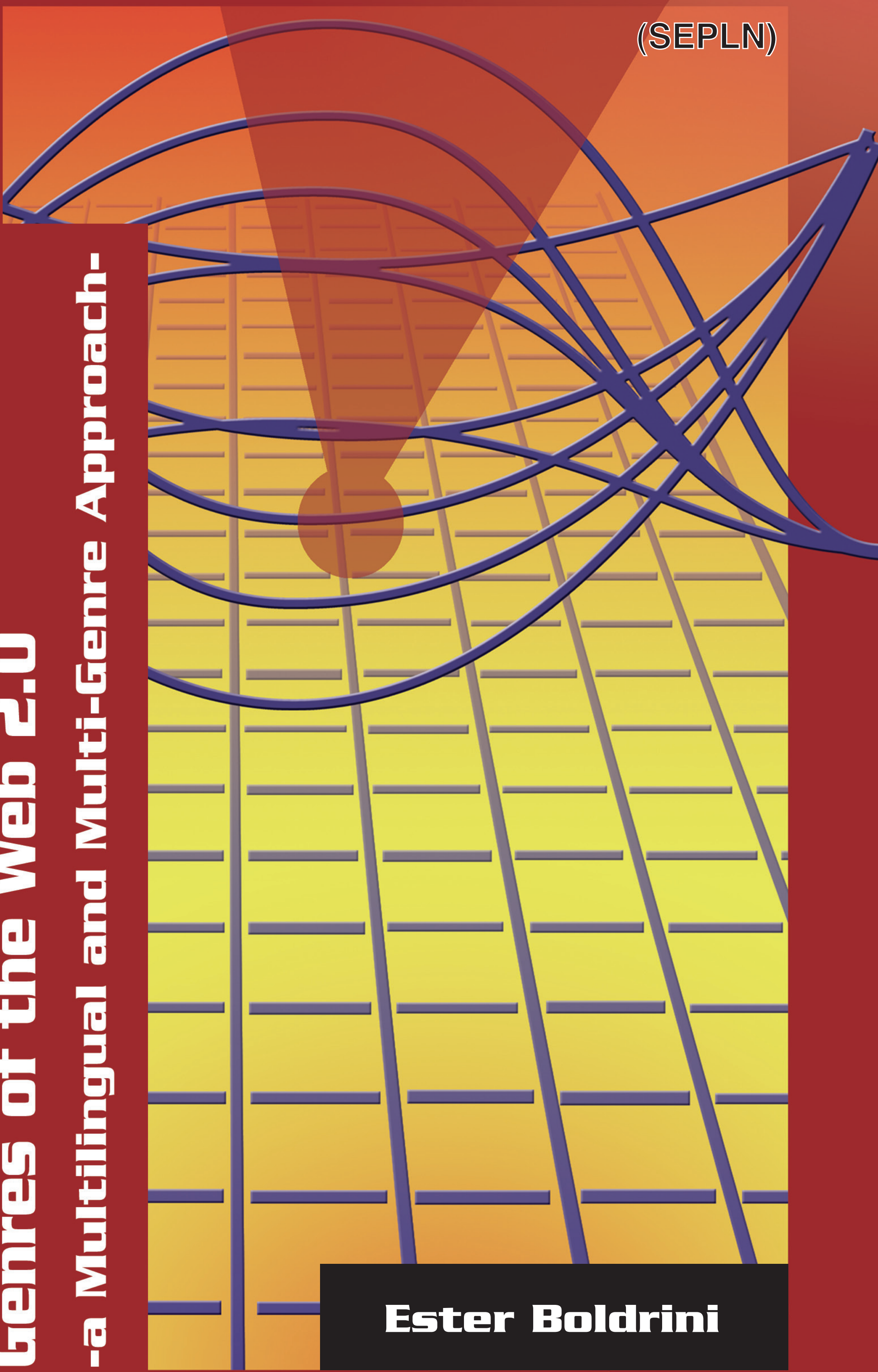
Colección de monografías: 14
Sociedad Española para el Procesamiento del Lenguaje Natural
(SEPLN)

EMOTIBLOG: A Model to Learn Subjective Information Detection in the New Textual Genres of the Web 2.0

-a Multilingual and Multi-Genre Approach-



Ester Boldrini

Colección de Monografías de la Sociedad Española
para el Procesamiento del Lenguaje Natural
(SEPLN). Número 14

Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)
http://www.sepln.org
secretaria.sepln@ujaen.es

Título: EMOTIBLOG: A Model to Learn Subjective Information Detection in the New Textual Genres of the Web 2.0 -a Multilingual and Multi-Genre Approach-
Autor: © Ester Boldrini
ISBN: 978-84-608-1977-6
Editores: L. Alfonso Ureña y Emilio Sanchís
Fotocomposición: COMPOBELL, S.L.

Prólogo

La Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) es una asociación científica sin ánimo de lucro creada en el año 1983 con el fin de promocionar y difundir todo tipo de actividades relacionadas con la enseñanza, investigación y desarrollo en el campo del procesamiento del lenguaje natural, tanto en el ámbito nacional como internacional.

Entre las actividades principales de la SEPLN figuran:

- La celebración de un congreso anual que sirve de punto de encuentro para los distintos grupos que trabajan en el área del procesamiento del lenguaje natural.
- La edición de la revista científica especializada Procesamiento del Lenguaje Natural de periodicidad semestral que cuenta con el Certificado de Revista Excelente de la Federación Española de Ciencia y Tecnología (FECYT).
- Un servidor Web (www.sepln.org) de referencia sobre procesamiento del lenguaje natural donde se encuentran en acceso abierto todas las publicaciones de la revista (journal.sepln.org).
- Una lista moderada de correo electrónico (SEPLN-L) que sirve como boletín de información periódica (quincenal) y como espacio de información y discusión para los miembros de la Asociación. La dirección para enviar cualquier comentario o aportación a la lista es sidsepln@si.ehu.es.
- Una Edición anual de Premios SEPLN a la Investigación en Procesamiento del Lenguaje Natural.
- A esta XIV Edición de los Premios SEPLN a la Investigación en Procesamiento del Lenguaje Natural se pudieron presentar a concurso trabajos monográficos de investigación originales e inéditos de cualquier extensión, escritos por un miembro de la SEPLN, y que no hubieran sido publicados o enviados a publicación con anterioridad a este concurso. Esta publicación presenta el trabajo premiado este año por la comisión evaluadora.

La Junta Directiva de la SEPLN, en nombre de la Sociedad, quiere dejar constancia aquí de la alta calidad de todas las obras presentadas a concurso en esta XIV Edición de los Premios SEPLN, y animar a todos sus miembros a la participación en sus futuras ediciones. Con la publicación de estas contribuciones en su Colección de Monografías, la SEPLN podrá aportar lo mejor de sus esfuerzos a la actualización y divulgación de la investigación en el campo del procesamiento del lenguaje natural.

Junio 2015
Sociedad Española para el Procesamiento del Lenguaje Natura

EMOTIBLOG: A Model to Learn Subjective Information Detection in the New Textual Genres of the Web 2.0 -a Multilingual and Multi-Genre Approach-

Ester Boldrini

A te papà…

Spero tu sia orgoglioso di me.

Grazie per avermi insegnato ad essere sempre onesta, leale e a mettere impegno in tutto quello che faccio.

Mi manchi tanto…troppo.

Agradecimientos

Este trabajo es sin duda el resultado de dos caminos y evoluciones paralelas: de mi vida profesional y también la personal y tengo la suerte de poder decir que han sido muchas las personas que han contribuido a que esto se hiciera realidad.

No cabe la menor duda de que esta Tesis no habría sido posible sin la ayuda inestimable, constante y paciente de mi fantástico Director de Tesis, Patricio Martínez-Barco. Su incalculable profesionalidad unida a un número impresionante de horas de trabajo, disponibilidad y paciencia infinitas, además de una personalidad muy modesta, sensible y humana han hecho que haya sido para mí una figura insustituible. Él me ha llevado de la mano estos años y esto me ha permitido disfrutar verdaderamente de la investigación, segura de tener a mi lado un profesional y un punto de referencia fijo con el cual contar y en el cual confío totalmente.

Después de haber dicho esto, mi agradecimiento va dirigido a la Universidad de Alicante y al Vicerrectorado de I+D+i que gracias a sus *Másteres Europeos* por un lado y a la *Beca de Iniciación a la Investigación,* por otro, me dieron la posibilidad de empezar este camino y de apasionarme a la investigación.

Mi más sincero agradecimiento es para el Departamento de Lenguajes y Sistemas Informáticos y en concreto al Grupo de Procesamiento del Lenguaje Natural y Sistemas de Información y a todos sus miembros que me han acogido desde el primer día como una más y nunca me han dejado sola solventándome mis miles de dudas y preguntas.

Un gracias especial va a Manuel Palomar, director del GPLSI que a pesar de sus muchos compromisos nunca ha dejado de confiar en mí y me ha apoyado constantemente preocupándose por el lado profesional y personal de mi vida, siempre.

Mis más sinceras gracias a Paloma Moreda, Andrés Montoyo, Rafael Muñoz que han estado siempre a mi lado, ayudándome en todo momento.

Agradezco a Alexandra Balahur, Elena Lloret, José Manuel Gómez y Javi Fernández por los muchos trabajos realizados juntos que me han enriquecido y hecho posible llevar a cabo esta Tesis.

Gracias a Paloma Moreda, Elena Lloret, José Manuel, a Mike Thelwall y Tommaso Caselli por sus consejos que han contribuido a que mi Tesis mejorara.

Como bien decía antes, la investigación ha formado parte de los últimos 4 años de mi vida, por lo tanto es inevitable que haya influenciado en a evolución de mi persona. Quizás esta sea la parte más difícil de escribir por tener que resumir tantas emociones en un espacio muy limitado.

Yo creo que el trabajo es una de las cosas más relevantes de la vida de una persona porque permite conocer a mucha gente, intercambiar ideas, adaptarse al otro, llegar a compromisos y estas son cosas que te permiten crecer y evolucionar.

Al fin y al cabo me he dado cuenta de que paso la mayoría de mis horas diarias con mis compañeros y puedo decir, y estoy muy orgullosa de ello, que tengo como dos equipos de compañeros de trabajo con los cuales he compartido esta fase de mi vida y espero poder compartir también todo lo que vendrá. Cada uno de ellos ha sido una pieza fundamental que me ha permitido sacar los dientes y seguir adelante sobre todo en momentos muy delicados y difíciles.

– A Manolo y Patricio, por haberme dicho de no preocuparme de nada, de irme a casa, de quedarme allí y de tomarme todo el tiempo que me hiciera falta.
– A Roberto, por los billetes de ida comprados y por haberme dicho ante el evento final del IPEA “nadie es imprescindible” y así haberme permitido estar donde tenía que estar y haberme dado la tranquilidad de no tener que arrepentirme de nada.
– A Paloma, Elena, Inma, Rossana, Michelle y Alicia amigas preciosas y raras de encontrar que me han dado aquel empujón para irme, y me han escuchado, llamado, aconsejado, dado su espalda para llorar y hecho mi trabajo cuando yo no podía.
– A Andrés y Rafa, por asegurarme que la vida me pondría cerca siempre un ángel para los momentos difíciles y para estar siempre pendientes de mí.
– A Lucía, que, aunque estando en Bruselas, ha sido como tenerla a mi lado.

– A Fran, por haberme dejado su Visa cuándo la necesitaba y dado su testimonio de que sí, se puede ir adelante.
– A José Manuel y Javi, por investigar ellos cuándo yo tenía la cabeza en otro lado.
– A Valeria y Simone, amigos inseparables que tanto me habéis escuchado.
– A mi madre, por no haberme hecho pesar nunca la lejanía y no cortarme nunca las alas, dándome la posibilidad y el apoyo para llegar hasta aquí.
– A Akram, con mucho amor, eres la persona que me completa, sin ti yo no lo habría conseguido…no hace falta que diga más…tú ya sabes…
– A Miriam, mi hermana, pieza imprescindible e insustituible en mi vida…
– Aunque os parezca raro, a Nerone, mi amigo fiel e inseparable por quitarme preocupaciones de la cabeza y por estar siempre a mi lado con mucha alegría.

A todos vosotros os quería dar las gracias y decir que habéis sido y sois mi rayo de sol en un momento muy, pero que muy oscuro. No sé si os dais cuenta de lo que habéis hecho. Os aseguro que habéis sido testigos de que es posible salir adelante y me habéis hecho ver que la vida no es tan negra como la veo ahora. Las palabras no pueden expresar cuanto os estoy agradecida.

Por último, decir que considero esta Tesis una etapa en mi formación y por lo tanto espero poder continuar este camino y sobre todo con todos y cada uno de vosotros.

Table of Contents

List of Tables

List of Figures

1. Introduction

In the past years the Web 1.0 consisted in a collection of documents linked and executed on the Internet. Tim Berners-Lee[1] and Robert Cailliau[2] created it while they were working at the CERN[3], the European organisation for Nuclear Research in Genera and published their results later in 1992. At that time, the opinionated data was not available or it was produced in a very small amount. Thus, when an individual needed to take a decision, he typically asked to his friends or family and also when a company wanted to know the opinions of its customers about its products they conducted surveys.

But today the situation has dramatically changed thanks to the creation and the massive usage of the Web 2.0 (Flew 2002):

> *we moved from personal websites to blogs and blog site aggregation, from publishing to participation, from web content as the outcome of large up-front investment to an ongoing and interactive process, and from content management systems to links based on tagging.*

Researchers were convinced that the web was acquiring more and more importance thanks to its new applications and sites. That is why, the *Web 2.0 Conference*[4] was established and its idea is generally associated with web applications that facilitate information sharing, interoperability, user-centered design[5] and collaboration in general on the World Wide Web.

---

1 www.w3.org/People/Berners-Lee/

2 www.robertcailliau.eu/Alphabetical/M/Me/Welcome.html

3 http://public.web.cern.ch/public/

4 www.web2summit.com/web2011

5 User-centered design (UCD) is a design philosophy where the end-user's needs, wants and limitations are a focus at all stages within the design process and development lifecycle.

This new network offers the possibility to interact employing a social media dialogue by means of the user-generated content[6] in a virtual community. The result is the birth and growth of blogs, forums, online reviews, etc. Their amount of data is increasing at an exponential rate, together with their predominant subjective content that reflects users' opinions about a wide range of topics (Cui, Mittal and Datar 2006) affecting people' in many aspects of their daily life.

According to Elis Pariser in The Filter Bubble[7]:

> *We are overwhelmed by a torrent of information: 900,000 blog posts, 50 million tweets, more than 60 million Facebook[8] status updates and 210 billion emails are sent off into the electronic ether every day*.

Moreover, Eric Schmidt[9] points out that:

> *if we record all human communication from the dawn of time to 2003, it' would take up about 5 billion gigabytes of storage space[10].*

The World Wide Web is a mine of language data with unprecedented richness and ease of access (Kilgarriff and Grefenstette 2003) and it has a great potential since users write about whatever is on their mind, thus creating data about almost everything and exploitable for a huge variety of real applications focused on different purposes.

One of the consequences of this evolution in communication is that research for analysing, interpreting, treating and exploiting subjective data is increasing in many areas of knowledge and in numerous cases those disciplines collaborate to create a more effective interdisciplinary perspective to solve the challenges subjectivity expression on the new textual genres poses.

This chapter presents the context of our research describing the textual genre

6 user-generated content, UGC is the term used to describe any form of content such as video, blogs, discussion from posts, digital images, audio files, and other forms of media that was created by consumers or end-users of an online system or service and is publically available to others consumers and end-users. User-generated content is also called consumer generated media

7 www.thefilterbubble.com

8 www.facebook.com/

9 http://www.google.com/about/corporate/company/execs.html#eric

10 http://dclibrary.wordpress.com/2011/06/08/overwhelming-amount-of-data/

and language we choose for our study, but also the motivation for that. Since it is a quite new research area, we will clarify the key terms we are going to use and we also underline the relevance of our research area by presenting the most relevant EU funding programmes established.

Finally, since in our study many are the disciplines that cooperate, in the following sections we provide a brief overview of such multidisciplinarity, describing the most relevant research carried out in Neuroscience, Cognitive Science, Psychology and Natural Language Processing, different disciplines that tackle subjectivity from different perspectives and for diverse purposes.

1.1 Subjective data studied from different prospectives

As we mentioned above, due to its great potential, the subjective information created in the social media context is acquiring more and more importance. Subjectivity (related to emotion, or to the so-called affected related phenomena) has traditionally been studied by disciplines such as, neuroscience, cognitive studies or psychology. It can be said that they are all focused on understanding the subjectivity production, recognition, reaction and classification, but each one from its own perspective.

Neuroscience

During the last 10 years, new research in neuroscience detected a close relationship between emotion and knowledge, to brain activity and consciousness (Taylor 1997). Davidson, Sherer and Goldsmith (2003) offered considerable advances in understanding the mechanisms by which the brain processes shape emotions and how human subjectivity can vary depending on a wide range of inquiry methods, such as neuroimaging techniques or laboratory paradigms designed to evaluate the cognitive and social constituents of emotion.

Damasio (1994), studied how the human brain recognizes emotional states and formulated the *somatic marker hypothesis.* This is a mechanism by which emotions guide (or bias) behaviour and decision-making, and positing that rationality requires emotional input. Damasio stresses on the importance of feeling that affect personal decisions. The intuitive signals that guide people in these situations consist in limbic-driven surges from the viscera, the somatic markers. Listening to their reactions, they can lead the patient to reject the negative course of action.

Cognitive Science

In the framework of cognitive studies, one of the most relevant research is the one carried out by LeDoux (1989). He is convinced that *emotion and cognition are mediated by separate but interacting systems of the brain*. In fact, the emotional system evaluates the biological significance of the stimuli from the external or internal world (thoughts, images and memories).

This evaluation takes place prior to conscious awareness, because only the results of the evaluation are made in a conscious way. LeDoux (1989) distinguishes between cognitive and emotion processes according to their consequences. For example, the computations that determine that a snake is a vertebrate, that it is biologically closer to an alligator than to a cow, and that its skin can be used to make belts and shoes, have very different consequences than the computations that determine that a snake is likely to be dangerous.

For LeDoux (1989), cognitive computations in the brain provide information about a stimulus and its relationship to other stimuli (knowledge about the world), while affective computations (that lead to emotions) provide information about the relation of the stimulus to the individual. Cognitive computations can also lead to other cognitions (*elaboration*– reasoning or thinking deeply about something). Affective computations lead to behavioural responses (e.g. avoidance), autonomic responses (increased heart rate, sweating, etc), and humoral reactions (changes in brain chemistry, such as an increase in adrenaline).

Psychology

The appraisal theory has been formulated by (¡De Rivera (1977), Frijda (1986), Ortony, Clore and Collins (1988) and Oatley and Johnson-Laird (1987) in the context of psychological studies and theories focused on the perception and production of models of emotions. The pillar of this theory is the assumption that emotions are extracted from our evaluations (*appraisals*) of events that cause a different reaction depending on the person. Thus, our evaluation of a situation causes an emotional response based on such evaluation and the nature of the emotional reaction can be foreseen analysing individual's response in similar situation/event. As we can deduce, different contexts can generate diverse emotional responses, as well as the same situation can trigger various affective reactions or equivalent reactions can be produced by different stimuli.

The appraisal criteria consist in the elements considered in the appraisal process (K. R. Scherer 1999). At present, there is no common approach and thus different versions of such theory have defined their own list of factors. However, K. Scherer (1988) demonstrated that the appraisal criteria proposed in such theories cover the same type of appraisals. He proposed five appraisal categories (*novelty, intrinsic pleasantness, goal significance, coping potential, compatibility standard*) that contain 16 appraisal criteria (*suddenness, familiarity, predictability, intrinsic pleasantness, concern relevance, outcome probability, expectation, conduciveness, urgency, cause: agent, cause: motive, control, power, adjustment, external compatibility standards, internal compatibility standards*). Moreover, he includes the values of these criteria in self-reported affect-eliciting situations to construct the vectorial model in the expert system GENESIS[11] (K. R. Scherer 1993).

Apart from classical disciplines such as neuroscience psychology, or cognitive studies, the subjective data has been studied in Computer Science and more concretely by its sub discipline called Natural Language Processing, among others. Thus the research we present in this work has been carried out from the point of view of this sub discipline briefly presented below.

Computer Science –Natural Language Processing

In addition to neuroscience psychology, or cognitive studies, the wealth of subjective data in the last 10 years there has been an explosion of interest among Natural Language Processing researchers who aim to develop knowledge models for representing the subjectivity expressed in texts that will help to automatically process subjective data.

The main difference between Natural Language Processing studies (the State of the Art will provide a more complete overview on this in Chapter 3) and the

11 www.genesisexpert.com

other disciplines previously presented is that it is focused on the creation of concrete applications and it is also an extremely interdisciplinary area of research which exploits theories and perspectives such as the one above mentioned, everything focused on designing new applications exploiting subjective data.

In fact, systems that are able to automatically discriminate between objective/ subjective discourses can be useful for the society in the broadest sense. For example, companies can employ them to detect the customer's opinions about their product or the ones of the competence, politics to predict the elections results, police to predict possible dangerous behaviours or situations, ect. However since topics discussed in the Web 2.0 are numerous, the number of possible applications is unpredictable.

Pioneer studies in this sense have been initiated in the 80s and 90s and Initial work in this area includes Turney (2002) and Pang, Lee and Vaithyanathan (2002) who applied different methods for detecting the polarity of product and movie reviews at document level or Snyder and Barzilay (2007) who classified the document polarity on a multi-way scale.

Pang, Lee and Vaithyanathan (2002) gave a step forward predicting star ratings on either a 3 or a 4 star scale and Snyder and Barzilay (2007) performed an analysis of restaurant reviews, predicting ratings for their various aspects. Initial approaches to detect subjectivity in text include the use of models simulating human reactions according to their needs and desires (Dyer 1987), fuzzy logic (Subasic and Huettner 2000), lexical affinity based on similarity of contexts, the basis for the build up of WordNet Affect (Strapparava and Valitutti 2004) and SentiWord-Net (Esuli and Sebastiani 2005), detection of affective keywords (Riloff and Wiebe 2003) and Machine Learning using term frequency (Pang, Lee and Vaithyanathan 2002); (Riloff, Wiebe and Phillips 2005).

1.2 TERMINOLOGY CLARIFICATION

One of the most challenging aspect of the branch of Natural Language Processing focused on the treatment of subjective data is that there has been to date no uniform terminology established for this relatively young field, where terms such as *emotion*, *sentiment*, *feeling*, *view*, ect. are employed in an interchangeably way.

The existence of this multiplicity of terms reflects the undeniable differences in the connotations that these terms carry in their original general-discourse usages. That is why a definition or a formalization of the basic terminology becomes essential to establish a common and unambiguous research framework.

Private states, sentiment, opinion, view, belief, conviction, persuasion

Wiebe (1994) one of the pioneer researchers in the treatment and exploitation of subjective data was influenced by theorists such as Banfield (1982) for defining her theory of *private states*.

She centred the idea of *subjectivity* on that of *private states* (which have been previously defined by Quirk (1985) and conceives them as:

> *the ones not open to objective observation or verification*. *For example a person may be observed to assert that God exists, but not to believe that God exists. Thus, belief is in this sense private*.

(Wiebe 1994) analysed private states in terms of their functional components: *experiencers* holding attitudes toward *targets*.

For example, in the private state in the sentence (1):

(1) *Jack loves his dog*

the experiencer is *John*, the attitude is *love*, and the target is *the dog*.

They created private state frames for three main types of private state expressions:

- Explicit mentions
- Speech events expressing them
- Expressive subjective elements

Thus, *opinions* (a view or judgement formed about something, not necessarily based on fact or knowledge), *evaluations* (a judgement about the amount, number, or value of something; assessment), *emotions* (a strong feeling deriving from one's circumstances, mood, or relationships with others), and *speculations* (the forming of a theory or conjecture without firm evidence) are grouped together into the category of the *private states*.

Apart from Wiebe (1994), other researchers defined the concept of private states.

According to Liu (2010) synonyms of private state can be considered: *opinion*, *view*, *belief*, *conviction*, *persuasion* and *sentiment (*with the connotations described above). According to him:

- Opinion implies a conclusion thought out yet open to dispute (each expert seemed to have a different opinion)
- View suggests a subjective opinion (very assertive in stating his views)
- Belief implies often-deliberate acceptance and intellectual assent (a firm belief in her party's platform)
- Conviction applies to a firmly and seriously held belief (the conviction that animal life is as sacred as human)
- Persuasion suggests a belief grounded on assurance (as by evidence) of its truth (was of the persuasion that everything changes)
- Sentiment suggests a settled opinion reflective of one's feelings (her feminist sentiments are well known)

Thus cording with K. R. Scherer (2005) the problem is how we can distinguish *emotions* from other affective phenomena such as *feelings, moods, or attitudes.* According with this researcher, using the term *feeling* (a single component that denotes the subjective experience process) as a synonym for *emotion* (the total multi-modal component process) produces confusions.

In the framework of the component process model, *emotion* is defined as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems as a consequence of the evaluation of an external or internal stimulus event (K. Scherer 1987 and 2001). The components of an emotion episode are the respective states of the five subsystems and the process consists of the coordinated changes over time.

> *Feelings integrate the central representation of appraisal-driven response organization in emotion reflecting the total pattern of cognitive appraisal and the motivational and somatic response patterning that underlies the subjective experience of an emotional episode (K. R. Scherer 2004).*

Scherer differentiates *emotion* (with feeling as one of its components) from other types of *affective phenomena*. Instances or tokens of these types, which can vary in degree of affectivity (i.e. *liking, loving, cheerful, contemptuous,* or *anxious*) are often called *emotions* in the literature but they should be distinguished from emotion. However there may be some overlap in the meaning of some words such as *preferences, attitudes, affective dispositions, and inter-personal stances*.

Scherer is convinced that: *the difficulty of differentiating emotion from other types of affective phenomena is a problem similar of defining the specificity of language in comparison with other types of communication systems.*

The term *sentiment* has been also employed with the connotation of automatic analysis of evaluative text and tracking of the predictive judgments in the works by Das and Chen (2001) and R. M. Tong (2001), researchers centred in market Sentiment Analysis.

In our research we will employ the terms of subjective data and subjectivity that includes the abovementioned terms and concepts defined by Scherer, Wiebe and Liu: *opinions, evaluations, speculations, views, beliefs, convictions, persuasions, sentiments and emotions.*

Sentiment Analysys and Opinion Mining

After having clarified the terminology we will employ in our study to refer to subjective data, the definition of the Natural Language Processing task that studies and treats subjectivity for concrete applications is essential in order to fully understand the purpose of our work. This discipline is Sentiment Analysis.

Liu (2010) defined Sentiment Analysis as:

> *Given a set of evaluative text documents D that contain opinions (or sentiments) about an object, opinion mining aims to extract attributes and components of the object that have been commented on in each document $d \in D$ and to determine whether the comments are positive, negative or neutral.*

In most of the previous research, the term Sentiment Analysis has been used in an interchangeably way with O*pinion Mining.* However, we believe that they cannot be considered as synonyms.

In fact, the first indicates the set of techniques to computationally treat subjective language, while the second one is focused on mining the subjective information for different purposes (not possible without the previous *Sentiment Analysis* process).

On the one hand, the popularity to the Sentiment Analysis terminology was given by Nasukawa and Yi (2003) who entitled their paper, *Sentiment analysis: Capturing favourability using natural language processing*, and another work by Yi, et al. (2003) titled *Sentiment Analyzer: Extracting sentiments about a given topic using natural language processing techniques*.

On the other hand, *Opinion Mining* firstly appears in a paper by Dave, Lawrence and Pennock (2003) presented at the *2003 WWW conference*[12] that gave popularity to such term within communities related to Web search or Information Retrieval.

According to Dave, Lawrence and Pennock (2003):

> *the ideal Opinion Mining* tool would *process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good).*

Esuli and Sebastiani (2006) definition of Opinion Mining is:

> *a recent discipline at the crossroads of Information Retrieval and Computational Linguistics, which is concerned not only with the topic a document treats, but also with the opinion it expresses*.

Much of the following research self-identified as *Opinion Mining* fits in these descriptions with its stress on extracting and analyzing judgments on various aspects of given items. However, the term has recently been employed more broadly including different types of analysis of evaluative text.

Thus, when a wider interpretation applies, Sentiment Analysis and Opinion Mining can be used to denote the same research field. However in this work we will employ them with their corresponding different meanings described above.

Both Sentiment Analysis and Opinion Mining are attracting huge interest from

12 www.2003.org/cdrom

different research areas, offering an extremely interdisciplinary context of work. This usefulness and interest in subjective data processing are also underlined by the growth of research laboratories and the European Commission funding programmes that we present in the section below.

1.3 Interest for sentiment analysis

sentiment Analysis is becoming extremely attracting for many research laboratories whose task is to build up real-life applications. Example of this can be the Affective Computing Research at MIT[13] by (Picard 1997), but also industrial ones can be AT&T[14].

Furthermore, the usefulness and importance *Sentiment Analysis* is demonstrated by the European Commission's numerous initiatives fostering such research.

We can mention the FP6 trans-European Network of Excellence HUMAINE 'Human–machine interaction network on emotions'[15] , the European projects PF-STAR 'Preparing future multisensorial interaction research'[16], AMI 'Augmented Multi-party Interaction'[17] and CHIL 'Computers in the Human Interaction Loop'[18].

Moreover, the 7th Framework Programme (FP7), fosters research on ICT also focused on subjective data treatment, together with the increasing number of workshops and special sessions on subjectivity in the new textual genres born with the Web 2.0 (on affective dialog, Embodied Conversational Agents, etc.).

Apart from the abovementioned initiatives that are the pioneers, the EC proposes a wide range of financing programmes focused on fostering research on Information and Communication Technologies and in most of the cases on subjects related with Natural Language Processing and Sentiment Analysis.

During these last years, the treatment of subjective data for the creation of concrete applications tested by target users and on the exploitation and dissemination of such products for the benefit of the EU society is a key topic of interest.

13 http://web.mit.edu/

14 www.att.com

15 www.emotion-research.fr

16 http://pfstar.itc.it/

17 http://amiprotocol.sourceforge.net/

18 http://chil. server.de/serlevts/is/101/

A) It is worth mentioning that research on ICT is not only financed on Theme 3 of the Cooperation programme of the 7th Framework Programme but it is also financed by means of:

- **The CAPACITIES Programme**: it supports the coherent development of policies, complements the Cooperation Programme, contributes to EU policies and initiatives to improve the coherence and impact of Member States policies, finds synergies with regional and cohesion policies, the Structural Funds, education and training programmes and the Competitiveness and Innovation Programme (CIP). (Research infrastructures[19] and International cooperation[20])

  - o **PEOPLE –Marie Curie[21]**: mobility and training of researchers. It gives high-level researchers the opportunity to carry out their own research teams in Europe. This action helps enhance the careers of these promising researchers by helping them attain research independence more rapidly.

  - o **Joint technology Initiatives JTIs[22]**: they aim to achieve greater strategic focus by supporting common ambitious research agendas in areas that are crucial for competitiveness and growth, assembling and coordinating at EU level a critical mass of research. They draw on all sources of R&D investment (public or private) and couple research tightly to innovation.

  - o **"PPP- Public-Private Partnerships":** As a remedy for the European financial and economic crisis 2008/09, a EU Economic recovery Plan 2010-2013 for Public-Private partnerships has been adopted. The three PPPs represent a powerful means for fostering research efforts in three large industrial sectors - automotive, construction and manufacturing - areas particularly affected by the economic downturn and

19 http://cordis.europa.eu/fp7/ict/e-infrastructure/home_en.html

20 http://cordis.europa.eu/fp7/ict/international/home_en.html

21 http://ec.europa.eu/research/mariecurieactions/

22 http://ec.europa.eu/research/participants/portal/page/cooperation?callIdentifier=ARTEMIS-2011-1

where innovation can significantly contribute towards a more green and sustainable economy. Synergies are planned with other FP7 themes to ensure higher impact. This is achieved notably with the three jointly funded Public-Private Partnerships (PPPs) of the European Economic Recovery Plan: Energy Efficient Buildings, Factories of the Future, and Green Cars. These PPPs are presented within the relevant ICT Challenges. They will, however, be called for separately in coordination with the other FP7 themes.

- **ERA- NET scheme[23]:** Its purpose is to foster the cooperation and coordination of research activities carried out at national or regional level in the Member States and Associated States by means of the networking of research activities conducted at national/regional level, and the mutual opening of national and regional research programmes. The scheme will contribute to improve the coherence and coordination across Europe of such research programmes. ERA-NET will also enable national systems to take on tasks in a collaborative manner (that they would not have been able to tackle independently). Both networking and mutual opening require a progressive approach. The ERA-NET scheme therefore has a long-term perspective that must also allow for the different way that research is organised in Member States and Associated States.

