, the staff
is nice and polite. (3) However the rooms were horrible
and the building is more than old, so definitely I will
never stay again.
```

Figure 1.5: The importance of discourse structure when analyzing opinionated texts

There are other important linguistic factors to take into account when automatically processing opinions, such as the presence of negations, quantifiers, modals and other more complex linguistic phenomena such as irony or sarcasm. The negation may affect the subject of the sentence (***No one*** *would like to do this*), the action (*Beckham does* ***not want to play again for Real***), or only a word (***no reason***), making the analysis of its effect problematic. When dealing with the effect of negation it is not only important to identify the negation tokens or words, but also to determine the scope of such negation (i.e., the words affected by it). For example, the negation in the sentence *Very comfortable bed and* <u>*not*</u> *too short* affects the words *too short*, while in the sentence *Kettle did* <u>*not*</u> *work* affects the word *work*. The negation can invert the polarity of the sentence affected by it, or diminish the strength of such polarity. For example, the negation in the sentence *I do* <u>*not*</u> *love you* clearly changes its polarity, while the sentence *The neighborhood is* <u>*not*</u> *bad* does not necessarily mean that it is a good neighborhood.

The use of quantifiers may also modify the polarity and intensity of a sentence, as it is not the same to say *This is a good party* than *This is a really good party*. While the negation typically inverts the emotional meanings of the words affected by it, quantifiers may amplify or diminish the intensity of such meanings and with different intensities (Quirk et al., 1985). The analysis of the quantifiers also implies the identification of quantification tokens and their scopes, using the grammatical role of the tokens (adverbs and adjective modifiers) and a quantifier list where each quantifier is assigned its own polarity and weight.

Modality is another important factor in the analysis of opinions, since it is important to distinguish between situations that could happen, situations that are happening or has happened, and situations that will happen. However, dealing with modality is not easy, as it has not always the same effect. In some cases a modal may change the polarity of a sentence like in *I should eat healthy*, or may not modify at all the polarity of the sentence like in *I would like to kill you*.

Detecting and dealing with irony and sarcasm are probably the most difficult tasks, as there are no conceptual models about how they are expressed, and only few very preliminary researches have studied these phenomena.

Finally, the frequent spelling errors and the specific vocabulary used in forums and blogs pose further problems when evaluating users´ opinions. Focusing on the English language, users are often non-native speakers, so that orthographical and grammatical errors are numerous. Other frequent errors are the omission of spaces between words, the incorrect use of punctuation marks and the incorrect ellipsis of subjects, as it can be shown in the example in figure 1.6 extracted from a hotel review from *booking.com*.

```
i thought the breakfast will be provided likealways
when u are in a hotel,unfortunately there is none
for breakfast. bed wasn.t made no free drink on
arrival no continental breakfast as advertised,
noisy broken tap in kitchen health n safety
non existent, dirty showers and toilets .
```

Figure 1.6: Ortographical and gramatical errors in product reviews

1.3. Main Contributions and Goals

In this thesis, a hybrid approach based on semantic analysis and linguistic rules to classify texts according to their polarity and intensity is presented. Traditional approaches to polarity recognition and intensity classification are usually formulated as a supervised machine learning problem. The features used as classification attributes are highly varied, and include the terms found in the review and their frequency, their part of speech, the syntactic dependencies among terms, or even the polarity of the terms (as found in an emotional lexicon). Supervised learning methods require labeled examples (i.e., product reviews with manually assigned polarities or ratings) which are widely and easily available on the web. However, these approaches usually neglect the linguistic component of the problem and the importance of semantics. Semantic-level analysis is necessary in order to solve word ambiguity. As already told, words may present different meanings according to their context, and correctly capture the meaning of each word is important in order to classify the polarity and intensity of a text. This problem may be solved by representing the text as a set of concepts instead of words, and using a word sense disambiguation algorithm to attach meanings to words. Therefore, an important contribution of this thesis is the use of a concept-based representation and different disambiguation strategies, as well as the study of the effect of word ambiguity in sentiment analysis-related tasks.

Most recent approaches to polarity and intensity classification address the problem by using the so called *polar expressions*. Polar expressions are words that contain a prior polarity. For example, *like* or *good* are positive polar expressions, while *hate* or *bad* are negative ones. In these systems, the text is translated into a final value calculated over the polar expressions which determines the polarity or intensity of the opinion. However, these works present the main drawback of assuming that polarity and intensity classification exclusively depends on the polar orientation of the words or concepts within the text, regardless of the sentiments or emotions they express. However, it is clear, for instance, that the words *cancer* and *cold*, though having a negative orientation, express different emotions: a *cancer* is usually associated with *fear* and *sadness*, while a *cold* is better associated with *displeasure* or *dislike*. This fine-grained emotional information may be necessary when classifying a text into positive or negative. Consider, for example, the sentences *Superb location, terrible room* and *Superb location,*

mediocre room from two hotel reviews. The two sentences have both positive and negative concepts, but the first one is clearly negative while the second one could be considered as positive for some potential customers. The problem above is even more evident when facing the intensity classification task. Our hypothesis is that representing the text as a set of emotions rather than polar expressions increases accuracy in polarity and intensity classification. To this end, the system makes use of an affective lexicon that attaches emotional meanings to concepts. Our aim is to test whether this emotion-based representation allows us to improve accuracy in polarity and intensity classification.

Finally, few approaches have studied the effect of more complex linguistic phenomena, such as negations, quantifiers and modality, in sentiment analysis. As already mentioned, these constructions may modify the emotional meaning of the text affected by them, and therefore need to be taken into consideration. Previous works consider negation as a simple polarity shifter that inverts the polarity of the polar expressions, while quantifiers, in turn, are considered as amplifiers or diminishers of the strength of such polarity. The few works that have dealt with modality in sentiment analysis understand modals as blockers that nullified the polarity of the polar expressions. However, we argue that representing the text as a set of emotions allows a more accurate modeling of negation, quantifiers and modality that may result in further classification improvements. In this line, another important goal of the present work is to analyze the most frequent uses of such modifiers in affective text from an emotional meaning perspective, their separate effect in polarity and intensity classification, and the interactions among them. The conclusions of this analysis are used to construct an emotional model of negation, quantification and modality. To the best of our knowledge, this is the first work to quantify the individual and combined effect of negation, quantification and modality in the context of sentiment analysis applications.

1.4. Thesis Road Map

This thesis is organized into six chapters. Each chapter is briefly discussed below.

Chapter 1. Introduction. This chapter provides a global vision of the

challenges faced in sentiment analysis. It also discusses the main contributions of this work.

Chapter 2. Summary of Previous Work. This chapter presents an overview of a selection of representative systems and approaches to sentiment analysis. It also introduces different resources used in sentiment analysis, and the few approaches that have addressed the problem of negation, quantification and modality.

Chapter 3. An Emotion-based Model for Polarity and Intensity Classification. This chapter presents the proposed approach for polarity and intensity classification based on emotional categories and that makes use of deeper linguistic analysis.

Chapter 4. Evaluation. This chapter presents the evaluation accomplished to assess the polarity and intensity classification method, and compare it to other state of art approaches.

Chapter 5. Discussion. This chapter discusses the results obtained in the evaluation.

Chapter 6. Conclusion and Future Work. This chapter summarizes the main conclusions derived from this work and provides future lines of work.

Chapter 2

Summary of Previous Work

2.1. The Emotion

Nowadays the most outstanding psychological theories on sentiment analysis are the *emotional dimensions theory* and the *emotional categories theory*. The first theory proposes the interpretation of human emotions throughout a set of emotional dimensions. This theory is based on James Russel studies (1980), where the human emotional space is presented as a circular bipolar structure that represents the pleasure and the activation. The emotional dimensions were employed in the development of the ANEW list (Affective Norms for English Words) according to the SAM standard (Self Assessment Manikin) (Bradley and Lang, 1999), where each English word is assigned a value, between 1 and 9, in each of the three dimensions proposed: pleasure, arousal and dominance. This work was supported by the idea that humans understand emotions as opposite poles.

On the other hand, the emotional categories theory understands the emotions as entities, primitive units with boundaries that can be countable (James, 1884). This idea can be traced to Rene Descartes (Anscombe and Geach, 1970), who proposed a set of primitive emotions from which the other emotions can be derived. In order to represent these categories, this theory makes use of the everyday language (i.e., *joy* and *anger* are emotional categories). The main handicap of this theory is the disagreement on the most adequate set of emotional categories. While some works argue that a small set of categories is the most suitable selection (Plutchik, 1980), other studies defend that a bigger hierarchical set of categories is necessary to enclose the rich human feelings (Parrot, 2001). This debate is also encouraged by

the specific purpose of the studies. For instance, Ekman (1992) argues that only six basic emotions (anger, disgust, fear, joy, sadness and surprise) are needed to analyze facial expressions, while Ortony et al. (1988) present a set of 22 emotions in their OCC standard to emotion synthesis in speech.

2.2. Sentiment Analysis

2.2.1. The Origins of Sentiment Analysis

The phrase *sentiment analysis* in NLP describes the task aiming to identify/classify text fragments that contain any kind of emotional or sentimental meaning. The sentiment analysis discipline is quite recent. We must go back only to the first decade of the 21st century to find the first approaches clearly focused on determining opinions, sentiments and feelings in text. However, its origins may be traced back to the last decade of the 20th century and the first works on metaphor and point of view interpretation, narrative and affect (Hearst, 1992; Sack, 1994; Wiebe, 1994; Wiebe and Bruce, 1995; Huettner and Subasic, 2000; Wiebe, Bruce, and O'Hara, 1999). During this time, the volume of published papers concerning sentiment analysis has grown exponentially, and today the discipline is one of the most prolific within the broader area of NLP. This impulse has been motivated by the good results obtained so far, due to the improvement of machine learning techniques and the availability of public resources to train and test them.

Due to the novelty of the discipline, there is still controversy on the terminology used to describe it, as well as the different tasks and applications that it involves. In fact, the first mention to sentiment analysis was presented by Turney (2002) and Pang et al. (2002) to refer to the task of classifying text according to its polarity, i.e. into positive or negative, depending on the sentiment expressed in it. This means that for a few years the phrase sentiment analysis was associated to the task of polarity recognition. However, as exposed by Pang and Lee (2008) in the main survey of the area, nowadays the term sentiment analysis is frequently used in a broader sense to cover a number of different sentiment-based tasks.

The phrase Opinion Mining was first coined by Dave and colleges (2003) when describing the perfect opinion mining system as the one capable of generating a search given a set of words and retrieving a set of query-related opinions classified into different grades of positivity and negativity. This task

had been usually associated to the areas of Information Retrieval and Web Mining, and the approaches proposed had been frequently related to the analysis of judgments on products. Nowadays, the name Opinion Mining is also used to refer to different tasks concerning the analysis of opinionated text and, as explained by Pang and Lee (2008), both Sentiment Analysis and Opinion Mining are commonly used indistinctly to refer to the entire discipline.

Finally, the reader may also find the phrase *subjectivity analysis* used as a near synonym of opinion mining and sentiment analysis to refer to the analysis of sentiment in text. This term is usually associated to the work of Janyce Wiebe (1990; 1994; 1995), who based on the idea of *privative states* defined by Quirk et al. (1985), that presents the subjectivity on the concepts of opinion, emotion, evaluation and speculation. However, subjectivity analysis has been mainly associated to the process of classifying text into objective or subjective, i.e. recognizing fragments of text that denote the writer subjectivity or opinion as opposed to those that expose facts. Wilson (2009) argues, clearly inspired by the theories of Wiebe, that sentiment analysis is a sub-task of subjectivity analysis, the later being the area that covers any task regarding the analysis of emotions and sentiments in text.

In this survey, we use the terms sentiment analysis and opinion mining interchangeably to refer to the entire discipline, while the term subjectivity analysis is considered as a sub-task of sentiment analysis and used to refer to the task of distinguishing opinionated text from objective text.

2.2.2. Subjectivity Classification

Subjectivity classification aims to investigate whether a text (typically, a sentence) reports an opinion or a fact. This leads to the distinction between two types of sentences: *objective* sentences, which express factual information, and *subjective* sentences, which express personal feelings, beliefs or opinions. Subjectivity classification is usually seen as a previous step to polarity and intensity classification, since it prevents the classifiers from considering irrelevant or even potentially misleading text. However, this task may be also used as an end in itself. Separating concrete data, such as, for instance, the weight of a digital camera or its price, from other subjective judgments, such as whether the reviewer likes or not its design or the customer services, may be of great use to companies when analyzing opinions

about their products.

Subjectivity classification is considered a difficult task, since subjectivity is quite context-sensitive, and has often proved to be more difficult than polarity classification (Mihalcea, Banea, and Wiebe, 2007). Indeed, this problem has been the focus of the 2006 Blog Track at TREC (Ounis et al., 2006). It is usually stated as a two-classes learning problem (subjective and objective) and faced using traditional supervised learning algorithms, such as Support Vector Machines (SVM) or Naïve Bayes classifiers (Wiebe, Bruce, and O'Hara, 1999; Riloff and Wiebe, 2003; Wilson, Wiebe, and Hwa, 2004). Hatzivassiloglou and Wiebe (2000), for instance, use a log-linear model to classify a sentence as subjective or not by judging from the orientation and gradability of the adjectives appearing in the sentence. Wiebe and colleges (2004) investigate and test different "clues of subjectivity", including low-frequency words, collocations, and adjectives and verbs identified using distributional similarity. Yu and Hatzivassiloglou (2003) use a Naïve Bayes classifier on a corpus consisting of Wall Street Journal to distinguish news and business articles (facts) from editorials and letters to the editor (opinions), achieving 97 % accuracy.

2.2.3. Polarity and Intensity Classification

Focusing on polarity and intensity classification, this task is usually formulated as a supervised ML problem with two classes (i.e., positive and negative), but sometimes considers a more fine-grained classification (e.g., strongly negative, negative, fair, positive and strongly positive; or equivalently, 1-5 stars). Traditional approaches consider the text as a bag of word frequencies, building a vector where the entries correspond to individual words and each word is assigned its frequency in the text or a binary value indicating its presence. Within these approaches, the most representative work is that of (Pang, Lee, and Vaithyanathan, 2002), where a comparison between three ML algorithms trained on the frequencies of positive and negative terms is presented. The authors conclude that unigram-based SVM classifiers can be efficiently used to classify movie reviews into positive and negative. More recently, Martineau and Finin (2009) improve this work by using a Delta TF-IDF function to score words, as a previous step to classify movie reviews. Others have also employed n-grams instead of individual words (Dave, Lawrence, and Pennock, 2003), or even more complex lexi-

cal features, such as phrases and information extraction patterns (Riloff, Patwardhan, and Wiebe, 2006).

Approaches based on word frequencies have the main drawback of being highly dependent on the application domain. It is obvious that such a classifier trained over a corpus of movie reviews is unlikely to obtain good results when tested on a corpus of digital cameras, since both domains use a very different vocabulary. Even restricting the use of the classifier to a single domain, its generalizability is limited. This is not an intuition, but it has been empirically demonstrated. Aue and Gamon (2005), for instance, showed that using ML methods trained over unigrams, bigrams and trigrams from one domain, barely outperforms the baseline when tested in another domain.

An alternative to word-based learning is sentiment-based learning. That is, instead of representing the text as a bag of words, the text is modeled as the set of emotions found in it. To this end, an affective lexical dictionary that attaches affective meanings to words is needed. In their simplest form, affective lexicons annotate words as negative, neutral or positive. The best known polarity lexicon is the General Inquirer (Stone et al., 1966). More complex resources annotate words within a set of emotions (e.g., *sadness*, *fear*, *happiness*, etc.). Examples of such resources are the ANEW word list (Affective Norms for English Words) (Bradley and Lang, 1999), the DAL dictionary (Whissell's Dictionary of Affect Language) (Whissell, 1989) and the LIWC Dictionary (Linguistic Inquiry and Word Count Dictionary) (Pennebaker, Francis, and Booth, 2001). This type of resources have been widely used in sentiment analysis tasks. Das and Chen (2001), for instance, use a general purpose dictionary along with a manually created lexicon specific to the finance domain to assign discrimination weights to terms based on their co-occurrence within positively vs. negatively labeled documents. Klenner et al. (2009) present a tool for polarity recognition that uses the General Inquirer to assign polarities to terms, and use 70 pattern-matching rules to predict the overall polarity for a set of product reviews and a set of textual testimonials from a depression group.

However, both the resources and the approaches above work with words instead of senses, disregarding the contextual meaning of such words, and the fact that a word may present various senses of which some of them could have different polarities. To illustrate this problem, consider, for instance, the sentences *Lucy was diagnosed with* ***cancer*** and *Lucy's zodiac*

sign is ***cancer***. The noun *cancer* in the previous sentences has two different meanings which also carry a very different polarity, but both meanings are represented by the same term, making word-based approaches unable to distinguish between them. This limitation may be overcome by using an affective lexicon based on meanings (or *senses*) instead of words, along with a word sense disambiguation algorithm (WSD) to determine the meaning of polysemous words by examining their context (see (Agirre and Edmonds, 2006) for a review on WSD techniques).