B) ICT basic research, financed by:

- **COST[24].** The ICT research area is best summarised as treating the processing, transmission, storage, retrieval, management, usage, and exchange of information and knowledge, with emphasis on fundamental aspects and pre-competitive technology development. New ideas and initiatives are welcome as well those with high interdisciplinary aspect.

23 http://ec.europa.eu/research/fp7/index_en.cfm?pg=eranet-projects

24 www.cost.esf.org/domains_actions/ict

– **IDEAS - ERC**[25] – without predefined subjects. The aim of the "Ideas" specific programme is to develop "exploratory research" to raise the level of excellence of research in Europe. The European Research Council (ERC) is has a key role in this programme.

o **ESF- European Science Foundation**[26]**.** Each year, the European Science Foundation announces a series of calls for proposals, which give the opportunity to propose collaborative research projects and networking activities with a European dimension. The calls span all fields of science through four main funding instruments, covering all types of scientific activities, from basic research and frontier science to networking and dissemination.

C) Research focused on the market/applied research:

– **CIP Programme**[27] **(ICT-PSP):** Information and Communication Technologies Policy Support fosters a wider uptake of innovative ICT based services and the exploitation of digital content across Europe by citizens, governments and businesses (in particular SMEs). The focus is placed on driving this uptake in areas of public interest while addressing EU challenges such as moving towards a low carbon economy or coping with an ageing society. The programme contributes to a better environment for developing ICT based services and helps overcome challenges such as the lack of interoperability and market fragmentation.

– **EUREKA – eurostars**[28]**.** Eurostars Programme is a European Joint Programme dedicated to the R&D performing SMEs and co-funded by the European Communities and 33 EUREKA member countries. It aims to stimulate these SMEs to lead international collaborative research and innovation projects by easing access to support and funding. It is fine-tuned to focus on the needs of SMEs, and specifically targets the development of new products, processes and services as well as the access to transnational and international markets.

25 http://erc.europa.eu/

26 www.esf.org/conferences/call

27 http://ec.europa.eu/cip/ict-psp/index_en.htm

28 www.eurostars-eureka.eu/

- **Ambient Assisted Living Joint Programme**[29]. Its objective is to enhance the quality of life of older people and strengthen the industrial base in EU by the use of Information and Communication Technologies (ICT). The motivation of the new funding activity is in the demographic change and ageing in Europe, which implies not only challenges but also opportunities for the citizens, the social and healthcare systems as well as industry and the European market. The concept of Ambient Assisted Living is understood as to extend the time people can live in their preferred environment by increasing their autonomy, self-confidence and mobility, to support maintaining health and functional capability of the elderly individuals, to promote a better and healthier lifestyle for individuals at risk, to enhance the security, to prevent social isolation and to support maintaining the multifunctional network around the individual, to support carers, families and care organisations, to increase the efficiency and productivity of used resources in the ageing societies.

D) Other initiatives

- **Pre-Commercial procurement (PCP)[30]:** it is an approach for procuring R&D services which enables public procurers to: share the risks and benefits of designing, prototyping and testing a limited volume of new products and services with the suppliers, without involving State aid, create the optimum conditions for wide commercialization and take-up of R&D results through standardization and/or publication.

E) European Cohesion programmes

FEDER with ICT section:

- Transnational Cooperation:
  - **MED programme[31]:** A EU transnational cooperation programme among the "Territorial Cooperation objective" of the EU Cohesion Pol-

29 www.aal-europe.eu

30 http://cordis.europa.eu/fp7/ict/pcp/home_en.html

31 www.programmemed.eu

icy. Partners from 13 countries including the whole Northern Mediterranean seacoast work together to strengthen the competitiveness, employment and sustainable development of this area. The transnational setup allows the programme to tackle territorial challenges beyond national boundaries, such as environmental risk management, international business or transport corridors.

- o **European South East (SUDOE)[32].** The public actors of the Spanish, French, Portuguese and British (Gibraltar) regions can contribute to the sustainable development of the Southern Western Europe Area developing transactional cooperation projects in innovation and environment, new information technologies and sustainable urban development. Working together, these regional actors contribute to the updating of the Southern Western Europe Area in line with the European Union in terms of development, employment and sustainable development. Interregional Cooperation.
- o **Interreg IVC[33].** The overall objective of the INTERREG IVC Programme is to improve the effectiveness of regional policies and instruments. A project builds on the exchange of experiences among partners who are ideally responsible for the development of their local and regional policies. The areas of support are innovation and the knowledge economy, environment and risk prevention. Thus, the programme aims to contribute to the economic modernisation and competitiveness of Europe. INTERREG IVC is linked to the objectives of Lisbon and Gothenburg agendas. Typical tools for exchange of experience are networking activities such as thematic workshops, seminars, conferences, surveys, and study visits. Project partners cooperate to identify and transfer good practices. Possible project outcomes include for example case study collections, policy recommendations, strategic guidelines or action plans.
- o **ESPON 2013 Programme[34]:** The applied research within the ESPON 2013 Programme aims at improving facts and evidence on European territorial

32 www.interreg-sudoe.eu/ESP

33 http://i4c.eu/accueil_en.html

34 www.espon.eu/main/Menu_Calls/Menu_Calls/Menu_PreAnnouncement/PreAnnouncementCallsAug11.html

structures, trends, perspectives and policy impacts. A particular focus is given to territorial potentials and challenges for a successful development of regions and cities of Europe. Cross-thematic applied research represents a major activity integrating existing thematic analysis and of new themes. A territorial approach is applied in these projects integrating relevant sectors accordingly. The applied research themes chosen deal with socio-economic as well as ecological issues that are always addressed in a territorial context, providing a European wide coverage of comparative information on regions and cities. The applied research also takes up territorial phenomena, such as urban structures, potential accessibility and urban sprawl, in order to enrich policy development with further elements relevant for territorial development and cohesion. The impact of EU policies is another area of applied research within ESPON. Projects support policy makers with information on impacts of concrete EU sector policies as well as tools for the ex-ante assessment of impacts of policy initiatives, in strategy documents and in EU Directives. The applied research actions are supported by a Knowledge Support System that plays the role of ensuring the scientific quality of such research. A pool of experts from all over Europe provides the basis for selecting experts for Sounding Boards following the projects throughout their lifetime and assessing the final research results.

– **Cross-border Cooperation:**
    - **Cuenca Mediterránea**[35] **Programme**. The multilateral cross-border cooperation "Mediterranean Sea Basin Programme" is part of the new European Neighbourhood Policy (ENP) and of its financing instrument (European Neighbourhood and Partnership Instrument - ENPI) for the 2007-2013 period: it aims at reinforcing cooperation between the EU and partner countries regions placed along the shores of the Mediterranean Sea. 14 participating countries, which represent 76 territories and around 110 million people, are considered eligible under the Programme: Cyprus, Egypt, France, Greece, Israel, Italy, Jordan, Lebanon, Malta, Palestinian Authority, Portugal, Spain, Syria and Tunisia.

35 www.enpicbcmed.eu

1.4 The new textual genres

in parallel with the explosion of the Web 2.0 the new textual genres such as forums, online reviews or blogs have increased at an exponential rate. They are employed by an extremely high number of uses all over the world originating a massive phenomenon in which people create a huge amount of subjective data in real time.

As any other textual genre, also forums, online reviews and blogs have their own feature that we briefly present below:

- **Forums** are online communities in which users can read but also post topics of common interest regarding a certain topic. They can be a useful tool for anyone doing business online who read the content or actively participates in the discussions. In fact analysing forum's archives can be effective to obtain a basic knowledge about a topic, but also a historical perspective on trends and opinions. A user can participate in forum with different modalities: member, moderator, or creator.
- **Online reviews** are critical texts that go beyond the simple price comparison sites offering numerous in-depth reviews and ratings on different products' features and are often submitted by site visitors. The shopping comparison sites like Kelkoo[36] has reviews, ratings and good introductory buyers guides. Other examples can be Review Centre[37], Reevoo[38] or Swotti[39].
- **Blogs** can be defined as a frequent, chronological publication of personal thoughts and Web links. Their content consists in a mixture of what is happening in a person's life and on the Web, thus they are like hybrid diary/guide sites. People maintained blogs long before the term was coined, but the trend gained its peak with the introduction of automated published systems, most notably Blogger[40] used daily by thousands of

36 www.kelkoo.es

37 www.reviewcentre.com

38 www.reevoo.com

39 www.buzztrend.com/es

40 www.blogger.com

people to simplify and speed-up the publishing process. Blogs are alternatively called web logs or weblogs.

The common feature of these new textual genres is that by means of them users generate a genuine and new language of public discourse and the amount of data raises each day exponentially.

In particular, the growing and predominant importance of blogs (if compared with forums and online reviews) is widely demonstrated by Technorati *State of the Blogosphere 2010*[41] that describes it as a consolidated textual genre, no longer in an upstart phase (as mentioned in the previous year survey).

Bloggers' use and engagement with various social media tools is increasing, making the demarcations between blogs, micro-blogs and social networks disappearing. User who create and post in blogs are called *bloggers* and they can be grouped in four different classes/profiles:

– **Hobbyists**, who blog just for fun are the majority in the blogosphere and represent the 64% of respondents.

– **Self-Employed** are the 21% of the total. They blog full time or occasionally for their own company or organization. 57% say they own a company and have a blog related to their business, while 19% report that their blog is their company. 65% say they manage their blog by themselves.

– **Part-Timers** are 13% of the blogosphere and they devote significant time to their blogs, with 61% of them who spends more than three hours blogging each week, and 33% updates their blog at least once a day.

– **Corporates**, 1% of respondents, blog full-time for a company/organization. However, only 24% of them report spending a full 40 hours per week blogging, and only half report that they receive incomes.

Regarding their profile, generally bloggers are a highly educated (half of them are graduated) and the average is that one blogger has three or more blogs from two or more years. This means that the blogosphere has an undeniable key position in today's society.

41 http://technorati.com/blogging/article/state-of-the-blogosphere-2010-introduction/

Generally, one single author creates a blog. However, there are exceptions for collective ones, where contributors post and debate short essays and opinion pieces. The comments are the users' point of view and their interpretations of an event/product, content and context. Regarding their common format, they consist in a series of entries posted into a single page in reverse-chronological order. As we mentioned above, they generally represent the personality of the author or reflect the purpose of the Web site that hosts the blog. Topics sometimes include brief philosophical musings, commentary on Internet and other social issues, and links to other related sites, especially those that support a point being made on a post[42].

Blogs are a more democratic form of expression than newspapers. However in some ways they are more exclusionary also because they address about a tenth of the people who use the Web. That is why the high, formal style of the newspaper may be nobody's native language, but at least it's a neutral voice that doesn't underline the speech of any particular group or class. On the contrary the blog style can be seen as an adaptation of an informal conversation.

In this context, Sentiment Analysis has a key role, since traditional systems are not useful enough to capture and process the new features the subjective discourse offers. This is why most of the Natural Language Processing resources and tools created until now have been built up thinking on the feature and needs of canonical and objective discourse (as the one employed in newspapers), thus they are not prepared for the linguistic variations the Web 2.0 presents.

In general, blogs are written in a mixture of styles where the most predominant is the informal one. The language we encounter is casual, familiar, and generally colloquial, thus more direct than a formal register with consequences such as abbreviations, short sentences, ellipses, as well as use of colloquialisms and sayings that depend on the user's profile and background (Gouws, et al. 2011).

Apart from this, we can also add that they have peculiar features, which distinguish them from other types of genres (Paquet 2002) such as personal editorship, a hyperlinked posting structure, frequent updates, free public access to the content via the Internet, and archived posts are the most important aspects to take into account. After having presented blogs, the next section is dedicated to describe the blog's language features in detail.

42 www.searchwindevelopment.techtarget.com/definition/blog

Main features of blog language

Due to the importance blogs achieve in the present context, it is important to be aware about their language features. The most significant are presented below.

– **Huge amount of data:**

Subjective data contained in the new textual genres is growing at an exponential rate, making impossible for users who are searching for specific summarised opinions about a certain topic to find it (or an overview of it) in a short time.

Due to this, one of the main important focus of Sentiment Analysis is to find effective methods and techniques to mine the information on the web, classify it according to its opinion holder (person who express the opinion), target (object of the discourse) and opinion polarity. After that, the next step will consist in presenting such information in a summarised/organised manner so that user can save time instead of manually checking and discriminating the information found.

– **Real time creation:**

The second challenge subjective data present is the fact that it is in continuous creation and updated in real time. On the one hand this feature is a good and beneficial aspect of the subjective information, since it allows the user to have at his disposal new and updated information for taking its decisions. However, on the other hand, this requires an active system of information retrieval, which explores the WWW data in a constant manner.

– **Wide range of topics and sources:**

The third feature that characterises subjective information is the fact that it is about a wide range of topics, since people write whatever is on their minds, thus generating an infinite number of topics with their associated subtopics.

This makes the discourse in most of the times unpredictable, requiring high performance systems able to detect the exact topic to which a specific private state is referring. In many cases we have also the situation that more topics are in the same sentence such as in example 2:

(2) I love the Samsung mobile, but IPhone is better.

As we can see in example (2) we have two targets in the same sentence and thus we require the systems to detect the topics and to understand the private state associated to each one of them.

A similar problem also applies for the multiple sources. In fact in many cases we have a mixture of sources (3).

(3) *My friend says he likes the iPhone but I prefer a Motorola.*

As we can see in example (3) two are the opinion holders in the sentence: *My friend* and *I.* Thus, similar to the mixture of target problem, the system needs to discriminate one source from another and after that associate to them their corresponding private states.

- **Mutilinguality:**

Due to the globalisation and the fact that the WWW reaches nearly all the places in the world and also due to the huge mobility of people, an additional challenge subjective analysis has to tackle is the multilingualism. English remains the most employed language. However Spanish, Arabic or Chinese are also widely employed and are gaining more and more users and others languages produce less data, but they also need to be treated.

- **Language Style:**

Strictly related with language, we also encounter the problem of style. The positive and attracting feature of language is that it is a powerful tool. *It shapes our understanding of others' mental states and has a pervasive online effect in emotion* (Lindquist 2009). It is a very precious and useful instrument but we also have to bear in mind that different languages imply different culture and interpretation of life. Thus in many cases the possible solution of Machine Translation is not enough.

Subjectivity can be expressed not only by single words but also by means of sayings and colloquialisms. They are unique for any language and also the same sentence in an equivalent context can have different meanings in two languages. Thus, this aspect represents a complex problem, since it does not only involve linguistic aspects but also the social context with all its implications.

- **Multimodal information:**

Multimodality is strictly related to style informality. With this term we are referring to the use of additional elements such as emoticons, which can give a special shape to the discourse that in the majority of the cases is mitigation. If we consider the following example (4):

(4) Your paper is a little confusing ;)

In this case the opinion holder is criticising the paper but by means of adding the emoticon he wanted to lower the intensity of the negative private state expressed. We can deduce that understanding the proper way to interpret the sentences including such elements is another key issue for a correct understanding and processing of subjective language.

Analysing the numerous challenges the subjective information brings, there is clear evidence that the language of the Web 2.0 is extremely challenging but at the same time a fascinating subject of research.

As we mentioned above, due to the complexity of the area, we will mainly concentrate on blogs since we believe that they represent a considerable amount of the Web 2.0 content with high effect on our society. Technorati survey in 2008 is tracking over 112.8 million blogs, a number that does not include all the 72.82 million Chinese blogs as counted by The China Internet Network Information Centre[43]. Blog statistics often concern the English language blogosphere and thus, we should not forget about the millions of other blogs that are not always included in such estimations.

Thus, they represent one of the most important sources of real-time, unbiased information, which can be exploited to develop many practical applications for diverse users with different profiles and needs.

Examples of such applications could be the business of brand image monitoring by means of which a company analyses the opinions of its clients and external people. Further options could be the detection of opinions regarding the competitors or the evaluation of client opinion on a product depending on their needs and experiences. Another scenario could be social one in which blogs can be exploited for monitoring attitudes for behavioural or psychological purposes or studies. Finally, an innovative way of exploiting such data could be the prediction of relevant social or economic tendencies by monitoring information. For example, the recent worldwide economic crisis could have been monitored via opinion gathering helping the people in charge to take the appropriate preventive measures.

43 www.cnnic.net.cn/en/index/

These are just a few examples of the potential options for real applications that the subjective data offers. As we can deduce, the number of situations in which Natural Language Processing techniques for subjective information can be employed is very high, due to the wide range of topics bloggers discuss daily.

For all of the abovementioned reasons, there has been increasing interest from researchers to develop methods to extract data from the subjective information available from this new source.

1.5 Subjectivity extraction and classification

As we can deduce from above, subjective data constitutes an essential source of information (Cui, Mittal and Datar 2006) and it is becoming a reference point for more and more people. That is why different authors have addressed the problem of extracting and classifying subjective data from different perspectives and at different levels, depending on a series of factors, which can be level of interest such as:

– **Overall/specific (5 and 6)**:

(5) *What do you think about iPhone?*

(6) *How is the iPhone screen?*

The first question is asking for a general opinion about a product, while the second seeks for an answer regarding a specific feature of the iPhone, its screen.

– **Querying formula (7 and 8)**:

(7) *Nokia E65*

(8) *Why do people buy Nokia E65?*

The first case it more general, while the second is explicitly asking for the peoples' opinion about this product. The question is the same but posed in different manners. Each one of the two examples could be seen as factual or opinionated queries.

– **Text type** (review on forum/blog/dialogue/press article)

Each textual genre has its own features and difficulties for its language treatment.

– **Manner of expression of opinion**:

Direct -using opinion statements- (9):

(9) I think this product is wonderful! This is a bright initiative

This is probably the simplest way of detecting private states since they are explicitly expressed thus being easy to detect.

Indirect -using affect vocabulary- (10):

(10) I love the pictures this camera takes!/Personally, I am shocked one can propose such a law!

In most of the cases here, the context is a key issue for the correct interpretation of the private state.

Implicit -using adjectives and evaluative expressions- (11):

(11) It's light as a feather and fits right into my pocket!

This is the most challenging case in which a language and context interpretation is needed.

While determining the overall opinion on a movie is sufficient to decide to watch it or not, when buying a product, people are interested in the individual opinions expressed about the different product features. Or for example, when discussing about a person, one can judge and give an opinion on the person's actions. Thus, the approaches taken can vary depending on the way in which a user asks for the data:

– **General formula (12):**

(12) What is your opinion about the iPhone 3G?

– **Specific question (13):**

(13) Why do people like X

In these cases the text source needs to be queried is crucial and we also have to group the different private states according to their polarity.

1.6 Conclusions

In this chapter we mainly presented the context of our research underlying why it is so important by mentioning its possible applications and also listing and describing the numerous EU funding programmes that are in charge of fostering such topic.

Since it is a quite young research area, we clarified the terminology that we will use during this research and also specified why we decided to work with three languages, why we choose them and why we selected blogs as textual genre for our research.

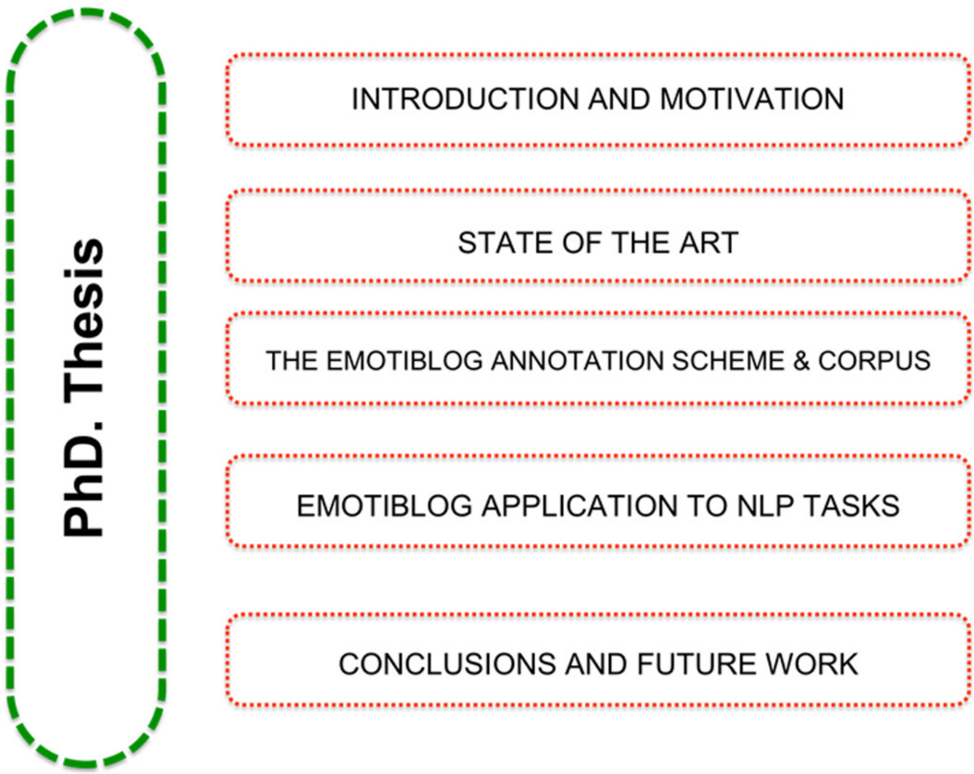
From the challenges we described through the section above, it is evident that an effective treat, process and interpretation of subjective data requires the use of specialised system training, tuning and tested within the different text spheres (and with these different elements).

In order to lessen and counteract the challenges mentioned, underpinned by the large quantity of subjective data available and the lack of resources to exploit such precious information, the general purpose of the research we carried out is focused on to:

- **Encounter** a fine-grained annotation schema able to capture the linguistic shapes of affective expression in non-traditional textual genres and above all in blogs.
- **Annotate**, if not existing, a collection of blog posts using the fine-grained annotation schema.
- **Evaluate (intrinsically)** the robustness of the annotation scheme creating Machine Learning models using the annotated elements and measuring the impact of each of its elements.
- **Test (extrinsically)** the efficiency of the model and annotated collection by means of applying it to the following Natural Language Processing tasks: such as Opinion Mining, Opinion Question Answering and Automatic Summarisation.

In order to have a clear view of how our work has been carried out, what are the steps we took and what makes our work contributing to the state of the art, the following section presents the main outline of this Thesis presenting briefly the content of each chapter.

Outline



As we can see from above, this work has been divided into 10 chapters. Each one has a parallel structure composed by an introduction, the development of the chapter and some conclusions.

After the introduction of our research in chapter 1, the following chapters are:

– **Chapter 2** presents the State of the Art in Sentiment Analysis and its application to the Natural Language Processing tasks with which we will work: Opinion Mining, Question Answering and Opinion Summarisation.

- **Chapter 3** presents in detail our EmotiBlog-Annotation-Model and the annotation process we carried out.
- **Chapter 4** describes the intrinsic evaluation we carried out to test the robustness and validity of the EmotiBlog-Annotation-Model model with its elements and thus to test if the annotation would be reliable and easy to apply.
- **Chapter 5** continues with the intrinsic evaluation of the EmotiBlog-Annotation-Model and it will focus on feature selection experiments for the three languages that compose EmotiBlog: English, Spanish and Italian. With this our aim is to check if the EmotiBlog annotation is useful to train Machine Learning algorithms.
- **Chapter 6** is an introductory chapter previous to the extrinsic evaluation. It presents the tools, corpus and methods for text processing we will employ later in the following chapters. Our aim here was to clarify some technical issue before describing the experiments in which our resource was employed.
- **Chapter 7** presents the experiments we carried out in the Opinion Mining, Question Answering and Automatic Summarisation focused on checking if the EmotiBlog-Annotation-Model and Annotated-Corpus can be a useful resource to contribute to the improvement of the performance dealing with such tasks.
- **Chapter 8** summarises the final conclusions of our research work and presents our future work.
- The **Spanish summary** offers an general overview of our work underlying our motivation, the experiments we carried out, the improvement we bring to the state of the art in Sentiment Analysis and our conclusions.

2. State of the art

Due to the increasing interest in Sentiment Analysis the last few years have seen a considerable explosion in this field of research. As if for factual data tools and methods have been developed and are quite tested, the effective analysis and treatment of subjective data still represent an important challenge to overcome.

Even if it is a relatively new research area, many are the works already done and in this chapter our idea is to provide an exhaustive overview above all on the resources and tools to analyse the subjective content as well as very quick snapshots of the application of such resources to the Natural Language Processing tasks we will employ for the extrinsic evaluation of EmotiBlog-Corpus: Opinion Mining, Opinion Question Answering and Automatic Summarisation.

As we mentioned in the introductory section resources for sentiment analysis can be grouped according to different criteria and in our work, in order to be clear and understand each resource potential and features, we decided to classify them depending on their:

- **Level of sentiment detection** granularity
- **Creation process** (manual or semi/automatic)
- **Language**, **domain** and **size**

After that, we also present the resources employed in Opinion Mining, Opinion Question Answering and Opinion Summarisation with the objective of analysing what has been done and what we can improve with our research.

The final part of the chapter summarises the most important weak points and aspects to improve in this research framework.

2.1 Levels of sentiment detection granularity

Many are the resources created for Sentiment Analysis. The first criteria by which we present them is the granularity of the subjective information they are able to capture. The resources presented below are listed from the finer to the coarse-grained.

The pioneer study and pillar resource for subjectivity analysis is represented by the **Multi-Perspective Question Answering (**MPQA) corpus[44] (Wiebe, Wilson and Cardie 2005); (Wiebe and Wilson 2005); (Wilson and Wiebe 2003), a collection of English newspapers extracts. It is composed by 10000 sentences annotated with their polarity feature using the homonymous annotation schema. In this syntactic labelling scheme the subjectivity detection is based on the theory on private states (Wiebe 1994).

According to authors, each subjectivity expression includes its source, target and properties with their corresponding intensity, significance and the type of attitude. For example, for the private state in the sentence (14):

(14) Mark loves his dog

The experiencer is "*Mark*," the attitude is "*love*," and the target is "*his dog*."

They create private state frames for three main types of private state expressions in text:

- Explicit mentions of private states
- Speech events expressing private states
- Expressive subjective elements

The second annotated type of element is the objective speech event, which includes annotations of objective events, with no shadow of subjectivity. They are considered as facts.

The **Cornell movie-review**[45] collection was first introduced by Pang and Lee (2004). It is composed by different datasets, which include automatically derived labels. Sentiment polarity datasets (1000 positive and 1000 negative processed reviews + 5331 positive and 5331 negative processed sentences/ snippets). Sen-

44 www.cs.pitt.edu/mpqa/databaserelease

45 www.cs.cornell.edu/people/pabo/movie-review-data

timent-scale datasets (a collection of documents whose labels come from a rating scale). Subjectivity dataset (5000 subjective and 5000 objective processed sentences).

The **NTCIR[46] multilingual corpus** has been used for the Multilingual Opinion-Analysis Task (MOAT) at NTCIR6, 2006. It consists in a collection of twenty topics and the training data contains annotations regarding opinion, opinion holder and sentiment polarity, as well as relevance information for a set of pre-defined topics.

In the **SemEval 2007 Task 18 –Affective Text** (Strapparava and Mihalcea 2007), the task participants had at their disposal a training set composed by 1000 news headlines provided by the task organizers. Emotions (e.g. joy, fear, surprise) and/or for polarity orientation (positive/negative).

SentimentWortschatz (SentiWS) (Remus, Quasthoff and Heyer 2010) is a publicly available German-language resource for sentiment analysis, opinion mining etc. It lists positive and negative sentiment bearing words weighted within the interval of [-1; 1] plus their part of speech tag, and if applicable, their inflections. The current version of SentiWS (v1.8b) contains 1650 negative and 1818 positive words, which sum up to 16406 positive and 16328 negative word forms, respectively. It not only contains adjectives and adverbs explicitly expressing a sentiment, but also nouns and verbs implicitly containing one.

Another relevant resource in the field is **WordNet Affect** (Strapparava and Valitutti 2004), an extension of WordNet Domains, including a subset of synsets suitable to represent affective concepts correlated with affective words. It has been developed semi-automatically by authors who assigned affect labels to words in WordNet and then expanded the lists using WordNet relations such as synonymy, antonym, entailment, and hyponymy. It includes semantic labels based on psychological and social science theories (Ortony, Clore and Collins 1988), (Elliot 1992) and (Ekman, Basic Emotions 1999) valence (positive or negative), arousal (strength of emotion). The 2004 version covers 1314 synsets, 3340 words.

SentiWordnet (Esuli and Sebastiani 2006) is freely available for research purposes with a Web-based graphical user interface. Each WORDNET[47] synset is as-

46 http://research.nii.ac.jp/ntcir/index-en.html

47 http://wordnet.princeton.edu

sociated to three numerical scores Obj(s), P os(s) and Neg(s), describing how objective, positive, and negative the terms contained in the synsets are. The method used to developing it is based on the quantitative analysis of the glosses associated to synsets, and on the use of the resulting vectorial term representations for semi-supervised synsets classification. The three scores are derived by combining the results produced by a committee of eight ternary classifiers, all characterized by similar accuracy levels but different classification behaviour.

The **General Inquirer** is a system created by (Stone, et al. 1966), that provides English-language content analysis using both the "Harvard" and "Lasswell"[48] general-purpose dictionaries as well as any dictionary categories developed by the user. It consists in a computer-assisted approach for content analyses of textual data not designed ad-hoc for subjectivity analysis. It contains 11788 sense-disambiguated words, out of which the subjective ones are annotated with polarity, strength and according to axes of emotion. The resource was created manually.

One relevant work as far as lexical resources is concerned, is the **Opinion Finder** lexicon (subjectivity clues) (Wilson, Hoffmann, et al. 2005) that performs subjectivity analysis, automatically identifying when opinions, sentiments, speculations and other private states are present in text. This resource aims to identify subjective sentences and to mark various aspects of the subjectivity, such as the source of the subjectivity and words that are included in phrases expressing positive or negative sentiments. It was built starting with the grouping of the subjectivity clues in Riloff and Wiebe (2003) and then enriched with polarity annotated subjective words taken from the General Inquirer and the lexicon proposed by Hatzivassiloglou and McKeown (1997)-8000 words-.

Micro-WNOP (Cerini, et al. 2007) is a corpus composed by 1105 WORDNET synsets divided into three parts: Common part. (110 synsets, which the 5 evaluators have evaluated all together to align their evaluation criteria). Group 1 (part. 496 synsets which have been evaluated by a group of three evaluators- Each evaluator has performed this part of the evaluation independently from the other ones). Group 2 (part. 499 synsets which have been evaluated by the remaining two evaluators- Each evaluator has performed this part of the evaluation independently from the other one). Two criteria have been adopted in the construction of the corpus: Opinion relevance (the corpus should

48 www.wjh.harvard.edu/~inquirer/lasswell.htm

contain enough synsets which are relevant respect to the opinion topic) and WORD-NET representativeness (the POS of the synsets in the corpus should be representative of the distribution of the synsets among the four POS).

The **Emotion Triggers** (Balahur and Montoyo 2008) are words or concepts expressing an idea which, depending on the reader's interest, cultural, educational and social factors, leads to a possible emotional interpretation of the text content. They are lexicons, which contain single words, whose polarity and emotions are not necessarily those, which are annotated within the resource in a larger context. The underlying difference between the abovementioned studies and our work resides in the fact that we annotate larger text spans to be able to consider the undeniable influence of context.

The **ISEAR corpus** (Scherer and Wallbott 1997) consists in a collection of phrases where people describe a situation when they felt a certain emotion. It is a real-life self- expressed emotion collection.

The **CINEMO** corpus (Brendel, Zaccarelli and Deuvillers 2010) of French emotional speech provides a richly annotated resource to help overcome the lack of learning and testing speech material for complex (blended or mixed emotions). The protocol for its collection was dubbing selected emotional scenes from French movies. 51 speakers are contained and the total speech time amounts to 2 hours and 13 minutes and 4k speech chunks after segmentation. Extensive labelling was carried out in 16 categories for major and minor emotions and in 6 continuous dimensions.

Annotation Scheme and Gold Standard for Dutch Subjective Adjectives (Maks and Vossen 2010) (Isa Maks, Piek Vossen). Gold standard for Dutch subjective adjectives. whether it expresses an opinion or attitude, or is factual.

The University of Glasgow built up the **TREC test collection**[49], consisting of blog posts (100649 blogs have been selected) over a range of topics. This collection differs from the standard Web test collections since no new blogs entries were added to the corpus after the first day of the crawl process. The blogs to be included in the collection were pre-determined before the outset of the fetching phase. Blogs came from several sources: top blogs (70701), splogs (17969)

49 http://ir.dcs.gla.ac.uk/test collections/ access to data.html

and other blogs (11979). The content of the Blogs06 collection was fetched over an eleven-week period (6th December 2005 - 21st February 2006). Assessments include relevance judgments and labels as to underline if posts contain relevant opinions and about the polarity of the opinions (positive, negative, or a mixture of both). (Ounis, et al. 2006).

Jijkoun and Hofmann (2009), created a **gold standard for Dutch subjectivity words**. The data set includes 1916 adjectives, which are annotated for 3 polarity categories (positive/negative/neutral) by 2 annotators.