2.2.4. Corpora and Lexicons

The development of sentiment analysis applications strongly relies on machine learning techniques and dictionary based approaches. Therefore, any system that attempts to correctly identify the affective meaning of a text will need, at least, two types of NLP resources: an annotated corpus to train and test the system, and an affective lexical dictionary that attaches affective meanings to words. According to different emotional theories (James, 1884; Anscombe and Geach, 1970; Russell, 1980), a wide range of resources have been developed, each of them supported by a psychological study and most of them specific for a given task.

As far as annotated corpora are concerned, perhaps the most popular collection is the Movie Review Corpus (Pang, Lee, and Vaithyanathan, 2002). This corpus contains a set of 2000 documents extracted from a website of movie reviews, where 1000 documents are classified as positive and 1000 documents are classified as negative, according to the opinions in the text about the film being described. This was one of the first corpora proposed for polarity detection and is intended for the analysis of sentiments at the document level. However, in the last years, the introduction of more sophisticated techniques based on linguistic and semantic features has moved the research interest to the analysis at the sentence level, as many researchers defend that the sentence should be the primary unit in sentiment analysis. To deal with sentence-level analysis, Pang and Lee published the Sentence Polarity Dataset (Pang and Lee, 2005). This dataset contains all the sentences within the Movie Review Corpus automatically extracted and assigned a polarity. Brooke (2009) uses a set of corpora, also based on the analysis at the document level, composed of reviews of different products (books, phones, restaurants, etc.), where each review is classified into positive or negative.

Many others have proposed similar corpora (Macdonald and Ounis, 2006; Thomas, Pang, and Lee, 2006; Snyder and Barzilay, 2007).

Focusing on the sentence-level analysis, a well-know corpus is the one proposed for the SemEval 2007 Affective Text task (2007), where a set of 1250 sentences from news headings are manually tagged with six basic emotions (*anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise*). Each emotion is scored between 0 and 100, where 0 indicates that the emotion is not present in the headline, and 100 indicates that the maximum amount of emotion is found in the headline. This corpus also includes a valence value in the interval [-100, 100] for each sentence, which expresses the negative (-100), neutral (0) or positive (100) intensity of the sentence. Another corpus for sentence level analysis is presented by Hu and Liu (2004) and consists of reviews from five electronic products, where each sentence is manually labeled with different degrees of positivity and negativity according to the opinions expressed in it. Also, each sentence is associated with a product feature from a pre-defined list.

Finally, it is important to mention the MPQA corpus (Wiebe, Wilson, and Cardie, 2005; Wilson, Wiebe, and Hoffmann, 2009), one of the most used and cited in sentiment analysis. The corpus consists of 692 documents from different news sources, and is annotated at the phrase level with several tags such as polarity, subjectivity or attitude. The idea of using phrases instead of sentences is based on the hypothesis that the sentimental or emotional meaning in a sentence is usually represented only by a set of words, which in many cases express different polarity or meaning mixed with other words. In fact, this corpus is the textual representation of the annotation scheme proposed in (Wiebe, Wilson, and Cardie, 2005) and (Wilson, Wiebe, and Hoffmann, 2009), which can be used to develop new corpora.

As far as emotional lexical resources are concerned, these resources are usually inspired in the two psychological theories presented in section 2.1. Concerning the emotional dimensions theory, the most popular lexicons are the ANEW word list (Affective Norms for English Words) (Bradley and Lang, 1999) and the DAL dictionary (Whissellś Dictionary of Affect Language) (Whissell, 1989). The first one consists of a list of words scored within three emotional dimensions: *pleasure*, *arousal* and *dominance*, according to the SAM standard (Self Assessment Manikin) (Bradley and Lang, 1994); while the second one contains 8742 words rated by people for their *acti-*

vation, *evaluation* and *imagery*. Regarding the emotional categories theory, the LIWC Dictionary (Linguistic Inquiry and Word Count Dictionary) (Pennebaker, Francis, and Booth, 2001) provides a set of 2290 words and stems, classified in one or more categories, such as *sadness*, *negative emotion* or *overall affect*.

It is worth mentioning other lexicons that have been specially developed to assist the polarity classification task, so that they simply contain words labeled as positive or negative. The General Inquirer (Stone et al., 1966) lexicon is one of the most widely used in this task, and consists of terms manually labeled into positive and negative according to their semantic orientation. Hatzivassiloglou and McKeown (1997) also propose a list of positive and negative adjectives, where their polarity or semantic orientation has been automatically detected.

The main limitation of these lexicons is the use of words or stems instead of concepts, disregarding the meaning of the words and the context in which they are used. On the contrary, the WordNet Affect database (Strapparava and Valitutti, 2004) provides a list of 911 WordNet synsets labeled within a hierarchical set of emotional categories. SentiWordNet (Esuli and Sebastiani, 2006b) also follows the idea of labeling meanings instead of words, but in this lexicon the WordNet synsets are assigned three values representing the degree of positivity, neutrality and negativity, the sum of these three values being 1.

2.3. Negation in Sentiment Analysis

The role of negation in sentiment analysis has been broadly studied in the past few years (Na et al., 2004; Whitelaw, Garg, and Argamon, 2005; Andreevskaia and Bergler, 2007; Jia, Yu, and Meng, 2009). First approaches in sentiment analysis, as stated, considered the text as a bag of words (Pang, Lee, and Vaithyanathan, 2002), building a vector where the entries correspond to individual words and each word is assigned its frequency in the text or a binary value indicating its presence. Others have also employed n-grams instead of individual words (Dave, Lawrence, and Pennock, 2003).

These simple approaches do not contain any explicit knowledge of emotional expressions, and negation need to be modeled in a very naive fashion. Das and Chen (2001), for instance, simply attach *NOT* to words occur-

ring close to negation terms, so that in the sentence *I didn't like the movie* the term *like* is converted into *like-NOT*. Concerning the scope of negation, some assume it to be those words between the negation token and the first punctuation mark (Pang, Lee, and Vaithyanathan, 2002), while others consider a fixed number of words after the negation token (Hu and Liu, 2004). However, since not all words near a negation token are affected by it, these approaches fail to realize, for instance, that the term *best* in the sentence *No doubt this hotel is the best* is not affected by the negation. As a result, as stated by Wiegand et al. (2010), this type of representation is linguistically inaccurate, and thus, it leads to a very poor impact of negation modeling compared to plain bag of words in which negation is not considered (Pang, Lee, and Vaithyanathan, 2002).

More recent approaches represent the text as a set of *polar expressions*, which are words that contain a prior polarity. Polar expressions are commonly obtained from a polarity lexicon that annotates words or concepts as negative, neutral or positive, such as the General Inquirer (Stone et al., 1966) and SentiWordNet (Esuli and Sebastiani, 2006b). In this context, negations are considered as polarity shifters, so that negated polar expressions are regarded as unnegated polar expressions with the opposite polarity. In this way, for instance, the sentences *I liked the movie* and *I didn't like the movie* present opposite polarity.

The most popular example of this kind of approaches is the one proposed by Polanyi and Zaenen (2006). They assign positive scores to positive polar expressions and negative scores to negative polar expressions. If a polar expression falls into the scope of a negation, then the polarity score is simply inverted. For instance, if *successful* is assigned a positive valence of *+2*, then *not successful* is assigned *-2*. A very similar work may be found in (Kennedy and Inkpen, 2006). These authors conclude that taking negations into account contribute to improve performance in a document-level polarity classification task. However, the exact improvement that is achieved is not reported. Therefore, no data about the real effectiveness of such approaches is available.

Also based on the idea of negation as a polarity modifier, Wilson et al. (2005; 2009) model negation as three types of features on polar expressions: *negation features*, for instance, if a negation occurs in a fixed window of four words preceding the polar expression; *shifter features*, i.e. the presence

of different polarity shifters such as *little*; and *polarity modification features*, which are words that are not explicit negations but modify polar expressions such as *disappointed* or *hope*. These features are added to a feature set comprising bag of words and counting of polar expressions and used as the input to a supervised ML algorithm to classify polarity at the expression level. This approach has the main limitation of not detecting the scope of negation properly.

More advanced is the approach presented by Choi and Cardie (2008). Their working assumption is that the words or constituents within an expression interact with each other to yield the overall polarity of the expression. The polarity of the words and constituents are obtained from a polarity lexicon, and a set of inference rules is used to derive a polarity for the entire expression from the polarity of the constituents. These rules also restrict the scope of negation to specific constituents instead of using the scope of the entire target expression.

The approaches above have several limitations. First, all of them consider that negation always inverts the polarity of the opinion expressed, but sometimes negation also diminishes the strength of such polarity. For instance, the token *not* in the sentence *Football is not my favorite sport* does not necessarily means that the speaker hates football. Second, none of them determines the exact scope of negation, but simply approximates it. Third, they assume that polarity classification exclusively depends on the polar orientation of the words or concepts within the text, regardless of the sentiments or emotions that they express.

2.4. Quantifiers in Sentiment Analysis

Quantifiers are special terms that change the degree or intensity of the emotions affected by them. They may amplify or diminish the strength of such emotions. For example, in the sentence *The hotel is terribly good*, the phrase *terribly good* is more positive than just *good* alone. In contrast, the word *less* in the sentence *It is less funny than I was expecting* diminishes the intensity of *funny*. Frequently, quantifiers are referred to as *intensifiers/amplifiers* or *downtoners* (Whitelaw, Garg, and Argamon, 2005; Andreevskaia and Bergler, 2007; Meena and Prabhakar, 2007).

The most common way of identifying quantifiers is the use of lists of

words which play a specific grammatical role in the sentence (adverbs and adjective modifiers). These lists normally contain a fixed value for all positive words and another value for all negative words. Polanyi and Zaenen (2006), for instance, use a list of tokens to detect the presence of quantifiers and calculate their effect by adding or subtracting a "point" to/from the base valence of a term. In this way, the intensifier *very* in the phrase *very suspicious*, where *suspicious* is assigned a valence of *-2*, strengthen this valence to *-3*; while the downtoner *little* in *little efficient* weakens the force of *efficient* from *+2* to *+1*.

Similarly, Kennedy and Inkpen (2006) assign all positive terms a value of *+2*. If they are preceded by an intensifier in the same clause then they are given a value of *+3*. If they are preceded by a diminisher in the same clause then they are given a value of *+1*. Negative sentiment terms are given a value of *-2* by default and *-1* or *-3* if preceded by a diminisher or an intensifier, respectively. The enhanced version of their method uses as features bigrams containing a combination of intensifier or diminisher with another feature word, so that, for instance, the bigram *very good* is represented by the feature *int_good*, where *int* indicates any intensifier; while the bigram *barely good* is represented by the feature *dim_good*, where *dim* indicates any diminisher.

Wilson, Wiebe, and Hoffmann (2009) model quantifiers as two binary features. The *preceded by intensifier* feature is true if the preceding word is an intensifier. The *self intensifier* feature is true if the clue instance itself is an intensifier. A word is considered to be an intensifier if it appears in a list of intensifiers and if it precedes a word of the appropriate part of speech.

However, all these works present an important limitation: they assign a fixed value to all intensifiers and diminishers, but it is clear that not all quantifiers amplify or diminish emotions to the same extent. For instance, the sentence *The movie is terribly funny* is undoubtedly more positive than the sentence *The movie is quite funny*, and this is due to the different intensities of the amplifiers *terribly* and *quite*. The same occurs with the downtoners *hardly* and *almost* in the sentences *I hardly love you* and *I almost love you*. Therefore, it seems that each quantifier should be assigned a value based on its intensity.

2.5. Modality in Sentiment Analysis

Modal operators allow to make a distinction between events or situations that are asserted to have happened, are happening or will happen (*realis* events) and those which might, could, should, ought to, or possibly occurred (*irrealis* events) (Polanyi and Zaenen, 2006). Different definitions and categorizations of modality have been proposed in the literature, each of them based on different aspects such as semantic, linguistic or philosophy. For example, von Fintel (2006) defines modality in relation to the expression of possibility and necessity, classifying it according to its meaning in *alethic modality*, *epistemic modality*, *deontic modality*, *bouletic modality*, *circumstantial modality* and *teleological modality*. Palmer (2001) proposes a classification of modality based on the speaker's point of view and distinguishes between *propositional modality*, categorized in *epistemic* and *evidential*, and *event modality*, categorized in *deontic* and *dynamic*. Finally, Baker et al. (2010) classify modality based on the indications perceived by the speaker or listener in *factivity*, *evidentially* and *sentiment*.

Research of modality in computational linguistics has been usually associated, and even compared, to the study of speculation or uncertainty in texts. This has been motivated by the studies about certainty in biomedical texts (Light, Qiu, and Srinivasan, 2004; Velupillai, 2010; Goldstein and Uzuner, 2010), where the main goal is to determine the fragments of texts that commit to the truth of a proposition or not. In fact, modality has been mainly interpreted from this perspective and broadly denoted as *hedging* ("words whose job is to make things more or less fuzzy" (Lakoff, 1973)) or *factuality* ("information conveying whether the events mentioned in text correspond to real situations in the world or, instead, to situations of uncertain status" (Sauri and Pustejovsky, 2009)). In this line, Light et al. (2004) present a preliminary analysis of how uncertainty is expressed in bioscience papers, and estimate that about 11 % of the sentences in the abstracts can be classified as speculative. Light et al. also test how automatic classification can deal with the categorization of sentences into speculative or not, achieving a very high accuracy in all the experiments (≈90 %). Medlock and Briscoe (2007) present a system to automatically classify sentences into speculative or not speculative, where the main contribution is the precise description of the task, along with the definition of what is and what is not a hedge. The detection of speculation in bioscience papers has become

an important research issue since it is proved to be useful in the extraction of biological relations from scientific papers (Van Asch, Morante, and Daelemans, 2010).

From a linguistic point of view, modality may be defined as the attitude or point of view from the speaker towards the state of world. Carter and McCarthy (2006) highlight two main meanings of modality. The first one is to do with probability, possibility and certainty, with the assessments of potential facts and with deductions based on these assessments. The second one involves getting things done or trying to control the course of events, usually expressing obligation and whether or not something is necessary, forbidden, desirable or permitted. Carter and McCarthy expose that the most important modal forms are represented by the modal verbs, which also divide between core modal verbs (*can, could, may, might, will, shall, would, should* and *must*) and semi-modal verbs (*dare, need, ought to* and *used to*). The authors also include as modal forms other expressions such as *had better, be meant to, be obliged to*, *be supposed to* and other verbs that act as modals in some situations such as *hope, manage, suppose, seem, wish* or *want*.

Modality is claimed to be an important factor in sentiment analysis (Polanyi and Zaenen, 2006), but it has only recently attracted the interest of the research community. As the authors point out, a modal may neutralize the polarity of the terms behind its scope, as it happens in the sentence *Breakfast could be much better considering the room rate*, where the modal verb *could* neutralizes the positivity of the phrase *much better*. The uncertainty introduced by modality can also affect the emotional meaning of the text, however, as stated by Baker et al. (2010), the uncertainty generally does not change this meaning since the affective connotation is positive or negative regardless of the event is a fact or not. The sentence *I would like to kill you* clearly expresses that the action is not done, but even in this situation the negativity of the sentence is clear. In fact, from an affective analysis perspective, modality usually changes the polarity and intensity of the text when the modal form is used to express desires, needs or complains, which is close to the second meaning of modality described by Carter and McCarthy (2006).

Very few approaches have dealt with modality in sentiment analysis systems. A pioneer study of modality in the context of a subjective analysis

task may be found in (Wiebe, Bruce, and O'Hara, 1999), which were implemented in (Wilson, Wiebe, and Hoffmann, 2009). In this work, modality is modeled as a binary feature that indicates the mere presence of modal forms in the text, and added to the feature set used as input to the ML algorithm. Brooke (2009) also analyzes the use of modality in opinionated text, but the study focuses on determining the fragments of text that contain irrealis terms in order to block the polarity values assigned to other terms after them. However, to the author's knowledge, no previous work has studied the actual impact of modality in sentiment analysis tasks, nor the effect of the different modal forms on the emotions and sentiments within their scopes has been previously examined.

Chapter 3

An Emotion-based Model for Polarity and Intensity Classification

In this chapter, we present an emotion-driven system for polarity and intensity classification. The system is based on supervised learning and makes use of semantic analysis and complex linguistic rules. The method performs word sense disambiguation in order to work with concepts rather than terms, and uses the knowledge in an affective lexicon to label these concepts with emotional categories. In this way, the input text is mapped into a set of emotional categories which represent the affective meaning of the text.

As several previous works have noted, the presence of negation, quantifiers and modality may affect the polarity and intensity of a text. The negation may change or diminish the polarity of a sentence, while the quantifiers may amplify or diminish the intensity. Modality may change, diminish or even not affect the polarity and intensity of the sentence. Thus, once the emotional categories have been identified, the system deals with the effect of such modifiers on the emotional categories under their scope. Finally, the input is translated into a vector of emotional categories which will be the input of a machine learning classifier.

The system accomplishes the process through seven steps: (1) pre-processing, (2) concept identification, (3) emotion identification, (4) solving negation, (5) solving quantifiers, (6) solving modality and (7) classification. Each of this steps is explained in detail in the following sections.

3.1. Pre-processing

In the first step, the system tags each word with its part of speech, splits the text into sentences and generates the syntactic tree for each sentence. To this aim, the GATE[1] architecture and the Stanford Parser[2] are used. In particular, the Annie English Tokeniser, Hash Gazetter, RegEx Sentence Splitter and the Stanford Parser modules in GATE are used to analyze the input.

As usual with NLP systems, this pre-processing step is accomplished in order to collect all the linguistic information that will be needed in the following steps. The part of speech and the grammatical category of words are needed to disambiguate and map them to concepts from the lexical database. The syntax tree is needed to detect and solve the effect of negation, quantifiers and modality.

3.2. Concept Identification

Next, all the terms found in the text are mapped onto their adequate senses in WordNet (Miller, 1995). To this end, the *Lesk* (Lesk, 1986) word sense disambiguation algorithm, as implemented in the WordNet::Sense-Relate perl package (Patwardhan, Banerjee, and Pedersen, 2005), is used. The disambiguation is carried out only over the words belonging to the grammatical categories of noun, verb, adjective and adverb, as only these categories can present an emotional meaning. As a result, we get the stem and sense in WordNet of each word, and this information is used to retrieve its synset. Besides, to enrich the emotion identification step, the hypernyms of each concept are also retrieved from WordNet.

The benefits of working with concepts instead of terms can be seen in the sentence *Test to predict breast cancer relapse is approved.* The concept *cancer* presents five senses in WordNet, and only one of them is related to a "malignant growth or tumor", while the others refer to "astrology" and the "cancer zodiacal constellation". Obviously, without a WSD algorithm, the wrong synset may be considered, and a wrong emotion may be assigned to the concept.

[1] General Architecture for Textual Engineering. http://gate.ac.uk/
[2] Stanford Parser. http://nlp.stanford.edu/software/lex-parser.shtml

3.3. Emotion Identification

The aim of the emotion identification step is to map the WordNet concepts previously identified to concepts in an affective lexicon, as well as to retrieve from this lexicon the emotional meaning of each concept.

Therefore, we need a lexicon that attaches emotional meanings to concepts from WordNet. To this aim, we have developed the SentiSense affective lexicon[3]. SentiSense classifies 2214 WordNet synsets in 14 emotional categories (see table 3.1). SentiSense also includes an antonym relationship between categories. Only relationships between emotional categories with a strongly opposite meaning are created, such as *like-disgust* and *joy-sadness*, as shown in table 3.1. The purpose of this antonym relationship is two-fold: first, it contributes to handle negations and modal forms; and second, it can be used to automatically expand the affective lexicon.

Category	**Antonym**	**Category**	**Antonym**
joy	sadness	hope	-
fear	calmness	compassion	-
surprise	-	ambiguous	-
anger	calmness	sadness	joy
disgust	like	calmness	-
love	hate	like	disgust
expectancy	-	hate	love

Table 3.1: Emotional categories in SentiSense and antonym relation among them

SentiSense has been developed semi-automatically in a two-phases process as in (Strapparava and Valitutti, 2004): first, 1200 WordNet synsets were manually labeled with an emotional category; second, these synsets were automatically extended with their synonyms, antonyms and derived adjectives using the corresponding relations in WordNet. Table 3.2 shows the number of synsets per emotional category and grammatical role.

As shown in table 3.3, each entry contains the WordNet synset identifier, its part of speech, its gloss or definition in WordNet and the emotional category assigned to it.

We hypothesize that the hypernyms of a concept entail the same emotions than the concept itself, but the intensities of such emotions decrease as we move up the hierarchy (i.e., the more general the hypernym becomes,

[3]Note that any other affective lexicon based on WordNet synset (e.g., WordNet Affect) may be used in this step.

Category	Nouns	Verbs	Adjectives	Adverbs
like	146	70	184	25
compassion	11	3	8	2
joy	63	30	52	19
anger	23	11	24	6
hate	2	4	11	1
expectancy	74	35	46	9
love	30	4	28	6
sadness	62	30	73	25
calmness	18	6	35	10
ambiguous	24	15	20	3
fear	104	43	38	14
hope	32	14	14	6
surprise	13	9	15	3
disgust	222	134	254	51
Total	824	408	802	180

Table 3.2: Distribution of synsets among different emotional and grammatical categories in SentiSense

```
<SentiSenseCorpus>
    <Concept synset="SID-00152712-A" pos="adjective"
    gloss="lacking cordiality..." emotion="disgust"/>
    <Concept synset="SID-00050667-R" pos="adverb"
    gloss="in a joyous manner..." emotion="joy"/>
    <Concept synset="SID-03430539-N" pos="noun"
    gloss="a weapon that discharges..." emotion="fear"/>
    <Concept synset="SID-02571914-V" pos="verb"
    gloss="come upon or..." emotion="surprise"/>
</SentiSenseCorpus>
```

Table 3.3: An extract of the SentiSense emotional lexicon (in XML format)

the less its emotional intensity is). Following this hypothesis, when no entry is found for a given concept in the affective lexicon, the emotional category associated to its nearest hypernym, if any, is used to label the concept. However, only a certain level of hypernymy is accepted, since an excessive generalization introduces some noise in the emotion identification. This parameter has been empirically set to 3. Previous experiments have shown that, upper this level, the working hypothesis becomes unreliable.

Several reasons led us to develop SentiSense, instead of using an existing resource (e.g., WordNet Affect or SentiWordNet). First, there is an issue with the granularity of representations of emotional categories. We consider the set of emotional categories in WordNet Affect to be excessively broad for our purpose, while the categories in SentiWordNet are excessively coarse-grained (i.e., positive, negative and neutral). In contrast, we consider the categorization in SentiSense to be suitable for our purposes while it is well-supported by most accepted emotional theories, such as the proposed by

(Plutchik, 1980), (Ortony and Turner, 1990), (Ekman, 1992) and (Parrot, 2001). Second, there is an issue of labeling ambiguity. We have detected a good number of synsets in WordNet Affect (113 out of 911) that have been labeled more than once, and with different emotional categories (e.g., the synset "a#00117872 *angered, enraged, furious, infuriated, maddened*" is labeled with three different categories: *anger*, *fury* and *infuriation*), making it difficult to discern which of them is more appropriate in each situation.

3.4. Solving Negation

Once the concepts are tagged with their emotional categories, the system deals with negated forms. Negation is commonly studied in sentiment analysis, and other NLP disciplines, and it may affect both polarity and intensity classification. In particular, negation may change the polarity of the text affected by it, or just diminish the strength of such polarity.

In this work, we focus on *explicit negations* (Choi and Cardie, 2008) (i.e., those introduced by negation tokens such as *not*, *never*). In contrast, we consider the so called *soft negators*, such as the adverb *rarely* and the adjective *less*, to act as quantifiers.

In order to detect the presence of explicit negations, a list of negation cues or signals is needed. We use a list of negation tokens extracted from different previous works (Councill, McDonald, and Velikovich, 2010; Morante, 2010). The list is presented in table 3.4 and includes common spelling errors such as the omission of apostrophes.

No	None	Non	Nor	Nothing	Neither	Nobody
Never	Nowhere	Not	n't	Don't	Dont	Doesn't
Doesnt	Didn't	Didnt	Haven't	Havent	Can't	Cant
Cannot	Couldn't	Couldnt	Needn't	Neednt	Wasn't	Wasnt
Weren't	Werent	Daren't	Darent	Hadn't	Hadnt	Isn't
Isnt	Aren't	Arent	Oughtn't	Oughtnt	Wouldn't	Wouldnt
Hasn't	Hasnt	Mightn't	Mightnt	Mustn't	Mustnt	Shouldn't
Shouldnt	Won't	Wont	Shan't	Shant		

Table 3.4: List of negation signals

However, it must be noted that some negation words may be also used in other expressions without constituting a negation of the opinion/emotion expressed, as in the sentence *Not only the room was dirty but the staff was very rude*. Therefore, when the negation token belongs to such expressions,

this is not processed as a negation. Examples of this type of expressions are *not only*, *not just*, *not merely*, *not even*, *not to mention*, *nothing if not* and *no wonder*.

The list of negation signals serves to detect the presence of explicit negations, but does not help to delimit the scope of such negations. The scope of a negation is determined by using the syntax tree of the sentence in which the negation arises, as generated by the Stanford Parser. To this end, we find in the syntax tree the first common ancestor that encloses the negation token and the word immediately after it, and assume all descendant leaf nodes to the right of the negation token to be affected by it. This process may be seen in figure 3.1, where the syntax tree for the sentence *Reception staff had limited English for an international hotel and were* <u>*not*</u> *very helpful* is shown. In this sentence, the method identifies the negation token *not* and assumes its scope to be all descendant leaf nodes of the common ancestor of the words *not* and *very* (i.e., *very helpful*).

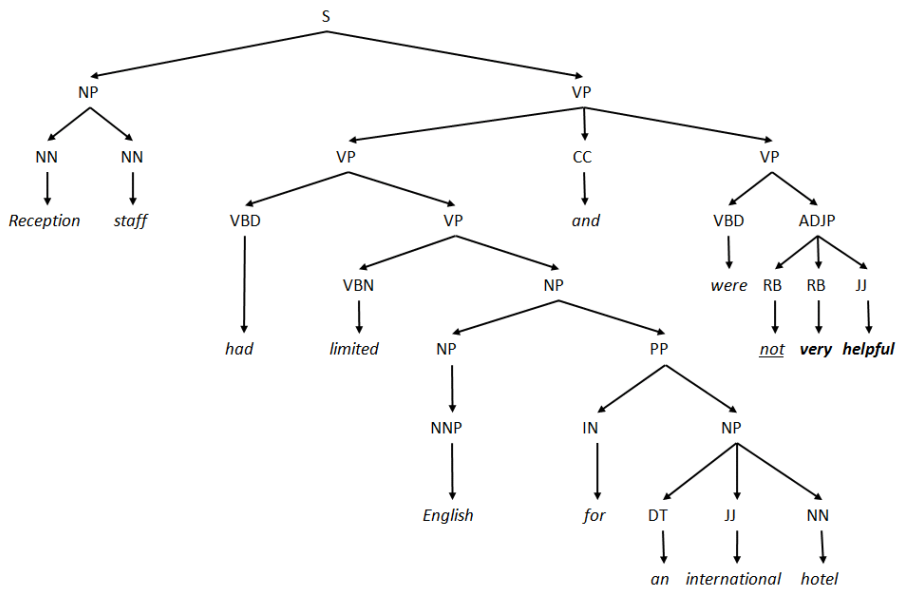


Figure 3.1: Syntax tree for the sentence *Reception staff had limited English for an international hotel and were not very helpful*. The negation signal *not* is underlined. The terms within the negation scope are shown in bold type

However, assuming all descendant of the first common ancestor as the negation scope may produce errors in special cases such as subordinate sentences or those with special punctuation marks. In order to avoid this, the negation scope detection method includes a set of rules to delimit the scope

in those cases. These rules are based on specific tokens which usually mark the beginning of a different clause (i.e., *because*, *until*, *why*, *which*, etc.). For example, in the sentence *I didn't need any transportation because the hotel place is good* the method detects as candidate negation scope the text *need any transportation because the hotel place is good*, but since the token *because* introduces a subordinate clause which will be out of the negation scope, the appropriate rule is fired and the scope is restricted to the words *need any transportation*. The syntax tree for this sentence is shown in figure 3.2. However, since some of these delimiters are ambiguous, their part of speech tags are used to disambiguate them. These delimiters, along with their appropriate part of speech, are shown in table 3.5.

Tokens	**POS**
so, as, because, that, if until, since, unless, before	IN
while, about, than, despite	IN
who, what, whose	WP
Why, when, where	WRB
which	WDT
however	RB
that	RB
“,”, - , :, (,), !, ?	-

Table 3.5: List of negation scope delimiters

This scope modeling is based on the one proposed by Jia et al. (2009), which has proved to consistently outperform simpler methods using a fixed window size, a text span until the first occurrence of a polar expression following the negation word or the entire sentence.

Even more important than delimiting the scope of a negation is to determine the actual impact or effect of such negation on the meaning of the text affected by it. As already told, negation in sentiment analysis has been typically considered as a simple polarity shifter, so that it just inverts the polarity of the polar expressions that are found within its scope (Agrawal and Siddiqui, 2009). In this model, we propose a more sophisticated modeling of negation based on the assumption that a negation changes the emotional meaning of the concepts affected by it. Consider, for instance, the following sentences:

No deaths in the accident.

I've never been afraid of the future.

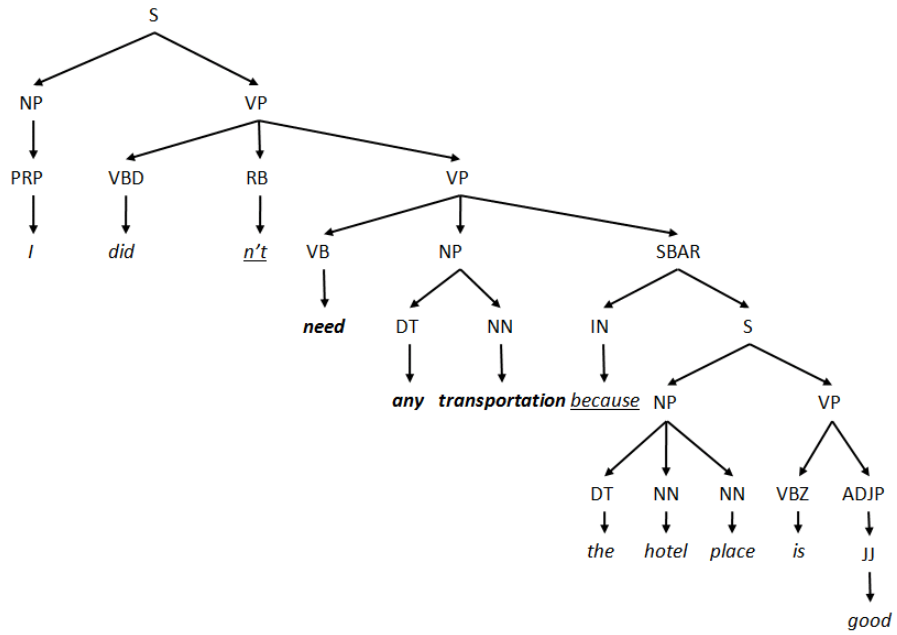


Figure 3.2: Syntax tree for the sentence *I didn't need any transportation because the hotel place is good*. The negation signal *n't* and the negation scope delimiter *because* are underlined. The terms within the negation scope are shown in bold type

On the one hand, since the term *deaths* in the first sentence denotes *sadness* and *fear*[4], the fact that there have been no deaths inspires *happiness* and *calmness*. On the other hand, since the term *afraid* denotes *fear* and falls into the scope of the negation, the second sentence expresses *calmness*. Therefore, for each emotion in the text that falls into the scope of a negation, our method changes it by its antonym. Thus, the affective lexicon must include an antonym relationship between emotions. If no antonym is found for the emotion in the lexicon (because the emotion itself has not a clear antonym), then the concept is labeled with no emotion. The strength of the antonym emotion is also diminished in a value α which needs to be empirically determined. The reason for diminishing such strength for the emotions derived from negations is that the authors consider that a negation changes the emotional meaning of a concept but usually in a lower percentage.

To finish with the study of negation, table 3.6 gives a sample of sentences with negated expressions obtained from different product reviews.

[4]We use the emotional categories in the SentiSense affective lexicon to assign emotions to the concepts in the text.

There was no way to open the window or balcony door.
None of the window switches are lit at night, same goes to power locks.
Neither the bed, nor the pillows are very comfortable.
I have nothing else good to say about this camera.
We picked a non smoking room and the floor and the room smelt badly.
If you're looking to buy a really good 5.8GHz system, then seriously consider this one if you don't mind the price.
We have never experienced any interference on the phone by the other devices.