Another annotation scheme and corpus for **subjectivity versus objectivity** classification, as well as polarity determination at sentence level was developed by Yu and Hatzivassiloglou (2003) in a semi-automatic manner. The authors start from a set of 1336 seed words, manually annotated by Hatzivassiloglou and McKeown (1997) which extended by measuring co-occurrence between the known seed words and new words. The hypothesis on which the authors based their approach is that positive and negative words have the tendency to co-occur more than it is expected by chance. As measure for association, the authors employ log-likelihood on a corpus that is tagged at the part-of-speech level.

In this paper, we propose **GermanPolarityClues** (Waltinger 2010), a new publicly available lexical resource for sentiment analysis for the German language. The manually finalized GermanPolarityClues dictionary offers thereby a number of 10141 polarity features, associated to three numerical polarity scores, determining the positive, negative and neutral direction of specific term features. Learning methods exhibits for both languages the best performance (F1: 0.83-0.88).

This paper presents **Q-WordNet** (Agerri and García-Serrano 2010), a lexical resource consisting of WordNet senses automatically annotated by positive and negative polarity. Polarity classification amounts to decide whether a text (sense, sentence, etc.) may be associated to positive or negative connotations.

Congressional floor-debate transcripts[50]: (Thomas, Pang and Lee 2006) is composed by speeches as individual documents together with labels (generated automatically) for the cases in which the speaker is for or against the legislation discussed in the debate, allowing for experiments with this kind of senti-

50 www.cs.cornell.edu/home/llee/data/convote.html

ment analysis, Indications of which "debate" each speech comes from, allowing for consideration of conversational structure, Indications of by-name references between speakers, allowing for experiments on agreement classification if one assigns gold-standard agreement labels from the support/oppose labels assigned to the pair of speakers in question.

Comlex (Macleod, Grishman and Meyers 1994) is a dictionary containing 38000 words for English. It also includes a wide number of attitude adverbs namely entries for approximately 21000 nouns, 8000 adjectives and 6000 verbs, all of which are marked with a rich set of syntactic features and complements. This resource was developed in the framework of the Proteus Project[51] at New York University at the Linguistic Data Consortium[52]. It contains fine-grained syntactic information.

Customer review datasets[53] introduced by Hu and Liu (2004), is a collection of reviews about five electronics products downloaded from Amazon[54] and Cnet[55]. The sentences have been manually labelled as to whether an opinion is expressed, and in positive case the feature from a pre-defined list is being evaluated. An addendum with nine products is also available[56]. A comparative-sentence dataset is available on request.

Review-search results sets[57]. This corpus, used by Pang, Lee and Vaithyanathan (2002), includes the top 20 results returned by the Yahoo! search engine in response to each of a set of the 69 queries containing the word "review". The queries were drawn from the free list of real MSN users' queries released for the 2005 KDD Cup competition. The search-engine results in the corpus are annotated as to whether they are subjective or not.

The **Whissell's Dictionary of Affect in Language** (Sweeney and Whissell 1984); (Whissell and Dewson 1986); (Whissell and Charuk 1985) contains affec-

51 http://nlp.cs.nyu.edu/index.shtml

52 www.ldc.upenn.edu

53 www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip

54 www.amazon.com

55 www.cnet.com

56 www.cs.uic.edu/~liub/FBS/Reviews-9-products.rar

57 www.cs.cornell.edu/home/llee/data/search-subj.html

tive norms for English Words (Bradley and Lang 1999) and Sentiment-bearing adjectives by (Hatzivassiloglou and McKeown 1997) Automatically-created Lists.

Mathieu (2006) built up a **computational semantic lexicon of French verbs** of **feeling, emotion, and psychological states** is presented here, as well as FEELING, a software program using this lexicon to provide an interpretation and to generate paraphrases. Semantic representations are described by means of a set of feature structures. Sixty newspapers "letters to the Editor" were taken as a domain for the evaluation of this work.

Economining[58] is a site, hosted by the Stern School at New York University[59], which includes three sets of data: Transactions and price premiums as well as feedback postings for merchants from Amazon.com. automatically derived sentiment scores for frequent evaluation phrases at Amazon.com. These formed the basis for the work reported in (Ghose, Ipeirotis and Sundararajan 2007), which focuses on interactions between sentiment, subjectivity, and economic indicators.

Multiple-aspect restaurant reviews[60] (Snyder and Barzilay 2007), is composed by 4488 reviews, both in raw-text and in feature-vector. Each review gives an explicit 1-to-5 rating for five different aspects — *food, ambiance, service, value, and overall experience* — along with the text of the review itself, all provided by the review author. A rating of five was the most common over all aspects, and creators report that 30.5% of the 3488 reviews in their randomly selected training set had a rating of five for all five aspects, although no other tuple of ratings was represented by more than 5% of the training set.

Multi-Domain Sentiment Dataset[61] (Blitzer, Dredze and Pereira 2007), consists of product reviews from several different product types taken from Amazon.com, some with 1-to-5 star labels, some unlabeled from many product types (domains). Some domains (books and DVDs) have a high number of reviews. Others (musical instruments) have only a few hundred. Reviews contain star ratings (from 1 to 5 stars).

58 http://economining.stern.nyu.edu/datasets.html

59 www.stern.nyu.edu

60 http://people.csail.mit.edu/bsnyder/naacl07

61 www.cis.upenn.edu/~mdredze/datasets/sentiment

Table I.
Resource classification by annotation granularity
(from finer-grained to coarser-grained)

NAME	ANNOTATION	REFERENCE
MPQA	• Objective speech event • Subjective speech event – source – target – properties with their intensity, significance and attitude	(Wiebe and Wilson 2005)
Cornell movie-review	• Sentiment polarity datasets – positive – negative) • Sentiment-scale datasets – rating scale • Subjectivity dataset – subjective – objective	(Pang and Lee 2004) (Pang, Lee and Vaithyanathan 2002)
The NTCIR[62] multilingual corpus	• Opinion • Opinion holder • Sentiment polarity • Relevance information (using a set of pre-defined topics)	http://research.nii.ac.jp/ntcir/index-en.html
SemEval 2007 Task 18 – Affective Text	• Emotions – (e.g. joy, fear, surprise) • Polarity orientation	(Strapparava and Mihalcea 2007)
SentimentWortschatz	• positive and negative sentiment bearing words – weighted within the interval of [-1; 1] – their part of speech tag – their inflections	(Remus, Quasthoff and Heyer 2010)
WordNet Affect	• Semantic labels • Valence – positive or negative • Arousal (emotion strenght)	(Strapparava and Valitutti 2004)
Sentiwordnet	• objective, • positive/negative	(Esuli and Sebastiani 2006)
General Inquirer	• subjective – polarity, – strength – axes of emotion	(Stone, et al. 1966)

62 http://research.nii.ac.jp/ntcir/index-en.html

NAME	ANNOTATION	REFERENCE
Opinion Finder	• source of the subjectivity • words included in phrases expressing positive/negative sentiments	(Wilson, Hoffmann, et al. 2005)
Micro-WNOP	• Opinion relevance • WORDNET representativeness	(Cerini, et al. 2007)
Emotion triggers	• polarity • emotions	(Balahur and Montoyo 2008)
ISEAR corpus	• Emotion	(Scherer and Wallbott, The ISEAR Questionnaire and Codebook 1997)
CINEMO	• Major and minor emotions	(Brendel, Zaccarelli and Deuvillers 2010)
Gold Standard for Dutch	• opinion or • attitude • factual	(Maks and Vossen 2010)
TREC test collection	• relevant opinions • Polarity – positive, negative, mixture of both.	http://ir.dcs.gla.ac.uk/test collections/ access to data. html
Gold standard for Dutch subjectivity words	• positive • negative • neutral	(Jijkoun and Hofmann 2009)
	• subjectivity • objectivity classification, • polarity determination	(Yu and Hatzivassiloglou 2003)
GermanPolarityClues	• positive, negative neutral	(Waltinger 2010)
Q-WordNet	• Positive, negative	(Agerri and García-Serrano 2010)
Congressional floor-debate transcripts	• For or against the legislation discussed	(Thomas, Pang and Lee 2006)
Comlex	• attitude adverbs	(Macleod, Grishman and Meyers 1994)
Customer review datasets	• whether an opinion is expressed – feature from a pre-defined list	(Hu and Liu 2004)
Review-search results sets	• subjective or not	(Pang, Lee and Vaithyanathan 2002)
Whissell's Dictionary of Affect in Language	affective norms for English Words and Sentiment-bearing adjectives	(Sweeney and Whissell 1984)
Computational semantic lexicon of French verbs	• Feeling • Emotion • Psychological states	(Mathieu 2006)

NAME	ANNOTATION	REFERENCE
Economining	• automatically derived sentiment scores	(Ghose, Ipeirotis and Sundararajan 2007)
Multiple-aspect restaurant reviews	– 1-to-5 rating for five different aspects	(Snyder and Barzilay. 2007)
Multi-Domain Sentiment Dataset	– 1-to-5 star labels	(Blitzer, Dredze and Pereira 2007)

2.2 The creation process

Manual creation

The resources, which have been created employing a manual process, are:

- *The* **General Inquirer** *is a system created by (Stone, et al. 1966)*
- **Comlex** *(Macleod, Grishman and Meyers 1994)* dictionary
- **Multi-Perspective Question Answering** (MPQA) corpus[63] (Wilson and Wiebe 2003) (Wiebe and Wilson 2005) (Wilson, Hoffmann, et al. 2005)
- *The* **Opinion Finder** *lexicon (subjectivity clues) (Wilson, Hoffmann, et al. 2005)*
- *(Somasundaran, et al. 2006)* **annotation scheme for manual labelling of opinion categories in meetings**
- **Annotation Scheme and Gold Standard for Dutch Subjective Adjectives** *(Maks and Vossen 2010)*
- *(Jijkoun and Hofmann 2009)*, created a **gold standard for Dutch subjectivity words.**
- *The* **CINEMO** *corpus of French emotional speech. (Brendel, Zaccarelli and Deuvillers 2010)*.

Semi and Automatic creation

The resources created in a semi or automatic way are presented below:

- *(Yu and Hatzivassiloglou 2003)* annotation **scheme and corpus**

63 www.cs.pitt.edu/mpqa/databaserelease

- **Economining**[64]
- **WordNet Affect** *(Strapparava and Valitutti 2004)*
- **SentimentWortschatz (SentiWS)** *(Remus, Quasthoff and Heyer 2010)*.
- **GermanPolarityClues** *(Waltinger 2010)*
- **Q-WordNet** *(Agerri and García-Serrano 2010)*

TABLE II. provides an overview of the resources divided according to their creation process.

Table II.

Resource creation method (manual, automatic/semi automatic)

NAME	MANUAL	SEMI/ AUTOMATIC	REFERENCE
MPQA	X		(Wiebe, Wilson and Cardie 2005)
Cornell movie-review		X	(Pang and Lee 2004), (Pang, Lee and Vaithyanathan 2002)
SentimentWortschatz		X	(Remus, Quasthoff and Heyer 2010)
WordNet Affect		X	(Strapparava and Valitutti 2004)
Sentiwordnet		X	(Esuli and Sebastiani 2006)
General Inquirer	X		(Stone, et al. 1966)
Opinion Finder	X		(Wilson, Hoffmann, et al. 2005)
Micro-WNOP	X		(Cerini, et al. 2007)
ISEAR corpus			(Scherer and Wallbott, The ISEAR Questionnaire and Codebook 1997)
CINEMO	X		(Brendel, Zaccarelli and Deuvillers 2010)
Gold Standard for Dutch	X		(Maks and Vossen 2010)
TREC test collection		X	http://ir.dcs.gla.ac.uk/test collections/ access to data.html

64 http://economining.stern.nyu.edu/datasets.html

NAME	MANUAL	SEMI/ AUTOMATIC	REFERENCE
Gold standard for Dutch subjectivity words	X		(Jijkoun and Hofmann 2009)
		X	(Yu and Hatzivassiloglou 2003)
GermanPolarityClues	X		(Waltinger 2010)
Q-WordNet		X	(Agerri and García-Serrano 2010)
Congressional floor-debate transcripts	X		(Thomas, Pang and Lee 2006)
Comlex	X		(Macleod, Grishman and Meyers 1994)
Customer review datasets	X		(Hu and Liu 2004)
Review-search results sets	X		(Pang, Lee and Vaithyanathan 2002)
Whissell's Dictionary of Affect in Language	X		(Sweeney and Whissell 1984)
Computational semantic lexicon of French verbs	X		(Mathieu 2006)
Economining	X	X	http://economining.stern.nyu.edu/datasets.html
Multiple-aspect restaurant reviews		X	(Snyder and Barzilay 2007)
Multi-Domain Sentiment Dataset			(Blitzer, Dredze and Pereira 2007)
Lexicon of appraisal terms	X		(Somasundaran, et al. 2006)

2.3 Language, domain and size

After having analysed the resources in terms of emotion detection granularity and creation process, our next classifications is based on the criteria of their domain, language and size.

As we can see in the TABLE III., most of the resources have been created for English and the more prominent domain is the general one. There are exceptions of works done for French, Dutch or even Chinese and Japanese, but unfortunately they are in a very small number if compared with the English resources. In a similar way the most recurrent domain is the general one, however some resources are about products or restaurant.

Table III.

Resources language, domain and size

NAME	LANGUAGE	DOMAIN	SIZE	REFERENCE
MPQA	English	General	10000 sen.	(Wiebe, Wilson and Cardie 2005)
Cornell movie-review	English	Movies	2000 reviews 10662 sen. 10000 sen.	(Pang and Lee 2004), (Pang, Lee and Vaithyanathan 2002)
The NTCIR multilingual corpus	English, Chinese, Japanese	20 topics	6000 sen.	http://research.nii.ac.jp/ntcir/index-en.html
SemEval 2007 Task 18 –Affective Text	English	General	1000 sen. 1000 sen. test	(Strapparava and Mihalcea 2007)
ISEAR corpus	English	Real life	7000 sen.	(Scherer and Wallbott, The ISEAR Questionnaire and Codebook 1997)
CINEMO	French	General (movies scenes)	4k speech chunks	(Brendel, Zaccarelli and Deuvillers 2010)
TREC test collection	English	Different topics	100649 blogs	http://ir.dcs.gla.ac.uk/test collections/ access to data.html
	English	General	1336 seed words	(Yu and Hatzivassiloglou 2003)
Congressional floor-debate transcripts	English	Politics	38 debates (train), 10 (test), 5 develop.	(Thomas, Pang and Lee 2006)
Customer review datasets	English	5 electronic products	n.a.	(Hu and Liu 2004)
Review-search results sets	English	Review	20 results (Yahoo!) from 69 queries with "review"	
Economining	English	Transaction,price premiums, feedback postings for merchants	n.a.	
Multiple-aspect restaurant reviews	English	Restaurants	4488 rev.	(Snyder and Barzilay. 2007)
Multi-Multi MultiDomain	English	Products	n.a.	(Blitzer, Dredze and Pereira 2007)

As we can see in TABLE III., most of the resources have been created for English, however, due to the need of resources in other languages, there have been attempts focused on mapping them in other languages.

Kim and Hovy (2006) studied language-mapping methods for subjectivity lexicons. They used a machine translation system and after that subjectivity analysis system that was created for English.

Mihalcea, Banea and Wiebe (2007) employed cross-language projection to learn multilingual subjective language. Using the Opinion Finder lexicon (Wilson et al., 2005) together with two bilingual English-Romanian dictionaries they translated each word. Concerning the collocation translation, they translated each word and after that filtered those translations that occur a minimum of three times on the Web.

Banea, Mihalcea, et al. (2008) carried our different experiments: they automatically translate the MPQA annotations into Romanian. Then, they used the automatically translated entries in the Opinion Finder lexicon to label a set of sentences in Romanian. Finally they inserted the translation direction and checked if the assumption that subjective language can be translated. In this way new subjectivity lexicons can be produced for languages with no such resources.

Banea, Mihalcea and Wiebe (2008) apply bootstrapping techniques to build up a subjectivity lexicon for Romanian. They used a set of subjective entries as seeds employing electronic bilingual dictionaries and a words training set. They started with a set of 60 words (noun, verb, adjective and adverb) from the translations of words of the Opinion Finder lexicon. Then, they filtered the translations using a similarity measure with the original words, based on Latent Semantic Analysis (LSA) techniques Deerwester, et al. (1990) scores.

Banea, Mihalcea and Wiebe (2010) translated the MPQA corpus into five languages. After that they expanded the feature space used in a Naïve Bayes classifier using the same data translated to 2 or 3 more languages. By means of expanding the feature space with data from other languages they obtained results almost as well as training a classifier for just one language on a large set of training data.

2.4 Resources and methods applied in nlp tasks

After having presented the resources and grouped them according to different criteria, another aspect to take into account is the manner in which researchers employed such resources together with other techniques and methods to improve the performance of the system dealing with Natural Language Processing tasks.

As we already mentioned, Sentiment Analysis is the task in charge of classifying the opinions about a specific topic and expressed by a source according to their polarity and sentiment. Having said this, an aspect to take into account is the fact that, depending on the users' needs, different levels of analysis are appropriate.

In fact, if we decide to go to the cinema to watch a movie and we check the people opinion on the web, in most of the cases the overall polarity would be enough to decide to go or not. However in case we need to buy a camera we would probably need to know the opinions about the different product features to take our final decision.

We can deduce that polarity classification becomes crucial for the development and high performance of systems for many Natural language Processing tasks focused on the treatment of subjective content. It needs of resources to have at disposal larger data set for their system training and testing.

We can say that among the ones we presented, the most widely employed resources are the following:

- The MPQA corpus
- The NTCIR multilingual corpus
- The SemEval 2007 corpus
- The ISEAR corpus
- WordNet Affect
- The TREC test collection
- SentiWordNet
- General inquirer
- The Opinion Finder
- Micro-WNOP

Apart from the resources above, researchers also employ different techniques to solve those tasks. Below we summarise the most significant techniques and resources used for the tasks of Polarity Classification (document, sentence and finer-grained levels), Opinion Question Answering and Opinion Summarisation.

Polarity Classification at different levels

Pang and Lee (2008) discriminated polarity classification into Classification using the representation of text as feature vectors where entries correspond to terms, either as count of frequencies (using tf-idf), or counting the presence or absence of a certain opinion word; and use of information related to the part of speech of the sentiment words and use of specialized Machine Learning algorithms for the acquisition of such words (adjectives, verbs, nouns, and adverbs).

Riloff, Wiebe and Phillips (2005) concentrated on the acquisition of nouns with sentiment. They used dependency parsing considering the dependency relations as features of Machine Learning algorithms. For the tasks in which sentiment on a certain topic must be extracted, the features used in machine learning for sentiment classifications were modified to contemplate information on the topic or named entities mentions related with such topic.

Koppel and Shtrimberg (2004) researched on good vs. bad news classification and this approach was considered similar as the sentiment classification task.

Sentiment Analysis has been studied at a **DOCUMENT LEVEL** for movies, book reviews etc. The starting point is the assumption that each review is about one single object (a product for example) and contains opinion from a single opinion holder.

Turney (2002) computed the single opinion words sentiment polarity (of the movie reviews) by means of a set of seed adjectives whose polarity was previously known and calculating the Pointwise Mutual Information score between the word to classify and the known word using the number of hits obtained by querying the two words together with the NEAR operator on the AltaVista search engine. The final review score is the sum of the polarities of each opinionated word in the review. Sentences have been previously filtered according to patterns bases on the presence of adjectives and adverbs.

Dave, Lawrence and Pennock (2003) extract patterns of opinion from a corpus of already graded reviews.

Other researchers focused on polarity classification at document level:

Pang, Lee and Vaithyanathan (2002) employed Naïve Bayes Machine Learning using unigram features demonstrating the fact that the employing unigrams outperforms the use of bigrams and of sentiment-bearing adjectives.

Mullen and Collier (2004) concluded that classifying sentiment using Support Vector Machines with features computed on the basis of word polarity, semantic differentiation computed using synonymy patterns in WordNet, proximity to topic features and syntactic relations outperforms n-gram classifications.

Pang and Lee (2003) classified reviews into a larger scale of values rather than into only positive and negative and employed SVM machine learning with similarity features. They compared the outcome with the number of stars given to the review.

Chaovalit and Zhou (2005) performed a comparison between different methods of supervised and unsupervised learning based on n-gram features and semantic orientation obtained by using patterns and dependency parsing.

Goldberg and Zhu (2006) presented a graph-based approach to sentiment classification at a document level. They represented documents as vectors, computed on the basis of presence of opinion words and linked each document to the most similar ones. Finally, they classified documents on the basis of the graph information using SVM machine learning.

Works focused on **SENTENCE LEVEL** are:

Yu and Hatzivassiloglou (2003) employed sentence level sentiment analysis with the aim of separating facts from opinions in a Question Answering scenario.

Other authors used subjectivity analysis to detect sentences from which patterns can be deduced for sentiment analysis using a subjectivity lexicon. (Hatzivassiloglou and Wiebe 2000)

(Wiebe and Riloff 2006)

(Wilson, Wiebe and Hwa 2004).

Kim and Hovy (2004) try to find the positive, negative and neutral sentiments expressed on a specific topic and the source of the opinions. Authors created sentiment lists using WordNet and then selected sentences that contained both the opinion holder as well as opinion statements and computes the sentiment of the sentence with a window of different sizes on the target, as harmonic and, geometrical mean of the sentiment scores assigned to the opinionated words.

Finer-grained research include works at **FEATURE LEVEL** such as:

Feature-based opinion mining is defined by Hu and Liu (2004), previously defined by Dave, Lawrence and Pennock (2003) - as the task of extracting the features of the object and the opinion words used in texts in relation to its features, classifying the opinion words and produce a final summary (result of the computing of the percentages of positive versus negative opinions expressed on each of the features).

Recently, authors have shown that performing very fine or very coarse-grained sentiment analysis may cause problems for the final application, since many times the sentiment is expressed within a context. This is what motivated McDonald, et al. (2007) who proposed an incremental model for sentiment analysis, starting with the analysis of text at a very fine-grained level and adding up granularity to the analysis (the inclusion of more context) up to the level of different consecutive sentences. They showed that this approached improved the sentiment analysis performance.

Table IV summarises the techniques mentioned in this section.

TABLE IV.

Polarity classification research

TECHNIQUE	AUTHOR
Polarity classification discrimination	(Pang and Lee 2008)
Good vs- bad news	(Koppel and Shtrimberg 2004)
AT DOCUMENT LEVEL	
Sentiment polarity of the individual opinion words using a set of seed adjectives	(Turney 2002)
Opinion patterns	(Dave, Lawrence and Pennock 2003)
Naïve Bayes Machine Learning using unigram features	(Pang, Lee and Vaithyanathan 2002)

TECHNIQUE	AUTHOR
Sentiment classification using SVM with features (word polarity, semantic differentiation computed using synonymy patterns in WordNet, proximity to topic features and syntactic relations outperforms n-gram classifications)	(Mullen and Collier 2004)
Reviews classification into a large scale of values	(Pang and Lee 2003)
Comparison between different methods of supervised and unsupervised learning	(Chaovalit and Zhou 2005)
Graph-based approach at a document level	(Goldberg and Zhu 2006)
AT SENTENCE LEVEL	
Separation of fact from opinions in a QA scenario	(Yu and Hatzivassiloglou 2003)
Sentences detection from which patterns deduced for sentiment analysis, based on a subjectivity lexicon	(Hatzivassiloglou and Wiebe 2000) (Wiebe and Riloff 2006) (Wilson, Wiebe and Hwa 2004)
FINE-GRAINED	
Analysis of text at a very fine-grained level and adding up granularity to the analysis	(McDonald, et al. 2007)

Opinion Question Answering and Opinion Summarisation

Opinion Question Answering and Opinion Summarisation are two Natural Processing tasks that, if performed effectively, could be extremely useful to find the required information among the huge quantity of subjective information available avoiding loosing much time for manually discriminating interesting information or not depending on the user's needs.

For both tasks, most of the state of the art has been focused on the development of systems for the treatment of factual data. (Quarteroni, et al. 2007) for Opinion Question Answering and (Kabadjov, Balahur and Boldrini 2009); (Steinberger, et al. 2007); (E. Hovy 2005); (Erkan and Radev 2004) for Automatic Summarisation.

However, due to the present context, the user's need is more focused on the analysis of subjective data, extremely difficult to manage. In fact in the case of Question Answering, answers can be longer than one simple sentence and the simple sentence understanding is an extremely tedious.

Subjective data is also challenging to manage for Automatic Summarisation systems in which there is the discrimination of useful pieces of the text, but after that the interpretation of the data is essential and in many cases users express their opinion in a non-canonical way and this can provoke the risk of error propagation.

In order to give a complete overview of the most significant research carried out in the framework of these three key Natural Language Processing tasks, a summary table in which we mention the research for each task and the corresponding evaluation campaign together with a brief description of the most relevant participating systems is presented in TABLE V.

Table V.

Overview in research in Sentiment Analysis

APPROACH	**AUTHOR**
OPINION QUESTION ANSWERING	
(Stoyanov, Cardie and Wiebe 2005) (Pustejovsky and Wiebe 2005)	Peculiarities of opinion questions
(Cardie, et al. 2004)	Opinion summarization to support a Multi-Perspective QA system, to identify the opinion-oriented answers for a given set of questions
(Yu and Hatzivassiloglou 2003)	Separated opinions from facts and summarized them as answer to opinion questions
(Kim and Hovy 2006)	Identified opinion holders, which are a key component in retrieving the correct answers to opinion questions
TAC 2008	
The Alyssa system (Shen, et al. 2007)	A SVM classifier trained on the MPQA, English NTCIR8 data and rules based on the subjectivity lexicon.

APPROACH	AUTHOR
(Varma, et al. 2008)	Query analysis to detect the polarity of the question using defined rules. Opinion filtering from fact retrieved snippets using a classifier based on Naïve Bayes with unigram features, assigning for each sentence a score that is a linear combination between the opinion and the polarity scores.
PolyU system (Li, et al. 2008)	The sentiment orientation of the sentence using the Kullback-Leibler divergence measure with the two estimated language models for the positive vs. negative categories.
The QUANTA system (Fangtao, et al. 2008)	Detected the opinion holder, the object and the polarity. It uses a semantic labeller based on PropBank and manually defined patterns. For the sentiment classification, they extract and classify the opinion words. For the answer retrieval, they score the retrieved snippets depending on the presence of topic and opinion words and choose as answer the top ranking results.
NTCIR 7 MOAT	
	The majority of the participants employed ML approaches using syntactic patterns learned on the MPQA corpus
OPINION SUMMARISATION	
Fine-grained, feature-based opinion summarization definition	(Hu and Liu 2004)
	(Stoyanov and Cardie 2006)
	(Saggion and Funk 2010)
	(Saggion, Lloret and Palomar 2010)
OPINION PILOT TRACK AT THE TEXT ANALYSIS CONFERENCE	
Most participants added new features (sentiment, positive/negative sentiment, positive/negative opinion) to account for the presence of positive opinions or negative ones	CLASSY (Conroy and Schlesinger 2008); CCNU (He, et al. 2008); LIPN (Bossard, Généreux and Poibeau 2008); IIITSum08 (Varma, et al. 2008) Italic (Cruz, et al. 2008)

2.5 Conclusions

In this second chapter we presented the most relevant tools and resources created in the framework of Sentiment Analysis. In order to have a clear view on what has been done, we grouped all of them according to the criteria of level of sentiment detection granularity, creation process (manual or semi/automatic) and language, domain and size.

Furthermore we focused our attention on presenting the Natural Language Processing tasks in which this discipline is currently applied that are Opinion Mining, Question Answering, and Automatic Summarisation and while applicable we mentioned the competitions organised and the most relevant participations until now.

We believe that this classification of resources is crucial to have first of all an overview of what has been done, but also to understand which is the contribution we bring with our research. We are also convinced that in research one of the most important aspects that make a concrete work innovative is the fact that we have to start from what is done after having made an exhaustive analysis of the previous research and understood the weak and strong points of previous research we can acquire clear ideas to improve the state of the art, but always trying to exploit as much as possible the previous research. This will make the work of a high quality.

After having performed this deep analysis, we can say that there is an evident lack of resources built up for languages other than English. Moreover most of the work done concentrated on studying texts from newspaper articles and analysed the subjectivity expressed in them in a very coarse-grained way.

There is a need to build up knowledge models able to understand the Human Language and discriminate it between objective and subjective and after that classify it. According to our opinion, after years of coarse-grained analysis mainly focused on English, corpora in other languages labelled in a fine-grained way are crucial to have a deep insight of the linguistic expressions employed to express subjectivity in the new textual genres of the Web 2.0.

Having taken into account this context, the resource we present, the Emoti-Blog-Annotation-Model has been designed to overcome the abovementioned challenges: lack of resources in languages other than English, fine-grained annotation and multi-level (document, sentence, element level) subjectivity annotation and also corpora composed by the extracts of new textual genres texts.

As a result of the conclusions drawn from the state of the art and a reflection on the pending issues, next chapter describes how we built up the EmotiBlog-Annotation-Model and how we collected the EmotiBlog-Corpus and labelled it. The annotation scheme is described in detail and also annotation examples are provided in order to understand better its structure. With this, our objective is to present the product of our research and to underline our contribution to the improvement of the State of the Art.

3. Emotiblog

Every corpus search and collection initiates with a linguistic challenge to overcome and its data is employed to analyse a linguistic phenomenon or to test a hypothesis stated by a researcher. In our case the main cause that motivated us to collect the EmotiBlog-Corpus was the desire to contribute to the improvement of the lack of fine-grained resources to work on Sentiment Analysis above all in languages other than English and with the new textual genres.

The World Wide Web is a mine of language data with unprecedented richness and ease of access (Kilgarriff and Grefenstette 2003) however in order to carry out studies of different nature, different corpora (with diverse features) are needed. Even if for many studies traditional collections such as the British National Corpus have been employed, in our case we need a genuine corpus extracted from one of the most predominant new textual genres produced with the growth of the Web 2.0: blogs.

This section presents the EmotiBlog-Annotation-Model and the EmotiBlog-Corpus collection and annotation processes. The annotation model is described together with real examples extracted from the corpus. A special stress is dedicated to the difficulties of the annotations and on the main principles we followed both for the corpus collection, as well for its labelling. More information on that can be also found in the APPENDIX I, where the brief annotation guidelines are provided.

3.1 The Emotiblog-Corpus

The EmotiBlog-Corpus consists in a collection of blog posts manually extracted from the Web during 2009. We collected it manually in order to assure an extremely precise corpus in terms of topic appropriateness and reliability of the sources. In fact, we wanted to be sure to have a corpus in which each and every blog post topic was related to our macro topics and also that we selected only this specific textual genre.

By carrying out this process, our objective was to create a unique collection and we can say that the main features that distinguish the EmotiBlog-Corpus from other corpora employed in Sentiment Analysis research are: its multilinguality, the fact that it is multi-domain (different topics) and fine-grained labelling.

Before starting the collection process, we carried out an analysis of some topics of news that were producing a high level of interest from bloggers and this is why we selected as topics the last USA elections. However, in order to carry out comparable studies and to have the possibility to extend our dataset for some experiments we also selected blog posts about the Kyoto Protocol and the elections in Zimbabwe.

These two topics are also present in the MPQA corpus (Stoyanov, Cardie and Litman, et al. 2004) and thus they could be useful for us to expand the annotated dataset for the elements the models have in common. Even if the MPQA is composed by newspaper articles, it is written in a different style. However after having carried out an empirical analysis of both corpora, we can deduce that there is a high probability to find terms in common since the topic is the same. Thus employing both corpora could be the appropriate technique to have more annotated elements and thus more data to train and test our Machine Learning system.

We collected blog posts produced in 2009 about the 3 topics presented and described in TABLE VI.

TABLE VI.

EmotiBlog corpus domains description

TOPIC	DESCRIPTION
The Kyoto Protocol	Opinions about USA citizens but also from all over the world about the Kyoto Protocol and the Bush policy on that
The elections in Zimbabwe	Opinions regarding the elections carried out in Zimbabwe and about the “president” of this country
The last USA elections	Opinions and expectations about the candidates for the USA presidency

As we mentioned above we selected the *Kyoto Protocol* and the *elections in Zimbabwe* because in this way the MPQA could be exploited to have more data to train and test our Machine Learning system. In fact, even if the MPQA is a corpus composed by newspaper articles and the EmotiBlog-Corpus by blog posts there is high probability to have words ad terms with similar connotations/meanings in common if we select the same topic.

After that, we added the USA election topic, since at the time of the corpus making up we were in a context in which this subject was fostering a tremendous number of private states from the different political parties and the public opinion in general because of the Obama's presence in the list of candidates. For this reason we believed that blog posts extracted about this issue could have a high percentage of subjectivity expressions.

During this work the different corpora will be named as follows:

- EmotiBlog-Corpus: the collection of blog posts including the three topics
- EmotiBlog-Corpus-Annotated: the collection of blog posts including the three topics annotated with the EmotiBlog-Annotation-Model
- EmotiBlog-Kyoto: The collection of blog posts about the Kyoto Protocol
- EmotiBlog-Kyoto-Annotated: The collection of blog posts about the Kyoto Protocol annotated with the EmotiBlog-Annotation-Model
- EmotiBlog-USA: The collection of blog posts about the USA elections
- EmotiBlog-USA-Annotated: The collection of blog posts about the USA elections annotated with the EmotiBlog-Annotation-Model
- EmotiBlog-Zimbabwe: The collection of blog posts about the elections in Zimbabwe
- EmotiBlog-Zimbabwe-Annotated: The collection of blog posts about the elections in Zimbabwe annotated with the EmotiBlog-Annotation-Model

In each case, the language will be specified (English, Spanish or Italian).

As we mentioned above, the multilinguality represents one of the most significant features that differentiates the EmotiBlog-Corpus from other existing corpora available for Sentiment Analysis. In fact with the aim of overcoming such problem of resources scarcity in languages other than English, the data we collected is in three languages: English, Spanish and Italian.

We chose English since it is the world most spoken language and Spanish because it is one of the biggest languages and most spoken in the world. Furthermore, Italian has also been chosen since we believe there is enough Natural Language Processing resources created that allow a study in the framework of Sentiment Analysis that would be comparable with English and Spanish.

We believe that having at least three working languages will allow us to carry out a high-quality multilingual work. In fact using only two, would have produced a not highly reliable study due to the huge variability of human language. Having at our disposal three languages we can better extract the linguistic phenomena employed in different cultures for expressing the subjectivity, thus avoiding the casual similarity between pairs of languages. Taking into account more that 2 options is the key to build up a reliable model for learning subjectivity patterns.

We collected 30000 words for each topic and language (the same amount of data for each language and topic) to have a balanced corpus and thus be able to carry out comparable experiments with the three languages.

Thus the result is a multilingual and multi-domain corpus composed by 270000 words, as shown in TABLE VII.

TABLE VII.
EmotiBlog-Corpus topics, size and languages

TOPIC	SIZE	LANGUAGE
The Kyoto Protocol	30,000 words	English, Spanish and Italian
The elections in Zimbabwe	30,000 words	English, Spanish and Italian
The last USA elections	30,000 words	English, Spanish and Italian
TOTAL		270,000 words

The corpus collection process

The corpus collection has been carried out basing on the principles that Lüdeling, Evert and Baroni (2006) describe in their paper titled *Using Web data for linguistic purposes.* According to the authors, *depending on the linguistic question or problem at hand, a researcher has to identify the data he needs*.

Our main motivation was the fact that there is no multilingual, multi topic and fine-grained annotated corpus composed by blog posts in English, Italian and Spanish composed by new textual genres extracts and thus our desire was to contribute to the improvement of the lack of resources to work on Sentiment Analysis above all in languages other than English.

In order to collect the EmotiBlog-Corpus we followed the principles listed below (Lüdeling, Evert and Baroni 2006):

– A *qualitative description* of the items to be found. We defined the textual genre and topics we wanted to collect in order to have a coherent and consistent corpus as a final result.
– A *stable corpus* (at least for the duration of the data acquisition so that the experiments can be replicated by other researchers). The texts we selected are from blog posts, one of the most relevant textual genres.
– The necessary *linguistic annotation* so that the items of interest can be located easily. This has been achieved by means of the annotation using the EmotiBlog-Annotation-Model.

When compiling the corpus one of our most important key issues was the reproducibility, understood as the features that makes possible to carry out the same experiments using the same corpus.

In case of using corpus composed by texts from the traditional textual genres, it is also possible to test the reproducibility of the results by means of repeating the experiment using a different corpus collected according to the same criteria (building up a second comparable corpus). However, in the case of blog posts this is unfeasible and can be simulated by dividing the corpus in different parts and the results obtained on one of them can be tested on the remaining ones.

Thus, we believe that being able to validate and reproduce scientific findings is essential for any quantitative study, whose relevance depends on the correctness and interpretability of the published procedures and results. While validation of experiments is in most cases trivial for traditional corpora, in our case the web is in constant update. As a consequence it is impossible to replicate an experiment with blog posts in an exact way at a later time. Some pages will have been added, some updated and some deleted since the original experiment.

Last but not least, we have to bear in mind that the accuracy of a corpus search depends on the range and the quality of the linguistic annotation (including pre-processing steps such as identification of word and sentence boundaries) all aspects that we describe in the next sections.

After having taken into account the abovementioned factors we collected our EmotiBlog-Corpus and analysed the information contained in it in an empirical manner to be able to propose a linguistic annotation that would be adequate, satisfactory, complete and easy to perform.

The next section describes how we built up the EmotiBlog-Annotation-Model, the elements that compose it, but it also explains the annotation process step by step, as well as the problem encountered during the labelling. This is essential in order to know the EmotiBlog-Annotation-Model from inside and thus to understand which is our contribution and how we can exploit our resources: Both the annotation model and annotated corpus.

3.2 The emotiblog-annotation-model

After having collected our corpus, the next step consisted in defining an exhaustive annotation model to label it. The creation of the EmotiBlog-Annotation-Model was mainly inspired by the MPQA (Stoyanov, Cardie and Litman, et al. 2004) and it is the result of a deep and empirical analysis of what we encountered in our blog posts collection.

From what we encountered, we can say that blogs are written in a non traditional style such as for example the one employed in the newspaper articles, thus our main objective before defining the model was to create a list of the linguistic strategies bloggers employed to express their subjectivity in this textual genre.

We detected that in general bloggers write in a more spontaneous way, than in newspapers and as we also explained in the introductory section, many are the challenges for interpreting their language such as mixture of sources and targets but also a wide use of sayings and collocations that in order to be properly interpreted must be treated as a global expressions and interpreted in a certain context. In fact the sense of a saying and collocation is not given by the sum of words that compose them, since it is an overall sense.

When we analysed the language employed and the linguistic elements used for expressing subjectivity, we understood that a fine-grained model was needed in order to be able to capture not just the basic expressions of subjectivity. Thus, we collected all the recurring linguistic elements used to give the subjectivity shadow to the text and we built up the first version of our annotation model.

As we explained above, subjective information is a general concept, which can be expressed by employing different linguistic strategies depending on the language and culture of each blogger.

After having analysed our corpus, we proposed a first version of EmotiBlog-Annotation-Model that is presented below.

Annotation levels –document, sentence, element-

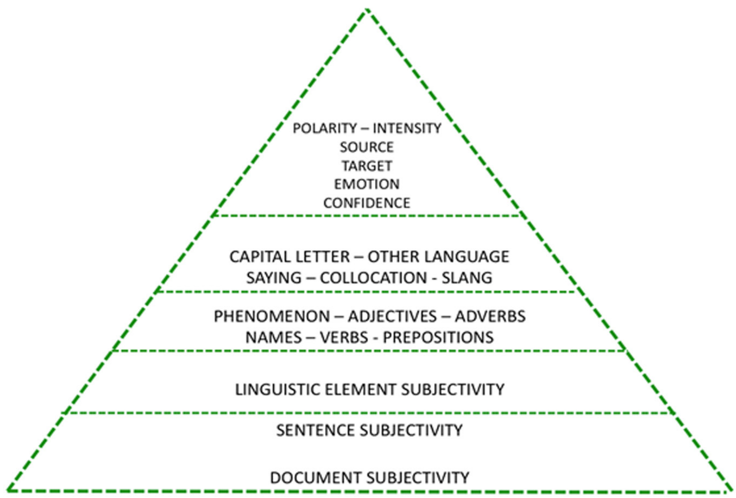


Figure I: The EmotiBlog-Annotation-Model structure

The figure above shows the basis structure of the EmotiBlog-Annotation-Model whose main structure is the discrimination between objective and subjective discourse and after that the annotation is done at different levels and taking into consideration the linguistic elements which give the subjectivity to the text.

The basic motivation beyond the EmotiBlog-Annotation-Model is that, after an exhaustive and deep analysis of the State of the Art in Sentiment Analysis, we discovered an urgent need for detecting subjectivity at multiple stages: document, sentence and element level and this is demonstrated by the examples (15 and 16).

> (15) **Good** post about Kyoto! I think the problem **lies** with the media **mis-intepreting** the threats of global warming which **confuses** the constituents who vote for people that do not have it on their agenda.

> (16) I **love** it when people say it will hurt the US Economy, especially business leaders. **Apparently** they do not have the foresight to realize how much **$$$$$$$** they could make if we switch to a greener capitalist system.

The extracts 15 and 16 are an example of blog posts about the decision of the US not to sign it. Analysing them and also many other equivalent posts, we reached the conclusion that different levels of annotation are necessary.

The document level annotation is needed to have a global document analysis, thus a general annotation about the entire document. This overall score will be the result of the sum of the sentences it is composed. After that, the element annotation is also fundamental. In fact the subjective shadow is given to the sentence and document by the single linguistic element by which they are composed (Underlined in the extracts 15 and 16).

As we can see they can be of different nature, such as adjectives, verbs, adverbs, etc. This post is composed by three sentences, each one with some subjective elements. We believe the three-level annotation is essential because it will allow a higher level of exploitability and application of the resource created. We are convinced that this kind of annotation will be suitable for studies focused at document, sentence and element level and we will check if the employment of some or the total number of the EmotiBlog-Annotation-Model has a key role for the EmotiBlog- Corpus-Annotated in the Natural Language tasks we choose for the extrinsic evaluation of our resource.

Objective vs. Subjective Discourse

After having clarified the multiple levels of annotation EmotiBlog-Annotation-Model allows, the next crucial issue is the discrimination between objective vs. subjective discourse.

One of the most distinguishing feature of EmotiBlog-Corpus-Annotated is the fact that the entire corpus is annotated by means of discriminating the sentences that present factual data (objective discourse) from the ones containing private states[8] subjective discourse). This can be seen in the post below:

> (17) The perils of arguing against an idiot. For nearly six years now Democrats have been calling George W. Bush a moron. He can't put a coherent sentence together, he has admitted only one mistake (taking responsibility for Katrina screwups) to date, and he seems to be screwing up Iraq, which pisses me off because I think that the democratization of Iraq would have been a very good thing.
>
> But forget about all of that for a minute. Texans probably think that W sounds just fine, he did admit a mistake, and President's make foreign policy snafus. It happens. The fact is that Democrats and Republicans are locked in an ideological battle. Each believes that implementing their policies will make the country and the world a better place, while enacting the opponent's platform will lead to a country filled with either slack-jawed, bumbling, racist hicks being ruled by four rich oil corporations or, alternatively, communist, possibly homosexual hippy deadbeats with STDs.

The extract (17) is an example of co-occurrence of both objective and subjective discourse that coexist in the same post. Thus after having analysed our corpus we concluded that the annotation of the totality of sentences would be useful to carry out experiment regarding the inter-annotator agreement in terms of objective/subjective discourse that it seems a simple distinction.

Intensity and Polarity

As we mentioned above, EmotiBlog-Annotation-Model is meant to be fine-grained. Before entering in detail with the description of each element this section defines and explains our interpretation of the concepts of polarity and intensity we apply to our resource.

Intensity is defined in the Oxford Dictionary[65] as the quality of being intense; thus intense is: *1) of extreme force, degree, or strength: the job demands intense concentration the heat was intense an intense blue (of an action) highly concen-*

65 http://oxforddictionaries.com/

trated: a phase of intense activity or 2) having or showing strong feelings or opinions; extremely earnest or serious: an intense young woman, passionate about her art a burning and intense look.

Thus, according to our interpretation, the intensity corresponds to the strength of the subjectivity (polarity) that is being expressed. This attribute is a key element of our EmotiBlog-Annotation-Model, since it focuses on measuring the strength of our subjective element.

As language users, we unconsciously perceive differences in the intensity in different private states in a natural way. For example, *love* and *like* are two different verbs expressing a positive emotion, but with different intensities and thus are used for transmitting different private states. In fact, the first one is less intense that the second emotion.

We can deduce that recognising the intensity of the subjective elements would mean doing two processes: discriminating between objective/subjective and the assigning the intensity label in case we are analysing a subjective element.

Polarity is another key elements of EmotiBlog-Annotation-Model. According to the Oxford Dictionary this term is defined *as the property of having poles or being polar: it exhibits polarity when presented to a magnetic needle/ the relative orientation of poles; the direction of a magnetic or electric field. The state of having two opposite or contradictory tendencies, opinions, or aspects: the polarity between male and female.*

In this work we use the sense of positive or negative sentiment being expressed by a word and in case of neutral polarity we will have an objective discourse.

It Is worth mentioning that we can distinguish between polarity in general and polarity in a specific context. In the first case the value will be invariable such as for example, *beautiful* and *ugly*. As we know, *beautiful* is generally used in a positive manner for something that we like, while *ugly* is employed when we want to describe something unpleasant.

If we take into consideration the contextual polarity, the sense we assign to a specific linguistic element is strictly related and associated to the context in which it is used, that in our case will be the sentence or the blog post.

Generally the strategies employed for inverting the polarity can be sarcasm or irony between others, apart from the classical modifiers or negation. Examples of this phenomenon can be seen below (18):

(18) Bush is a perfect president rejecting the Kyoto Protocol…

We can see that if we take into account the context, the linguistic element *comfortable* with a positive polarity assumes a negative shadow.

(19) And the Protocol was designed to allow, and even to encourage, fraud. Not only have signatories fiddled their 1990 emissions to allow themselves the right to emit more in 2010 than they did in 1990; many of them have set up "**cap-and-trade**" schemes, such as that which you have proposed, and have then fiddled the operation of the schemes. The European dictatorship, for instance, allowed each of its satrapies to trade quantities of emissions that exceeded their current total emissions by a **comfortable** margin. That is why the European "cap-and-trade" scheme collapsed.

Extract (18 and 19) are a clear example of contextual polarity. In fact the *"…"* are used with the purpose of changing the adjective meaning and if we look the definition of the adjective comfortable we obtain the following definition: 1 (especially of clothes or furnishings) providing physical ease and relaxation. 2 as large as is needed or wanted thus, the prior polarity of such adjective is positive for default, but in the above case it changes because of the influence of the context.

Another example of contextual polarity is presented in the extract (20) below.

(20) It **amazes** me how much president Bush gets what he wants. I'ts hard to come up with any policy that he wanted and didn't get. I keep hearing liberals call him stupid but he has won every political battle he faced. I personally don't like most of his policies but this myth that he is not smart is completely false. If there are any stupid people it's the Nancy Pelosi led Democrats who cannot beat a president with a ~20% approval rating.

The verb *to amaze* generally means: *surprise is when (someone) greatly; fill with astonishment.* However its polarity is highly influenced by the context. In this case (21) the polarity is negative, however in many other cases it can be extremely positive such as in:

(21) Your present is **Amazing**!! Thank you so much

In (21) we can see that the punctuation is reinforcing the adjective intensity.

Subjectivity Classification

Human expression of subjectivity is extremely complex and difficult to classify in rigid rules. In general it can be said that subjectivity is employed to express positive or negative reactions to external as well as internal stimuli. According to modern psychology, emotion, behaviour and cognition influence each other.

Thus, each subjective status affects human motivation, nervous function, learning, physical acts, physiological arousal and communication with other people. *Sadness*, for example, causes a person to cry and withdraw from social circles, while *surprise* causes sigh and raise people's eyebrows, while anger provokes trembling and aggressive behaviour.

Numerous researchers worked on emotion definition and classification. Robert Plutchik, Paul Ekman, Wallace Friesen, Carrol Izard and Silvan Tomkins are among the names that have made significant contributions to the study and classification of human emotions. They defined certain emotions as basic.

For example, according to Ekman (1999) the basic ones are: *sadness, happiness, anger, fear, disgust, and surprise*. These emotions **combine in different ways and form other emotions**, including *compassion, boredom, embarrassment, rage, hunger, and more*.

According to Ekman and Friesen (1969), there are 9 characteristics, which distinguish **basic emotions** from one another and from other affective phenomena. These features are presented in the table below:

Table VIII.

Nine distinctive features of emotions

N	FEATURE DESCRIPTION
1	Distinctive universal signals
2	Presence in other primates
3	Distinctive physiology
4	Distinctive universals in antecedents events
5	Coherence among emotional response
6	Quick onset
7	Brief duration
8	Automatic appraisal
9	Unbidden occurrence

Some of these characteristics (1,3,4) distinguish one emotion from another, while the others are useful to discriminate emotions from other affective states, such as moods, emotional traits, attitudes, etc.

Further research carried out by Plutchik and Ekman confirmed the evolutionary nature of emotions, a topic previously discussed in detail by Charles Darwin. Agreeing with Darwin, Plutchik believed that emotions evolved for the sake of human survival and reproduction.

Ekman also agreed with Darwin and after studying an isolated tribe in Papua New Guinea, he concluded that some emotions are universal and innate. Moreover, research proves that emotions affect and shape the essence of life for mankind and it is no wonder that psychology, neuroscience, ethics, sociology, and metaphysics, among other fields, all deal with the study of human emotions.

As in many disciplines, theorists disagree and Ortony and Turner (1990) carried out a wide range of research focused on the identification of basic emotions.

Different researchers centred their studies on classifying emotion and its expression. The most relevant are presented in the table below.

Table IX.

Different emotion classifications

RESEARCHER	EMOTION CLASSIFICATION
(Plutchik 1980)	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
(Arnold 1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
(Ekman, Friesen and Ellsworth 1982)	Anger, disgust, fear, joy, sadness, surprise
(Frijda 1986)	Desire, happiness, interest, surprise, wonder, sorrow
(Gray 1985)	Rage and terror, anxiety, joy
(C. E. Izard 1977)	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
(James 1884)	Fear, grief, love, rage
(McDougall 1926)	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
(Mowrer 1960)	Pain, pleasure

RESEARCHER	EMOTION CLASSIFICATION
(Oatley and Johnson-Laird 1987)	Anger, disgust, anxiety, happiness, sadness
(Panksepp 1982)	Expectancy, fear, rage, panic
(Tomkins, Affect theory 1984)	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
(Watson 1930)	Fear, love, rage
(Weiner and Graham 1984)	Happiness, sadness

In TABLE X. we present a deeper list of emotions as described in (Parrott 2001), where they were classified into a short tree structure.

Table X.

Parrot's emotion classification

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment

Primary emotion	Secondary emotion	Tertiary emotions
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
	Sympathy	Pity, sympathy
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

In order to annotate subjectivity expressions we selected the categories Scherer proposed (K. R. Scherer 2005). Before describing them it is worth underlying that his work is focused on the definition and classification of emotion in conversation. However, after having analysed the previous work done and described above, we decided that the Scherer's classification was more adequate to work with blogs. In fact, as we already explained in the introductory section, the blog style could be compared with a conversation and thus we deduced that a classification for the conversational context would be more appropriate.

Moreover, as we can see in the tables above, emotion classification carried out by other researchers is quite limited in terms of number of categories and because we wanted to build up a fine-grained annotation model, as a starting point we decided to use and expand the Scherer's classification.

According to this author,

> *the inherent fuzziness and the constant evolution of these language categories as well as inter-language, inter-cultural, and inter-individual differences make it difficult to define central working concepts in the universal, invariant, and consensual fashion generally required by a systematic scientific approach.*

In the framework of the component process model, emotion is defined as an *episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism* (K. R. Scherer 2001) (K. Scherer 1987).

Emotions are generally elicited by stimulus events. A central aspect of the component process definition of emotion is that the eliciting event and its consequences must be relevant to major concerns of the organism.

Events and their appraisal can change rapidly because of updated information and re-evaluations. As appraisal influences the responses in the interest of adaptation, the consequence is that the emotional response is also likely to change rapidly. Emotions prepare adaptive action tendencies and their motivational underpinnings. We can deduce that they have a strong effect on emotion-consequent behaviour, often changing the ongoing behaviour sequences and generating new goals and plans.

According to Scherer, it is impossible to give a definitive number/list of emotions. Researchers developing emotion theories, inspired by Darwin, have suggested different numbers of basic emotions (Ekman 1972), (C. Izard 1971), (Tomkins and McCarter 1964). Most of these are utilitarian emotions as defined above and play a crucial role in adapting to frequently occurring and prototypically patterned types of significant events in the life of organisms.

As a consequence, emotions like *anger, fear, joy, and sadness* are relatively frequently experienced (with anger and joy outranking all others; see the quasi-representative actuarial survey reported by K. R. Scherer (2005). Given the aspects of frequency, Scherer suggested calling these frequent emotions ***modal*** rather than *basic*, given that there is little consensus as to the meaning and criteria for how basic is to be defined (K. R. Scherer 2005).

Obviously, the small number of modal emotions (between 6 and 14 depending on the theorists) is hardly representative for the range of human emotionality. Thus Scherer suggested having recourse to the study of folk concepts of emotion in order to solve the problem of the number and nature of discriminable types of emotions. If in the evolution of languages, certain types of distinctions between different kinds of emotional processes have been considered important enough for communication to generate different words or expressions, these distinctions cannot be ignored.

Different researchers tried to apply this distinction (Levi 1984); (Lutz 1988); (J. A. Russell 1991); (Russell, Qualter and McGuigan 1995); (Wierzbicka 1999). The problem is to map the fuzzy and complex semantic fields of the emotion concepts onto the scientific definitions. This is particularly important as in distinguishing emotions the task is to examine fine-grained differences, analysing all of the components of the emotion processes.

Emotion terms can be rated by native speakers of different Natural Languages with respect to a number of items for each of the design features. This would include items on the eliciting event, the type of appraisal the person is likely to have made of the event and its consequences, the response patterns in the different components, and the behavioural impact (action tendencies) generated, as well as the intensity and duration of the experience.

In addition to the examination of subtle differences in the meanings of different emotion terms and providing similarity-of-profile data that can be used to statistically identify the relationships between members of emotion families and the overall structure of the semantic space for emotions, such data for different languages inform us about potential cultural and linguistic differences in emotion expression.

This aspect, apart from the scientific interest (Breugelmans, et al. 2005); (Fontaine, et al. 2002), is crucial to have comparability of instruments for intercultural studies. The major advances in recent years regarding the measurement of individual components such as appraisal are from (Scherer et al., 2001), brain mechanisms (Davidson, Sherer and Goldsmith 2003) physiological response patterns (Stemmler 2003), and expressive behaviour (Harrigan, Rosenthal and Scherer 2005).

Thus, having taken into account the abovementioned studies our conclusion was that the Scherer's classification was the best that fits for the blog posts textual genre. Thus, in order to present an exhaustive emotion classification and to make this subdivision proper and effective division, we were inspired by K. R. Scherer (2005) who created an alternative dimensional structure of the semantic space for emotions. The graph below represents the mapping of the term J. A. Russell (1983) uses for his claim of an emotion circumflex in two-dimensional valence by activity/arousal space (upper-case terms).

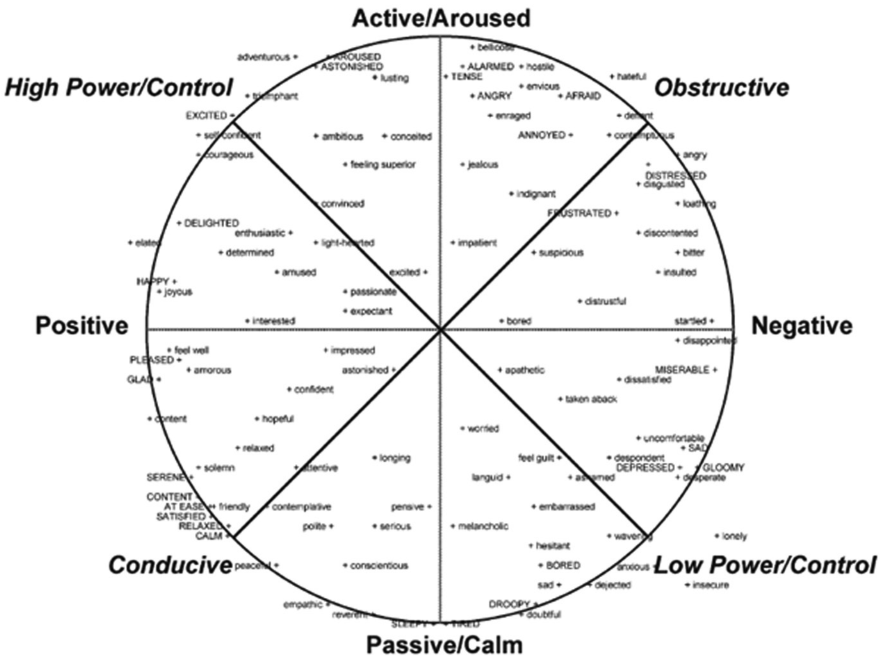


Figure II: Alternative dimensional structures of the semantic space for emotions

The figure above shows the mapping of the terms J. A. Russell (1983) uses as markers for his claim of an emotion circumflex in two-dimensional valence by activity/ arousal space (upper-case terms). Onto this representation Scherer added the two-dimensional structure based on similarity ratings of 80 German emotion terms (þ, lower-case terms, translated to English) from an earlier study that demonstrated the fact that semantic space may be organized by appraisal criteria (K. R. Scherer 1984). The plus (þ) signs indicate the exact location of the terms in a two-dimensional space. This simple superposition yields a remarkably good fit and it also shows that adding additional terms makes Russell's circumplex less of an obvious structural criterion – to obtain a perfect circle in a multidimensional scaling analysis seems to require the inclusion of non-emotion terms, as in the case of "sleepy, tired, and droopy" to mark the low arousal pole (as implicitly acknowledged by J. A. Russell (1991).

More importantly for the present purposes, a 458 rotation of the axes corresponds to an explanation of the distribution of the terms in a two-dimensional space composed by goal conduciveness and coping potential. Feelings that are members of any one specific emotion family can be expected to vary most among each other with respect to intensity (e.g. irritation–anger–rage), which, as argued above, may correlate with but is not the same as physiological arousal. It was therefore decided to map the intensity dimension as the distance of an emotion category's position in the goal conduciveness-coping potential space from the origin (Reisenzein 1994).

In order to create a graphically intuitive presentation, members of each emotion family were represented as a set of circles with increasing circumference. Moreover, the number of emotion families was limited to 4 per quadrant, yielding a total of 16 (which seems reasonable considering that the upper limit of the number of "basic emotions" is often considered to be around 14). The choice of the concrete families was also in large part fostered by what are generally considered to be either basic or fundamental emotions or those frequently studied.

We started from this classification, grouping sentiments into positive and negative, but we divided them as high/low power control, obstructive/conductive and active/passive. Further on, we distributed the subjectivity within our list into the Scherer slots creating other smaller categories included in the abovementioned general ones.

The result of this division is shown in TABLE XI.

Table XI.

EmotiBlog-Annotation-Model subjectivity status categories

GROUP	EMOTIONS
Criticism	Sarcasm, irony, incorrect, criticism, objection, opposition, scepticism.
Happiness	Joy, joke.
Support	Accept, correct, good, hope, support, trust, rapture, respect, patience, appreciation, excuse.
Importance	Important, interesting, will, justice, longing, anticipation, revenge.
Gratitude	Thank.
Guilt	Guilt, vexation.
Fear	Fear, fright, troubledness, anxiety.
Surprise	Surprise, bewilderment, disappointment, consternation.
Anger	Rage, hatred, enmity, wrath, force, anger, revendication.
Envy	Envy, rivalry, jealousy.
Indifference	Unimportant, yield, sluggishness.
Pity	Compassion, shame, grief.
Pain	Sadness, lament, remorse, mourning, depression, despondency.
Shyness	Timidity.
Bad	Bad, malice, disgust, greed.

TABLE XI. presents a complete list of the emotions we selected to be part of Emoti-Blog-Annotation-Model. After having added to the Scherer's classification a larger list of emotions, we grouped all of them into subgroups in order to help the evaluation process. We decided to have an extreme fine-grained list of subjective status but at the same time we group these into sub groups for the experiments. In fact emotions chosen by the annotator from the same subgroup will have less negative impact when calculating the inter-annotation agreement. For example if 2 annotator labels the same linguistic element as *bad* or *disgust* this annotation will obtain higher results that if they label *envy* and *bad* pertaining to two different subgroups.

The EmotiBlog-annotation- process – the Model Elements –

After having analysed the basic principles of the EmotiBlog-Annotation-Model and its structure, this section will be dedicated to present and describe the annotation model with the help of concrete examples extracted from the real corpus and thus, the annotation process.

Two experienced annotators labelled the corpus with the GATE[66] tool in a separate manner. In order to assure a high-quality annotation, the annotation consisted in two mail steps.

The first one in which the annotators labelled a small number of blog posts and then they put in common the annotation in order to check if they had understood the model structure correctly and to share doubts regarding some cases of subjectivity.

Then, after having clarified the problematic aspects, they labelled the entire corpus in Spanish and after the testing abovementioned and described in the following section, they annotated the English part and the Italian part of the Kyoto blog posts corresponding to the Kyoto protocol.

It is worth underlying that the elements the EmotiBlog-Annotation-Model contemplates are the results of a deep and empiric analysis of our blog posts collection. In fact after having collected the texts we analysed carefully the language employed and especially the linguistic elements as well as expressions that were used to confer the subjective touch. Basing on this analysis we built up the annotation model.

As we already explained, the EmotiBlog-Annotation-Model has been designed to allow the annotation at document, sentence and element levels and it also discriminates between objective and subjective speech. In each case the annotator has the possibility to assign and label the corresponding subjective elements (i.e. *adjective*) and specify their attributes (i.e. *polarity, intensity, modifier or not, emotion, etc*) that we will see in detail later on.

66 http://gate.ac.uk/overview.html

The complete list of EmotiBlog-Annotation-Model elements is presented in the Table below:

Table XII.

General view of the EmotiBlog-Annotation-Model

LEVEL	ELEMENT
Objective Speech event	
Subjective Speech (phenomenon)	Adjectives
	Adverbs
	Prepositions
	Verbs
	Names
	Capital letter
	Onomatopeic
	Punctuation
	Saying, collocation, slang, other language
Cross post	Coreference

In our research, we started from the annotation model proposed by Wiebe, Wilson and Cardie (2005), for the Multi Perspective Question Answering (MPQA) corpus, which is constructed upon the General Architecture for Text Engineering (GATE[67]) framework. Since our annotation model is different, we had to create new annotation files. They were built using XML schemas. The files had a structure defining the elements, containing the possible attributes each of those can have and their corresponding restrictions of type and value.

The definition of the annotation elements was done in a modular manner, so that the annotation schemes could be easily changed and adapted to newly iden-

67 http://gate.ac.uk/

tified phenomena in the corpus at hand. Figure III shows the breakdown of an element. In this case the adjective.

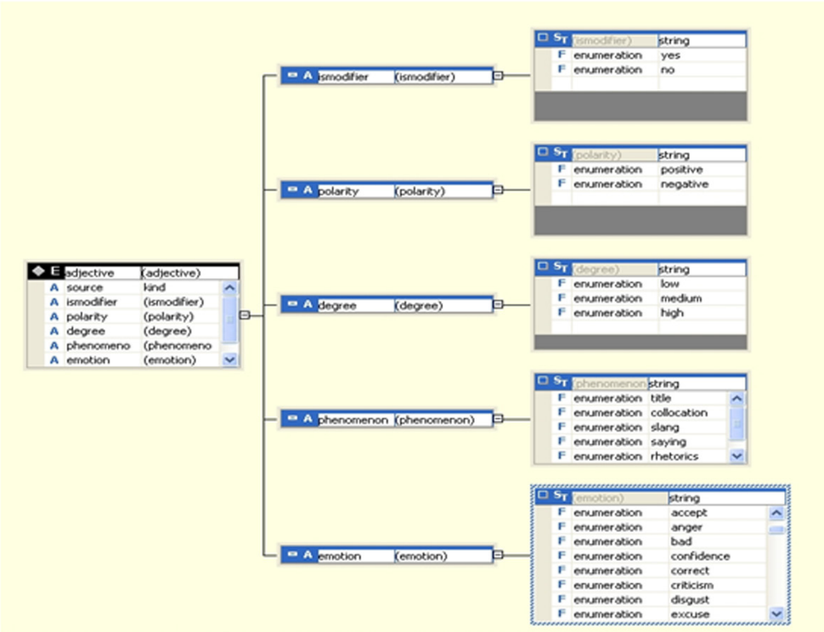


Figure III: Breakdown of one EmotiBlog-Annotation-Model element

As we can see, TABLE XII. is divided into different sections:

– Objective sentence

(22) <objective-speech-event gate:gateId="39" confidence="high" target="idiosyncrasies" source="w">**There are also other accounting idiosyncrasies, such as those discussed here: http://sciencepolicy.colorado.edu/prometheus/uk-emissions-4353** </objective-speech-event>

When we encounter an objective speech, we have to indicate its *source*, *target* and also the *annotator's confidence* for labelling this specific sentence. In our example the target are the idiosyncrasies, the source is "*w*" (the blogger) and the labeller's confidence is *high* (22).