Table 3.6: Sample of opinionated sentences with negated expressions

3.5. Solving Quantifiers

Quantifiers are special tokens that may amplify or diminish the emotional intensity of the words affected by them. The analysis of quantifiers is similar to the analysis of negation. The detection of the quantifier tokens is carried out using a quantifier list adapted from Brooke (2009), where each quantifier is assigned a percentage value based on its intensity. The tokens in the list only act as quantifiers when they play a certain grammatical role in the sentence (adverbial or adjectives modifiers), so each token is associated its corresponding grammatical role. An example of this can be shown in the sentences *The rooms where so beautiful* and *The rooms where small so we did not like them*, where the token *so* only acts as a quantifier in the first sentence. The quantifier list is shown in table 3.7. In this list, positive percentages correspond to quantifiers that behave as intensifiers (i.e., increase the intensity of the emotions affected), while negative percentages correspond to quantifiers that behave as diminishers (i.e., decrease the intensity of such emotions).

As with the scope of negation, the scope of the quantifiers is determined using the syntax tree of the sentence. To delimit the scope when the sentence presents more than one clause, the list of delimiters shown in table 3.5 is used.

Once the quantifiers are detected and their scopes are determined, all the concepts enclosed by them are marked with the corresponding quantifier percentages, which will be used to amplify/diminish the emotional intensity of such concepts. To illustrate this process, consider the sentence:

> *We also got the rate that included breakfast, which was very convenient for us and less expensive than going somewhere else.*

This sentence presents two quantifiers: *very*, that amplifies the emotions

Term	**%**	**Term**	**%**	**Term**	**%**
Very	75	Small	-50	So	75
Great	60	Really	80	More	35
Much	35	Little	-75	Bit	-35
Some	40	Less	-55	Barely	-75
Hardly	-85	Almost	-20	Slightly	-75
Relatively	-50	Somewhat	-60	Fairly	65
Pretty	75	Rather	75	Quite	35
Perfectly	75	Obviously	75	Certainly	75
Completely	85	Definitely	95	Absolutely	95
Highly	75	Truly	75	Especially	70
Particularly	45	Significantly	45	Awfully	75
Totally	85	Fully	85	Entirely	75
Strongly	55	Extremely	95	Incredibly	85
Terribly	75	Immensely	75	Such	55
Exceptionally	75	Exceedingly	85	Vastly	75
Super	75	Thoroughly	65	Tremendously	85
Ridiculously	-65	Extraordinarily	75	Total	75
Huge	75	Tremendous	85	Complete	75
Absolute	75	Incredible	75	Utter	75
Big	65	Bigger	35	Biggest	75
Extra	20	Major	35	High	55
Higher	45	Highest	65	Real	20
Minor	-75	Low	-75	Lower	-65
Lowest	-85	Few	-75	Fewer	-65
Fewest	-85	Lot	75	Difficult	-65
Only	-25				

Table 3.7: List of quantification tokens

affected by it, and *less*, which diminishes them. The scope of *very* comprises the words *convenient for us*, and as the concept *convenient* is labeled with the emotion *like*, then the concept is marked with the percentage of the quantifier (75 %). The scope of *less* encloses the word *expensive*, and as the concept *expensive* is labeled with the emotional category *disgust*, the concept is marked with the percentage of the quantifier (-65 %).

Finally, table 3.8 gives a sample of sentences with quantifiers obtained from different product reviews.

The characters are not very well described, but only vaguely outlined.
Breakfast is expensive - much better to go buy some croissants in the morning.
The steering is perfectly weighted, quick, and sharp.
All the staff were extremely helpful.
This memory is cheapest, lowest quality on the market.
Our room and bathroom were cramped and we could barely fit our suitcases.
Braking tends to feel a bit mushy at highway speeds.
It was a noisy place, we could hardly sleep.

Table 3.8: Sample of opinionated sentences with quantifiers

3.6. Solving Modality

In order to detect the presence of modality in text, we use the linguistic classification of modal forms proposed by Carter and McCarthy (2006). This classification distinguishes between three main categories of modal forms: *modal verbs*, *modal expressions* and *other verbs*. The modal verbs are, in turn, classified into *core modal verbs* and *semi-modal verbs* (see table 3.9). Starting from this classification, we conducted an initial analysis to see how modal terms are used in affective texts. To this aim, we analyzed the use of modals in various product review corpora and decided to focus on the analysis of core modal verbs, since we found that semi-modal verbs, modal expressions and other verbs with modal meaning rarely appear in the corpora, and that they usually do not have any effect on the emotional meaning of the concepts within their scope. This occurs, for instance, in the sentence *I got worse and <u>needed</u> to go back to hospital*, where the semi-modal verb *need* does not modify the emotional meaning of the sentence.

Core modals		**Semi-modals**
Can	Would	Dare
Will	Should	Need
Could	Must	Ought to
May	Might	Used to
Shall		

Table 3.9: List of modal verbs

Focusing on core modal verbs, we found three main patterns in their use which entail different effect on the emotions affected: *core modal verb*, *core modal verb + to be*, *core modal verb + present perfect (have -ed)*. Thus, we will study these three patterns separately.

The detection of the modal verbs is addressed in a similar way than with the previous modifiers, using a list of modal tokens where each token has associated its appropriate part-of-speech (i.e., verb) to avoid ambiguity errors (e.g., the use of *may* as a month). Each modal verb has associated one or more rules representing the three patterns mentioned above.

The scope of the modal verbs is delimited in the same way than the scope of negations and quantifiers. However, the effect of modality from an affective point of view is not as homogeneous as the effect of the other modifiers. It is also important to notice that developing a system able to capture the exact emotional meaning of the different modal forms and their

effect should include a high degree of pragmatic knowledge which is out of the scope of the current research in computational linguistics. This is the main reason why the few approaches that have dealt with modality in sentiment analysis consider modal forms as simple polarity blockers (Polanyi and Zaenen, 2006; Brooke, 2009).

Modality has many different interpretations, but most of them are related to the probability or uncertainty of an event or situation. This may suggest that modality could be a polarity shifter, as there is no certainty on the event or situation. However, from an emotional point of view, the uncertainty of an event does not usually affect the overall emotional meaning of a text, as it can be shown in the sentence *We would obviously stay there again when we return.*

Another important meaning of modality, as exposed by Carter and McCarthy (2006), refers to its use to express desires, needs or obligations about events that have happened, are happening or will happen. This use of modality is, perhaps, the most frequent in affective text and the one that really affects the emotional polarity and intensity of the text, as it may be observed in the sentences *It would have been nice to have another window*, *The staff at the reception could be more pleasant* or *The Hotel should have been more organized.*

Therefore, the various uses of modality determine different effects on the emotional meaning of the text:

1. **Effect I:** No effect, when the modal expresses a low degree of uncertainty.

2. **Effect II:** A decrease in the intensity of the emotions within its scope, when modality is used to express desires or a high degree of uncertainty.

3. **Effect III:** The reversal of the emotions along with the decrease of their intensity, when the modal is used to express needs, obligations or disagreements.

After a careful analysis of the use of the different core modal verbs in affective texts, we have associated each modal form to one of the possible effects above. The main conclusions of this analysis, which are summarized in table 3.10, are:

- It seems that people mainly use two core modal verbs, *could* and *should*, to express needs, obligations and disagreements. These verbs usually serve to communicate the people's perspective of how events should be, thus changing the emotional meaning of the affective text within their scope. Examples of these uses of *could* and *should* are:

  Internet could be free.

  They should offer a better quality toilet paper.

  However, it is important to note that *could* alone is more frequently used as the past tense of the verb *can* and barely expresses modality, as in the sentence *I couldn't go to your party because I was sick.*

- Desires are mostly expressed using the modal verb *would*, whose meaning is more related to the suggestion of how things should be done better, decreasing the emotional intensity of the affective terms. An example of this use of *would* is found in the sentence:

  It would be nice if there is a kettle in the room.

  However, the pattern *would + have -ed* usually denotes disagreement, so that it changes the emotional meaning of the text and diminishes its intensity, as in the sentence *Tea and coffee making facilities in the rooms would have been good.*

- The core modals *must* and *might* rarely change the emotional meaning of the text, and mainly express a little uncertainty of the events, as it may be seen in the following sentences:

  Pictures online must be a different hotel.

  There is a shopping mall next to the hotel where you can get anything you might need.

  However, when *might* is used within the pattern *might + have -ed*, it usually expresses disagreement and so it changes and diminishes the emotional meaning of the affective terms, as in the sentence *A touch of romance might have been nice for the story.*

- The verb *may* usually expresses uncertainty but in a higher degree than *must* or *might*. Thus, the intensity of the emotional meaning of the text affected by this form is decreased, as in the sentences:

  Too little sleep may mean too fat kids.

  The surroundings may be good for some people but not me.

- The analysis of the core modal verbs *can*, *will* and *shall*[5] has revealed that these verbs are generally used to express situations with a very low degree of uncertainty, and that hardly affect the emotional meaning of the text. This can be seen in the sentences:

  You can hear the toilet flushing.

  If you run into problems this hotel will do everything possible to help.

  I shall prevent the slightest shadow of partiality.

- We have also examined the semi-modal verbs *dare*, *need*, *ought to* and *used to* in order to determine if their use affect the emotional meaning of the text, and reached the same conclusion as for verbs *can*, *will* and *shall*.

We have also studied if negation may change the effect of a modal verb and found that negation does not change this effect. To illustrate this, consider the sentence *Breakfast couldn't have been better (**like**) for the price paid*. We first process the modal without taking into account the negation. The modal form *could* changes the meaning of the affective terms within the phrase *better (**like**) for the price* to their antonym emotional meanings (***disgust***), and diminishes their intensities. However, since *could* is negated, once its effect is solved the effect of negation need to be taken into account. This changes again the meaning of *better (**disgust**) for the price* to its antonym (***like***).

To end with, table 3.11 gives a sample of sentences with modal forms obtained from different product reviews.

[5]Here we refer to the use of *shall* to describe future events.

Modal form	**Effect**
Could	No effect
Could + be	Antonym emotion + diminish
Could + have -ed	Antonym emotion + diminish
Should	Antonym emotion + diminish
Should + be	Antonym emotion + diminish
Should + have -ed	Antonym emotion + diminish
Would	No effect
Would + be	Diminish
Would + have -ed	Antonym emotion + diminish
Must	No effect
Must + be	No effect
Must+ have + ed	No effect
May	Diminish
May + be	Diminish
May + have -ed	Diminish
Might	No effect
Might + be	No effect
Might + have -ed	Antonym emotion + diminish

Table 3.10: Modal verbs

We could get new towels at the desk.
The bathroom could have been cleaner.
The staff should communicate more with each other.
I would recommend my friends to stay in this hotel.
A traditional American breakfast with eggs and toast would have been nice to have.
The surroundings may be good for some people but not me.
If you pay in euros you might receive the change in Romanian currency equivalent.
This must have been the worst room in the building.

Table 3.11: Sample of opinionated sentences with modal forms

3.7. Classification

In the last step, the system uses the information gathered in the previous steps to translate the text into a *Vector of Emotional Intensities (VEI)*, which will be the input to a machine learning algorithm. The VEI is a vector of 14 positions, each of them representing one of the emotional categories of the SentiSense affective lexicon. The VEI is generated as follows:

- For each concept, C_i, labeled with an emotional category, E_j, the weight of the concept for that emotional category is set to $1,0$.

- If no emotional category was found for the concept, and it was assigned the category of its first labeled hypernym, E_{hyper_j}, then the weight of the concept for the hypernym emotional category is computed as

follows:

$$weigth(C_i, E_{hyper_j}) = \frac{1}{depth(hyper_j) + 1} \tag{3.1}$$

- If the concept is affected by a negation and the antonym emotional category, E_{anton_j}, was used to label the concept, then the weight of the concept for the antonym category is multiplied by 0,6. Different tests have been carried out to set this parameter, and the 0,6 value got the best results. As already told, the reason for using a weight lower than 1 for the emotional categories derived from negations is that negation changes the emotional meaning of a concept but usually in a lower percentage.

- If the concept is affected by a quantifier, then the weight of the concept for its corresponding emotional category is increased/decreased in the quantifier percentage, as shown in equation 3.2.

$$weigth(C_i, E_j) = weigth(C_i, E_j) \times (1 + quantifier\ percentage) \tag{3.2}$$

- If the concept is affected by a modal form, then the weight of the concept for its emotional category is calculated as follows (equation 3.3), where the values 0,5 and 0,6 were empirically set.

$$weigth(C_i, E_j) = \begin{cases} weigth(C_i, E_j) & \text{if if the modal is of type I} \\ weigth(C_i, E_j) * 0,5 & \text{if if the modal is of type II} \\ weigth(C_i, E_{anton_j}) * 0,6 & \text{if the modal is of type III} \end{cases} \tag{3.3}$$

- Finally, the position in the VEI of the emotional category assigned to the concept is set to the weight of the concept previously calculated.

The sentence *Cleaning starts very early even on Sunday, furnitures are moved, very noisy, not possible to relax and sleep*, for instance, produces the VEI *[1,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 0,0 - 2,35]* (*like, compassion, joy, anger, hate, expectancy, love, sadness, calmness, ambiguous, fear, hope, surprise, disgust*). In this sentence, the concepts *cleaning* and *relax* are labeled with the emotional category *like*, and the concept *noisy* is labeled with the emotional category *disgust*. Since the concept *relax* falls into a negation scope, its emotional category is changed to its antonym (i.e., *disgust*), and the system adds 0,6 to that VEI position. Besides, the concept *noisy* is amplified by the quantifier *very*, whose weight

is 75 %, so the *disgust* category in the VEI is increased in 0,75. Thus, the final value of the *like* VEI position is 1,0, while the final value of the *disgust* position is 2,35 (0,6 from *relax* + 1,75 from *very noisy*). As already mentioned, this VEI is used as the input to a ML algorithm that classifies the sentence according to its polarity and intensity.

Chapter 4

Evaluation

In this chapter, the evaluation methodology is presented, as well as the results of the different experiments performed to evaluate the system. These results are also compared with those obtained by two state-of-art approaches. With these experiments, we aim to evaluate the system performance in the polarity and intensity classification tasks, both at the sentence and document level. We also want to determine if this performance may be improved by dealing with emotional modifiers and solving word ambiguity.

4.1. Evaluation Methodology

4.1.1. Evaluation Metrics

All experiments are performed using 10-fold cross validation. The results reported are averages over the 10 folds. Performance is measured in terms of accuracy, recall, precision and F-measure (Baeza-Yates and Ribeiro-Neto, 1999). Accuracy is the proportion of instances correctly classified. Recall is the percentage of all instances of the class C that are correctly classified. Precision is the percentage of instances classified as class C that are class C in truth. F-measure is the harmonic mean of recall and precision.

4.1.2. Evaluation Collections

This section presents different datasets that have been used to evaluate the system, both in the polarity and intensity classification tasks. The two first collections (the SemEval 2007 dataset and the SFU corpus) are publicly available. The SemEval 2007 dataset is used to carry out the experiments

for polarity and intensity classification at the sentence level, while the SFU corpus is used to evaluate polarity classification at the document level. The third one, the HotelReview corpus, has been developed as part of this thesis and allows us to evaluate both polarity and intensity classification at the sentence and document level.

SemEval 2007: Affective Text Task

The SemEval 2007 collection for the Affective Text task[1] consists of a training set and a test set of 250 and 1000 news headlines respectively, extracted from different news sites. Each sentence is labeled with a value between -100 and 100, where -100 means a highly negative emotional intensity, 100 means highly positive and 0 means neutral.

For the polarity classification task, we map the 1000 headlines in the test set to two classes (negative = [-100, 0), positive = [0,100]), obtaining 526 headlines for the negative class and 474 headlines for the positive class. For the intensity classification task, we map the 1000 headlines to three classes as in the SemEval challenge (negative = [-100, -50), neutral = [-50, 50) and positive = [50, 100]), and to four classes (strongly negative = [-100, -50), negative = [-50, 0), positive = [0, 50) and strongly positive = [50, 100]). The 3-classes distribution results in 251 instances for the negative class, 598 for neutral and 151 for positive. The 4-classes distribution results in 251 instances for the strongly negative class, 279 for negative, 319 for positive and 151 for strongly positive.

SFU Review Corpus

The Simon Fraser University review corpus (SFU)[2] contains 25 positive and 25 negative reviews of 8 different categories of products (400 reviews in total), such as books, cars, computers, cookware, hotels, movies, music and phones, extracted from *Epinions.com*[3]. The classification into positive and negative is based on the "recommended" or "not recommended" tag that the reviewer provided. This dataset is intended for polarity classification evaluation at the document level and consists of 200 positive documents and 200 negative documents.

[1] SemEval 2007 Affective Text task. http://www.cse.unt.edu/~rada/affectivetext/
[2] SFU review corpus. http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html
[3] Epinions.com. http://www10.epinions.com/

HotelReview Corpus

The HotelReview[4] corpus is intended for the polarity and intensity classification tasks at the document level and for polarity classification at the sentence level. To create the dataset, we collected 25 reviews per hotel for 60 different hotels (1500 reviews) from *booking.com*[5]. Each review contains the following information:

- The city where the hotel is located, the reviewer nationality, the date when the review was written and the type of reviewer from a set of 7 categories, such as *solo traveler*, *young couple* and *group*.