– Subjective sentence

(23) <phenomenon gate:gateId="5" target="Kyoto Protocol" category="phrase" degree="high" source="Rebecca Tandy" polarity="negative" emotion="irony> **Mr. Protocol really likes treaties,**

so we got him this treaty instead of a card, so if you could just—all the other countries have already signed it"</phenomenon>

The example (23) is subjective and a normal phrase. The labeller marks it as subjective, but we will have to include a series of additional elements with their corresponding attributes we will detail in the next section. In fact when a sentence is considered subjective, we have to detect different aspects. The first one is what kind of text we are analysing (*phrase, saying, collocation, slang*, *other language ect.)* and after that which is the element that makes it subjective (*adjectives, adverbs, prepositions, verbs, nouns, capital letter, or punctuation)* with their corresponding attributes.

– COREFERENCE

Coreference is considered in this study only at a cross-post level. We decided to include it because of the nature of blogs. As we mentioned in the Introductory Section, blogs are like a conversation among different participants (bloggers) and thus, there is an extremely high possibility to have one blogger that is answering or commenting about the previous post o some previous ones. That is the reason why the EmotiBlog-Annotation-Model contains the possibility to detect correference but only at cross-post level. We decided not to include the correference detection in general to the entire corpus, since we believe it would be not so beneficial for our studies in which we do not want to include noise. This aspect could be taken into consideration for future works (26).

(24)<coreference gate:gateId="9" antecedent="previous writer" phenomenon="phrase" type="pronominal" source="w"> **You** </coreference> are right, but think about the CI emissions of countries like China.

The example 24 above represents a clear example of correference at cross post level. The antecedent is the previous writer and the source ifs the writer of the post we are analysing, while the type is pronominal correference.

EmotiBlog-Annotation-Model Elements with their Attributes

After having introduced the EmotiBlog-Annotation-Model elements, this section is dedicated to have a deep insight of each element that compose the model as well as their attributes. TABLE XIII. in the following page presents a complete overview of EmotiBlog-Annotation-Model.

Table XIII.

Detailed view of the EmotiBlog-Annotation-Model elements and their attributes

LEVEL	ELEMENT	ATTRIBUTE											
		CONFIDENCE	SOURCE	TARGET	POLARITY	DEGREE	PHENOMENON	MODIFIER	TYPE	MODE	ANTECEDENT	EMOTION	COMMENT
Objective Speech	objective-speech-event	X	X	X									
	Phrase, title	X	X	X	X	X						X	X
	Saying, collocation, slang, other language	X											X
	Adjectives	X	X	X	X	X	X	X				X	X
	Adverbs	X	X	X	X	X	X	X				X	X
Subjective Speech	Prepositions	X	X	X	X	X	X	X				X	X
	Verbs	X	X	X	X	X	X	X		X		X	X
	Nouns	X	X	X	X	X	X					X	X
	Capital letter	X	X	X	X	X	X					X	X
	Punctuation	X	X	X	X	X	X					X	X
	Onomatopeic[68]	X	X	X	X	X	X					X	X
Cross post	Coreference	X	X						X		X		X

TABLE XIII. contains the totality of the EmotiBlog-Annotation-Model annotation elements and attributes with their description. As we can see from above, there are common attributes for nearly the totality or for most of the EmotiBlog-Annotation-Model elements, while other such as *antecedent* are specific for an element.

68 This element has been removed alter the feature impact experiments for not being relevant (see xxx)

The first one is the *confidence* whose values can be*: high-medium-low* and depending on the level of annotator's confidence he will choose one value or the other.

Another common element is the *polarity* that can be *positive or negative* and its *intensity* can vary between *high - medium - low*.

PHENOMENON

Subjective speech (Phenomenon)	Type	Phrase, title	Confidence, comment, level, emotion, polarity, source and target

When we detect a subjective sentence we have to decide if it is a normal phrase, such as in (25)

(25) <phenomenon gate:gateId="5" confidence="high" target="Kyoto Protocol" category="phrase" degree="high" source="Rebecca Tandy" polarity="negative" emotion="irony> **Mr. Protocol really likes treaties, so we got him this treaty instead of a card, so if you could just—all the other countries have already signed it"**</phenomenon>

Or if it is a title such as in (26)

(26) <phenomenon gate:gateId="1"confidence="high" category="title" target="Bush" source="w" polarity="negative" degree="high" emotion="criticism">**Bush Pulls U.S. Out of Kyoto Treaty**</phenomenon>

In both cases (phrase or title) we have to specify their attributes, for example, taking sentence (26) as example we can see that the annotator is confident, it is a title which target of the discourse is Bush and the source is the writer/blogger "*w*".

Moreover, the sentence polarity is *negative*, with *high intensity* and the emotion in the case of (26) is *criticism*.

The EmotiBlog-Annotation-Model also allows the detection of sayings, collocations, slang and other languages expressions:

Subjective Sentence (phenomenon)	saying, collocation, slang, other language	Confidence, comment, level, emotion, polarity, source and target

(27) Concerning the Kyoto Protocol, <phenomenon gate:gateId="18"confidence="high" category="saying" target="Kyoto Protocol" source="w" polarity="negative" degree="medium" emotion="sarcasm"> **the game is not worth the candle** </phenomenon> according to Bush.

Sentence (27) is an example of *saying*. We decided to label them as single elements because their meaning is global and it is not the result of the summing up of the words that compose the expression.

According to the Oxford dictionary, a saying (27) is *a collection of short, pithy expressions identified with a particular person, especially a political or religious leader*, while a collocation (30) is *the habitual juxtaposition of a particular word with another word or words with a frequency greater than chance.*

We also detect cased of slang (28) that is *a type of language consisting of words and phrases that are regarded as very informal, are more common in speech than writing, and are typically restricted to a particular context or group of people* or expressions in other languages with a special subjective charge (28). In each case we add the additional.

(28) The refusal to sign the Kyoto protocol came <phenomenon gate:gateId="3"confidence="high" category="collocation" target="Kyoto Protocol" source="w" polarity="negative" degree="high" emotion="bad">**like a bolt from the blue** </phenomenon> for many Americans.

Expression in languages different from the one the blogger is writing, such as in (29) are also detected.

(29) The Bush strategy has been planned <phenomenon gate:gateId="23" confidence="high" confidence="high" category="other language" target="Kyoto Protocol" source="w" polarity="negative" degree="high" emotion="bad">**ad hoc** </phenomenon> for obtaining popular support.

We decided to contemplate these elements since they are frequently employed by bloggers and they contain a high subjective charge. Moreover they are highly interesting since they are culture and also contextual dependent. In this way, labelling such expressions in the 3 languages we obtain a database with genuine expressions exploitable for many interesting and useful studies focused on practical applications.

After having described the sentence level, we enter into the sentence labelling the linguistic elements, which give the subjectivity to the text.

ADJECTIVES

Subjective Sentence	Adjectives	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target

(30) Your <adjective target=”Kyoto Protocol” gate:gateId=”10” confidence=”high” phenomenon=”phrase” degree=”medium” polarity=”negative” emotion=”criticism” source=”w” ismodifier=”yes”>**pointless** </adjective>devotion to the <adjective target=”Kyoto Protocol” gate:gateId=”11” phenomenon=”phrase” degree=”medium” polarity=”negative” emotion=”criticism” source=”w” ismodifier=”yes”>**pointless**</adjective> Kyoto protocol

As we can see in the example above, we are annotating an adjective that is used two times. In both cases the annotator is confident about his annotation. The adjective is *negative* with a *high intensity* and expressing *criticism*. It is a modifier of the noun that follows it (30).

ADVERBS

Subjective Sentence	Adverbs	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target

(31) Bush <adverb target=”Kyoto Protocol” gate:gateId=”127” confidence=”high” phenomenon=”phrase” degree=”medium” polarity=”negative” emotion=”criticism” source=”w” ismodifier=”yes”> **finally** </adverb> signs Kyoto Protocol

The adverb *finally* is considered to be a *modifier*, with *negative polarity* and *intense level*. The source is the *blogger* and the target of the discourse is *Bush* (31).

PREPOSITIONS

Subjective Sentence	Prepositions	Confidence, comment, level, emotion, phenomenon, polarity, source and target

(32) Bush refuses the Kyoto Protocol <preposition target=”Bush” gate:gateId=”9” confidence=”high” phenomenon=”phrase” degree=”medium” polarity=”negative” emotion=”criticism” source=”w”> **against** </preposition> my willingness to take part in it

The preposition against clearly indicates that it is a *modifier* with a *negative polarity*, *medium intensity* and expressing *criticism* (32).

VERBS

Subjective Sentence	Verbs	Confidence, comment, level, emotion, phenomenon, polarity, mode, source and target

(33) It was the Clinton Gore team that <verb gate:gateId=”5” target=”Global Warming” phenomenon=”phrase” source=”w” tense=”Indicative” emotion=”criticism” polarity=”negative” intensity=”medium”>**blew**</verb> the biggest opportunity the US had ever had for leadership during the 1997 Kyoto agreement talks.

The verb *blew* used by the *writer* is *negative* with *medium polarity* and expressing *criticism* (33).

NOUNS

Subjective Sentence	Nouns	Confidence, comment, level, emotion, phenomenon, polarity, and source

(34) So don’t believe all this <noun gate:gateId=”15” phenomenon=”phrase” confidence=”high” source=”w” target=”Global Warming” emotion=”bad” polarity=”negative” intensity=”high”> **nonsense**</noun> about waiting for the next president to sort it out.

The writer is labelling the noun as *negative*, with *high polarity* and expressing the idea of something *bad* (34).

CAPITAL LETTER

Subjective Sentence	Capital letter	Confidence, comment, level, emotion, phenomenon, polarity, source and target

(35) This president is a <capital letter gate:gateId=”18” confidence=”high” phenomenon=”phrase” source=”w” target=”Bush” emotion=”anger” polarity=”negative” intensity=”high”>**DICTATOR**</capital letter>

In the example (35) the capital letter is employed to express *anger* with *high intensity and negative polarity.*

PUNCTUATION

Subjective Sentence	Punctuation	Confidence, comment, level, emotion, phenomenon, polarity, source and target

(36) Carbon emissions are increasingly dramatically <punctuation gate:gateId=”34” confidence=”high” phenomenon=”phrase” source=”w” target=”Global Warming” emotion=”fear” polarity=”negative” intensity=”high”>**!!!**</punctuation>

In the example (36) the triple exclamation mark is employed by the *writer* to express *fear*, with *negative polarity* and *high intensity*.

GLOBAL EXAMPLE

After having presented each single element of the EmotiBlog-Annotation-Model and explained how to label it, the example 37 shows an entire blog post annotated following the explanation above and the annotation guide in Appendix I.

(37)

<link>

http://thedawgrun.blogspot.com/2007/09/bush-pulls-us-out-of-kyoto-treaty-uh-er.html

</link>

<overall_sentiment>criticism</overall_sentiment>

<text1><phenomenon gate:gateId=”45” confidence=”high” target=”Bush” category=”title” degree=”high” source=”w” polarity=”negative” emotion=”criticism”>**Bush** <phenomenon gate:gateId=”46” confidence=”high” target=”Bush” **category**=”saying” degree=”high” source=”w” polarity=”negative” emotion=”criticism”>**Pulls U.S. out**</phenomenon> **of Kyoto Treaty -** <punctuation gate:gateId=”47” confidence=”high” target=”Bush” phenomenon=”title” degree=”high” source=”w” polarity=”negative” emotion=”sarcasm”>**. . .**</punctuation><phenomenon gate:gateId=”48” confidence=”high” target=”Bush” category=”saying” degree=”high” source=”w” polarity=”negative” emotion=”sarcasm”>**wait a second**</phenomenon> <punctuation gate:gateId=”49” confidence=”high” target=”Bush” phenomenon=”title” degree=”high” source=”w” polarity=”negative” emotion=”sarcasm”>**...**</punctuation></phenomenon>

<objective-speech-event gate:gateId=”50” confidence= “high” target=”Glenn Reynolds” source=”w”>**Glenn Reynolds at Instapundit points out a little whoopsie on the part of the Associated Press:**</objective-speech-event>

<objective-speech-event gate:gateId=”51” confidence= “high” target=”Kyoto protocol” source=”Glenn Reynolds”>**“REVISIONIST HISTORY: The Associated Press gets it wrong on Kyoto again: "Readers with a long memory may recall that the United States never adopted the Kyoto Protocol because the Clinton administration never submitted it for ratification to the Senate.**</objective-speech-event> <objective-speech-event gate:gateId=”52” confidence= “high” source=”w” target=”Glenn Reynolds”>**The Clinton administration never submitted it to the Senate for ratification because in July 1997 the Senate voted 95-0 to adopt a resolution stating that**</objective-speech-event> <objective-speech-event gate:gateId=”53” confidence= “high” source=”Clinton” target=”Kyoto protocol”>**’the United States should not be a signatory to any protocol to, or other agreement regarding, the United Nations Framework Convention on Climate Change of 1992, at negotiations in Kyoto.’**</objective-speech-event><objective-speech-event gate:gateId=”54” confidence= “high” source=”Clinton” target=”Kyoto protocol**”>”Yet according to AP, the U.S. was a party to Kyoto until Bush unilaterally pulled us out.”**
</objective-speech-event>
</text1>
</f3>

</paragraph>

Example (39) is an extract of blog post annotation. We labelled the different elements, which give the subjectivity to the text, but also objective sentences, as explained above.

Before entering in detail in the annotation, we can see from above that in the sentences one tag that has not been defined within the model appears. gate:gateId="xx". This is the automatically generated ID number that GATE assigns to the annotations.

Concerning the subjectivity annotation, the first two aspects we would like to underline are the fact that the link where the post has been extracted is the first information we give. In our opinion this is a crucial issue if we want to expand our corpus, but also for other purposes such as reproducibility. After that, the overall subjectivity is indicated and after this the real annotation starts. Since it is a real and quite extended example, we can find part of the cases we presented above annotated following the EmotiBlog guidelines.

We would like to underline that as we mentioned, we have a three-levels annotation: document, sentence and element and each of them has its own attributed listed in TABLE XIII.

We take the opportunity of having the possibility to show an entire blog post annotated to mention the most important principle upon which the EmotiBlog annotation has been carried out and this is the **consistency**. In our annotation guide (Apendix I) we specified that for example if we give to a document the intensity *high,* then the sentences that compose it we cannot label them as *low*. Or for example if in one sentence we label the emotion *anger* that the linguistic element inside the sentence cannot be marked as *happiness* and the same occurs with the *intensity level*. This seems a trivial aspect, but we decided to explain it since during our annotation process, when training annotators we detected some mistakes provoked by such aspects of the annotation.

3.3 Conclusions

In this chapter we presented the EmotiBlog-Annotation-Model, the fine-grained annotation schema and the EmotiBlog-Corpus, a collection of blog posts in English, Spanish and Italian and about three topics: the Kyoto Protocol, the election in Zimbabwe and the last USA elections. We described the EmotiBlog-Annotation-model structure, entering in detail with the explanation of each element with the corresponding attributes. After that we described how we collected our corpus and gave concrete examples of each element annotation.

As first step, we specified the annotation model definition process together with its basic structure and we then entered in detail explaining each model element together with its attributes and how we used such element for the annotation process.

Finally, the last part of this chapter has been dedicated to present a real and entire blog post annotated with EmotiBlog-Annotation-Model. In fact after having presented each element we considered necessary to provide the reader an overview to see "EmotiBlog in action".

We believe that the EmotiBlog-Corpus we created by labelling our corpus with the EmotiBlog-Annotation-Model represents a step forward previous research. In fact it consists in a collection of blog posts in different languages and about diverse topics. Moreover, the annotation scheme employed to label it is finer-grained than previous work allowing capturing a higher number of linguistic elements that give the subjectivity to the text. Last but no least, it contemplates aspects of the language such as collocations and sayings, thus creating additional knowledge about the languages the corpus is composed by and many possibilities of further research.

After having created our resource, next chapter presents the experiments carried out to check if EmotiBlog-Annotation-Model is a clear and if the annotation it generates is reliable. For this purpose next chapter presents the intrinsic evaluation we carried out by means of measuring the agreement among annotators to check the level of reliability of the annotation and also to test if the model has been designed in a good way and it is easy employable.

4. Measuring the emotiblog-annotation-model reliability

After having described the corpus creation, definition of the EmotiBlog-Annotation-Model and also the annotation process as well as the tool used, this chapter describes how the model is evaluated from inside, thus with an intrinsic method.

Our purpose here is to check if its annotation is feasible and easy to perform by the annotators, since our objective is to understand if the EmotiBlog-Annotation-Model is reliable and thus a suitable to help systems to learn how to detect the subjectivity in the new textual genres in a multilingual framework.

The annotation has been carried out in different steps. The first language annotated was Spanish because if compared with English, it has a more complex sentence structure and thus we wanted to be sure the EmotiBlog-Annotation-Model elements and attributes were enough exhaustive to detect the subjectivity expressions an a deep way in a complex structured language.

We decided not to label the three languages in a parallel manner because our idea was to perform different stages of labelling and testing.

These gradual annotation steps and testing phases are listed below and also represented in a visual way in the graph above:

- We label the Spanish part of the EmotiBlog-Corpus
- We calculate the Inter Annotator's agreement
- We perform General Feature Classification for the Spanish labelled corpus
- We carry out feature selection by dimensionality reduction for the Spanish labelled corpus for the polarity detection task
- We label the English part of the EmotiBlog-Corpus

- We evaluate the English corpus after the reclassification
- We improve the EmotiBlog-Annotation-Model
- We annotate the Italian part of EmotiBlog-Corpus
- We carried out the experiments on the Italian EmotiBlog-Corpus

As we can deduce from the list above, we preferred a gradual work. We start with the annotation of the Spanish corpus and the calculation of the Inter-annotator agreement. As we will see in the following sections, we performed firstly this experiment because we wanted to measure the level of reliability of the annotation model in terms of coherence and also depending on the needs of our corpus.

By evaluating the inter-annotator's agreement we had the possibility to check if the model is easy to employ and if it is adequate for our needs of fine-granularity but at the same time clear to be applied without special doubts by the annotators.

After having calculated the inter-annotator's agreement and checked its suitability for our needs, we perform a general feature classification for the Spanish labeled corpus and also a feature selection experiments on the EmotiBlog-Kyoto-Annotated in Spanish with the aim of checking the performance of our classification system in general but specially for the polarity task.

Apart from that, we measure the percentage of impact of each EmotiBlog-Annotation-Model element to see if all of them were needed and useful for the classification purpose.

The next step consisted in labeling the EmotiBlog-Kyoto in English and after that we evaluated the annotation after the reclassification, measuring the impact of each EmotiBlog-Annotation-Model element in the English collection.

In this way we produced the final version of the EmotiBlog-Annotation-Model with which we labeled the Italian part and made experiments for measuring the performance of the classification also with this language.

4.1 The inter annotator's agreement evaluation

After having collected the corpus, two experienced annotators labelled the Spanish-Kyoto blog posts using the EmotiBlog-Annotation-model The annotators followed the instructions of the annotation guidelines (see Apendix I) and made a short training after having labelled some posts to check if the EmotiBlog-Annotation-Model was clear enough. After this process, the next step to be taken was to measure the inter-annotator agreement between their annotations.

By means of calculating this value our purpose is to check if the EmotiBlog-Annotation-Model structure, and idea in general is appropriate, clear and robust for the needs of the textual genre we are considering and for the needs of subjectivity expression.

According to our idea, the EmotiBlog-Annotation-Model should be:

- *Appropriate* in the sense that it should be useful to satisfy the needs of the subjectivity expression in the textual genres
- *Clear* in the sense that if the annotators can clearly understand the main ideas behind the scheme and thus how to label after having carried out a small training and read the annotation guidelines.
- *Robust* in the sense that the overall annotation is coherent and consistent (as explained in section 3)

Two annotators (A and E) labelled 100 texts independently, a total of 30.000 words and we took into consideration the Spanish corpus about the ratification of the Kyoto Protocol. Thus, after the annotation process we measure the inter-annotator agreement of the different elements and attributes that compose the EmotiBlog-Annotation-Model with the purpose of checking if both annotators agree on which expressions should be marked.

At the beginning our idea was to employ the most traditional and widely used measure for the inter-annotator agreement evaluation: the kappa value (when statistic classes are present), and the observed agreement (when non statistic classes are present).

Generally, the kappa is computed according to Cohen method (Cohen 1960); (Carletta 1996); (Artstein and Poesio 2008):

$$Kappa = \frac{observed\ proportion\ of\ agreement - chance\ expected\ proportion\ of\ agreement}{1 - chance\ expected\ proportion\ of\ agreement}$$

After having tried to apply the kappa value we realised that this measure was not the appropriate for our annotation granularity, since it does not allow the evaluation of each element with their corresponding attributes, given that the boundaries we considered in the annotation model were highly variable. Thus we decided to use the following measure:

$$agr(a||b) = \frac{|A\ matching\ B|}{|A|}$$

The elements and attributes we evaluated are those, which make up the model, and they are listed below. The criteria used for matching is that the annotator's labelling should overlap and have the same annotated orientation, intensity and pertain to the same emotion category:

Table XIV.

Inter-Annotator agreement results for Spanish

Annotation	a	b	a||b	b||a	average
Noun	A	E	0.783	0.753	0.765
Adjective	A	E	0.782	0.613	0.681
Verb	A	E	0.863	0.742	0.802
Adverb	A	E	0.831	0.764	0.794
Preposition	A	E	0.862	0.672	0.763
Punctuation	A	E	0.784	0.891	0.832
Capital letter	A	E	0.663	1	0.831
Other Language (English)	A	E	0.273	1	0.632
Other Language (Latin)	A	E	0.662	0.662	0.661
Phrase	A	E	0.524	0.662	0.592
Objective	A	E	0.762	0.734	0.745
Total average					0.736

Looking at the results of TABLE XIV., we can deduce that the elements with the best performance are *capital letter* and *punctuation* and the ones, which obtained the lowest average, are *phrase* and *English*. In fact the detection of other language expressions with subjective shadow is extremely dependent of the culture of the annotator. In fact if for example one expression for an Italian could be sarcastic, for a Spanish person it cannot have the same subjectivity shadow. With the remainder of the elements we obtained results that are between 0.76 and 0.80 approximately.

This could be due to the fact that because the elements, which better performed are the ones more easily detectable without any knowledge. The problem comes when we look at the other elements of the model that are more dependent on the context of the blog post. We believe this could be one of the most relevant causes origins of such results.

Moreover, we would like to underline the fact that, as foreseen, due to the EmotiBlog-Annotation-Model granularity, the evaluation process was extremely complex. In fact, for example when annotators identify the same expression in the text, they could differ in their marking of the expression boundary, as well as in other elements and attributes such as the emotion intensity producing a results lowering. And even more, there was no guarantee that the annotators will identify the same set of subjectivity expressions.

However, looking at the percentages obtained, we can say that the total average of agreement is 0,736, better if compared with the state of the art (0,71 (Wilson 2008) a research in which the inter annotator agreement was calculated for the MPQA annotation scheme, less fine-grained than EmotiBlog-Annotation-Model.

In order to gain a better understanding on the relevance of each element of the EmotiBlog-Annotation-model we will perform some feature selection experiments. We will measure the importance of each element and we will also check if a fine-grained annotation model is beneficial for Machine Learning systems.

4.2 Conclusions

This chapter has been dedicated to present the experiments we carried out to calculate the inter annotator agreement for the labelled part of the Spanish corpus about the Kyoto Protocol. This value can be defined as the degree to which multiple human annotators arrive at the same annotations when confronted with the same Natural Language text.

We obtained results better than the baseline demonstrating that the annotation scheme is clear and the model does not present any basic deficiencies, thus it is useful to label the subjectivity in the new textual genres on different languages.

After having carried out an intrinsic evaluation and having demonstrated that the EmotiBlog-Annotation-Model can be used and its elements are understandable by the annotators and that they are comprehensive for subjectivity expressions in different languages and for the new textual genres, chapter 5 will present the feature selection experiments. We want to check if the EmotiBlog-Corpus-Annotated obtains positive results for the classification experiments and thus if it is useful and valid to train and test Machine Learning system.

5. Feature selection experiments

After having evaluated the labelling of the EmotiBlog-Kyoto-Annotated in Spanish by means of the inter-annotator agreement, the next step consists in checking if the EmotiBlog-Annotation-Model elements are useful (since it is a fine-grained model) and thus if the extended list of elements and their attributes has a positive effect on a Machine Learning classifier.

As we mentioned above, this intrinsic test will be done gradually for each language. In this way we start with Spanish, then with English and after having measured the impact of the different elements of the EmotiBlog-Annotation-Model we refined it and we then labelled and tested the EmotiBlog-Kyoto-Annotated in Italian.

We decided to start with Spanish since it has a more complex structure than English and in this way if we obtain good results for the Spanish, we suppose the also the English experiments will be similar in terms of results, and after having refined the model we test again with a complex language, Italian.

To this end, we carried out a set of gradual preliminary Machine Learning experiments in the Spanish annotated corpus to have the possibility to carry out a comparative study for the three languages that compose the EmotiBlog-Corpus.

It is worth mentioning that the classification techniques we employ to test the three languages are the most widely used. In fact our focus is to apply them to test EmotiBlog-Annotation-Model as first step. However in the future we will enter in detail in the Machine Learning algorithms to reach a semi automatic annotation of subjectivity and thus we will compare different Machine Learning algorithms and techniques.

The algorithms we choose are SVM and Multinomial Naïve Bayes. We decided to employ them and not carrying out a deep and detailed comparison of different

Machine Learning approaches because in this way we can compare our results with similar research carried out in Sentiment Analysis (see Related Work section). Moreover, due to the fact that the EmotiBlog-Corpus is unique, we employ the well-known 10-fold cross-validation method in order to have a bigger dataset.

Furthermore, the techniques we use are robust, due to their extended usage, and also their versatility. Using such approaches we believe our work will be more comparable.

Taking into account the fact that our annotation scheme is extremely fine-grained, our aim here is to check if this high–level of granularity contributes to an improvement of the State of the Art. We want to the validity of the annotation with EmotiBlog-Annotation-Model for classification purposes and also to measure the impact of each element to be able to refine our resource and propose the final version.

5.1 Feature classification for spanish

As a first step we will describe a set of preliminary Machine Learning experiments implemented with the Spanish part (30000 words) of our annotated corpus about the Kyoto Protocol (EmotiBlog-Kyoto-Annotated) to evaluate the consistency of our model and the impact on the extended list of elements.

We first perform Machine Learning experiments checking which is the global system performance and after that we enter in detail in the EmotiBlog-Annotation-Model elements to measure the impact they have in such classification.

As we explained above, we decided to use the Spanish part because it is well known that the Spanish syntactic structure is more complex than English, thus we started with it in order to test EmotiBlog-Annotation-Model in the most complex context.

Our procedure consisted in using each sentence as an individual instance. They have been extracted from the corpus annotations, since each sentence is labelled separately for the classification task, thus the sentence terms correspond to the features. In our system each feature represents a word or a set of them, and can be used as they were found in the text or using their stem, depending on the experiment.

They have been extracted splitting each sentence into individual words (by means of a tokeniser) and its polarity is the category.

Thus, the sentences and their polarity have been extracted from the corpus annotations, while the terms have been obtained splitting the sentences into words. The difference between each experiment consists on the set of terms used.

The figure below shows the EmotiBlog-Kyoto-Annotated in Spanish occurrences of positive, negative and objective sentences. As we can see, there are present in a different percentage since the corpus we collected constitutes a real example of genuine collection of blog posts. Our main objective is to work with real examples in order to be sure our model will be useful for real-life applications.

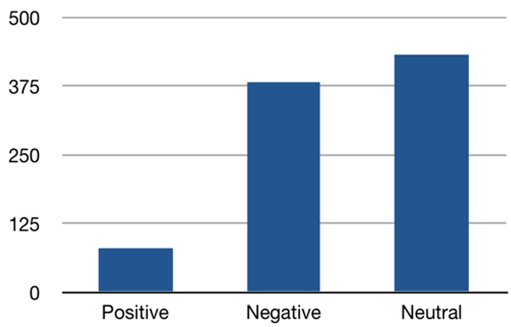


Figure IV: EmotiBlog-Corpus-Kyoto occurrences of positive, negative and objective sentences

The blog were above all about the Bush's decision not to sign the Kyoto Protocol and thus we have a slight minority of examples of sentences presenting subjective data and the majority of them are objective, explaining facts around the subject of the Kyoto Protocol and the Bush government, while the rest express negative subjective content.

For the experiments implementation we used the Weka[69] implementation of the widely used Support Vector Machine (Gamon 2004); (Vapnik 1995)and Multinomial Naïve Bayes (Lewis and Gale 1994) (Sebastiani 2002) algorithms. We chose this tool since it consists in a collection of Machine Learning algorithms for data mining tasks

69 http://www.cs.waikato.ac.nz/ml/weka/

and the algorithms can either be applied directly to a dataset or called from your own Java code. Moreover, it contains tools for data pre-processing and classification between others and it is widely-used for developing new Machine Learning schemes. Moreover, we chose these algorithms because much of the research in Sentiment Analysis demonstrated their effectiveness for this type of tasks. The first due to its robustness against noise and the second because of its simplicity and efficiency. We also employed stemming techniques in some experiments (the Snowball[70] implementation for the Spanish language) to obtain the roots of the words we take into account.

As starting point, we extracted a bag of words from the corpus without taking into account the annotated elements. We will use this approach as baseline because it is the simplest one and it does not take into account fine-grained annotations.

Here it is worth mentioning that Spanish is a language that uses accents, but users in informal registers may decide to use them or not. After having carried out an empiric study of our corpus we can deduce that bloggers often neglect to put them. For this reason we decided to eliminate all of them in our experiments.

In some tests we also eliminated stop words and negation words. In fact, the stop words do not add meaning to the text and the negation because we will conduct a special study on negation in the future.

However, although the negation does not add any meaning by itself and does not contain semantic information (Yang and Pedersen 1997) we check how its inclusion or exclusion could affect the results. Although the polarity changes depending on the position of these words inside the sentence, in our preliminary experiments our aim was to evaluate how much their presence change the results.

Moreover, as mentioned in Section 3, we would like to underline the fact that there are some words that acquire a different meaning when are grouped together, such as *sayings* and *collocations*. They are properly annotated in the corpus as global expressions (See EmotiBlog-Annotation-Model description Chapter 3), thus in some of our experiments we consider those groups of words as single features to check if these elements have a relevant impact for the subjectivity detection.

70 http://snowball.tartarus.org/

Finally the *stemming* consists in the reduction of the inflected or derived words to their stem, base or root form.

The abbreviations employed in the tables blow that present the results of our experiments are the following:

- BL (baseline)
- RS (removing stopwords)
- RN (removing negation)
- SC (sayings and collocations as single features)
- ST (stemming)

TABLE XV. presents the results in terms of accuracy and F-measure for the MNB and SVM Machine Learning algorithms for each option in Spanish.

Table XV.

Results in terms of accuracy and F-measure for the MNB and SVM Machine Learning algorithms for each option in Spanish

Combinations	ML Features	MNB		SVM	
		Accuracy	F-measure	Accuracy	F-measure
BL	941	0.647	0.592	0.685	0.644
RS+RN	877	0.566	0.477	0.654	0.610
RS	878	0.532	0.420	0.625	0.572
SC	875	0.588	0.511	0.663	0.620
ST	**819**	**0.672**	**0.625**	**0.714**	**0.683**
RS+RN+ST	764	0.594	0.516	0.661	0.618
RS+ST	765	0.622	0.556	0.689	0.652
SC+ST	781	0.617	0.554	0.694	0.659

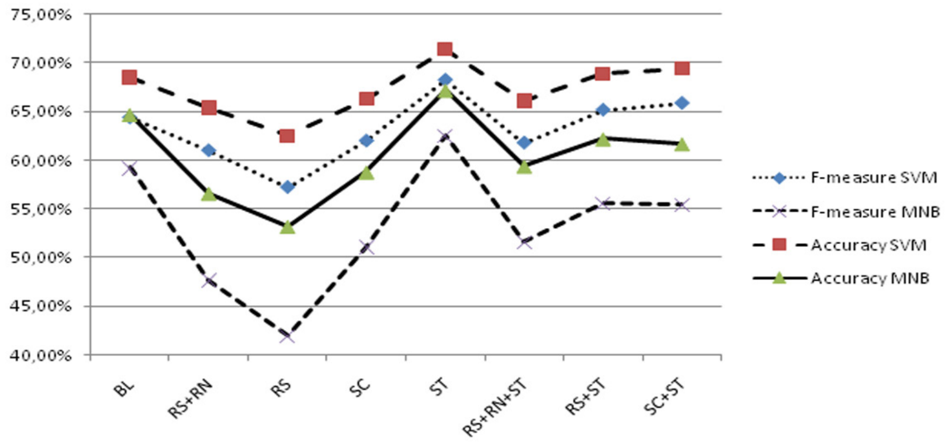


Figure V: Comparative results in terms of accuracy and F-measure for the MNB and SVM Machine Learning algorithms for each option in Spanish

TABLE XV. presents the results obtained in our preliminary experiments. The values obtained show that our starting point reaches an accuracy of 68.5% and an F-measure of 64.4% using SVM.