- A score in 0-10 describing the overall opinion of the reviewer. This score is not given by the reviewer, but automatically calculated by *booking.com* from the rates assigned by the reviewer to 6 aspects: *Hotel staff*, *Services/facilities*, *Cleanliness of your room*, *Comfort*, *Value for money* and *Location*. Unfortunately, these disaggregated scores are not available in the reviews.
- A brief free-text describing, separately, what the reviewer liked and disliked during the stay in the hotel.

We have observed that the overall score assigned to a review frequently bears no relation at all with the text describing the user opinion about the hotel, so that two reviews with nearly the same score may reflect very different opinions. For instance, the two following reviews are assigned the score '6,5', but the second is clearly more negative than the first:

- *Good location. Nice roof restaurant - (I have stayed in the baglioni more than 5 times before). Maybe reshaping/redecorating the lobby.*
- *Noisy due to road traffic. The room was extremely small. Parking awkward. Shower screen was broken and there was no bulb in the bedside light.*

To overcome this drawback, we asked two annotators to assign a first category within the set *[Excellent, Good, Fair, Poor, Very poor]* and a second category within the set *[Good, Fair, Poor]* to each review based on the

[4]This collection is available for research purposes. http://nil.fdi.ucm.es/index.php?q=node/456.
[5]Booking.com. http://www.booking.com/

text describing it. To solve inter-judge disagreement, all the reviews with conflicting judgments were removed. Finally, we randomly selected 1000 reviews. The final distribution of the reviews is 200 for each class in the 5-classes categorization; and 349, 292 and 359, respectively, in the 3-classes categorization.

To evaluate the polarity classification task at the sentence level, we consider all sentences in the "like" section as positive, and all sentences in the "dislike" section as negative. As a result, we get 4340 sentences, 2187 negative and 2153 positive. Regarding polarity classification at the document level, the text in the "like" section of a review is considered as a positive document, while the text in the "dislike" section is considered as a negative document. Therefore, we get 1585 documents, 733 negative and 852 positive. An example of hotel review is shown in table 4.1.

<**HotelReview** ***idDoc***="D_8" ***hotelID***="H_2" ***hotelLocation***="Paris" ***reviewerCategory***="Young couple" ***reviewerNationality***="Belgium" ***date***="February 10, 2010" ***score***="9.3" ***5_classes_intensity***="Good"> ***3_classes_intensity***="Good">

<**PositiveOpinion**> I liked the location, breakfast was nice as well as Tea Time Buffet, that was really nice. Parking on weekends is free (on the street, and it's safe). We got to the room and it was very smelly (cigarettes) so we asked and changed and got a nice room. I'd recommend this hotel definitely. But hey... it's not a 4 star hotel...</**PositiveOpinion**>

<**NegativeOpinion**>Staff is nice except for one receptionist (a man) at night, he was not helpful at all, I asked him for directions and he said there's nothing I can do if you don't know Paris. Anyways, everybody else was nice.</**NegativeOpinion**>

</**HotelReview**>

Table 4.1: An example of hotel review from the *HotelReview* corpus

4.1.3. Baselines

Our system is compared with two state-of-art approaches. The first approach is the one proposed in (Pang, Lee, and Vaithyanathan, 2002) and is based on support vector machines (SVM) over bags of unigrams. We briefly explain it here: Let $u_1, ..., u_m$ be a set of m unigrams that can appear in a review. Let $f_i(r)$ be the number of times u_i occurs in the review r. Then, r is represented by the review vector $R = f_1(r), ..., f_m(r)$ and used as the input to the LibSVM Weka[6] classifier. The set of m unigrams is limited to unigrams appearing at least 4 times in the corpus. In spite of its simplicity,

[6]Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/

this approach turned out to be more effective than more complex approaches using features such as bigrams, part-of-speech tagging and word positions, and might actually be considered somewhat difficult to beat.

The second one is a traditional approach based on polar expressions. To this aim, we have modified the system proposed in this thesis to work with polar expressions instead of emotions. To do this, we mapped the emotional categories of the SentiSense affective lexicon to three real values, +1,0, 0,0 and -1,0, so that +1,0 represents emotions with positive connotations, 0,0 stands for neutral or ambiguous emotions, and -1,0 denotes negative emotions. This mapping is shown in table 4.2.

Category	**Value**	**Category**	**Value**
joy	+1,0	hope	+1,0
fear	-1,0	compassion	0,0
surprise	0,0	ambiguous	0,0
anger	-1,0	sadness	-1,0
disgust	-1,0	calmness	+1,0
love	+1,0	like	+1,0
expectancy	0,0	hate	-1,0

Table 4.2: Mapping of emotional categories from SentiSense to polarity values

Concerning the adaptation of the system to deal with polarity values instead of emotions, the six first steps of the algorithm (i.e., pre-processing, concept identification, emotion identification, etc.) remain unchanged, except for negation and modality processing, where the antonym emotion is not retrieved. In the classification step, each concept that was tagged with an emotional category is translated into the appropriate polarity value. If the concept was tagged with the emotion of its hypernym, then the polarity value is weighted according to equation 3.1. If the concept is affected by a negation, then its polarity value is simply inverted. If the concept falls into the scope of a quantifiers, its polarity value is weighted according to equation 3.2. Similarly, the effect of modal forms is solved according to equation 3.3, except that the polarity is inverted rather than using the antonym emotion.

Finally, the polarity values assigned to all concepts in the text are added. For the polarity classification task, if the value obtained is positive, then the text is classified as positive. If such a value is negative, then the text is classified as negative. For the intensity classification task, the added value is mapped to the interval [-100,100] and the text is classified according to the partitioning of this interval in intensity classes.

To illustrate this process, consider again the sentence *Cleaning starts very early even on Sunday, furnitures are moved, very noisy, not possible to relax and sleep*, which was mapped to the emotions {*like* (cleaning), *disgust* (noisy) and *like* (relax)}. The set of emotions is mapped to the set of polarity values {1,0 (cleaning), -1,0 (noisy) and 1,0 (relax)}. Since the concept *relax* falls into a negation scope, its polarity value is inverted and multiplied by 0,6 (-0,6). Besides, the concept *noisy* is amplified by the quantifier *very*, whose weight is 75 %, so that its polarity value is multiplied by 1,75 (-1,75). Thus, the final polarity value for the sentence is 1,0 (cleaning) + (-1,75) (noisy) + (-0,6) (relax) = -1,35. The sentences is, therefore, classified as negative.

4.1.4. Machine Learning Methods

To determine the best ML algorithm for polarity and intensity classification, several Weka classifiers were trained and evaluated using 10-fold cross validation. For the sake of briefness, we only show here the results of the three best performing classifiers: a logistic regression model (Logistic), a support vector machine (LibSVM) and a functional tree (FT). The best outcomes for these algorithms were reported when using their default parameters in Weka.

4.1.5. Experiments

We conducted different experiments in order to answer the following research questions:

1. What is the system's performance in the polarity classification task?

2. What is the system's performance in the intensity classification task?

3. Is it possible to enhance classification by properly dealing with negations, quantifiers and modality? If so, there exist synergies between the modifiers?

4. Does the presence of word ambiguity affect the system's performance and may it be improved by employing WSD techniques?

4.2. Evaluating Polarity Detection

The first group of experiments aims at evaluating the method performance in the polarity classification task, both at the sentence and document level.

4.2.1. Sentence Level Polarity Detection

We first evaluate the polarity classification task (i.e., *positive* vs. *negative*) at the sentence level, using two different corpora: the SemEval 2007 news dataset and the HotelReview corpus. To this end, the news dataset has been mapped to a -100/100 classification (-100 = [-100, 0), 100 = [0,100]). Regarding the HotelReview dataset, all sentences in the "like" section are considered positive, while all sentences in the "dislike" section are considered negative.

Table 4.3 shows the results as average accuracy of these experiments. In this table, the best results for each dataset are shown in bold. The best accuracy is achieved by the LibSVM classifier in the SemEval dataset (75,1 %), and by Logistic in the HotelReview dataset (85,0 %). However, all ML techniques behave similarly.

Method	**SemEval**		**HotelReview**	
	Acc.	F-1	Acc.	F-1
LibSVM	**75,1**	**75,0**	84,8	84,7
Logistic	74,3	74,3	**85,0**	**84,9**
FT	74,3	74,3	84,9	**84,9**
Polar expressions	73,2	74,7	74,7	78,0
Unigrams	52,2	35,9	77,6	77,6

Table 4.3: Average results for different classifiers in the sentence polarity prediction task

Concerning comparison with the baselines, when evaluating the *Unigrams* approach on the SemEval 2007 collection, the accuracy decreases sharply, obtaining an accuracy of 52,2 %. This classifier categorizes all instances as negative, so that the F-measure decrease considerably. In contrast, this same classifier behaves much better when evaluated over the HotelReview corpus (77,6 % accuracy and F-measure). Regarding the *Polar expressions* approach, the difference in accuracy for the SemEval 2007 dataset is not marked (73,2 % accuracy), being the F-measure even better than for Logistic and FT classifiers. Concerning the HotelReview corpus, the results

of the *Polar expression* approach considerably decrease (74,7 % accuracy), being similar to those obtained by the *Unigrams* approach.

4.2.2. Document Level Polarity Detection

We next test our approach in the document polarity classification task. To this purpose, we evaluate the method using two different datasets: a collection of hotel reviews (the HotelReview corpus) and a collection of reviews from different types of products (the SFU corpus). For the HotelReview corpus, the text in the "like" section of a review is considered as a positive document, while the text in the "dislike" section is considered as a negative document.

The results are shown in table 4.4. The best classifier is Logistic, with 93,6 % accuracy for the HotelReview corpus and 73,2 % accuracy for the SFU corpus. However, in this task the differences with the *Unigrams* approach decrease with respect to the sentence level classification task for both datasets (see table 4.3).

Method	**HotelReview**		**SFU**	
	Acc.	F-1	Acc.	F-1
LibSVM	93,1	93,1	62,5	60,7
Logistic	**93,6**	**93,6**	**73,2**	**73,3**
FT	93,3	93,3	71,2	71,2
Polar Expressions	88,4	89,2	62,5	68,3
Unigrams	85,6	85,5	66,7	66,7

Table 4.4: Average results for different classifiers in the document polarity prediction task

The differences with respect to the *Polar expressions* approach for the HotelReview corpus are not as marked ($\approx 5\,\%$), as for the sentence level ($\approx 10\,\%$), but are quite noticeable for the SFU corpus ($\approx 11\,\%$).

4.3. Evaluating Intensity Classification

The second group of experiments aims at evaluating the method in the intensity classification task both at the sentence and document level.

4.3.1. Sentence Level Intensity Classification

Apart from identifying polarity, we also want to examine the ability of our system to determine the emotional intensity of the sentences. To this aim, we define two intensity distributions in the SemEval 2007 corpus. For the first distribution, we map the news dataset to 3-classes: negative [-100, -50), neutral [-50, 50) and positive [50, 100]. For the second distribution, we map the dataset to 4-classes: strongly negative [-100, -50), negative [-50, 0), positive [0, 50) and strongly positive [50, 100]. We can see in table 4.5 that, as the number of intensity classes increases, the results decrease progressively, since the task becomes more difficult.

Method	**2-classes**		**3-classes**		**4-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
LibSVM	**75,1**	**75,0**	63,2	55,7	45,8	41,1
Logistic	74,3	74,3	63,7	59,3	**46,2**	**43,6**
FT	74,3	74,3	**64,1**	**59,5**	45,0	42,5
Polar Expressions	73,2	74,7	53,7	50,4	32,1	38,2
Unigrams	52,2	35,9	59,7	44,7	32,1	15,6

Table 4.5: Average results for different classifiers in the sentence intensity prediction task (SemEval 2007 dataset)

Concerning the *Unigrams* approach, it surprisingly performs better in the 3-classes task than in the 2-classes one, but still worse than the three variants of our method. Regarding the 4-classes distribution evaluation, our system also reports promising results, which are significantly better than those of the *Unigrams* approach, whose performance decreases considerably with regard to the 3-classes task.

The performance of the *Polar expressions* approach is quite similar to that of the system proposed in this thesis for the 2-classes classification task, as stated, but decrease sharply, even more than the *Unigrams* approach, for the 3-classes. The performance for the 4-classes prediction task is quite similar to that obtained for the *Unigrams* approach, which differences with respect of the better classifier for the proposed method are remarkable. This behavior suggests that the *Polar expressions* approach properly capture the polarity of the text, but is not able to correctly understand its intensity.

4.3.2. Document Level Intensity Classification

We next evaluate our approach in the document intensity prediction task. For these experiments, we use the HotelReview corpus. As shown in section 4.1.2, each document within this corpus is assigned a rate in a 5-point scale (i.e., *Excellent*, *Good*, *Fair*, *Poor*, *Very Poor*) and a rate in a 3-point scale (i.e., *Good*, *Fair*, *Poor*). Table 4.6 summarizes the average results.

Method	**2-classes**		**3-classes**		**5-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
LibSVM	93,1	93,1	68,4	68,7	45,9	44,7
Logistic	**93,6**	**93,6**	**70,8**	**70,8**	**50,0**	**48,3**
FT	93,3	93,3	68,6	68,5	46,2	45,0
Polar Expressions	88,4	89,2	47,2	55,7	29,8	35,6
Unigrams	85,6	85,5	54,4	46,5	30,9	25,4

Table 4.6: Average results for different classifiers in the document intensity prediction task (HotelReview corpus)

It may be observed that performance considerably decreases with respect to the 2-classes task when increasing the number of classes ($\approx 20\,\%$ for 3-classes and $\approx 40\,\%$ for 5-classes), but is still significantly better than that achieved by the *Unigrams* approach ($\approx 30\,\%$ for 3-classes and $\approx 50\,\%$ for 5-classes).

The performance of the *Polar expressions* approach is quite similar to that of our method but decreases sharply when increasing the number of classes ($\approx 40\,\%$ for 3-classes and $\approx 60\,\%$ for 5-classes). These results confirm our hypothesis that the *Polar expressions* approach is not able to properly capture the intensity of the text.

4.4. Evaluating the Effect of Emotional Modifiers

We pursue two primary goals with these experiments. The first is to evaluate the importance of processing emotional modifiers, (i.e., negation, quantification and modality), in polarity and intensity classification. The second is to compare our emotion-driven strategy to the *Polar expressions* approach, described in section 4.1.3, which is able to deal with these modifiers. We only show the results of the best performance classifier, *Logistic*.

To evaluate if dealing with emotional modifiers improves polarity and intensity classification, we investigate the performance of our system with

and without solving modifiers. Furthermore, we want to analyze the influence of modifiers both at the sentence level (i.e., when classifying sentences according to their polarity and intensity) and document level (i.e., when classifying documents according to these criteria). We perform these experiments over two different corpora: the SemEval 2007 collection for the sentence level experiments and the HotelReview corpus for the document level experiments.

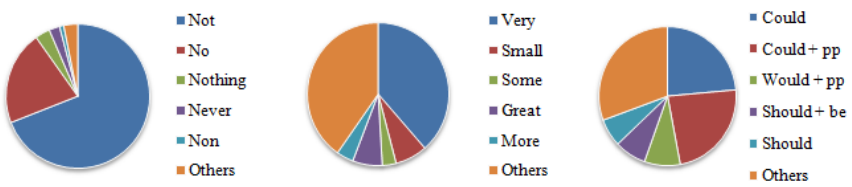
Table 4.7 shows the percentage of sentences/documents in the SemEval and HotelReview datasets that contain negations, quantifiers and modal verbs, respectively. Looking at this table, we see that the presence of emotional modifiers is significantly higher in the HotelReview dataset than in the SemEval dataset. The explanation seems to be in the different type of documents. In particular, modifiers are less frequent in news headlines than in product reviews. Given the low percentage of sentences with modifiers in the SemEval collection (around 4 %), we expect the effect of dealing with modifiers in this dataset to be negligible. In contrast, the effect of processing modifiers in the HotelReview corpus is expected to be noticeable. Regarding the relative frequency of the different modifiers, quantifiers are the most frequent in both datasets, followed by negations and, finally, modal verbs.