With MNB the results are about 6% and 10% lower in terms of accuracy and F-measure respectively, while the best results were obtained using stemming techniques achieving an accuracy of 71.4% and an F-measure of 68.3%. In fact the use of stemmer improves the results in every experiment we performed with this approach.

As we can also see in TABLE XV. the use of a full list of stopwords influences negatively the results. This is probably due to the fact that since they do not add any significant meaning to the sentence, they generate noise.

An additional aspect we would like to underline is that the fact that the negation is a relevant feature. We deduce this because its inclusion raises the results between 4% and 7% with respect to the same experiment but carried without this feature.

Globally, analysing the ratings obtained, we can say that they are promising and thus they encourage us to continue working on the model.

5.2 Feature selection for spanish

In order to evaluate the features described in the EmotiBlog-Annotation-Model section, we also conducted experiments employing feature selection by dimensionality reduction focused on the polarity classification task, and more specifically, applying the Information Gain (IG) (Lewis and Gale 1994) algorithm and focusing on the polarity classification task.

The goal of this dimensionality reduction is to obtain the best features for the polarity classification to improve and simplify the classification task. In fact we believe that the polarity classification is the less complex and thus in order to evaluate each element of the EmotiBlog-Annotation-Model we should assure a high performance for the polarity classification that constitutes the basis and after that enter in detail with the elements classification. Apart from that, we also have to bear in mind that the polarity classification includes a well-performed classification of the elements, result of the global polarity, thus it represents a key issue of the EmotiBlog-Annotation-Model validity check.

In these experiments we employ the same corpus used above for the feature selection, thus the 30,000 words of the EmotiBlog-Kyoto-Annotated.

We apply the global feature selection, which measures the relevance of each term of the corpus taking into account the three categories of polarity. The results obtained are shown in TABLE XVI.

Table XVI.

Results of the global feature selection for Spanish

Percentage of Selected Features	ML Features	MNB		SVM	
		Accuracy	F-Measure	Accuracy	F-Measure
80%	752	0.487	0.322	0.496	0.341
85%	799	0.485	0.320	0.525	0.399
90%	846	0.488	0.325	0.539	0.429
95%	893	0.510	0.371	0.598	0.528
99%	931	0.564	0.470	0.650	0.602

As shown in Figure IV the corpus we are using is unbalanced because it is a real sample of blog posts, thus the categories do not have a similar amount of sentences, especially for the positive polarity) and the non-neutral polarities (especially positive ones) have fewer instances, so the frequency of the positive terms decreases noticeably.

After having carried out these experiments and seen that the more feature we use, the best results we obtain, we also wanted to check the impact of each EmotiBlog-Annotation-Model element, an extremely crucial aspect for us since we are proposing an extreme-fine grained model for the annotation of the subjectivity.

Even if we want to improve the state of the art creating a model more detailed which is able to capture the major number of features employed by the blogger to express its subjective information, we have to be sure that this high fine-granularity is employable in automatic classification.

The most important aspect to take into account here is that we need high percentages of classification in order to be able to carry out the annotation of the rest of the EmotiBlog-Corpus in a less time-consuming way. That is why we evaluated the impact of each element in the system. Results are presented in the table below.

Table XVII.

Impact of the EmotiBlog elements on the system for Spanish

ELEMENT	EFFECT
verb	2.998%
phrase	2.664%
adjective	2.244%
noun	1.756%
preposition	0.338%
pronoun	−0.323%
onomatopoeic	−0.784%
adverb	−0.914%

TABLE XVII. shows the percentage of improvement for each element. It represents the proportion of experiments that have been improved by including each feature (their impact).

It has been calculated by taking the number of experiments improved by adding an element and diving this by the number of experiments that did not experiment an improvement.

Looking at the table above, we can say that the majority of the elements have a beneficial effect on the system. However pronoun, onomatopoeic and adverb gave negative results.

This measurement of the impact of each element is very useful because, depending on our needs, we can decide what to include for the Machine Learning system training depending on our requirements.

5.3 Reclassification for english

After having carried out the annotation of the EmotiBlog-Corpus in Spanish, calculated the inter-annotator's agreement and carried out feature selection experiments for the Spanish part, we labelled the EmotiBlog-Corpus in English about the Kyoto Protocol and we carried out the cross-fold evaluation.

TABLE XVIII. shows the results obtained for English after the reclassification of our corpus in terms of precision, recall and F-measure, whilst TABLE XIX. illustrates the impact of the EmotiBlog-Annotation-Model elements for the English corpus. After having carried out the experiments with the Spanish language our purpose now is to check the validity of the EmotiBlog-Kyoto-Annotated in English. Since it is a language with a completely different syntactic structure, we want to check how the system classifies the English part of the corpus and also to measure the impact of the EmotiBlog-Annotation-Model elements because our idea is to provide a multilingual resource. In this way we can test our annotation model with two different languages and thus, depending on the results obtained, decide the definitive list of elements.

Table XVIII.

Reclassification results for English

	Precision	Recall	F-Measure
Subj.	0.922	0.754	0.830
Obj.	0.756	0.72	0.738
Posit.	0.721	0.82	0.767
Neg.	0.924	0.924	0.924
Neut.	0.956	0.985	0.970

Table XIX.

Effect of the EmotiBlog-Annotation-Model elements on the system for English

ELEMENT	EFFECT
phrase	2.951%
verb	0.560%
pronoun	0.337%
adjective	0.221%
noun	−0.177%
onomatopoeic	−0.278%
preposition	−0.283%
adverb	−0.525%

The reclassification results are more satisfactory if compared with previous results.

TABLE XIX. shows that concerning the English language the elements that have beneficial effect on the system are less than for Spanish language.

Comparing the impact obtained for Spanish and English we decided to refine the model. Even if adverbs obtain negative results for the polarity classification task our idea is to keep them in the model because we believe that they would be useful for a finer-grained classification for example for the intensity classification task. We also decided to keep in the model the elements that give for one language positive results and for another negative because their summing was positive, but we deleted the onomatopoeic element because experiments proved they are not necessary and relevant for the classification process.

5.4 Sentence-level classification for italian

In order to check if EmotiBlog-Annotation-Model is indeed a useful resource for all the languages contemplated (English, Italian and Spanish), in the subsequent experiments we tested the ability to detect and classify opinion in blog posts in Italian. Due to annotation reasons we cannot carry out the same experiments about the effect of each feature, thus we did the cross-validation of the sentence-level opinion classification.

In line with the previous tests for English and Spanish, we used the refined EmotiBlog-Annotation-Model to label the EmotiBlog-Kyoto in Italian.

Subsequently, each of the annotated elements was extracted from the corpus. In the following experiments we tested, through a ten-fold cross-validation, if we are able to correctly classify the sentences in the blog posts in Italian.

In order to achieve this, we computed the Lesk (Salton and Lesk 1971) similarity score between each of the sentences in the corpus and each of the annotated elements. Thus, each sentence was represented as a vector of features, each of these corresponding to the similarity scores to all annotated elements.

Subsequently, we performed two types of classification. The first one was geared towards assessing the accuracy of distinction among subjective and objective sentences. The second classification was conducted among the 3 considered classes of sentiment polarity: positive, negative and objective.

The result of the cross-validation is presented in TABLE XX.

Table XX.
Cross-validation of the sentence-level opinion classification for Italian

	Precision	Recall	F1
Subj.	0.731	0.563	0.636
Obj.	0.861	0.675	0,754
Posit.	0.712	0.732	0,722
Neg.	0.894	0.951	0,922

As we can see in TABLE XX. the results for the experiments with the part of the EmotiBlog-Kyoto-Annotated in Italian classification stay in line with the previous evaluations.

Nevertheless, a slight drop in performance when classifying the blog sentences can be observed, both among subjective and objective, as well as among the three classes of sentiment polarity. The explanation for this phenomenon is the higher language variability in this subset of the corpus, which was observable from the feature vectors' scarcity.

5.5 Conclusions

In this chapter we carried out an intrinsic evaluation focused on testing if EmotiBlog-Annotation-Model is a consistent annotation scheme and to verify if the extended list of elements and their attributes has a positive effect for the automatic classification.

Thus we carried out a gradual set of preliminary Machine Learning experiments carried out in the Spanish, English and Italian parts of our annotated corpus in order to have an idea of the validity of EmotiBlog-Annotation-Model in classification experiments and its usefulness for working with tree languages, allowing multilingual and comparable studies.

Employing the most widely used Machine Learning techniques we selected the Support Vector Machine and Multinomial Naïve Bayes algorithms to check the validity of the annotation with EmotiBlog-Annotation-Model for classification purposes and to measure the impact of each EmotiBlog-Annotation-Model element.

The results obtained are promising and experiments have shown that this model is appropriate for the training of Machine Learning models for the multilingual Sentiment Analysis task.

As we mentioned in the introductory section, Sentiment Analysis is the first step toward achieving applications that exploit this language analysis and employ it for concrete purposes. In fact the efficient treatment of subjective language is an essential basis for a satisfactory performance of any application. As a consequence the next chapter of this work will present the tools and techniques we employ to carry out the extrinsic evaluation of EmotiBlog-Annotation-Model and corpus by means of checking if its inclusion could positively affect the performance of systems dealing with the tasks of Opinion Mining, Opinion Question Answering and Opinion Summarisation.

6. Tools and techniques for the extrinsic evaluation

As we explained in the introductory section, after having described the EmotiBlog-Annotation-Model we carried out the intrinsic evaluation calculating the inter annotator agreement and performing feature selection experiments with the purpose of testing the coherence, robustness and adequacy of the model and thus the usefulness of our resource and the reliability of the annotation produced by the EmotiBlog-Annotation-Model.

After that, the next part of this work is dedicated to carry out an extrinsic evaluation of the EmotiBlog-Corpus-Annotated by means of exploiting its labelling to improve the state of the art of some concrete Natural Language Processing tasks.

Many tasks could have been selected as example, however for this work we decided to take into consideration three of them:

- Opinion Mining: Given a set of evaluative text documents D that contain opinions (or sentiments) about an object, opinion mining aims to extract attributes and components of the object that have been commented on in each document $d \in D$ and to determine whether the comments are positive, negative or neutral (Liu 2010).
- Question Answering for opinionated content: an information retrieval application whose aim is to provide inexperienced users with a flexible access to information, allowing them writing a query in natural language and obtaining not a set of documents that contain the answer, but the concise answer itself (Vicedo, et al. 2003).
- Automatic Summarisation of subjective texts: methods developed in this field try to replace human summarisers by producing summaries using automatic means (Orasan 2006).

We decided to use the abovementioned tasks for different reasons. Concerning Opinion Mining we decided to test the EmotiBlog-Annotation-Model over this task since it is undoubtedly related with the treatment of subjective data. As we have seen in chapter 3, where the state of the art is presented, much work has been done but the resources created are most of them for English and they allow a coarse grained annotation, apart from the fact that they mostly concentrated on the treatment of the newspaper textual genre. Thus, our purpose is to check the impact of employing our resource on this pioneer task.

Concerning Question Answering dealing with opinionated data, it is well known that most of the research on this discipline has developed systems for factual questions, and the association of subjective information with Question Answering still present pending challenges due to the scarcity of adequate resources and methods to properly tackle this task. That is why we decided to employ the EmotiBlog-Annotation-Model to check if its employment can improve the performance of Question Answering systems dealing with opinionated data.

Automatic summarisation is the other task we decided to work with because it is different from the previous one. Here we wanted to check how the EmotiBlog-Annotation-Model and Corpus could be employed in order to make possible the summarisation of subjective content. This is a young discipline and much work has to be carried out in order to reach a high performance of summarisation systems dealing with opinionated content. Our purpose here is to test if our resource is a key factor for the effective treatment of opinionated content. Until now most of the work done has concentrated on the summarisation of factual data, thus the systems created are not able to properly manage other kind of information and the consequence is that if we have for example the text (38):

(38) A **good post**, and **pretty fair**.
But **what you don't observe** is that the EU, like the US, has basically followed a business-as-usual path on emissions. (You'll want to recheck your assertion on the comparison of EU vs. US population growth.)
There are also other accounting idiosyncrasies, such as those discussed here:
http://sciencepolicy.colorado.edu/prometheus/uk-emissions-4353
The bottom line is that if you really think that the EU experience is a success story, rather than a cautionary tale about the real chal-

lenges of reducing emissions even with strong political support, then **I am surprised**, as you usually take a more critical look at the data than demonstrated here.

I love it when people say it will hurt the US Economy, especially business leaders. Apparently they do not have the foresight to realize how much **$$$$$$$** they could make if we switch to a greener capitalist system.

As we can see from the example above, this blog post contains both factual and opinionated data. The information in bold is the most relevant opinionated data and if we employ a normal summarisation system, it will not be able to capture and summarise its content and the resulting summary would be just taken from the factual data. That is why we decided to exploit our resource in this task: to provide the summarisation system the opportunity to have a resource to properly interpret the subjective text and thus to be able to reflect it in the final summary.

Apart from that the tasks we chose are mainly focused on building up real-time applications for the exploitation of the huge amount of subjective information available on the Web 2.0 in real time.

Thus, in order to carry out our experiments, we mainly employed:

- The EmotiBlog-Corpus-Annotated with the EmotiBlog-Annotation-Model (its total feature list or a partial selection of them, depending on each case and purpose)
- The EmotiBlog-Corpus-Annotated enriched with different lexical resources, thus creating new corpora
- Other available corpora that could be useful to compare our results

Moreover, when carrying out the experiments we employed different algorithms to train our system depending above all on which aspects of EmotiBlog-Corpus-Annotated we want to exploit more.

As we presented in the chapter dedicated to the EmotiBlog-Annotation-Model in-depth description (chapter 3), the labelling this annotation scheme provides is extremely fine-grained. That is because, after having carried out a deep research on the state of the art, we drew the following conclusions:

- There is no model able to detect a such wide range of linguistic elements, which the user employs to shape the expression of subjectivity

- Most of the work already done mainly concentrates on the English language
- Previous research mainly focused on newspaper articles and other traditional textual genres, thus there is a lack of treatment of the new textual genres which now are a precious source of valuable information and influence people's behaviour

Thus, taking into account the factors mentioned above, our main goal from now on is to test the EmotiBlog-Annotation-Model validity as fine-grained, multilingual and appropriate for the new textual genres.

With the aim of present the EmotiBlog-Annotation-Model contribution to each task in a clear way, the following section presents and describes the corpora, algorithms and language treatment we used to work in each task.

6.1 Opinion mining task

As presented above, in the framework of Opinion Mining, our purpose is to check if the EmotiBlog-Annotation-Model and EmotiBlog-Corpus are useful resources to be employed in multilingual Opinion Mining and if it can improve the results obtained by the previous research. In fact, this task needs an effective treatment of subjective data and thus it represents the immediate Natural Language Processing Task that would need a resource like the one we created. As shown in the state of the art section, previous approaches mostly concentrated on subjectivity versus objectivity classification, thus less attention was paid on annotating emotion on a fine-grained level.

Spanish

In order to perform the experiments on the Opinion Mining task for the Spanish language (we carried out such experiments in Spanish and English due to the fact that our hypothesis is that Spanish and Italian would obtain similar results due to their similar syntactic structure), we used:

- As set of sentences on the recycling topic
- EmotiBlog-Kyoto-Annotated, our labelled corpus of blog posts about the in Spanish

- A manually created a set of 150 sentences on recycling, (50 positive and 50 negative). We decided to use this topic because it is related with the Kyoto Protocol and thus, after having carried out an empiric study of both collections we concluded that they share many terms and thus the topics would be compatible to be used for this experiment. We used this topic because we also wanted to check the effectiveness of EmotiBlog-Kyoto-Annotated in related even if not identical topics.

According to Section 3, we obtained 1647 subjective and 1336 objective sentences with an agreement of 0.59 and 0.74 respectively and in this context we decided to use those sentences upon which we agreed on the phrases and whose annotation length was more than four tokens of the type noun, verb, adverb or adjective.

We processed the sentences with:

- POS in order to detect all possible relationships with the adjacent words
- Lemmatised with FreeLing in order to group together the different inflected forms of a word so they can be analysed as a single item

After that, we represented each sentence as feature vector composed by unigram features containing:

- Positive and negative categories of nouns, adverbs, adjectives, prepositions and punctuation signs (having 1 in the corresponding position of the feature vector for the words contained and 0 otherwise
- The number of bigrams/trigrams and 4-grams overlapping with each of the phrases we labelled as positive and negative or objective, respectively
- The overall similarity given by the number of overlapping words with each of the positive, negative or objective phrases from the corpus, normalized by the length of the given phrase.

(38) shows examples of two vectors for joy and anger:

```
0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,joy
0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,anger
```

Thus, at this step the EmotiBlog-Corpus-Annotated plays a key role since each sentence to be classified is compared to each of the annotations in the EmotiBlog-Corpus-Annotated, and the similarity score (computed with the Lesk distance (Salton and Lesk

1971)) was used as a feature in the SVM SMO classifier.

We decided to employ this classifier since it is extremely useful in those cases in which we have sparse examples and this version has proven to be more precise than the traditional SMO algorithm (Knebel, Hochreiter and Obermayer 2008). Due to the fact that the EmotiBlog-Annotation-Model is extremely fine-grained, there is no other corpus with the same elements, and thus our set was unique. SVM SMO is ideal in such cases, where we have sparse features. While labelling with the EmotiBlog-Annotation-Model, the annotator can detect many different linguistic elements, which also are described by multiple attributes and the result of this process with a pilot corpus such as EmotiBlog-Kyoto-Annotated can be the sparse values of the features annotated. In order to solve this factor, SVM SMO allows us to have a more compact and significant representation of the different features taken into account.

- We tested our method with the sentence classification among subjective and objective, for which the vectors contain subjective or objective as global values
- We performed a classification of subjective sentences into positive and negative, for which case the classification vectors contain the values positive and negative
- We performed a ten-fold cross validation of the corpus for each of the two steps

In order to carry out the classification (on new examples, 150 sentences on recycling) using all n-gram features of our test data,

- We run FreeLing[71] on the set of positive and negative sentences on recycling to lemmatize and tag each word on part of speech
- We represent sentences as a feature vector in the same manner as in the first conducted experiment, and we carried out two experiments on this data:
    - The first one aims to train a SVM SMO classifier on the corpus phrases pertaining to the “subjective” versus “objective” categories, and to test it on answers for the recycling topic pertaining to the positive or negative categories. In this way we will check if the EmotiBlog-Corpus-Annotated is a useful resource to automatically discriminate subjective/objective sentences

71 http://nlp.lsi.upc.edu/freeling/

- o The second consists in classifying the instances according to positive and negative classes. This is done with the aim of testing of EmotiBlog-Corpus-Annotated is useful to provide an exact discrimination between positive and negative instances.

English

The experiments we carried out in the framework of Opinion Mining for English are mainly focused to check if our resource is effective to measure the sentiment intensity on the EmotiBlog-Annotation-Model three-level scale: *high, low and medium*. We decided to perform the experiments aso in this language due to the fact that English is syntactically different if compared with Spanish, thus our aim here is to check if our resource is also useful for English. The consequence is that in these experiments we employ different corpora and techniques as described below.

For this purpose, the corpora we use are:

- A small collection of quotes (reported speech) from newspaper articles presented in Balahur, Steinberger and Kabadjov, et al. (2010) enriched with the manual fine-grained annotation from the EmotiBlog-Annotation-Model
- The collection of newspaper titles in the test set employed in the SemEval 2007 task number 14 – Affective Text (Strapparava and Mihalcea 2007)
- A corpus of self-reported emotional response – ISEAR (Scherer and Wallbott 1997)

We decided to use such resources since they are well known and widely used collection and thus this will allow us to have comparable results. Apart from that, they are labelled with some of the EmotiBlog-Annotation-Model elements and this means that they are the ideal corpus to make comparable studies and to check if the use of our EmotiBlog-Corpus-Annotated would mean an improvement of the baseline.

It is worth mentioning that if for the polarity we will use the three corpora, for the intensity classification task we will employ only the second corpus, since it is the only one in which the polarity values (between -100 and 0 and 0 and 100) are assigned to the titles.

The steps taken in our experiment are briefly listed below:

- Training Models creation:
  - o Extraction of the Named Entities contained in the annotations using Lingpipe[72] and merged with a "_" all the tokens pertaining to the Named Entities (when we have a Named Entity composed by different nouns), to treat it as a single Name Entity
  - o Merging under a single punctuation sign all the annotations of punctuation signs that had a specific meaning together
  - o Processing of the annotated data, using Minipar[73] for detecting the source and topic of the discourse that generally is expressed by Name Entities
  - o Computing, for each word in a sentence, a series of features (some of these used in Choi, et al. (2005): part of speech, capitalisation, opinionatedness, intensity, emotion, Syntactic relatedness with another opinion word and Role in 2-word, 3-word and 4-word annotations -in many cases a subjective word can have next to it other subjective words part of the expression or modifiers of the element we took into account-)
- Polarity Measurement
  - o We computed the length of the longest sentence in EmotiBlog-Kyoto-Annotated. The feature vector for each sentence contains the feature vectors of each of its component words, and 0s for the corresponding feature vectors of the words (which the current sentence has less than the longest annotated sentence).
  - o Finally, to each sentence as a feature we added binary features for subjectivity and polarity, the value corresponding to the intensity of opinion and the general emotion. These feature vectors are fed into the Weka[74] SVM SMO Machine Learning algorithm and a model is created (EmotiBlog I).

72 http://alias-i.com/lingpipe/

73 http://webdocs.cs.ualberta.ca/~lindek/minipar.htm

74 www.cs.waikato.ac.nz/ml/weka/

- o Thus, a second model (EmotiBlog II) is created by adding to the collection of single opinion and emotion words annotated in EmotiBlog-Kyoto-Annotated, the Opinion Finder lexicon[75] and the opinion words found in MicroWordNet[76], the General Inquirer[77] resource and WordNet Affect (Strapparava and Valitutti 2004).

– Intensity Measurement

- o As we can deduce from above, different corpora means having to solve many challenges due to their different kinds of annotation. The most relevant difference, apart the fact that the other corpora have been labelled in a very coarse-grained way is the fact that researchers employ different polarity degree scales.
- o As we know from chapter 3, EmotiBlog-Annotation-Model contemplated three-degrees of polarity intensity: high – medium –low. We decided to employ such levels because we believe they would be less ambiguous for example if compared with a 1-100 scale but also with a 0-5 scale. Having just three options, the annotator will have less doubt at the time of labelling the polarity intensity.

Different labelled corpora (with various annotation schemes) means having to find a way in order to compare our scale with theirs. In fact, SemEval scores between -100 and 0 and 0 to 100 for the polarity of the titles) thus we mapped the values contained in the Gold Standard of the task into 3 categories:

– [-100, -67] is high (value 3 in intensity) and negative (value -1 in polarity)
– [-66, 34] medium negative
– [-33, -1] is low negative
– [1 and 100] are mapped in the same manner to the positive category.
– 0 was considered objective (since it does not express any subjectivity) and its intensity will be 0.

75 www.cs.pitt.edu/mpqa/opinionfinderrelease/

76 www-3.unipv.it/wnop/

77 www.wjh.harvard.edu/~inquirer/

6.2 Question answering task

In the framework of Question Answering task, our aim is employing the EmotiBlog-Corpus-Annotated, but also other existing resources, to understand if our resource improves the performance of systems dealing with some specific problems of Question Answering such as the identification of expected polarity, opinion retrieval, opinion analysis and answer retrieval in English and Spanish. Here also, we decided to employ here two languages because of their different syntactic structure and we believe that the results obtained for the Spanish language would be similar also for Italian. Moreover, research carried out in Opinionated Question Answering has only implemented a sentence-level approach, while in EmotiBlog-Corpus-Annotated we have document, sentence and element annotation levels (a relevant aspect to improve the multiple sentences answer retrieval).

The corpora employed in the framework of the Question Answering task evaluation are:

- EmotiBlog-Kyoto-Annotated in English and Spanish
- The TAC 2008 Opinion Pilot test collection (part of the Blog06 corpus), composed by documents with the answers to the opinion questions given on 25 targets containing documents on a multitude of subjects

The main difference between the two corpora is that EmotiBlog-Kyoto-Annotated is monothematic, in fact only posts about the Kyoto Protocol compose it, while the TAC 2008[78] corpus contains documents on a multitude of subjects.

As we already explained in Section 3, the first distinction the EmotiBlog-Annotation-Model allows is between objective and subjective speech. Subsequently, a finer-grained annotation is employed for each of the two types of data.

In order to perform our Opinionated Question Answering experiments, we build up a question collection (F: factual queries and O: opinionated questions) whose answers annotated with EmotiBlog-Annotation-Model are the subset of opinion questions in English and Spanish:

78 http://www.nist.gov/tac/2008/

Table XXI.

Set of mixed opinionated and factoid question in English and Spanish

NUM	TYPE		QUESTION
1	F	F	What international organization do people criticize for its policy on carbon emissions? *¿Cuál fue uno de los primeros países que se preocupó por el problema medioambiental?*
2	O	F	What motivates people's negative opinions on the Kyoto Protocol? *¿Cuál es el país con mayor responsabilidad de la contaminación mundial según la opinión pública?*
3	F	F	What country do people praise for not signing the Kyoto Protocol? *¿Quién piensa que la reducción de la contaminación se debería apoyar en los consejos de los científicos?*
4	F	F	What is the nation that brings most criticism to the Kyoto Protocol? *¿Qué administración actúa totalmente en contra de la lucha contra el cambio climático?*
5	O	F	What are the reasons for the success of the Kyoto Protocol? *¿Qué personaje importante está a favor de la colaboración del estado en la lucha contra el calentamiento global?*
6	O	F	What arguments do people bring for their criticism of media as far as the Kyoto Protocol is concerned? *¿A qué políticos americanos culpa la gente por la grave situación en la que se encuentra el planeta?*
7	O	F	Why do people criticize Richard Branson? *¿A quién reprocha la gente el fracaso del Protocolo de Kyoto?*
8	F	F	What president is criticized worldwide for his reaction to the Kyoto Protocol? *¿Quién acusa a China por provocar el mayor daño al medioambiente?*
9	F	O	What American politician is thought to have developed bad environmental policies? *¿Cómo ven los expertos el futuro?*
10	F	O	What American politician has a positive opinion on the Kyoto protocol? *Cómo se considera el atentado del 11 de septiembre?*
11	O	O	What negative opinions do people have on Hilary Benn? *¿Cuál es la opinión sobre EEUU?*
12	O	O	Why do Americans praise Al Gore's attitude towards the Kyoto protocol and other environmental issues? *¿De dónde viene la riqueza de EEUU?*

NUM	TYPE		QUESTION
13	F	O	What country disregards the importance of the Kyoto Protocol? *¿Por qué la guerra es negativa?*
14	F	O	What country is thought to have rejected the Kyoto Protocol due to corruption? *¿Por qué Bush se retiró del Protocolo de Kyoto?*
15	F/O	O	What alternative environmental friendly resources do people suggest to use instead of gas en the future? *¿Cuál fue la posición de EEUU sobre el Protocolo de Kyoto?*
16	F/O	O	Is Arnold Schwarzenegger pro or against the reduction of CO2 emissions? *¿Qué piensa Bush sobre el cambio climático?*
17	F	O	What American politician supports the reduction of CO2 emissions? *¿Qué impresión da Bush?*
18	F/O	O	What improvements are proposed to the Kyoto Protocol? *¿Qué piensa China del calentamiento global?*
19	F/O	O	What is Bush accused of as far as political measures are concerned? *¿Cuál es la opinión de Rusia sobre el Protocolo de Kyoto?*
20	F/O	O	What initiative of an international body is thought to be a good continuation for the Kyoto Protocol? *¿Qué cree que es necesario hacer Yvo Boer?*

As you can see in the table above, we created factoid and opinionated queries for English and for Spanish; however, there are some that could be defined between factoid and opinion and the system can retrieve multiple answers after having selected for example, the polarity of the sentences in the corpus.

The steps we carry out in our experiments are:

– In order to perform an effective question analysis:
  - o We need to determine both the Expected Answer Type (EAT) of the question – as in the case of factoid ones, as well as new elements, such as Expected Polarity Type (EPT)
  - o However, opinions are directional, thus they suppose the existence of a source and a target to which they are addressed.

As a consequence, we introduce two new elements in the question analysis:

- The Expected Source (ES)
- The Expected Target (ET)

These two elements are selected by applying Semantic Roles and choosing the source as the agent in the sentence and the direct object (patient) as the target of the opinion.

- The expected answer type (EAT) (e.g. opinion or other) is determined using Machine Learning using SVM SMO, by taking into account the interrogation formula, the subjectivity of the verb and the presence of polarity words in the target Semantic Roles.
- In the case of expected opinionated answers, we also compute the Expected Polarity Type – by applying Sentiment Analysis on the affirmative version of the question. An example of such a transformation is: given the question *"What are the reasons for the success of the Kyoto Protocol?"*, the affirmative version of the question is *"The reasons for the success of the Kyoto Protocol are X".*

Candidate snippet retrieval

In the answer retrieval stage, we employ four strategies:

1 Using the JIRS (JAVA Information Retrieval System) IR engine (Gómez, Rosso and Sanchis 2007) to find relevant snippets. JIRS retrieves passages (of the desired length), based on searching the question structures (n-grams) instead of the keywords, and comparing them. At this stage we use the EmotiBlog-Kyoto-Annotated to help the system to learn how to classify the answers thus its annotation has a key role in the process.

2 Using the *Yahoo!*[79] search engine to retrieve the first 20 documents that are most related to the query. Subsequently, we apply Latent Semantic Analysis on the retrieved documents and extract the words that are most related to the topic. Finally, we expand the query using words that are very similar to the topic and retrieve snippets that contain at least one of them and the ET.

79 http://es.yahoo.com/

3 Generating equivalent expressions for the query, using the DIRT paraphrase collection (Lin and Pantel 2001) and retrieving candidate snippets of length 1 and 3 (length refers to the number of sentences retrieved) that are similar to each of the new generated queries and contain the ET. Similarity is computed using the cosine measure. Examples of alternative queries for *"People like George Clooney"* are *"People adore George Clooney", "People enjoy George Clooney", "People prefer George Clooney".*
4 Enriching the equivalent expressions for the query in 3, with the topic-related words discovered in 2., using Latent Semantic Analysis.

Polarity and topic polarity classification

In order to determine the correct answers from the collection of retrieved snippets, we must filter only the candidates that have the same polarity as the question Expected Polarity Type.

For polarity detection (in both question and answer) we use:

– A combined system employing SVM SMO on unigram and bigram features trained on the NTCIR MOAT 7[80] data and an unsupervised lexicon-based system to have more data at our disposal.
– In order to compute the features for each of the unigrams and bigrams, we compute the tf-idf scores.

The unsupervised system uses the Opinion Finder lexicon to filter out subjective sentences – that contain more than two subjective words or a subjective word and a valence shifter (obtained from the General Inquirer resource).

Subsequently, it accounts for the presence of opinionated words from four different lexicons – Micro WordNet (Cerini, et al. 2007), WNAffect (Strapparava and Valitutti 2004) Emotion Triggers (Balahur and Montoyo 2008) and General Inquirer (Stone, et al. 1966).

Joint topic-polarity analysis:

– We first employ Latent Semantic Analysis to determine the words that are strongly associated to the topic.

80 http://research.nii.ac.jp/ntcir/permission/ntcir-7/perm-en-MOAT.html

– Consequently, we compute the polarity of the sentences that contain at least one topic word and the question target.
– *Finally, answers are filtered using the SemRol system for Semantic Role labelling described in (Moreda 2008)*. Subsequently, we filter all snippets that have the required target and source as agent or patient.

6.3 Automatic summarisation

The experiment for the Automatic Summarisation task are dedicated to check if the EmotiBlog-Annotation-Model brings any improvement to the Opinionated Summarisation task applied to blog posts on a selection of topics, focused on providing the user with a summary (of different compression rate) of positive and negative opinions about a specific topic employing the EmotiBlog-Annotation-Model.

For the Automatic summarisation related experiments we used:

– 51 blogs extracted from the Web (299.568 words). The blog posts are written in English, have the same structure and have been annotated a selection of the EmotiBlog-Annotation-model: polarity, degree, source and topic (and also subtopics).

After having collected the corpus, we labelled it using some of the EmotiBlog-Annotation-Model elements presented below:

– Polarity (Positive, negative)
– Level (Low, medium, high)
– Source (name)
– Target (name)

Extract (39) is an example of annotation:

(39) **<topic>**economic situation**</topic>**
<topic2>government**</topic2>**

<topic3>banks**</topic3>**

<new> Saturday, May 9, 2009 My aim in this blog has largely been to give my best and most rational perspective on the reality of the economic situation. I have tried (and I hope) mostly succeeded in avoiding emotive and partisan viewpoints, and have tried as far as possible to see the actions of politicians as misguided. Of late, that perspective has been slipping, for the UK, the US and also for Europe.

<phenomenon gate:gateId=”1” target=”economic crisis” degree1=”medium” category=”phrase” source=”Cynicus

Each of the elements indicated in the list above has been selected because they provide important information that is relevant to the task at hand. The polarity has the function of indicating if the opinion expressed in the sentence is positive of negative.

Moreover, we labelled the data at the opinion level, choosing the level of polarity intensity between low, medium or high. Finally, we specified the source of the discourse in order to be able to detect who said what, and the target of the sentence, so as to understand the topic of the discourse.

We decided not to include all the elements of EmotiBlog-Annotation-Model to avoid noise. The result of the annotation process is a gold standard, which will be used to evaluate some of the aspects of the generated summaries. The subjective sentences are annotated with polarity and its intensity but also with the source and the target of the discourse.

We would like to stress upon the fact that we have the option to indicate more than one topic (topic + subtopic(s)). We decided to contemplate cases of multiple topics only if they are relevant in the blog. In this case (39), the main topic is the economic situation, while the secondary ones are the government and banks.

After having defined the topics, the first paragraph contains objective information and thus, we do not label it; we therefore annotate the following sentence that contains subjective information. As you can see, the economic crisis is the target. Finally, the polarity of the sentence is negative, the intensity level of this polarity is medium and the author is Cynicus Economicus.

Opinionated sentence detection

The first step we took in our approach was to determine the opinionated sentences, assign each of them a polarity (among positive and negative) and a numerical value (1-3 as in EmotiBlog-Annotation-Model: high-medium-low) corresponding to the polarity strength (the higher the negative score, the more negative the sentence and similarly, the higher the positive score, the more positive the sentence).

Given that we are faced with the task of classifying opinion in a general context, we employed a simple, yet efficient approach, presented in (Balahur, Steinberger and van der Goot, et al. 2009) exploiting the annotation of the EmotiBlog-Corpus to classify the sentence polarities. Moreover, in order to have a more extensive database of affect-related terms, we used:

– WordNet Affect (Strapparava and Valitutti 2004)
– SentiWordNet (Esuli and Sebastiani 2006)
– MicroWNOp (Cerini, et al. 2007)

Each of the employed resources was mapped to four categories, which were given different scores: positive (1), negative (-1), high positive (4) and high negative (-4). These values performed better than the usual assignment of only positive (1) and negative (-1) values.

– First, the score of each of the blog posts was computed as sum of the values of the words identified; a positive score leads to the classification of the post as positive, whereas a final negative score leads to the system classifying the post as negative.
– Subsequently, we performed sentence splitting using Lingpipe[81] and classified the obtained sentences according to their polarity, by adding the individual scores of the affective words identified.

We have used the combined resources, which have proven to classify in a more balanced manner (Balahur, Steinberger and van der Goot, et al. 2009). The measure of the intensity scores can also be used as an indication of the sentence importance and can it thus constitute a criterion for summarization, as shown in (Balahur, Lloret, et al. 2008).

After having clarified the resources and the methods we employ in order to carry out the extrinsic evaluation of our EmotiBlog-Annotation-Model, the following sections will present the way in which we exploited EmotiBlog-Annotation-Model and Corpus in the Opinion Mining, Opinion Question Answering and

81 http://alias-i.com/lingpipe/

Opinion Summarisation tasks and the conclusions we can draw about the role of our resource and how it impacts on the baselines obtained in the state of the art.

6.4 Conclusions

This chapter is an introduction to the extrinsic evaluation we carried out to the EmotiBlog-Annotation-Model and Corpus. Before entering in detail with our experiments we explained why we decided to work with Opinion Mining, Opinion Question Answering and Opinion Summarisation and we underlined what we wanted to check about the EmotiBlog in each of these tasks.

After that, we pointed out that in order to carry out our tests we employed different resources, techniques and text processing in each case. The most important reason for that is our idea is that, in order to check the EmotiBlog-Annotation-Model usefulness in the tasks we have to compare the results obtained with other previous experiments. We should also employ other created resources in order to compare their performance and use a mixed approach in order to exploit as much as possible the work done; in this way we can enrich the resource we created.

After having presented the resources, tools and processes we will use in the EmotiBlog-Annotation-Model test, next chapter presents the application of our resource to the Opinion Mining task.

7. Opinion mining, opinion question answering and opinion summarisation

After having described and explained the tools and resources employed to carry out the extrinsic evaluation of the EmotiBlog-Annotation-Model, this chapter will be devoted to illustrate the EmotiBlog-Annotation-Model contribution to the Natural Language Processing tasks of Opinion Mining, Opinion Question Answering and Opinion Summarisation. Our purpose with such experiments is checking if EmotiBlog is a useful resource for systems performing these tasks and thus if it can bring a significant improvement to them.

Opinion Mining can be defined as *a recent discipline at the crossroads of Information Retrieval and Computational Linguistics, which is concerned not only with the topic a document treats, but also with the opinion it expresses* (Esuli and Sebastiani 2006).

As we can deduce, this task needs the treatment of subjective data and thus it represents the immediate Natural Language Processing Task that would need a resource like the one we created. As a consequence we decided to test our EmotiBlog-Annotation-Model firstly with such task. Our purpose here is to check if the EmotiBlog-Annotation-Model and annotated-Corpus is a useful resource to be employed in multilingual Opinion Mining and if it can improve the results obtained by the previous research.

As shown in the state of the art section, in chapter 3, previous approaches focused on corpus annotation, mostly concentrated on subjectivity versus objectivity classification, thus less attention was paid on annotating emotion on a fine-grained level.

After having carried our empirical study on the corpus we collected we reached the conclusion that a finer-grained level of granularity was essential and the first step to reach it is the contemplation of three annotation levels. That is why the EmotiBlog-Annotation-Model allows:

– Document
– Phrase
– Element

levels of annotation. They are related one with the others, and thus this dependence between components of the discourse is useful to construct similarity models for training Machine Learning algorithms that process different values of n-grams, as well as sentences as a whole.

In this chapter we propose and evaluate a method for subjectivity polarity classification, based on n-gram and phrase similarity features used with Machine Learning taken from the annotated EmotiBlog-Corpus. We check if the information provided by the EmotiBlog-Annotation-Model is useful for Machine Learning systems that need to exploit information related with subjectivity and discourse structure.

Thus, our evaluation will be two fold:

– By cross-fold validation of the subjective and objective phrases in the annotated EmotiBlog-Corpus
– Using as alternative resource, a corpus of negative and positive opinions on recycling, created ad hoc but annotated with the EmotiBlog-Annotation-Model

The use of n-grams has to be taken into account as additional as well as the annotation of the single words. Thus the strategy we aim to implement here to check the validity of EmotiBlog-Corpus is to exploit its different levels of annotation.

Moreover, we would like to stress the fact that we decided to work with the Spanish and English corpora since we assume that in the case that the approach we take for Spanish obtains positive results, the same approach can be used with similar results for the Italian corpus since both languages share a very similar syntactic structure. Thus our aim here is to use English and Spanish due to their substantial differences at syntactic level.

Research in opinion-related tasks has gained importance in the past years, but there are still many aspects that require analysis and improvements, especially for approaches that combine Sentiment Analysis with other Natural Language Processing tasks such as Question Answering.

The TAC 2008 Opinion Pilot task[82] and the subsequent research performed on the competition data have demonstrated that answering opinionated questions is significantly different from the equivalent tasks in the context but dealing with factual data and this fact was also confirmed by recent work by (Kabadjov, Balahur and Boldrini 2009). That is why we selected Opinionated Question Answering as a task for testing our resource.

After having checked the usefulness of the EmotiBlog-Annotation-Model for the Opinion Mining Task, we also check if it can be employed as a useful resource to help to improve the performance of Question Answering systems dealing with opinionated data, quite different from the ones that have to treat only factual information.

In the this case we employ our EmotiBlog-Corpus-Annotated, but also other existing resources, to understand if it can improve the performance of systems dealing with some specific problems of Question Answering such as the identification of expected polarity (EPT, expected source – ES and expected target –ET-), opinion retrieval (at the level of one and three-sentences long snippets, using topic-related words), opinion analysis (using topic detection and anaphora resolution) and answer retrieval in English and Spanish. Again we decided to employ here two languages because of their different syntactic structure and we believe that the results obtained for the Spanish language would be similar also for Italian.

An additional motivation for testing the EmotiBlog-Corpus-Annotated in the framework of this task is the fact that although previous approaches opinion questions have longer answers than factual ones, the research done in Opinionated Question Answering so far has only considered a sentence-level approach, while in EmotiBlog-Corpus-Annotated we have three levels of annotation: document, sentence and element, an aspect we can exploit to improve the retrieval of multiple sentences answer.

82 www.nist.gov/tac/2008/summarization/op.summ.08.guidelines.html

The last Natural Language Processing task we decided to take as example to test the EmotiBlog-Annotation-Model usefulness is Opinionated Summarisation applied to blog posts on a selection of topics. Its main purpose is to provide the user with a summary (of different compression rate) of positive and negative opinions about a specific topic employing the EmotiBlog-Annotation-Model.

In order to carry out this test, we labelled a collection of a corpus of blog posts together with the comments given on them (threads) in English about different topics, at the level of opinion, polarity and post/comment, as well as sentence importance using the EmotiBlog-Annotation-Model. We decided to select five macro topics that are economy, science and technology, cooking, society, and sport.

After having collected the corpus, we employed a partial version of EmotiBlog-Annotation-Model to avoid noise. In fact the EmotiBlog-Annotation-Model is fine-grained, but for the first step of our experiments we only needed some of the elements the taken from the annotation scheme.

After that, exploiting the annotation we trained the system and automatically classified the polarity at a sentence and also at a document level and we propose a method to summarize similar opinions grouped for topics.

7.1 Opinion mining experiments in spanish

After having carried out the inter-annotator agreement, we in this experiment we consider only the sentences upon which we agreed on the phrases whose annotation length was not more than four tokens of the type noun, verb, adverb or adjective. We processed them as explained in section 6 and represent them as a feature vector.

As briefly touched above, our idea of using n-grams has been considered in order to exploit as much as possible the EmotiBlog-Annotation-Model structure and the information it provides. In fact, the three levels allow the annotation of the overall document subjectivity, after that the sentence and for each sentence we label the linguistic elements which give subjectivity to the text. They can be more than one in the same sentence and in most of the cases that are related one with the other.

The result is that, during the annotation process we are carrying out two processes in one: the annotation of the subjectivity, but also the detection of the connection between different words of the same sentence.

Thus we decided to exploit as much as possible this syntactic information we have can obtain from the EmotiBlog-Annotation-Model and train the Machine Learning system with the single words, but also using n-grams to take advantage of the EmotiBlog-Annotation-model structure that describes the relation between annotated elements.

Example of this correlation can be the example (40) below:

(40) I ***love*** *your* ***pretty*** *dog*

As showed above, it is common to find an *adverb*+ *adjective*+ *noun* with a subjective connotation that is linked with the connotation of the following/previous linguistic element. Moreover, for example the polarity of a sentence will be composed by the polarity of the subjective elements within such sentence.

Thus, even if it is not a fixed rule, in some cases the syntactic structure can help us to detect more subjective elements around the one automatically detected and this it could be positive in many cases even if we have to be aware that this strategy cannot be used to detect the totality of subjective elements in a text.

Each sentence is classified and compared to each of the annotations in the EmotiBlog-Corpus, and the similarity score.

After that, we tested our method with the sentence classification (subjective and objective).

The results, taking into account precision and recall are presented in TABLE XXI.

As a baseline we have a 33% of possibility of good classification. In fact we use 3 classes (positive, negative and objective) with the same number of examples, thus we have a one out of three chances to classify it in the right way.

Table XXII.

Classification using ten fold cross validation

	PRECISION	RECALL	F1
Subj.	0.988	0.632	0.771
Obj.	0.682	0.892	0.687
Posit.	0.799	0.511	0.623
Neg.	0.892	0.969	0.929

Subj/Obj represents the classification of phrases among subjective and objective and the Pos/Neg stands for the sentences classification according to their polarity: positive or negative.

This classification was performed on new examples, the 150 sentences on recycling presented in section 6.

In these experiments, we test the importance of the annotating affect in texts at the token level. Thanks to the EmotiBlog-Corpus-Annotated, we have a large number of nouns, verbs, adverbs and adjectives labelled as positive or negative and also at the emotion level.

We used these words when classifying examples using n-grams, with n ranging from 1 to 4. In order to test their importance, we removed the vector components accounting for their presence in the feature vectors and re-classified, both at the level of objective versus subjective, as well as at the positive, negative level. In TABLE XXIII. we can see the results obtained.

Table XXIII.

Classification results using all n-grams and n-grams, n>2

	PRECISION	RECALL	F1
Classification results using n-grams, n>2			
Subj.	0.977	0.619	0.758
Obj.	0.442	0.954	0.604
Posit.	0.881	0.769	0.821
Negat.	0.923	0.962	0.942

Classification results using n-grams, n>2			
Subj.	0.933	0.601	0.731
Obj.	0.432	0.743	0.546
Posit.	0.834	0.642	0.726
Negat.	0.902	0.910	0.906

As we can notice from the results of using all n-grams, if we employ the annotated elements, it is easier to distinguish the subjective sentences. We believe this is due to the fact that we train on subjective n-grams. As far as the positive and negative classification is concerned, the results are both high, as well as balanced, proving the correctness of our approach.

As we can see, removing single words with their associated polarities from the training data resulted in lower scores. Therefore, the fine-grained annotation the EmotiBlog-Annotation-Model provides helps at the time of training the Opinion Mining system and is well worth the effort.

7.2 Experiments in english

Due to the fact that EmotiBlog-Annotation-Model contains annotations for individual words, as well as for multi-word expressions and at a sentence level, and the fact that they are labelled with polarity and emotion, our next experiments aim to show how the annotated elements can be used as training for Opinion Mining and Polarity Classification tasks, as well as for emotion detection in English.

Moreover, since the EmotiBlog-Annotation-Model also allows the annotation of the polarity intensity, this allows us to carry out an experiment focused on automatically determining the sentiment intensity as it is measured on the EmotiBlog-Annotation-Model: on a three-level scale -*high, low and medium*-.

By performing such experimentation, our main aim is to check if the EmotiBlog-Corpus-Annotated in English can be successfully exploited to automatically extract subjectivity expression and its intensity.

The first step of our test consists in creating the training models as described in the previous section.

After having created the training model we evaluated them by performing different tests (Balahur and Montoyo 2010).

The evaluation of the polarity and intensity classification tasks was carried out by employing the EmotiBlog I and II constructed models on two test sets – the JRC quotes[83] and the SemEval 2007 Task Number 14 test set (Strapparava and Mihalcea 2007).

Since the quotes often contain more than one sentence, we considered the polarity and intensity of the global quote as the most frequent result in each class, corresponding to its constituent sentences.

Table XXIV.

Results for polarity and intensity classification using the models built from the EmotiBlog annotations

Test Corpus	Eval. type	Precision	Recall	F1
JRC quotes I	Polarity	32.131	54.09	40.314
	Intensity	36.002	53.2	42.943
JRC quotes II	Polarity	36.421	51.001	42.945
	Intensity	38.731	57.812	46.386
SemEval I	Polarity	38.572	51.323	44.043
	Intensity	37.394	50.941	43.129
SemEval II	Polarity	35.833	58.682	44.496
	Intensity	32.342	50.413	39.404

83 http://langtech.jrc.ec.europa.eu/JRC_Resources.html

Table XXV.

SEMEVAL System results for emotion annotations

Emotions & Systems	Fine	Coarse		
	R	Prec	Rec	F1
Anger				
Swat	24.51	12.00	5.00	7.06
UA	23.20	12.74	21.6	16.03
Upar7	32.33	16.67	1.66	3.02
Disgust				
Swat	18.55	0.00	0.00	-
UA	16.21	0.00	0.00	-
Upar7	12.85	0.00	0.00	
Fear	Fear	Fear	Fear	
Swat	32.52	25.00	14.40	18.27
UA	23.15	16.23	26.27	20.06
Upar7	44.92	33.33	2.54	4.72
Joy				
Swat	26.11	35.41	9.44	14.91
UA	2.35	40.00	2.22	4.21
Upar7	22.49	54.54	6.66	11.87
Sadness				
Swat	38.98	32.50	11.92	17.44
UA	12.28	25.00	0.91	1.76
Upar7	40.98	48.97	22.02	30.38
Surprise				
Swat	11.82	11.86	10.93	11.78
UA	7.75	13.70	16.56	15.00
Upar7	16.71	12.12	1.25	2.27

As we can see from above, the results shown in TABLE XXV. show a significantly high improvement over the ones obtained in the SemEval task in 2007, presented in TABLE XXI. where the fine-grained evaluations were conducted using the Pearson measure of correlation between the system scores and the gold standard scores, averaged over all the headlines in the data set.

In the coarse-grained evaluation each emotion was mapped to a 0/1 classification:

- (0 = [0,50)
- 1 = [50,100]
- each valence was mapped to a -1/0/1 classification (-1 = [-100,-50], 0 = (50,50), 1 = [50,100])

The table shows both the fine-grained Pearson correlation measure and the coarse-grained accuracy, precision and recall figures and *R* represents the margin within which the result is correct.

After having analysed the tables of our experiments and compared it with the performance of the systems tested in the framework of the SemEval competition we can deduce that systems performing the opinion task did not have at their disposal the nor annotated EmotiBlog-Kyoto annotated nor its enrichment.

Another explanation of their lower performance is the fact that but also because they did not use Machine Learning on a corpus comparable to EmotiBlog-Kyoto (as seen from the results obtained when using solely the EmotiBlog I corpus).

We would like to underline the fact that we were able to evaluate the polarity intensity with the SemEval corpus because it is he only one with this data. Thus from this consideration we can also draw the conclusion that the EmotiBlog-Corpus-Annotated improves the state of the art, because apart for being finer-grained.

In the second experiment, we tested the performance of emotion classification using the two models built using EmotiBlog-Annotation-Model on the three corpora – JRC quotes, SemEval 2007 Task No.14 test set and also the ISEAR corpus.

The JRC quotes are labelled using EmotiBlog-Annotation-Model. However, the remaining two are annotated with a small set of emotions – 6 in the case of the SemEval data (joy, surprise, anger, fear, sadness, disgust) and 7 in ISEAR (joy, sadness, anger, fear, guilt, shame, disgust). Moreover, the SemEval data contains more than one emotion per title in the Gold Standard, therefore we consider as correct any of the classifications containing one of them. R refers to the margin of correctness of the obtained results. In fact, if you fix an extremely fine value, the probability of correctness would be too small.

R represents the margin within which the result is considered as correct. It has been proposed in the framework of the SEMEVAL competition, thus we decided to maintain it because in this way we obtain results that are comparable with already existing and consolidates competitions and thus we are able to understand the effective performance of our system and check if we improve the actual state of the art.

In order to unify the results and obtain comparable evaluations, we assessed the performance of the system using the alternative dimensional structures defined in Figure II that consists in establishing some categories of subjectivity with inside different shadows of such subjectivity and thus even if we label with two subjective states but in the same category this will have less negative impact. Those not overlapping with the category of any of the 8 different emotions in SemEval and ISEAR are considered as “Other” and are not included either in the training, nor test set.

As we can see, the “Emotions” category contains the following subjective status: joy, sadness, anger, fear, guilt, shame, disgust, and surprise.

Table XXVI.

Results for emotion classification using the models built from the EmotiBlog annotations

Test corpus	Eval. type	Precision	Recall	F1
JRC quotes EmotiBlog Model I	Emotions	24.723	15.082	18.735
JRC quotes EmotiBlog Model II	Emotions	33.651	18.981	24.272
SemEval EmotiBlog Model I	Emotions	29.032	18.893	22.890
SemEval EmotiBlog Model II	Emotions	32.984	18.453	22.665
ISEAR EmotiBlog Model I	Emotions	22.312	15.012	17.948
ISEAR EmotiBlog Model II	Emotions	25.624	17.831	21.029

Analysing our internal results for the emotion categories we can deduce that the best values we obtained for emotion detection were for the *anger* category, where the precision was around 35 percent, for a recall of 19 percent.

The worst results obtained were for the ISEAR category of "shame", where precision was around 12 percent, with a recall of 15 percent. We believe this is due to the fact that the latter emotion is a combination of more complex affective states and it can be easily misclassified to other categories of emotion.

Moreover, from the error analysis performed, we realized that many of the affective phenomena presented were more explicit in the case of texts expressing strong emotions such as "joy" and "anger", and were mostly related to commonsense interpretation of the facts presented in the weaker ones.

We also observed that the texts pertaining to the news category obtain better results, most of all news headlines. This is due to the fact that such texts, although they contain few words, have a more direct and stronger emotional charge than direct speech (which may be biased by the need to be diplomatic, find the most suitable words etc.).

Finally, the error analysis showed that emotion that is directly reported by the persons experiencing it is more "hidden", in the use of words carrying special meaning or related to general human experience.

However, the results in all corpora are comparable, showing that the approach is robust enough to handle different text types.

All in all, the results obtained using the fine and coarse-grained annotations in EmotiBlog-Annotation-Model increased the performance of emotion detection as compared to the systems in the SemEval competition, thus improving the state of the art.

7.3 Question answering experiments

After having analysed the question we wanted to take into account, we detected the candidate snippet retrieval and perform the topic-polarity classification of snippets.

At this stage, we evaluate our approaches on both the EmotiBlog question collection, as well as the TAC 2008 Opinion Pilot test set. We compare them against the performance of the (A. Balahur, E. Boldrini, et al. 2009) system but also with the best TAC 2008 system (Copeck, et al. 2007) and (Varma, et al. 2008) the worst one (second approximation in (Balahur, Lloret, et al. 2008) scoring systems (as far as F-measure is concerned).

For both the TAC 2008 and EmotiBlog sets of questions, we employ the Semantic Roles system in SA to determine the Expected Source, Expected Topic and Expected Polarity Type.

Subsequently, for each of the two corpora, we retrieve 1-phrase and 3-phrase snippets. The retrieval of the of the EmotiBlog-Corpus candidate snippets is done using query expansion with Latent Semantic Analysis and filtering according to the Expected Topic. Further on, we apply Sentiment Analysis using the approach described in Section 6 and select only the snippets whose polarity is the same as the determined question Expected Polarity Type.

TABLE XXVI. presents the results obtained for English and for Spanish. We indicate the id of the question (Q), the question type (T) and the number of answer of the Gold Standard (A). Moreover, we present the number of the retrieved questions by the traditional system (TQA) and by the opinion one (OQA).

We take into account the first answer, the first 5 and 10 first answers, until 50.

The results are presented in TABLE XXVII. TABLE XXVIII.

Table XXVII.

Results for English

Q	T	A	Number of found answers							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
1	F	5	0	0	0	2	0	3	4	4
2	O	5	0	0	0	1	0	1	0	3
3	F	2	1	1	2	1	2	1	2	1
4	F	10	1	1	2	1	6	2	10	4
5	O	11	0	0	0	0	0	0	0	0
6	O	2	0	0	0	0	0	1	0	2

Q	T	A	Number of found answers							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
7	O	5	0	0	0	0	0	1	0	3
8	F	5	1	0	3	1	3	1	5	1
9	F	5	0	1	0	2	0	2	1	3
10	F	2	1	0	1	0	1	1	2	1
11	O	2	0	1	0	1	0	1	0	1
12	O	3	0	0	0	1	0	1	0	1
13	F	1	0	0	0	0	0	0	0	1
14	F	7	1	0	1	1	1	2	1	2
15	F/O	1	0	0	0	0	0	1	0	1
16	F/O	6	0	1	0	4	0	4	0	4
17	F	10	0	1	0	1	4	1	0	2
18	F/O	1	0	0	0	0	0	0	0	0
19	F/O	27	0	1	0	5	0	6	0	18
20	F/O	4	0	0	0	0	0	0	0	0

Table XXVIII.

Results for Spanish

Q	T	A	Number of found answers							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
1	F	9	1	0	0	1	1	1	1	3
2	F	13	0	1	2	3	0	6	11	7
3	F	2	0	1	0	2	0	2	2	2
4	F	1	0	0	0	0	0	0	1	0
5	F	3	0	0	0	0	0	0	1	0

Q	T	A	Number of found answers							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
6	F	2	0	0	0	1	0	1	2	1
7	F	4	0	0	0	0	1	0	4	0
8	F	1	0	0	0	0	0	0	1	0
9	O	5	0	1	0	2	0	2	0	4
10	O	2	0	0	0	0	0	0	0	0
11	O	5	0	0	0	1	0	2	0	3
12	O	2	0	0	0	1	0	1	0	1
13	O	8	0	1	0	2	0	2	0	4
14	O	25	0	1	0	2	0	4	0	8
15	O	36	0	1	0	2	0	6	0	15
16	O	23	0	0	0	0	0	0	0	0
17	O	50	0	1	0	5	0	6	0	10
18	O	10	0	1	0	1	0	2	0	2
19	O	4	0	1	0	1	0	1	0	1
20	O	4	0	1	0	1	0	1	0	1

The retrieval of the TAC 2008 1-phrase and 3-phrase candidate snippets was done using JIRS. After having filtered with JIRS the possible candidates we calculate the candidate snippets polarity with different methods and we use the similarity with annotated sentences with EmotiBlog-Corpus -Annotated.

Subsequently, we performed different evaluations, in order to assess the impact of using different resources and tools. Since the TAC 2008 had a limit of the output of 7000 characters, in order to compute a comparable F-measure, at the end of each processing chain, we only considered the snippets for the 1-phrase retrieval and for the 3-phases one until this limit was reached.

1 In the first evaluation, we select the snippets that have the same polarity as the question Expected Polarity Type and the Expected Topic is found in the snippet. (i.e. *What motivates peoples negative opinions on the Kyoto Pro-*

tocol? The Kyoto Protocol becomes deterrence to economic development and international cooperation/ Secondly, in terms of administrative aspect, the Kyoto Protocol is difficult to implement. - same EPT and ET)

2 We also detected cases of same polarity but no Expected Topic, e.g. *These attempts mean annual expenditures of $700 million in tax credits in order to endorse technologies, $3 billion in developing research and $200 million in settling technology into developing countries* –EPT negative but not same ET.

3 In the second evaluation, we add the result of the Latent Semantic Analysis process to filter out the snippets from 1., containing the words related to the topic starting from the retrieval performed by Yahoo, which extracts the first 20 documents about the topic.

4 In the third evaluation, we filter the results in 2 by applying the *SemRol* system and setting the condition that the Expected Topic and Expected Source are the agent or the patient of the snippet.

Table XXIX.

Results for the TAC 2008 question set

System	F-measure
Best TAC	0.534
Worst TAC	0.101
JIRS + SA+ET (1 phrase)	0.377
JIRS + SA+ET (3 phrases)	0.431
JIRS + SA+ET+LSA (1 phrase)	0.489
JIRS + SA+ET+LSA (3 phrases)	0.505
JIRS + SA+ET+LSA+SR (1 phrase)	0.533
JIRS + SA+ET+LSA+SR (3 phrases)	**0.571**

BL (baseline), RS (removing stopwords), RN (removing negation), SC (sayings and collocations as single features) and ST (stemming).

From the results obtained, we can draw the following conclusions.

The first can be the ambiguity of the questions e.g. *¿De dónde viene la riqueza de EEUU?*. The answer can be explicitly stated in one of the blog sentences, or a system might have to infer them from assumptions made by the bloggers and their comments. Therefore, the answer is highly contextual and depends on the texts analyzed and there is also the need for extra knowledge on the concepts involved.

The hypothesis that Opinionated Question Answering requires the retrieval of longer snippets was confirmed by the improved results, both in the case of EmotiBlog-Corpus, as well as the TAC 2008 collection and in this case we had the conformation that EmotiBlog-Annotated-Corpus is a useful resource that can improve the retrieval of answers. As we can see in the table above, the approach using JIRS and the similarity with the EmotiBlog-Annotated-Corpus obtain the best results.

Secondly, opinion questions require the joint topic-sentiment analysis; as we can see from the results, the use of topic-related words in the computing of the affect influences the results in a positive manner and joint topic-sentiment analysis is especially useful for the cases of questions asked on a monothematic corpus and these elements are included in the EmotiBlog-Annotation-Model, as well as the target and source, relevant at the time of answer filtering, not only helping in the more accurate retrieval of answers, but also at placing at the top of the retrieval the relevant results.

Nonetheless, as we can see from the relatively low improvement in the results, much remains to be done in order to appropriately tackle OQA. As seen in the results, there are still questions for which no answer is found (e.g. 18). This is due to the fact that its treatment requires the use of inference techniques that are presently unavailable (i.e. define terms such as “improvement”).

7.4 Automatic summarisation experiments

After having collected the corpus, we labelled it using some of the EmotiBlog-Annotation-Model.

Extract (39) shows an example of annotation:

<topic>economic situation**</topic>**

<topic2>government**</topic2>**

<topic3>banks**</topic3>**

<new> Saturday, May 9, 2009 My aim in this blog has largely been to give my best and most rational perspective on the reality of the economic situation. I have tried (and I hope) mostly succeeded in avoiding emotive and partisan viewpoints, and have tried as far as possible to see the actions of politicians as misguided. Of late, that perspective has been slipping, for the UK, the US and also for Europe.

<phenomenon gate:gateId="1" target="economic crisis" degree1="medium" category="phrase" source="Cynicus Economicus" polarity1="negative" >I think that the key turning point was the Darling budget, in which the forecasts were so optimistic as to be beyond any rational belief**</phenomenon>**…

After the annotation process, the next step consisted in generating summaries from blogs, and, more specifically, from the posts about news, we used, as a core for the summarization process, the summarization approach proposed in (Lloret, 2011). However, as this system produces generic summaries, the blog posts had to be pre-processed and classified according to their polarity before producing the final summaries. At this step we would like to stress the fact that the EmotiBlog-Corpus-Annotated has been exploited to classify better the sentences according to their polarity (see section 6).

Once all subjective sentences have been classified, we grouped them according to their polarity, distinguishing between positive and negative. It is worth mentioning that, although the polarity of all blog sentences was determined, we only took into consideration the ones belonging to the comment posts and not in the initial news post of the blogs. This was motivated by the fact that the purpose of our summaries is to contain opinions stated by the users who have already read that news and want to express their thoughts in relation to it.

Having all sentences without noisy information, the next step was to run the summarization approach that employs textual entailment to remove redundant information, and computes word-frequency and noun-phrases length to detect relevant sentences within a document. The output of the system is an extract, which means that the most important sentences are extracted to produce the final summary (Lloret and Palomar 2009).

Two different summaries were produced for each blog:

– one with the positive opinions
– one with the negative ones

Finally, as a post-processing stage, we bound together the summaries belonging to the same blog to produce the final summary.

In the end, we generated 51 opinion summaries from different topics one corresponding to each blog of the corpus described in the previous sections.

Summary Evaluation

This is a complex step since different humans would produce diverse summaries, resulting in several possible correct summaries as gold standard. In (Donaway, Drummey and Mather 2008) it was shown how the result for a summary changed depending on which human summary was taken as reference for comparison with the automatic one.

We decided to focus more on the quality of the summaries rather than on its content according to the criteria of: *redundancy* (presence of repeated information), *grammaticality* (spelling or grammatical mistakes*) focus* (whether it is possible or not to understand the topic of the summary) and *difficulty* (the extent to which a human can understand a summary as a whole or not)-the DUC criteria.

The evaluation has been manually carried out by two potential users.

The tables below show the results obtained.

Table XXX.

Results of the evaluation for 10% , 15% and 20% compression ratio

10% COMPRESSION RATIO			
	Non Accept.	**Understand**	**Accept**
Redun.	26%	45%	29%
Gramm.	4%	22%	74%
Focus	33%	43%	24%

15% COMPRESSION RATIO			
Redun.	0%	6%	94%
Gramm.	2%	27%	71%
Focus	26%	29%	45%
20% COMPRESSION RATIO			
Redun.	4%	10%	86%
Gramm.	0%	55%	45%
Focus	14%	47%	39%

As we can see in TABLE XXX. we decided to create summaries at three different compression ratios (10%, 15% and 20%), in order to analyze the impact of the size of a summary (Hovy and Lin 1999). The different summary sizes would allow us to draw conclusions about the length of the summary and the qualitative evaluation.

Analysing the results obtained, we can notice that as far as the grammaticality criterion is concerned, the results show a decrease of grammaticality errors as the size of the summary lowers. We can see that the number of acceptable summaries varies from 74% to 45%, for a compression ratio of 20% and 10%, respectively. This is obvious, because the longer the summary, the more chances are for it to have orthographic or grammatical errors.

Due to the informal language used in blogs, we thought a priori that summaries would contain many spelling mistakes. Contrary to this thought, generated summaries are quite well written, only 4% of them, at most, being non-acceptable.