		Negations	**Quantifiers**	**Modals**
SemEval	Per sentence	3,64 %	4,23 %	3,24 %
HotelReview	Per sentence	20,85 %	33,36 %	6,9 %
	Per document	36,09 %	60,32 %	16,15 %

Table 4.7: Distribution of sentences/documents with negations, quantifiers and modal verbs, respectively, in the SemEval and HotelReview datasets

The charts in figures 4.1(a), 4.1(b) and 4.1(c) show the top five most frequent negations, quantifiers and modal verbs, respectively, along with their relative frequency. As expected, the most common negation token is *not* (69,1 %) followed by *no* (21,1 %). Thus, these two tokens represent more than 90 % of the negations in the corpora. The most frequent quantifier is *very*, representing over 40 % of the occurrences of all quantifiers. Concerning modal verbs, the most frequent one is *could*, either alone or accompanied by the present perfect ($\approx$ 47 % of all modal verbs).

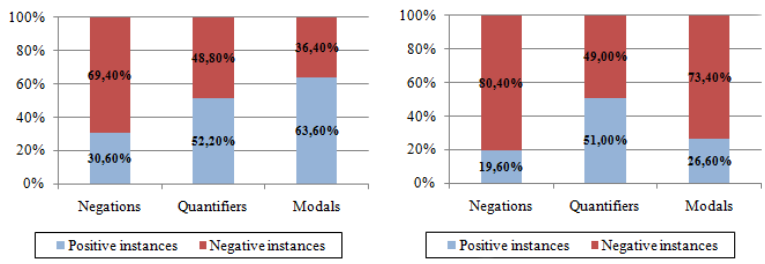
Moreover, figures 4.2(a) and 4.2(b) show the distribution of each modifier in polarity classes (positive instances and negative instances) for each dataset. The most interesting observation from these charts is that negated



(a) Negations (b) Quantifiers (c) Modal verbs

Figure 4.1: Charts showing the top five most frequent negations, quantifiers and modal verbs, respectively, along with their relative frequency in the datasets

forms are significantly more frequent in negative instances than in positive instances, and this occurs in all datasets. This seems to indicate that it is more common to negate positive emotions and opinions than negative ones. Quantifiers, in turn, are equally distributed throughout polarity classes, while the distribution of modals depends on the dataset.



(a) SemEval dataset (b) HotelReview dataset

Figure 4.2: Charts showing the distribution of modifiers across polarity classes

4.4.1. Negation

The aim of the first group of experiments is to quantify the impact of negation and the adequacy of the model proposed in two sentiment analysis tasks, polarity and intensity classification, both at the sentence and document level.

We first focus on sentence polarity and intensity classification, and evaluate the two systems (our emotion-driven approach and the *Polar expressions* approach) with and without processing negations, on the news headlines in the SemEval collection, using the same classes distributions than in the pre-

vious experiments. The results are shown in table 4.8.

Method	**Polarity**		**3-classes**		**4-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	72,8	72,8	63,0	58,4	45,3	42,9
Emotions + Negations	**73,2**	**73,1**	**63,3**	**58,7**	**46,0**	**43,3**
Polar expressions + No modifiers	72,9	74,4	**54,4**	**51,0**	**32,1**	**38,2**
Polar expressions + Negations	**73,2**	**74,7**	53,6	50,5	31,2	37,4

Table 4.8: Average accuracy and F-measure in the sentence polarity and intensity classification tasks with and without processing negation

Regarding polarity classification, it may be seen that both approaches report similar results, the improvement achieved by dealing with negation being also similar ($\approx$ 0,4 %). However, as the number of classes increases, the emotional approach obtains much better results ($\approx$ 15 % improvement in accuracy for the 4-classes task). The effect of negation also varies from one approach to another: in our approach, the treatment of negation is translated into an improvement of 0,5 % accuracy on average, while in the polar expressions-based strategy dealing with negation often decreases this accuracy ($\approx$ 0,7 % in the 3-classes task and $\approx$ 0,4 % in the 4-classes one).

These results indicate that, as we expected, even though the use of polar expressions to represent the text may suffice when facing the polarity detection task, it seems inadequate for more fine-grained tasks in which the text need to be classified in different classes of intensity. In such tasks, dealing with negation as a simple polarity shifter is also inadequate and the text is better modeled as a set of emotional categories.

It is also worth mentioning that the effect of negation in this experiment is almost negligible for both approaches, the reason being the low frequency of negations in the SemEval dataset.

We next examine the results of the document polarity and intensity classification tasks on the documents from the HotelReview collection. These results are shown in table 4.9.

Concerning the polarity classification task, the emotion-based strategy obtains significantly better accuracy compared to the polar expressions approach. Surprisingly, the improvement achieved by dealing with negated forms is higher in the latter than in the former (7,2 % accuracy *vs.* 3,5 %). However, as number of classes increases, the behavior is similar to that seen in table 4.8 for the sentence intensity classification task: on the one hand, the emotion-based approach gets considerably better results ($>$ 20 % im-

Method	**Polarity**		**3-classes**		**5-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	88,5	88,6	63,6	63,6	43,4	41,1
Emotions + Negations	**92,0**	**92,1**	**68,1**	**66,6**	**46,4**	**44,7**
Polar expressions + No modifiers	78,2	80,5	42,8	51,6	27,3	32,0
Polar expressions + Negations	**85,4**	**86,5**	**45,5**	**52,5**	**29,3**	**34,7**

Table 4.9: Average accuracy and F-measure in the document polarity and intensity classification tasks with and without processing negation

provement in accuracy for the 3-classes task and $> 15\,\%$ for the 5-classes task); on the other hand, accuracy improves when dealing with negations in the emotion-based approach in a higher rate than in the polar expressions approach.

Finally, by comparing the results in tables 4.8 and 4.9 we confirm our intuition that the effect of negation is more marked in the HotelReview dataset than in the SemEval dataset. This is due to the fact that more than 36 % of the documents in the HotelReview dataset present negated forms, while these forms only appear in 3,64 % of the sentences in the SemEval collection.

4.4.2. Quantifiers

The second group of experiments is directed to find out to what extent dealing with quantifiers may help to improve accuracy in polarity and intensity classification. Once again, we conduct experiments both at the sentence and document level, on the SemEval and HotelReview datasets, respectively.

Table 4.10 shows these results as average accuracies for the sentence polarity and intensity classification tasks. It may be observed that the improvement achieved by processing quantifiers is worthless in both approaches (0,2 % on average for the polar expressions strategy and 0,13 % on average for the emotion-based one). This was expected, since quantifiers in the SemEval corpus are infrequent. However, even though the percentage of sentences with at least one quantifier is greater than the percentage of sentences with negations (4,23 % *vs.* 3,64 %), the effect of negations is more significant than that of quantifiers. This is not surprising, since negations reverse the polarity of the emotions affected by them, while quantifiers only weigh their intensity or strength. It is worth mentioning that Kennedy and Inkpen (2006) reached a similar result and concluded that the consideration

of negation words is more important than that of diminishers.

Method	**Polarity**		**3-classes**		**4-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	72,8	72,8	63,0	58,4	**45,3**	**42,9**
Emotions + Quantifiers	**73,1**	**73,1**	**63,2**	**58,8**	**45,3**	**42,9**
Polar expressions + No modifiers	**72,9**	**74,4**	54,4	51,0	32,1	38,2
Polar expressions + Quantifiers	**72,9**	**74,4**	**54,5**	**51,2**	**32,6**	**38,8**

Table 4.10: Average accuracy and F-measure in the sentence polarity and intensity classification tasks with and without processing quantifiers

Concerning the document polarity and intensity classification tasks (see table 4.11), the results show that there is a consistent improvement in performance for both approaches and also for all distributions of polarity and intensity classes. However, as with the case of negation, the increase is higher when classifying documents rather than sentences ($\approx$ 1,6 % improvement in accuracy on average for both approaches, which is more than eight times the average improvement achieved in the sentence classification tasks). Once again, this was expected due to the different datasets used in the experimentation and the fact that the number of instances affected by quantifiers considerably varies from one dataset to another.

Method	**Polarity**		**3-classes**		**5-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	88,5	88,6	63,6	63,6	43,4	41,1
Emotions + Quantifiers	**89,2**	**89,3**	**65,8**	**63,7**	**45,3**	**43,4**
Polar expressions + No modifiers	78,2	80,5	42,8	51,6	27,3	32,0
Polar expressions + Quantifiers	**79,5**	**81,6**	**44,5**	**53,1**	**29,0**	**34,0**

Table 4.11: Average accuracy and F-measure in the document polarity and intensity classification tasks with and without processing quantifiers

4.4.3. Modality

We next investigate the effect of processing modality in polarity and intensity classification.

Concerning classification at the sentence level, table 4.12 shows that processing modality does not change accuracy when the polarity recognition task is faced using the polar expressions approach. However, performance does drop when polarity is classified in different classes of intensity (a decrease of 0,3 % for 3-classes of intensity and 0,2 % for 4-classes). In contrast,

the results of the emotion-based approach improve when modality is processed, regardless of the task addressed. Yet, the improvement is virtually negligible, around 0,16 % accuracy on average.

Method	**Polarity**		**3-classes**		**4-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	72,8	72,8	63,0	58,4	45,3	42,9
Emotions + Modals	**73,1**	**73,0**	**63,1**	**58,5**	**45,4**	**43,4**
Polar expressions + No modifiers	**72,9**	**74,4**	**54,4**	**51,0**	**32,1**	**38,2**
Polar expressions + Modals	**72,9**	**74,4**	54,1	50,6	31,9	37,9

Table 4.12: Average accuracy and F-measure in the sentence polarity and intensity classification tasks with and without processing modality

Regarding classification at the document level, it may be seen from table 4.13 that processing modality consistently improves accuracy across the different approaches and tasks. Again, the improvement is higher when classifying documents instead of sentences, and this is due to the higher number of modals that are found in the HotelReview dataset (see table 4.7).

Method	**Polarity**		**3-classes**		**5-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	88,5	88,6	63,6	63,6	43,4	41,1
Emotions + Modals	**89,8**	**89,8**	**64,1**	**63,9**	**43,7**	**41,3**
Polar expressions + No modifiers	78,2	80,5	42,8	51,6	**27,3**	**32,0**
Polar expressions + Modals	**79,7**	**81,7**	**42,9**	**51,8**	26,8	31,5

Table 4.13: Average accuracy and F-measure in the document polarity and intensity classification tasks with and without processing modality

4.4.4. All Modifiers

With the last set of experiments, we aim to answer the question *how well do the modifiers perform working all together?*. That is, we want to check if there exists synergy between modifiers so that it is possible to obtain better results when the modifiers are used together than when they are applied separately.

The results for sentence polarity and intensity classification processing all modifiers are given in table 4.14. The most interesting thing that this table reveals is that, as far as the polar expressions approach is concerned, using all three modifiers together does not improve accuracy compared to not using any modifier or using some of them separately for any of the tasks

that have been evaluated. In contrast, the results for the emotional approach and the three tasks systematically improve when all modifiers are processed. The improvement, however, is not as marked as we might expect due to the low percentage of modifiers that are found in the SemEval corpus.

Method	**Polarity**		**3-classes**		**4-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	72,8	72,8	63,0	58,4	45,3	42,9
Emotions + All modifiers	**74,3**	**74,3**	**63,7**	**59,3**	**46,2**	**43,6**
Polar expressions + No modifiers	72,9	74,4	**54,4**	**51,0**	**32,1**	**38,2**
Polar expressions + All modifiers	**73,2**	**74,7**	53,7	50,4	**32,1**	**38,2**

Table 4.14: Average accuracy and F-measure in the sentence polarity and intensity classification tasks with and without processing all modifiers

Concerning these same results for the document level, table 4.15 shows that, for both classification strategies, processing all modifiers outperforms the *No modifiers* baseline, in terms of accuracy and F-measure. More specifically, dealing with the three modifiers in the polar expressions approach allows for a 10,2 % improvement in accuracy for the polarity recognition task, but this improvement decreases as the task becomes more difficult (i.e., as the number of intensity classes increases). In fact, accuracy increases 4,4 % for the 3-classes task and just 2,5 % for the 5-classes task. With respect to the emotion-based approach, the advantage gained by dealing with all modifiers is similar for all classification tasks ($\approx$ 6,25 % accuracy on average). These improvements across both approaches seem to indicate that emotional modifiers are quite useful for polarity and intensity classification.

Method	**Polarity**		**3-classes**		**5-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
Emotions + No modifiers	88,5	88,6	63,6	63,6	43,4	41,1
Emotions + All modifiers	**93,6**	**93,6**	**70,8**	**70,8**	**50,0**	**48,3**
Polar expressions + No modifiers	78,2	80,5	42,8	51,6	27,3	32,0
Polar expressions + All modifiers	**88,4**	**89,2**	**47,2**	**55,7**	**29,8**	**35,6**

Table 4.15: Average accuracy and F-measure in the document polarity and intensity classification tasks with and without processing all modifiers

Finally, the most noticeable observation is that, especially in the emotion-based approach, the effect of combining modifiers is sometimes greater than the sum of their effects alone. Therefore, our hypothesis about the existence of synergies between emotional modifiers is confirmed.

4.5. Evaluating the Effect of Word Ambiguity

The last group of experiments has been conducted to examine the effect of word ambiguity. To this aim, we repeated the experiments for polarity and intensity classification both at the sentence and document level on the SemEval 2007 and the HotelReview datasets, using different WSD strategies. First, we use the Lesk algorithm to disambiguate the meaning of the words according to their context. Second, we simply assign to each word its first sense in WordNet. Third, we select these senses randomly.

The results are presented in tables 4.16 and 4.17. It may be seen that using WSD improves the accuracy obtained by all ML methods compared to the random mapping baseline. However, the differences between the Lesk and the 1st sense strategies are not marked, the first achieving an average improvement of less than 2 % accuracy over the second.

WSD	**Method**	**2 classes**		**3 classes**		**4 classes**	
		Acc.	F-1	Acc.	F-1	Acc.	F-1
Lesk	*LibSVM*	**75,1**	**75,0**	63,2	55,7	45,8	41,1
	Logistic	74,3	74,3	63,7	59,3	**46,2**	**43,6**
	FT	74,3	74,3	**64,1**	**59,5**	45,0	42,5
1st Sense	*LibSVM*	71,8	71,7	62,9	54,9	44,6	40,1
	Logistic	72,8	72,7	**63,1**	**58,4**	**45,4**	**43,6**
	FT	**73,0**	**73,0**	63,0	58,3	43,6	41,7
Random	*LibSVM*	63,2	61,9	60,1	49,6	**42,9**	39,5
	Logistic	**63,5**	62,7	**61,7**	**55,0**	42,2	**40,0**
	FT	63,3	**62,8**	61,2	54,4	40,8	38,1

Table 4.16: Average results for different classifiers in the polarity and intensity classification task at the sentence level when different WSD strategies are used

WSD	**Method**	**2 classes**		**3 classes**		**5 classes**	
		Acc.	F-1	Acc.	F-1	Acc.	F-1
Lesk	*LibSVM*	93,1	93,1	68,4	68,7	45,9	44,7
	Logistic	**93,6**	**93,6**	**70,8**	**70,8**	**50,0**	**48,3**
	FT	93,3	93,3	68,6	68,5	46,2	45,0
1st Sense	*LibSVM*	91,7	91,8	67,0	66,2	45,3	44,2
	Logistic	**91,9**	**92,0**	**67,9**	66,1	**46,8**	**44,7**
	FT	91,3	91,4	67,4	**67,4**	44,6	43,7
Random	*LibSVM*	80,8	80,8	56,4	51,8	34,9	33,9
	Logistic	81,8	81,7	**58,9**	55,3	**37,7**	35,6
	FT	**82,5**	**82,3**	56,1	**55,6**	35,9	**35,7**

Table 4.17: Average results for different classifiers in the polarity and intensity classification task at the document level when different WSD strategies are used

Chapter 5

Discussion

In this section, we analyze and discuss the results obtained in the different experiments presented in the previous chapter. First, we focus on polarity and intensity classification. Next, we study the effect of the emotional modifiers and word ambiguity on sentiment analysis.

5.1. Polarity Detection

The results in table 4.3 showed that our method compares favorably with previous approaches in both the SemEval and HotelReview datasets. These good results seem to confirm our hypothesis that a emotion driven approach combined with a deeper linguistic analysis considerably improves polarity classification compared to shallow approaches based on term frequencies or polar expressions.

The difference between our best classifier and the *Unigrams* approach is 22,9 % accuracy for the SemEval 2007 dataset, and 7,4 % accuracy for the HotelReview corpus. The remarkable difference between the performance of the *Unigrams* approach in the two datasets is due to the vocabulary used. The SemEval dataset comprises a set of news headlines covering very different topics, while the sentences in the HotelReview corpus focus on opinions about hotels, where the vocabulary used is quite homogeneous and it is better captured with the *Unigrams* approach. On the other hand, the difference between our system and the *Polar expressions* approach is not so marked for the SemEval 2007 dataset (1,9 % accuracy), while is quite significant for the HotelReview corpus (10,3 % accuracy). This disparity is due to the better behavior of our system in the HotelReview corpus. In

contrast, the *Polar expressions* approach obtains nearly the same results in both datasets.