Another important fact that can be inferred from the results is related to how the summaries deal with the topic.

Finally, regarding redundancy, results are not conclusive, since they experiment variations in size and degree of goodness, so we cannot establish any trend. What can be seen from the results is that the summaries of 20% size obtain the best results on average over the rest of the size.

This is due to the fact that this compression ratio achieves higher percentage (for the understand and accept degrees of goodness) in two (grammaticality and focus) out of the three criteria proposed. Only the 15 % compression ratio summaries obtained better results in the redundancy criterion.

On the other hand, as far as the difficulty criteria concerned, results are also encouraging. According to the evaluation performed, the longer the summaries, the easier they are to understand in general. Grouping the percentages of summaries, we obtained that 65%, 82% and 92% of the summaries of size 10%, 15% and 20%, respectively, have, either medium or low level of difficulty, which give us an idea of they could be understand as a whole without serious difficulties. Again, for this criterion, the 20% summaries achieve the best results; this has also been proven by previous researches, which demonstrated that this compression ratio is more suitable for an acceptable quality of summaries (Morris, Kasper and Adams 1992).

7.5 Conclusions

In this chapter we presented the experiments we carried out in the framework of Opinion Mining, Opinionated Question Answering and Automatic Summarisation aiming ate checking if our EmotiBlog-Annotation-Model and EmotiBlog-Annotated-Corpus could bring improvements to the performance of systems dealing with such tasks.

Concerning Opinion Mining, our purpose was to check if EmotiBlog-Annotation-Model and Annotated Corpus are useful resources for Opinion Mining systems even if it provides finer-grained information if compared with the one allowed by the previous annotation schemes described in the state of the art.

That is why we evaluated the Spanish -with a more complex syntactic structure than English- corpus through ten-fold cross validation and further on we described a method to mine opinion from user input in Spanish using n-gram and phrase level similarity with the annotated elements of the EmotiBlog-Corpus, obtaining high precision results. We explained the reason why we employed such methods, in order to potentiate the EmotiBlog-Annotation-Model structure and information it provides.

Last, but not least, we proved that using the fine-grained annotation we obtained better results that using only the coarse-grained ones and this demonstrated the fact that the EmotiBlog-Annotation-Model is useful as a fine-grained resource.

We also tested the performance of our EmotiBlog-Corpus-Annotated in English and after having presented and described our results, we can say that our resource has a beneficial effect on the Opinion Mining task and improves the results obtained for example in the SemEval competition.

Moreover, another conclusion we can draw from the experiments performed is the fact that the strategy we employed of adding additional resources that contemplate subjective words allowed us to improve the precision and recall.

From the obtained results there is evidence that our resource allows capturing the relevant linguistic phenomena for expressing subjectivity, also in textual genres other than blogs and it is appropriate for the training Machine Learning models for the task of Opinion Mining in languages with different syntactic structure.

In the framework of Opinionated Question Answering task, we presented and evaluated different methods and techniques with the aim of improving the task of Opinionated Question Answering with the help of the EmotiBlog-Corpus-Annotated.

From the evaluations performed using different Natural Language Processing resources and tools, we deduce that the EmotiBlog-Annotation-Model and Annotated-Corpus are crucial. We concluded that joint topic-sentiment analysis, as well as the target and source identification are crucial for the correct performance of this task and they are elements present in the EmotiBlog-Annotation-Model.

We have also demonstrated that the retrieval of longer answers than just one sentence is crucial and allows an improvement of the results. This is feasible also thanks to the EmotiBlog-Annotation-Model structure (document, sentence, element annotation) and the consequential EmotiBlog-Kyoto-Annotated. We employed it in order to show to the system how to retrieve the correct answers improving its performance.

We thus showed that opinion Question Answering requires the development and application of appropriate strategies and resources at different stages (recognition of subjective question, detection of subjective content of the question, source, and target and retrieving of the required data), elements that the EmotiBlog-Annotation-Model contemplates.

Finally, we also tested the EmotiBlog-Annotation-Model and EmotiBlog-Corpus-Annotated on the Opinion Automatic Summarisation. We collected a corpus of blogs together with the comments given on them. This is an English corpus about five topics: economy, science and technology, cooking, society, and sport.

After having collected the corpus, we labelled it using a partial version of EmotiBlog-Annotation-Model and we then automatically classified the polarity at sentence and also at a document level exploiting also the EmotiBlog-Corpus-Annotated and we propose the build up of summaries with different compression ratios (10%, 15% and 25%.).

The summaries contain positive and negative opinions, divided according to their corresponding polarity, and EmotiBlog-Corpus-Annotated played a crucial role for helping the system to discriminate between positive and negative.

We evaluated summaries taking into consideration different parameters: redundancy, grammaticality, focus and difficulty, obtaining encouraging results, thus proving that the EmotiBlog-Annotation-Model can be employed as a useful resource for the processing of the texts that will be summarised, allowing an efficient polarity classification.

The conclusion we can draw from the results obtained is that The EmotiBlog-Annotation-Model brings a substantial contribution to the task. Most of the work done in Automatic Summarisation has been done with factual data and systems have been working with factual data. Thus the EmotiBlog-Annotation-Model and its annotated corpus allow the system to have at disposal the data for the module of treating the opinionated data, allowing an improvement of the task.

8. Conclusions & future work

This work has been focused on the EmotiBlog-Annotation-Model and the EmotiBlog-Corpus-Annotated, the multilingual and multi domain resource we created to detect subjectivity in the new textual genres with the intention of contributing to the improvement of the Sentiment Analysis task.

Our main motivation was the huge amount of subjective data available on the Internet due to the wide employment of the new textual genres born with the Web 2.0. As it has been demonstrated, this data has an undeniable influence on people's behaviour and it can be exploited for many real-life applications. This is the reason why we concentrated on the Sentiment Analysis task, the discipline in charge of treating the opinionated data as first step to achieve robust applications, which are able to exploit this data.

Our focus was on Sentiment Analysis on a multilingual level (English, Spanish and Italian) and we concentrated mainly in blogs, due to their demonstrated relevance in our society.

The contributions we bring with the present work are numerous, but they can be summarised as follows:

- We presented and analysed the relevance and usefulness of the new textual genres born with the Web 2.0 and above all blogs.
- We described in detail blogs main features, peculiarities and bloggers' profile.
- We carried out an in-depth analysis of the state of the art in Sentiment Analysis.
- Basing on the conclusions drawn from the previous point we deduced that there is an evident scarcity of corpora in languages other than English and composed by blog posts, thus,

  - o We created a multilingual corpus of blog posts in English, Spanish and Italian about three topics: the Kyoto Protocol, the elections in Zimbabwe, the USA elections.
- – From the State of the Art analysis we also concluded that there is shortage of fine-grained annotation models to capture the expression of subjectivity
  - o After a deep analysis of the corpus we collected, we built up the EmotiBlog-Annotation-Model, a fine-grained annotation scheme to detect subjectivity in these texts. In order to be the most adequate as possible to the needs of our corpus, we made an extensive analysis of the subjectivity definition and classification and choose the most suitable according to the needs of our dataset.
- – We labelled part of our corpus.
- – We carried out and intrinsic evaluation of the EmotiBlog-Corpus-Annotated and thus consequently of the EmotiBlog-Annotation-Model:
  - o Calculating the inter-annotator agreement to check if the model was clear enough and if it allows for an easy and unambiguous annotation by different annotators.
  - o Carrying out feature selection experiments to check if the annotation allowed a proper classification according to the EmotiBlog-Annotation-Model elements and attributes. We also measured the impact of the EmotiBlog-Annotation-Model in order to measure the effect of each element and refine the model.
- – We also performed an extrinsic evaluation in order to check if the EmotiBlog-Corpus-Annotated was a useful and beneficial resource to improve the performance of the key Natural Language Processing tasks dealing with subjective data at the moment. At this stage, apart from using our resource, we also employed different other corpora and some of them of different textual genres to check also if the EmotiBlog-Annotation-Model would be suitable also for other textual genres. The tasks in which we worked are listed below together with the EmotiBlog-Annotation-Model contribution:
  - o Opinion Mining: we exploited the EmotiBlog-Corpus-Annotated in order to train the Machine Learning system to correctly classify the sentences

of our documents depending on their subjectivity (subjective/objective), polarity (positive/negative) and intensity (low/medium/high) improving the baseline.

- o Opinion Question Answering: we employed our resource to help the system in learning how to classify the possible answers into positive and negative. The EmotiBlog-Annotation-Schema contemplates the necessary elements such as source, target, topic that are key aspects to improve the Opinionated Question Answering task, thus mixed with other resources can allow a considerable improvement of the baseline.
- o Opinion Summarisation: the EmotiBlog-Annotation-Model was used to label our ad-hoc collection of blog posts about different topics and also to train the classification system to distinguish the sentence polarities, needed for the summary generation. The results obtained show that the discrimination has been done in a proper way and thus the approach is correct and the EmotiBlog-Annotation-Model together with the information it provides is a valid resource that allows the treatment and thus inclusion of the subjective information, thus providing a step forward the state of the art, which has been treating objective data.

Table XXXI.
Better results obtained

OPINION MINING IN SPANISH			
Classification using ten fold cross validation			
	Precision	**Recall**	**F1**
Neg.	0.892	0.969	0.929
Classification results using all n-grams and n-grams, n>2			
	Precision	Recall	F1
Negat.	0.923	0.962	0.942
Negat.	0.902	0.910	0.906

OPINION MINING IN ENGLISH			
Results for polarity and intensity classification using the models built from the EmotiBlog annotations			
Corpus	**Precision**	**Recall**	**F1**
JRC quotes II INTENSITY	36.421	51.001	42.945
JRC quotes II POLARITY	38.731	57.812	46.386
Results for emotion classification using the models built from the EmotiBlog annotations			
Corpus	**Precision**	**Recall**	**F1**
SemEval EmotiBlog Model I	29.032	18.893	22.890
QUESTION ANSWERING IN ENGLISH			
System		**F-Measure**	
JIRS + SA+ET+LSA+SR (3 phrases)		0.571	
SUMMARISATION 10% COMPRESSION RATIO			
	Non Accept.	**Understand**	**Accept**
Redun.	26%	45%	29%
Gramm.	4%	22%	74%
Focus	33%	43%	24%

More concretely, in chapter 1 we made a deep study of our research framework. We presented the context of our research with the development of the Web 2.0 and the consequential growth of the new textual genres with all the implications they bring in our society. This represents a massive phenomenon and thus, due to the huge amount of subjective data available on the Web and the possible real time applications we can build up exploiting it, many are the disciplines that study how subjectivity is expressed. We described the main ideas of Neuroscience, Cognitive, Science, Psychology, but also Natural Language Processing, that is our research perspective. After that, we stressed on the importance of working with the new textual genres, we analysed the main ones and we justified why we decided to carry out our research mainly in blogs.

Another consequence of the huge explosion of this new research area, many are the terms used interchangeably and thus in order to clarify our use of such terminology we defined the difference between Sentiment Analysis and Opinion Mining. According to our point of view the first one is the step that comes before and that allows a high performance Opinion Mining process. In fact, with Sentiment Analysis the language is properly analysed and treated before being exploited for concrete purposes. Apart from that, we also clarified what we means when we employ the term *subjectivity* since this concept is strictly related to the EmotiBlog-Annotation-Model.

Furthermore, we stressed the importance of Sentiment Analysis studies, subject also extremely interesting for the EU that opens each year many different programmes focused on research on the information and Communication Technologies and included in this topic in the last few years, there is a strong stress on the importance of the creation of real-life application exploiting subjective data.

Finally we presented the milestones of our work and thus our research purposes.

In Chapter 2 we carried out we analysed in detail the State of the Art with a special focus on the creation of linguistic resources for Sentiment Analysis. We classified them according to different criteria: size, language, textual genre and domain creating comparative tables in order to show the aspects we would like to improve with our work. The following step consisted in briefly presenting the main research carried out in the framework of Opinion Mining, Question Answering and Opinion Summarisation.

Chapter 3 represents the main nucleon of this work and our main contribution to the improvement of the State of the Art presented previously. Here we described how we build up the EmotiBlog-Annotation-Model. We explained in detail the annotation levels EmotiBlog-Annotation-Model allows and we provided an in-depth description of the entire collection of the elements with their corresponding attributes providing examples in each case, thus entering in detail in the annotation process. Special emphasis was put on explaining why we chose such elements with their attributes and why we chose the values they have.

In chapter 4 we carried out part of the intrinsic evaluation on the EmotiBlog-Annotation-Model by calculating the inter annotator agreement with the objective of checking if the annotation it allows was clear, unambiguous ad easy to perform. The test has been carried out with two experienced annotators and on the EmotiBlog-Kyoto labelled in Spanish and we obtained positive results.

The second part of the intrinsic evaluation of the EmotiBlog-Annotation-Model was carried out in Chapter 5 in which we describe the results obtained in the feature classification experiments where we also we measure the impact of each element of the annotation model in order to understand which of them were beneficial for the classification purposes. After having checked the results, we refined the model producing its final version and the results obtained proved that our granularity approach is correct.

Before performing the extrinsic evaluation of the EmotiBlog-Annotation-Model and EmotiBlog-Corpus-Annotated, in chapter 6 we described all the resources, tools and procedures that we will employ in the following chapters using the EmotiBlog –Corpus-Annotated for the improvement of the three Natural Language Processing Tasks we selected.

In chapter 7 we start the extrinsic evaluation of the EmotiBlog-Annotation-Model and the EmotiBlog-Corpus-Annotated with the Opinion Mining Task. We exploited the EmotiBlog-Annotation-Model in order to train the Machine Learning system to train our system to correctly classify the sentences of our document depending on their subjectivity (subjective/objective), polarity (positive/negative) and intensity (low/medium/high) improving the baseline demonstrating that our resource is valuable for the classification process.

Moreover we entered in the task of Opinionated Question Answering using our resource to help the system to learn how to classify the possible answers into positive and negative. The EmotiBlog-Annotation-Schema has the necessary elements such as source, target, topic that are useful to improve the Opinionated Question Answering task, thus mixed with other resources can allow a considerable improvement of the baseline.

Finally, we carried out the extrinsic evaluation of our resource in the framework of the Opinion Summarisation task. In this case EmotiBlog-Annotation-Model was employed to label our ad-hoc collection of blog posts about different topics and also to train the classification system to distinguish the sentence polarities, needed for the summary generation. The results obtained show that the discrimination has been done in a proper way and thus the approach is correct and EmotiBlog-Annotation-Model and the information it provides is a valid resource that allows high-level results.

As general conclusion, we can say that there is no doubt about the fact that Sentiment Analysis is an extremely challenging task but at the same moment it is a fascinating area of research. Even if much work has been done, there is still big room for improvement.

The intrinsic evaluation showed that the fine-granularity we chose is appropriate and the annotation seems to be feasible without any significant problem. Moreover, the EmotiBlog-Annotation-Model and Corpus-Annotated have also demonstrated to be effective for English, Spanish and Italian, but we also expect that it could work with high results in other languages that share a similar syntactic structure.

Concerning the textual genre, even if our collection is taken from blog posts, we also had the possibility during our experiments to work with other textual genres as newspaper articles and we proved that the EmotiBlog-Annotation-Model is compatible with other textual genres.

From the extrinsic evaluation we performed, we can deduce that the EmotiBlog-Annotation-Model is beneficial in the three Natural Language tasks we selected. It is a key factor for the Opinion Mining task since it allows a better and more precise classification of sentence, element polarity and intensity.

The EmotiBlog-Annotation-Model employment is also strategic in the Question Answering task because it brings improvement of the baseline since it helps to correctly classify the answers and the possible snippets to retrieve the answers, but it also contemplates the key elements to improve the task (target, source, etc).

We also concluded that our resource has a positive impact on the Opinionated Automatic Summarisation since it provides a module for analysing the subjectivity in sentences making the normal opinion summarisation task adapt for subjective content and thus, it means that EmotiBlog is valid and a key point also for this task.

Much is the future work that has to be done. Our idea on that is to continue to focus on the linguistic perspective of this challenging research area. Concretely:

- Concerning the dataset:
  - We will complete the annotation of the EmotiBlog-Corpus in order to have more data to train and test our system.
  - We will take into consideration other new textual genres such as online reviews and forums.
- Regarding the languages:
  - We will like to introduce more language with different syntactic structure to check if EmotiBlog-Annotation-Model is also valid and also to enlarge our dataset.
- For the intrinsic evaluation:
  - We will carry out more in depth feature selection experiment testing different algorithms and find the best solution able to exploit the level of granularity the EmotiBlog-Annotation-Model offers.
- For the extrinsic evaluation:
  - We will enter in more detail in the tasks already explored to check if more data will improve the performance of systems using the EmotiBlog-Corpus-Annotated.
  - We are currently studying and employing the EmotiBlog-Annotation-Model to the recognition of the emotion cause events. More concretely in (Russo et al., 2011) we proposed a method to automatically identify linguistic contexts, which contain possible causes of emotions or emotional states from Italian newspaper articles (La Repubblica Corpus). Our methodology is based on the interplay between relevant linguistic patterns and an incremental repository of common sense knowledge on emotional states and emotion eliciting situations. Until now our approach has been evaluated with respect to manually annotated data (using for the subjectivity part the EmotiBlog-Annotation-Model). The results obtained so far are satisfying and support the validity of the methodology proposed.

9. Relevant pubblications

Annotation model

Boldrini, E., Balahur, A., Martínez-Barco, P., Montoyo, A. 2009. EmotiBlog: an Annotation Scheme for Emotion Detection and Analysis in Non-traditional Textual Genre. The 2009 World Congress in Computer Science, Computer Engineering, and Applied Computing.

Boldrini, E., Balahur, A., Martínez-Barco, P., Montoyo, A. 2009. *EmotiBlog: a finer-grained and more precise learning of subjectivity expression models.* In Proceedings of LAW IV, ACL.

Boldrini, E., Balahur, A., Martínez-Barco, P., Montoyo, A. 2009*. EmotiBlog: a fine-grained annotation schema for labelling subjectivity in the new-textual genres born with the Web 2.0.* In Proceedings of SEPLN.

Opinion mining

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco 2009. P. Cross-topic Opinion Mining for Real-time Human-Computer Interaction. ICEIS 2009.

Fernández, J., Boldrini, E., Gómez J. M., Martínez-Barco, P. 2011. Análisis de Sentimientos y Minería de Opiniones: el corpus EmotiBlog. In Proceedings of SEPLN 2011.

Kabadjov, M, A. Balahur, and E. Boldrini. "Sentiment Intensity: Is It a Good Summary Indicator?" In Proceedings of the 4th Language Technology Conference LTC, pp. 380-384, 2009.

Fernández, J., Boldrini, E., Gómez J. M., Martínez-Barco, P. 2011. Evaluating the Robustness of EmotiBlog for Sentiment Analysis and Opinion Mining. In Proceedings of RANLP 2011.

Question answering

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2009. Opinion and Generic Question Answering systems: a performance analysis. In Proceedings of ACL, 2009, Singapore.

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2009. Opinion Question Answering: Towards a Unified Approach. To appear in proceedings of the ECAI conference.

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2009. Fact versus Opinion Questions Classification and Answering: Challenges and Keys. The 2009 International Conference on Artificial Intelligence ICAI.

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2009. A Unified Proposal for Factoid and Opinionated Question Answering. In Proceedings of the COLING conference.

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2009. A Comparative Study of Open Domain and Opinion Question Answering Systems for Factual and Opinionated Queries. To appear in Proceedings of RANLP 2009.

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2009. Towards the Definition of Requirements for Mixed Fact and Opinion Question Answering Systems. In Proceedings of Topic Semantic Analysis. CIKM 2009.

Balahur, A., Boldrini, E., Montoyo, A., Martínez-Barco, P. 2010. The OpAL System at NTCIR 8 MOAT. In Proceedings of NTICR 8 MOAT.

Automatic summarisation

Balahur, A., Lloret, E., Boldrini, E., Montoyo, A., Palomar, M., Martínez-Barco, P. 2009. Summarizing Threads in Blogs Using Opinion Polarity. In Proceedings of Emerging Text Types Workshop. RANLP 2009.

Detection of emotion cause events

Russo, I., Caselli, T., Rubino, F., Boldrini, E., Martínez-Barco, P. 2011. EMOCause: An Easy-adaptable Approach to Extract Emotion Cause Contexts. In Proceedings of the WASSA 2011.

Resumen

Introducción

Actualmente, el uso masivo e intensivo de la Web 2.0 y las consecuencias que este fenómeno conlleva han dado lugar a gran incremento de la información subjetiva que cada día tiene más influencia en la vida y decisiones de los millones de usuarios de la red. Los investigadores están convencidos del hecho de que la Web está adquiriendo cada día más importancia gracias a las numerosas aplicaciones que se pueden producir a partir de la información que contiene.

El resultado de este proceso y una mayor implicación e los usuarios es la aparición de nuevos géneros textuales como blogs, foros o reseñas y por lo tanto un crecimiento exponencial de la información subjetiva que contienen, reflejando gustos, preferencias y opiniones de los millones de personas acerca de una amplia variedad de temas que afecta directamente a las decisiones y el comportamiento de los usuarios en muchos aspectos de sus vidas. Por tanto, la Web tiene un gran potencial para la creación de una serie de aplicaciones de varias tipologías y con grandes ventajas.

Una de las consecuencias de esta evolución en la manera de comunicar es el desarrollo de un tipo de investigación que intenta crear métodos efectivos para analizar, interpretar, tratar y explotar los datos subjetivos en continuo crecimiento desde un punto de vista interdisciplinario.

La subjetividad hasta ahora siempre ha sido estudiada por parte de disciplinas tradicionales como por ejemplo la neurociencia, la psicología o la ciencia cognitiva; pero en los últimos 10 años, gracias también a su crecimiento el procesamiento del lenguaje natural (PLN) es otra disciplina que lo analiza. Su finalidad es la de crear modelos de conocimiento para la representación de la subjetividad que pueda ayudar a procesar automáticamente la información subjetiva.

La principal diferencia respecto a las otras disciplinas es que el PLN está centrado en la creación de aplicaciones concretas para el aprovechamiento máximo de los datos subjetivos disponibles en la web y toma en consideración las conclusiones formuladas por las disciplinas que se han mencionado anteriormente. Sistemas que sean capaces de discriminar entre contenido objetivo y subjetivo pueden ser de gran ayuda para muchos actores como empresas, partidos políticos, o también para cualquier tipo de usuario.

Terminología

Centrándonos en el PLN, unos de los aspectos más desafiantes de esta subdisciplina es el hecho de que hasta ahora no se ha establecido una terminología uniforme y por lo tanto términos como emoción, sentimiento, sensación, punto de vista, etc. se emplean de manera intercambiable perdiendo así el matiz concreto que los diferencian. La existencia de tantos términos refleja los innegables matices distintos que los caracterizan. Por lo tanto, para poder delimitar nuestro ámbito de investigación es necesaria una aclaración de dicha terminología.

En nuestro trabajo tomaremos la definición empleada por Wiebe (1994), que a su vez se basó en la de Banfield (1985): *lo que no está abierto a la observación o comprobación subjetiva. Wiebe analiza los "private states" (estados privados) en términos de componentes funcionales: persona que experimenta alguna actitud hacia algo*. Según ella, existen tres tipos de "private states": menciones explícitas, actos de habla que los expresa, y elementos subjetivos. Entonces, opiniones, evaluaciones, emociones, especulaciones se agrupan en la categoría de "private states".

Nosotros emplearemos los términos de subjetividad y datos/información subjetiva incluyendo las definiciones de los términos de Wiebe (1994) y también las de Liu (2010) y Scherer (2004).

Los términos análisis de sentimientos y minería de opiniones también deben ser aclarados. Hasta ahora han sido usados en muchos casos de manera indistinta, pero según nuestra opinión, no se pueden considerar como sinónimos. De hecho, el primero indica una serie de técnicas para el tratamiento computacional del lenguaje subjetivo, mientras que el segundo se centra en la minería de la información subjetiva para distintas finalidades y no sería posible sin el proceso previo de análisis de sentimientos. Lo que es relevante también es que tanto el análisis de sentimientos como la minería de opiniones, son de gran interés para numerosas disciplinas fomentando así el trabajo interdisciplinario.

El gran interés en esta área de investigación queda patente por el nacimiento de numerosos grupos de investigación sobre este tema y también por la gran-cantidad de convocatorias públicas convocadas por la Comisión Europea que fomentan la investigación en este campo.

Géneros Textuales

En paralelo con la explosión de la Web 2.0 y el crecimiento de los nuevos géneros textuales, podemos decir que dos de los factores a tener en cuenta son el tipo de lenguaje producido y también la rapidez con la que la información crece diariamente.

Los blogs son uno de los nuevos medios de comunicación virtuales que más predominan en la web. Se pueden definir como plataformas en las cuales se publican cronológicamente pensamientos personales y su contenido consiste en una mezcla de lo que está pasando en la vida de una persona y en la web. Por lo tanto se pueden considerar como una especie de diario. En general, podemos decir que están escritos empleando estilos diferentes entre los cuales el informal es el predominante. El lenguaje que podemos apreciar es casual, informal y en general se podría categorizar como coloquial. Las consecuencias inmediatas del empleo de este tipo de lenguaje son abreviaciones, frases cortas, elipsis, uso de coloquialismos y frases hechas cuya elección depende del perfil del usuario.

La importancia de los blogs (comparados con foros y reseñas) está demostrada por la encuesta de Technorati[84] titulada *The state of the Blogosphere 2010* que los define como un género textual consolidado y que ya no se encuentra en un estado de iniciación (desde la encuesta del año anterior).

84 http://technorati.com/

Desafíos

En el contexto que acabamos de describir , el análisis de sentimientos tiene un papel fundamental dado que los sistemas tradicionales de PLN no están pensados para el tratamiento de la información subjetiva y las dificultades que ello conlleva. De aquí que la mayoría de los recursos desarrollados en el marco del PLN han sido para satisfacer las necesidades de los datos objetivos y su consecuencia es que no están preparados para el tratamiento de la información subjetiva ni los nuevos géneros textuales que aparecen asociados a la web 2.0.

Grandes cantidades de datos, creación en tiempo real, gran variedad de temas y fuentes, multilingualidad, estilos, o información multimodal, son los principales desafíos que presenta la información producida en el marco de la Web 2.0 y que dificultan los objetivos de esta área de investigación.

Como podemos deducir, la complejidad de los nuevos géneros textuales es tan alta y por lo tanto en nuestro trabajo nos centraremos principalmente en los blogs, puesto que representan una parte considerable de la información producida en la web 2.0. Una encuesta de Technorati del 2008 contabilizaba 112,8 Millones de blogs, sin tener en cuenta los 77,2 millones de blogs chinos existentes. Además las estadísticas sobre blogs en la mayoría de los casos se encuentran sobre los de lengua inglesa, así que a este cantidad deberíamos sumar también los producidos en otros idiomas.

Como podemos deducir, representan una de las fuentes más importantes de información producida en tiempo real, que puede ser explotada para el desarrollo de muchas aplicaciones de uso real atendiendo a las necesidades de múltiples actores.

De aquí deducimos que los datos subjetivos constituyen una fuente esencial de información y están llegando a ser el punto de referencia para muchas personas. Esta es la razón por la cual distintos investigadores en PLN han abarcado el problema de su tratamiento y explotación desde distintas perspectivas y a diferentes niveles dependiendo de una serie de factores. Estos pueden ser por ejemplo pregunta específica vs. general, tipo de fórmula interrogativa, tipo de texto, manera de expresar la subjetividad, etc. Por lo tanto, podemos ver que este tipo de investigación ha de abarcar todos estos aspectos para poder ser efectiva, de buena calidad y útil.

Información subjetiva

Los datos subjetivos constituyen una fuente esencial de información y están llegando a ser el punto de referencia para muchas personas cada día más. Esta es la razón por la cual distintos investigadores han abarcado el problema de su tratamiento y explotación desde distintas perspectivas y a diferentes niveles dependiendo de una serie de factores. Estos pueden ser por ejemplo pregunta específica vs. general. Tipo de fórmula interrogativa, tipo de texto, manera de expresar la subjetividad.

Por lo tanto podemos ver que la investigación sobre la información subjetiva de la web 2.0 ha de abarcar todos estos aspectos para poder ser efectiva, de buena calidad y útil.

Objetivos de la tesis doctoral

Después de los desafíos mencionadas en la sección anterior, resulta evidente el hecho de que un tratamiento, procesamiento, interpretación efectivos de la información subjetiva es imprescindible. Esto conlleva la creación de sistemas para el entrenamiento y el testeo de procesos en distintos géneros textuales y con varios niveles de granularidad.

Para poder contribuir a tal objetivo, este trabajo tiene como objetivos:

- Analizar y proponer un esquema de anotación a granularidad fina para poder capturar todos los matices del lenguaje y expresiones empleadas en los nuevos géneros textuales.
- Anotar, si es preciso, una colección de entradas de blogs usando el esquema de anotación propuesto.
- Evaluar intrínsecamente la estabilidad del modelo de anotación creando modelos de aprendizaje automático con los elementos anotados y medir el impacto de cada elemento.
- Evaluar extrínsecamente la eficiencia del modelo y del corpus anotado aplicándolo a varias tareas de PLN: minería de opiniones, búsqueda de respuestas y generación de resúmenes automáticos.

Estado de la cuestión

Dado el crecimiento exponencial de la cantidad de información subjetiva, en los últimos años ha habido un gran interés en la disciplina de análisis de sentimientos. Sin embargo, aunque se ha avanzado mucho en proponer aplicaciones que permitan tratar la información subjetiva, los datos subjetivos necesarios para estas aplicaciones todavía necesitan recursos y enfoques para que puedan ser identificados de manera efectiva, siendo todavía un reto en la actualidad.

Aunque se trata de un área de investigación relativamente reciente, hay muchos trabajos hechos. En las tablas que a continuación se muestran, queremos presentar una visión global de todos los enfoques y recursos existentes en la actualidad. Los recursos se caracterizarán por: nivel de granularidad (Tabla 1), proceso de creación (Tabla 2) y por último, lengua, dominio y tamaño (Tabla 3).

Tabla 1:

Recursos clasificados por su granularidad[85]

NOMBRE	ANOTACIÓN	REFERENCIA
MPQA	• Discurso objetivo • Discurso subjetivo – fuente – tópico – intensidad, actitud	(Wiebe and Wilson 2005)
Cornell movie-review	• Conjuntos de datos con polaridad – positiva – negativa) • Datos con escala de sentimientos – escala de valores • Datos subjetivos – subjectivo – objectivo	(Pang and Lee 2004) (Pang, Lee and Vaithyanathan 2002)
The NTCIR[85] multilingual corpus	• Opinión • Fuente de la opinión • Polaridad del sentimiento • Relevancia de la infor. (usando una colección de temas predefinidos)	http://research.nii.ac.jp/ntcir/index-en.html
SemEval 2007 Task 18 — Affective Text	• Emociones – (e.g. alegría, miedo, sorpresa, etc.) • Polaridad	(Strapparava and Mihalcea 2007)

85 http://research.nii.ac.jp/ntcir/index-en.html

NOMBRE	ANOTACIÓN	REFERENCIA
SentimentWortschatz	• Palabras que indican sent. Positivos y negativos positive – Medidos entre [-1; 1] – POS	(Remus, Quasthoff and Heyer 2010)
WordNet Affect	• Roles semánticos • valencia – positiva/negativa • Niveles de la emoción	(Strapparava and Valitutti 2004)
Sentiwordnet	• objetiva • positiva/negativa	(Esuli and Sebastiani 2006)
General Inquirer	• subjectiva – polaridad, – fuerza – ases de emoción	(Stone, et al. 1966)
Opinion Finder	• fuente • palabras de la frases que expresan emoción	(Wilson, Hoffmann, et al. 2005)
Micro-WNOP	• relevancia de la opinión • Representatividad en WORDNET	(Cerini, et al. 2007)
Emotion triggers	• polaridad • emoción	(Balahur and Montoyo 2008)
ISEAR corpus	• emoción	(Scherer and Wallbott, The ISEAR Questionnaire and Codebook 1997)
CINEMO	• Emociones principales y segundarias	(Brendel, Zaccarelli and Deuvillers 2010)
Gold Standard for Dutch	• opinión • actitud • factual	(Maks and Vossen 2010)
TREC test collection	• opinions relevantes • polaridad – positiva, negativa, ambas.	http://ir.dcs.gla.ac.uk/test collections/ access to data. html
Gold standard for Dutch subjectivity words	• positiva • negativa • neutral	(Jijkoun and Hofmann 2009)
	• subjetividad • objetividad • determinación polaridad	(Yu and Hatzivassiloglou 2003)
GermanPolarityClues	• positiva, negativa, neutral	(Waltinger 2010)
Q-WordNet	• Positiva, negativa	(Agerri and García-Serrano 2010)

NOMBRE	ANOTACIÓN	REFERENCIA
Congressional floor-debate transcripts