Table 4.3 also showed notable differences between the results achieved by our approach in the two corpora. A careful analysis has shown that one reason for this finding is the difference in the number of sentences not labeled with any emotion between both datasets. Even though the affective lexicon covers nearly the same percentage of concepts in both collections (20,5 % for HotelReview and 22 % for SemEval 2007), the number of sentences not labeled with any emotion in the SemEval 2007 dataset is 27 %, while in the HotelReview Corpus is 16 %. This happens mainly because a good number of sentences in the SemEval dataset do not have any emotional meaning *per se*. For instance, the sentence *Looking beyond the iPhone* does not express any emotion by itself, unless you are fan or detractor of Apple.

A second reason for the different results in both datasets could be the different size of the evaluation sets (1000 sentences for the SemEval collection *vs.* 4340 sentences for the HotelReview corpus). However, we have repeated the experiments for the HotelReview Corpus using 1000 sentences randomly selected as the evaluation set and found that the results are comparable to that obtained when using the whole corpus (83,5 % for *libSVM*, 85,2 % for *Logistic* and 84,5 % for *FT*).

It is also worth mentioning that a good percentage of the system's errors, when evaluated over the HotelReview dataset, come from (1) mislabeled instances in the training set (e.g., the sentence *Anyway, everybody else was nice* in a review is incorrectly placed in the <NegativeOpinion> section) and (2) very frequent spelling errors in the reviews. In contrast, these types of errors are not observed in the SemEval corpus, which was created manually.

Regarding polarity detection at the document level, the results in table 4.4 showed that the system behaves even better than at the sentence level for the HotelReview corpus. We think that this is due to the fact that the reviews frequently present neutral sentences that do not express any opinion (and so are not labeled with any emotion), but are necessarily classified as positives or negatives when working at the sentence level. In contrast, the number of whole reviews in the corpus that are not labeled with any emotion is not meaningful (3 %).

Table 4.4 also showed that the results of the method are lower when evaluated over the SFU collection. One reason for the variation across the

two corpora is the different coverage of their vocabulary in the affective lexicon. We have found that around 20,5 % of the concepts in the HotelReview corpus are mapped to an emotion, while only 9,5 % of the concepts in the SFU dataset are found in SentiSense. Another reason is related to the structure of the SFU corpus. This corpus contains 50 documents (25 positive and 25 negative) from 8 different products, and the instances appear ordered by product type in the dataset. Since the evaluation is performed using 10-fold cross validation, 40 instances of each product type are tested in each fold, and the 360 additional instances are used for training. Thus, the training set only presents 10 instances of the same product type as the instances in the evaluation set. In fact, this experiment may be shown as a kind of multi-domain polarity detection task, where the ML algorithm learns from instances of a domain to classify instances from other different domains.

Table 4.4 showed that our system performs better than previous works for both corpora. Comparing tables 4.3 and 4.4 we also realize that the *Unigrams* approach behaves considerably better when classifying documents rather than sentences. Since the documents are longer, it is expected to find more repeated terms across documents, which produces a more complete representation of the text to be classified. Regarding the *Polar Expressions* approach, its performance is quite similar to that of our system in the HotelReview corpus. However, the difference increases in the SFU dataset. One reason for this difference is, again, the less percentage of words labeled with an emotion in SentiSense.

5.2. Intensity Classification

We next analyze the results of the intensity classification task.

Concerning the intensity classification at the sentence level, the results obtained by the system are also encouraging (see table 4.5). As expected, the performance of the method decreases with the number of classes, since the difficulty of the task increases. In particular, the performance decreases in 11 % accuracy when moving from 2-classes to 3-classes, and in 17,9 % accuracy when moving from 3-classes to 4-classes. The main reason for this difference is that augmenting the number of classes makes the boundaries between them more diffuse.

The system also yields considerably better results than the *Unigrams* and

Polar Expressions approaches in both experiments (3-classes and 4-classes). It is interesting that the *Unigrams* approach behaves better in the 3-classes task than in the 2-classes one. The reason is that, in the 3-classes task, the *Unigrams* approach classifies all instances as neutral. Since the classes in the dataset are not balanced (251 instances for the negative class, 598 for neutral and 151 for positive), classifying all instances as neutral produces a high accuracy, but the F-measure is penalized. In contrast, the difference between accuracy and F-measure for our system is not marked, which means that all classes are equally learned.

The behavior of the *Polar expressions* approach is similar to that of our system. Again, the performance decreases with the number of classes. However, the difference in accuracy between the different experiments are much more marked: 19,5 % decrease in accuracy for the 3-classes comparing with the 2-classes and 21,6 % decrease for the 4-classes comparing with the 3-classes. These results suggest that the *Polar expressions* approach could be a good representation for polarity detection task but insufficient when dealing with the most complex intensity classification task.

Regarding the intensity classification task at the document level, the results are again better than those obtained at the sentence level (see tables 4.5 and 4.6), since the documents are richer in information. As expected, the accuracy decreases with respect to the 2-classes task. As found in the experiments at the sentence level, augmenting the number of classes makes the boundaries between them less marked, so that, even for humans, it may be difficult to decide which class corresponds to a given instance. In fact, it has been found that the corpus presents a number of reviews tagged with different classes, but determined as expressing the same emotional intensity by different judges. For instance, the two following reviews are classified in the corpus as *Very Poor* and *Poor* respectively, but have been considered by different judges to entail a similar degree of negativity:

Only the location was a positive aspect. The room was very small, in the basement of the hotel and we could hear people walking around all night long. The shower and the toilet closets were really uncomfortable and small as well.

Very near transport to London central. On arrival the apartment was very smelly of old spicy food and standard of cleanliness was poor. Entrance

to apartment (hallway) needs a paint job.

Even with this handicap, the behavior of the system for the intensity classification task is surprisingly good, classifying correctly around 3 in 4 document for the 3-classes task, and 1 in 2 documents for the 5-classes task.

Comparing these results with those of the *Unigrams* and *Polar expressions* approaches it has been found that, even if the performance of both approaches is good when facing the polarity classification problem, their accuracy decreases sharply in the 3-classes classification task (31,2 % for *Unigrams* and 41,2 % for *Polar expressions*), achieving ≈20 % accuracy less than our system. In the same way, the performance of both approaches is considerably worse than the performance of our system for the 5-classes prediction task (≈20 % accuracy), classifying correctly less than 1 in 3 documents. Again, these results suggest that both approaches behave well when classifying polarity, but are insufficient when aiming to classify a text according to its intensity.

5.3. The Effect of Emotional Modifiers

The results presented so far demonstrate that traditional approaches based on term frequencies and polar expressions are not sufficient when facing complex sentiment analysis tasks, such as classifying sentences and documents in different degrees of positivity and negativity, and it is therefore necessary a more fine-grained emotional representation for the text to be classified. This enhanced representation allows for improvements over 20 % accuracy in different classification tasks (see tables 4.5 and 4.6). Consequently, there is a need for a model of negation, quantification and modality able to define the effect of such modifiers on the emotions embedded in the text that are affected by them.

The experiments accomplished have revealed that, as already pointed out in the literature, the presence of negations, quantifiers and modals may modify the emotional meaning of a text, and that it is possible to improve polarity and intensity classification by correctly modeling and dealing with such constructions. However, to the author knowledge, no previous work has quantified the actual impact of modifiers in sentiment analysis tasks and the interactions among them. In this respect, the main findings of this study can be summarized in the following points:

- It has been found that, when the text is represented as a set of polar expressions, the effect of the modifiers is not uniform. In fact, while processing negations, quantifiers and modals increases classification accuracy in the polarity recognition task, this accuracy often decreases when facing the intensity classification task. As already mentioned, the reason seems to be that polar expressions are not enough to discriminate between different classes of intensity.

- Focusing on the method proposed in this work, it has been found that not all emotional modifiers have the same impact. This may be clearly observed in table 5.1, where the increases in accuracy and F-measure as compared to the *No modifiers* baseline are shown. Indeed, while the effect of negations is quite marked (3,7 % improvement in accuracy on average for the experiments on the HotelReview dataset), the effect of quantifiers and modals is much lower (1,6 % and 0,7 % improvement in accuracy, respectively). Therefore, it may be concluded that negation is the modifier that helps the most.

Classifier	**Polarity**		**3-classes**		**4 and 5 classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
SemEval dataset						
Emotions + Negations	0,4	0,3	0,3	0,3	0,7	0,4
Emotions + Quantifiers	0,3	0,3	0,2	0,4	0,0	0,0
Emotions + Modals	0,3	0,2	0,1	0,1	0,1	0,5
Emotions + All modifiers	1,5	1,5	0,7	0,9	0,9	0,7
HotelReview dataset						
Emotions + Negations	3,5	3,5	4,5	3,0	3,0	3,6
Emotions + Quantifiers	0,7	0,7	2,2	0,1	1,9	2,3
Emotions + Modals	1,3	1,2	0,5	0,3	0,3	0,2
Emotions + All modifiers	5,1	5,0	7,2	7,2	6,6	7,2

Table 5.1: Percentages of increases for accuracy and F-measure as compared to the *No modifiers* baseline when different combinations of emotional modifiers are tested

- We can see in table 5.1 that, as the number of intensity classes increases, the results decrease progressively, since the task becomes more difficult. However, as we might expect, the improvement achieved by processing emotional modifiers slightly increases with the number of classes. As already noted, augmenting the number of classes makes the boundaries between them less marked. In this context, correctly measuring the effect of modifiers on the emotions affected by them

becomes very important.

- Modifiers are not independent but interact with each other, so that sometimes it is even possible to achieve higher improvement when they are used jointly than the sum of the improvements obtained when they are considered independently. An explanation for this finding is that, when no negation and modality detection is performed, if the emotions assigned to certain concepts in the text are wrong (because these concepts are affected by a negation or a modal), then the quantifiers will be weighting the wrong emotions, and so, decreasing the system performance. To illustrate this, consider the following sentences:

  Neither the bed nor the pillows are very comfortable.

  Staff could be more friendly.

  The first sentence presents a negation (*neither...nor*) and a quantifier (*very*) that affect the concept *comfortable*. This concept is labeled with the emotion *like* in the SentiSense affective lexicon. If the negation is not processed, then the quantifier *very* will intensify this positive emotion in 75 % and the system will classify this sentence as highly positive. In contrast, the presence of the negation makes the concept *comfortable* to be assigned the emotion *disgust*, and thus the quantifier increases the weight of this negative emotion in 75 %. Similarly, not dealing with modals would make the second sentence to be classified as positive.

- This interaction between quantifiers and other modifiers explains the low improvement achieved when the quantifiers are processed separately (see table 5.1), even if the presence of quantifiers in the datasets is very common. Besides, it should be noted that, unlike negations, quantifiers do not change the emotional meaning of the text affected, but just increase or diminish its strength. In turn, the low effect of modals may be explained by its lower frequency in the datasets and the fact that some of them do not affect the emotions in their scopes.

- The experiments in table 4.14 showed that the effect of modifiers in classifying news headlines from the SemEval dataset is hardly noticeable, since the presence of modifiers in this type of texts is quite infre-

quent. Moreover, in spite of the high frequency with which modifiers appear in the HotelReview dataset, the improvement achieved is also less than expected (see table 4.15). In order to account for the true effect of modifiers in polarity and intensity classification, we have conducted a further experiment which consists in removing from the HotelReview corpus those reviews that do not contain, respectively, negations, quantifiers or modal verbs. Therefore, we now have three different subsets: *HotelReview-Negations*, *HotelReview-Quantifiers* and *HotelReview-Modals*, and we measure the effect of negations, quantifiers and modals, respectively, on each of these subsets. The results of these experiments are shown in table 5.2. It may be observed from this table that the results are similar to those obtained when the whole corpus is used, except for modality, which in this case produces an improvement of 3,2 % accuracy on average. These findings seem to indicate that the importance of modifiers depends on the type and domain of the documents being analyzed. While negation and speculation detection is critical in the classification of medical records, it plays a less important role in polarity and intensity classification of news headlines and product reviews.

Classifier	**Polarity**		**3-classes**		**5-classes**	
	Acc.	F-1	Acc.	F-1	Acc.	F-1
HotelReview-Negations						
Emotions	89,2	88,6	61,6	60,1	40,0	39,0
Emotions + Negations	92,8	92,6	66,4	66,1	43,3	43,0
HotelReview-Quantifiers						
Emotions	90,4	90,4	64,3	62,3	43,5	41,8
Emotions + Quantifiers	91,5	91,5	65,9	65,3	43,6	42,1
HotelReview-Modals						
Emotions	87,5	87,3	55,2	53,2	41,5	40,4
Emotions + Modals	91,8	91,7	58,1	56,3	44,0	42,6

Table 5.2: Average accuracy and F-measure in the document polarity and intensity classification tasks when only documents containing negations, quantifiers and modal verbs, respectively, are considered

- Finally, we want to check if the impact of modifiers depends on the overall polarity and intensity of the text in which they are used. To this end, table 5.3 shows the precision and recall for each class, when classifying the documents in the *HotelReview-Negations*, *HotelReview-Quantifies* and *HotelReview-Modals* subsets, respectively, into two and

three classes of polarity and intensity. One interesting thing that may be shown from this table is that negative instances (i.e., *Poor*) are better classified than positive ones (i.e., *Good*). Besides, when considering an intermediate class (i.e., *Fair*), this becomes the worst learned class. This seems to indicate that negative opinions are stronger than positive ones, at least in this corpus. However, it is with the positive class that modifiers seem to help the most.

Classifier	**2-classes**				**3-classes**					
	Good		**Poor**		**Good**		**Fair**		**Poor**	
	Pr.	Re.	Pr.	Re.	Pr.	Re.	Pr.	Re.	Pr.	Re.
HotelReview-Neg										
Emotions	78,4	61,6	91,1	95,9	50,0	36,4	48,6	52,3	72,0	80,6
Emotions + Neg	87,4	74,1	93,9	97,4	62,2	50,5	52,8	54,1	76,6	80,6
HotelReview-Quan										
Emotions	88,8	92,8	92,2	87,8	66,4	81,1	48,9	30,4	69,6	72,8
Emotions + Quan	90,5	93,2	92,7	89,7	69,0	82,9	52,3	35,2	72,7	77,9
HotelReview-Mod										
Emotions	79,0	72,1	90,2	93,1	48,6	52,2	33,3	22,5	67,8	79,6
Emotions + Mod	86,2	82,4	93,7	95,2	54,8	59,7	36,7	25,4	68,9	79,6

Table 5.3: Average precision and recall for each polarity and intensity class in the document polarity and 3-classes intensity classification tasks

5.4. The Effect of Word Ambiguity

Tables 4.16 and 4.17 showed that the use of a word sense disambiguation algorithm to solve text ambiguity improves the result of the polarity classification task, both at the sentence and document level. The WSD algorithms are able to more accurately identify the concepts that are being referred to in the document than when the random mapping approach is used, and this leads to the creation of a VEI vector that more accurately reflects the content of the text. The improvement achieved, however, is more remarkable when working with sentences than when the whole document is used for predicting polarity. This result is not surprising: the more information available for prediction, the less the effect of assigning the wrong meaning to a word.

Another interesting result is that, even though the use of word disambiguation improves the classification accuracy and F-measure, the improvement with respect to the first sense heuristic is less than expected. This is

due to the fact that the senses of the words in WordNet are ranked according to their frequency, and so the first sense of a word is also the most frequent one. Besides, the most frequent sense heuristic in WSD is usually regarded as a difficult competitor. In contrast, the improvement with respect to the random sense heuristic is quite remarkable.

Chapter 6

Conclusions and Future Work

This thesis presents a novel approach to emotional polarity and intensity classification. The system performs a more complex linguistic processing than most existing approaches, and makes intensive use of knowledge and contextual information to deal with common NLP problems such as word ambiguity and the presence of negations, quantifiers and modals in opinionated text. The method is based on mapping the text onto concepts from the WordNet lexical database and emotional categories from an affective lexicon. Unlike previous approaches, the text is modeled as a set of emotions instead of terms or polar expressions, which allows us to better capture the overall polarity and intensity of the text. Different WSD strategies are implemented in order to correctly understand the meaning of each word according to its context, and advanced syntactic techniques are used to identify negations, quantifiers and modals, their scope and their effect on the emotions affected by them.

An extensive evaluation has been performed to assess the system in two different tasks, polarity recognition (i.e., positive *vs.* negative) and intensity classification (i.e., different degrees of positivity and negativity), both at the sentence level and document level. To this purpose, three different datasets have been used: a corpus of news headlines, a set of hotel reviews, and a collection of user opinions from 8 different products. The results obtained outperform those achieved by other systems aiming at solving the same tasks, and demonstrate the feasibility of our approach.

Automatic detection of linguistic modifiers in affective text has recently attracted the attention of the sentiment analysis research community. Negations, quantifiers and modals may change the emotional meaning of the text affected by them, so that it is important to determine their scopes and effects in order to properly classify sentences and documents in different classes of polarity and intensity. However, the problem is much more complex than simply reversing, diminishing or amplifying the polarity of the expressions that fall into the scope of the modifiers, and requires to establish their effect on the emotions embedded in the text.

In the research presented in this thesis, we show that it is possible to improve polarity and intensity classification by representing the text as a set of emotions and tackling negations, quantifiers and modal forms as emotional modifiers. We compare our emotion-based model to the most widely accepted strategy based on polar expressions, and the results of the experiments reveal the benefit of our novel approach.

A focus of this research is measuring the real impact of modifiers in sentiment analysis-related tasks. To the best of our knowledge, no previous work has quantified this impact. We experiment with two tasks, polarity recognition and intensity classification, and evaluate the impact of modifiers in different datasets. Although processing modifiers produces significant improvements in all two tasks and datasets, the improvement is less than expected, which suggests that emotional modifiers are not as critical for sentiment analysis as they are for other disciplines and tasks (e.g., clinical text processing).

However, we find that not all modifiers affect emotions to the same extent. Thus, negation is the most influential phenomenon, whilst the effect of quantifiers and modals is less significant. The explanation for this finding is that, whereas negations invert the emotional meaning of the terms within their scope, quantifiers simply increase or decrease the intensity of such emotions. Modals, in turn, either not have any effect or decrease the emotional intensity, and only in a few cases reverse the emotions. Another interesting finding is that the effect of combining modifiers is usually more than the sum of their effects alone, which confirms the existence of synergies between emotional modifiers.

The results obtained also show that the use of word sense disambiguation algorithms in order to take to account the context of each word and represent

the text at the conceptual level improves the performance of the system. This confirms our assumption that working with concepts reduces the number of errors produced when assuming the wrong meaning of a word. These errors are frequent in term based approaches that make use of affective lexicons based on words or terms instead of concepts.

It is also important to notice that the good results obtained are achieved using an affective lexicon that only covers approximately the 20 % of the vocabulary in the evaluation datasets, or even less for the SFU corpus (9,5 %). This suggests that it is not necessary an affective lexicon that covers all the vocabulary of the input text. On the other hand, it will interesting to study the level of coverage that maximize's performance.

Although these results are promising, there are some issues that have to be further considered. First, in order to better classify texts (especially sentences) with low emotional meaning, it would be necessary to consider additional information about their context (i.e., the body of the news item, its section in the newspaper, the city where the hotel is placed, etc.). Second, it would be interesting to extend the method to deal with complex linguistic phenomena such as the presence of irony and sarcasm. Detecting and dealing with irony and sarcasm are probably the most difficult tasks, as there are no conceptual models about how they are expressed, and only few very preliminary researches have studied these phenomena. It would be also interesting to study the effect of the document structure in the polarity and intensity classification tasks, as some patterns of the discourse structure may affect the polarity. This patterns are more evident when working at the document level. For example, a common discourse structure is that in which the first sentences express a polarity, positive or negative, ending with a sentence expressing the opposite polarity and changing the whole polarity of the document. Another direction for future work will be to translate our model to different languages, specially to Spanish.

References

Agirre, E. and P. Edmonds. 2006. *Word Sense Disambiguation: Algorithms and Applications*, volume 33 of *Text, Speech and Language Technology*. Springer, 1 edition.

Agrawal, S. and T. J. Siddiqui. 2009. Using syntactic and contextual information for sentiment polarity analysis. In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, pages 620–623. ACM.

Andreevskaia, A. and S. Bergler. 2007. CLaC and CLaC-NB: Knowledge-based and corpus-based approaches to sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 117–120. Association for Computational Linguistics.

Anscombe, E. and P. Geach. 1970. *Descartes Philosophical Writings*. Nelson: The Open University.

Aue, A. and M. Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of the International Conference on Recent Advances in NLP*.

Baeza-Yates, R. and B. Ribeiro-Neto. 1999. *Modern Information Retrieval*. New York: ACM Press, Addison-Wesley.

Baker, K., M. Bloodgood, B. Dorr, N. W. Filardo, L. Levin, and C. Piatko. 2010. A modality lexicon and its use in automatic tagging. In *Proceedings of Language Resources and Evaluation Conference*, pages 1402–1407. European Language Resources Association.

Balahur, A., E. Lloret, E. Boldrini, A. Montoyo, M. Palomar, and P. Martínez-Barco. 2009. Summarizing threads in blogs using opinion polarity. In *Proceedings of the Workshop on Events in Emerging Text Types*, pages 23–31. Association for Computational Linguistics.

Bradley, M. M. and P. J. Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry*, 25(1):49–59.

Bradley, M.M. and P.J. Lang. 1999. Affective norms for english words (ANEW): Instruction manual and affective ratings. Technical report c-1, The Center for Research in Psychophysiology, University of Florida.

Brooke, J. 2009. A semantic approach to automated text sentiment analysis. Master's thesis, SIMON FRASER UNIVERSITY.

Carenini, G. and J. C. K. Cheung. 2008. Extractive vs. NLG-based abstractive summarization of evaluative text: the effect of corpus controversiality. In *Proceedings of the Fifth International Natural Language Generation Conference*, pages 33–41. Association for Computational Linguistics.

Carenini, G., R.T Ng, and A. Pauls. 2006. Multi-document summarization of evaluative text. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 305–312.

Carter, R. and M. McCarthy. 2006. *Cambridge grammar of english. A compresive guide spoken and written english grammar and usage.* Cambridge University Press.

Choi, Y. and C. Cardie. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 793–801. Association for Computational Linguistics.

ComScore. 2007. Online consumer-generated reviews have significant impact on offline purchase behavior. Retrieved December 15, 2010, from http://www.comscore.com/press/release.asp?press=1928.

Councill, I., R. McDonald, and L. Velikovich. 2010. What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59. University of Antwerp.

Das, S. and M. Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific Finance Association Annual Conference*.

Dave, K., S. Lawrence, and D. 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*, pages 519–528. ACM.

Ekman, P. 1992. Are there basic emotions? *Psychological Review*, 99(3):550–553.

Elango, P. 2005. Coreference resolution: A survey. Technical report, University of Wisconsin Madison.

Esuli, A. and F. Sebastiani. 2006a. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*.

Esuli, A. and F. Sebastiani. 2006b. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, pages 417–422.

Goldstein, I. and O. Uzuner. 2010. Does Negation really matter? In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 23–27. Association for Computational Linguistics.

Hatzivassiloglou, V. and K. R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, pages 174–181. Association for Computational Linguistics.

Hatzivassiloglou, V. and J. M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th Conference on Computational linguistics*, pages 299–305. Association for Computational Linguistics.

Hearst, M. 1992. Direction-based text interpretation as an information access refinement. In Paul Jacobs, editor, *Text-Based Intelligent Systems*. Lawrence Erlbaum Associates, pages 257–274.

Hitlin, P and L. Rainie. 2004. The use of online reputation and rating systems. *Pew Internet & American Life Project Memo*.

Hu, M. and B. Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM International Conference on Knowledge Discovery and Data Mining*, pages 168–177. ACM.

Huettner, A. and P. Subasic. 2000. Fuzzy typing for document management. In *ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes*, pages 26–27.

Ide, N. and J. Véronis. 1998. Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24:2–40.

James, W. 1884. *What is emotion?*, volume Mind 9.

Jia, L., C. Yu, and W. Meng. 2009. The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, pages 1827–1830. ACM.

Jindal, N. and B. Liu. 2006. Identifying comparative sentences in text documents. In *Proceedings of the 29th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 244–251. ACM.

Kennedy, A. and D. Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22:2006.

Kim, S-M. and E. Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics*. Association for Computational Linguistics.

Kim, S-M. and E. Hovy. 2006. Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 483–490. Association for Computational Linguistics.

Klenner, M., S. Petrakis, and A. Fahrni. 2009. A tool for polarity classification of human affect from panel group texts. In *Proccedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1 –6.

Lakoff, G. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2:458–508.

Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual International Conference on Systems Documentation*, pages 24–26. ACM.

Light, M., X. Y. Qiu, and P. Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *Proceedings of the HLT-NAACL. Workshop: BioLINK, Linking Biological Literature, Ontologies and Databases*, pages 17–24. Association for Computational Linguistics.

Liu, B. 2010. Sentiment analysis and subjectivity. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL. ISBN 978-1420085921.

Liu, B., M. Hu, and J. Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web*, pages 342–351. ACM.

Macdonald, C. and I. Ounis. 2006. The TREC Blogs06 collection : Creating and analysing a blog test collection. *DCS Technical Report Series*.

Martineau, J. and T. Finin. 2009. Delta TFIDF: An improved feature space for sentiment analysis. In *Proceedings of the Third International Conference on Weblogs and Social Media*.

Medlock, B. and T. Briscoe. 2007. Weakly supervised learning for hedge classfication in scientific literature. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 992–999. Association for Computational Linguistics.

Meena, A. and T. V. Prabhakar. 2007. Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. In *Proceedings of the 29th European Conference on Information Retrieval Research*, pages 573–580. Springer-Verlag.

Mihalcea, R., C. Banea, and J. Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 976–983. Association for Computational Linguistics.

Miller, G. A. 1995. WordNet: a lexical database for English. *Commun. ACM*, 38:39–41.

Morante, R. 2010. Descriptive analysis of negation cues in biomedical texts. 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 Conference on International Language Resources and Evaluation*. European Language Resources Association (ELRA).

Na, J. C., H. Sui, C. Khoo, S. Chan, and Y. Zhou. 2004. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. In I. C. McIlwaine (Ed.), editor, *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference*, pages 49–54.

Nicolov, N., F. Salvetti, and S. Ivanova. 2008. Sentiment analysis: Does coreference matter? In *Proceedings of the Symposium on Affective Language in Human and Machine*, pages 37–40.

Ortony, A., G. Clore, and A. Collins. 1988. *The cognitive structure of emotions*. Cambridge University Press.

Ortony, A. and T. J. Turner. 1990. What's basic about basic emotions? *Psychological Review*, 97(3):315–331.

Ounis, I., M. De Rijke, C. Macdonald, G. Mishne, and I. Soboroff. 2006. Overview of the TREC 2006 blog track. In *Proccedings of the Text Retrieval Conference, TREC. NIST*.

Palmer, F. R. 2001. *Mood and modality*. Cambridge University Press.

Pang, B. and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, pages 271–278.

Pang, B. and L. Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 115–124. Association for Computational Linguistics.

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

Pang, B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 79–86.

Parrot, W.G. 2001. *Emotions in social psychology: Essential readings.* Psychology Press.

Patwardhan, S., S. Banerjee, and T. Pedersen. 2005. SenseRelate::TargetWord: A generalized framework for word sense disambiguation. In *Proceedings of the ACL on Interactive Poster and Demonstration Sessions*, pages 73–76. Association for Computational Linguistics.

Pennebaker, J. W., M. E. Francis, and R. J. Booth. 2001. *Linguistic Inquiry and Word Count (LIWC): LIWC2001.*

Plutchik, R., 1980. *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, chapter A general psychoevolutionary theory of emotion, pages 3–33. New York: Academic.

Polanyi, L. and A. Zaenen. 2006. Contextual valence shifters. In W. Bruce Croft, James Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *The Information Retrieval Series.* Springer Netherlands, pages 1–10.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A comprehensive grammar of the English language.* Longman.

Riloff, E., S. Patwardhan, and J. Wiebe. 2006. Feature subsumption for opinion analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 440–448. Association for Computational Linguistics.

Riloff, E. and J. Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 105–112. Association for Computational Linguistics.

Russell, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Sack, W. 1994. On the computation of point of view. In *Proceedings of AAAI*, page 1488.

Sauri, R. and J. Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43:227–268.

Snyder, B. and R. Barzilay. 2007. Multiple aspect ranking using the good grief algorithm. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 300–307.

Stone, P. J., D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. 1966. *The General Inquirer: A computer approach to content analysis*. MIT Press.

Strapparava, C. and A. Valitutti. 2004. WordNet-Affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 2004, pages 1083–1086.

Thomas, M., B. Pang, and L. Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 327–335.

Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the Association for Computational Linguistics*, pages 417–424.

Van Asch, V., R. Morante, and W. Daelemans. 2010. Towards improving the precision of a relation extraction system by processing negation and speculation. In *Proceedings of the 4th Symposium on Semantic Mining in Biomedicine*, pages 164–165.

Velupillai, S. 2010. Towards a better understanding of uncertainties and speculations in Swedish clinical text: analysis of an initial annotation trial. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 14–22. Association for Computational Linguistics.

Von Fintel, K., 2006. *Encyclopedia of Philosophy*, chapter Modality and Language. MacMillan Reference.

Whissell, C.M, 1989. *Emotion: Theory, Research, and Experience*, chapter The dictionary of affect in language, pages 113–131. New York: Academic Press.

Whitelaw, C., N. Garg, and S. Argamon. 2005. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pages 625–631. ACM.

Wiebe, J. M. 1990. Identifying subjective characters in narrative. In *Proceedings of the International Conference on Computational Linguistics*, pages 401–408.

Wiebe, J. M. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.

Wiebe, J. M. and R. Bruce. 1995. Probabilistic classifiers for tracking point of view. In *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pages 181–187.

Wiebe, J. M., R. F. Bruce, and T. P. O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 246–253. Association for Computational Linguistics.

Wiebe, J. M., T. Wilson, R. Bruce, M. Bell, and M. Martin. 2004. Learning subjective language. *Computational Linguistics*, 30:277–308.

Wiebe, J. M., T. Wilson, and C. Cardie. 2005. Annotating expressions of opinions and emotions in language. volume 1.

Wiegand, M., A. Balahur, B. Roth, D. Klakow, and A. Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68. Association for Computational Linguistics.

Wilson, T., J. M. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354. Association for Computational Linguistics.

Wilson, T., J. M. Wiebe, and P. Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433.

Wilson, T., J. M. Wiebe, and R. Hwa. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of the 19th National Conference on Artifical intelligence*, pages 761–767. AAAI Press.

Yu, H. and V. Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 129–136. Association for Computational Linguistics.

Apendix A

Publications

A.1. Polarity and Intensity Classification

1. Carrillo de Albornoz J, Plaza L, Gervás P. 2010. Improving Emotional Intensity Classification using Word Sense Disambiguation. In Journal on Research in Computing Science, volume 46. Special issue: Natural Language Processing and its Applications, pages 131-142.

2. Carrillo de Albornoz J, Plaza L, Gervás P. 2010. A Hybrid Approach to Emotional Sentence Polarity and Intensity Classification. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL 2010), pages 153-161. Uppsala, Sweden.

3. Carrillo de Albornoz J, Plaza L, Gervás P, Díaz A. 2011. A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating. In Proceedings of the 33rd European Conference on Information Retrieval (ECIR 2011), pages 55-66. Dublin, Ireland.

4. Plaza L, Carrillo de Albornoz J. 2011. Sentiment Analysis in Business Intelligence: A survey. Book chapter in: Customer Relationship Management and the Social and Semantic Web: Enabling Cliens Conexus. IGI-Global. Editors: R. Colomo-Palacios, J. Varajao and P. Soto-Acosta.

A.2. Practical Applications of Sentiment Analysis

1. Carrillo de Albornoz J, Garmendia L, Gervás P. 2008. Sistema automático de decisión del plano de cámara en videojuegos y entornos

virtuales. In Proceedings of the Segundo Congreso Internacional de Matemáticas en la Ingeniería y la Arquitectura, pages 219-230.

2. León C, Carrillo de Albornoz J, Gervás P. 2008. A Framework for Building Creative Objects From Heterogeneous Generation Systems. In Proceedings of the 5th International Joint Workshop on Computational Creativity. Madrid, Spain.

3. Carrillo de Albornoz J, Gervás P. 2009. Automated Camera Management Based on Emotional Analysis of the Film Script. In Proceedings of the Computer Graphics, Visualization, Computer Vision and Image Processing (IADIS Intenational Conference), pages 41-48. Algarve, Portugal.

A.3. Other Natural Language Processing Applications

1. Plaza L, Carrillo de Albornoz J, Prados J. 2010. Sistemas de Acceso Inteligente a la Información Biomédica: una Revisión. In Revista Internacional de Ciencias Podológicas, volume 4, number 1, pages 7-15.

A.4. Other Artificial Intelligence Applications

1. Gómez-Albarrán M, Bautista S, Carrillo de Albornoz J. 2008. Personalized Access and Students'Coauthoring in Repositories of Learning Objects: The Case of a Repository of Programming Examples. In Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies.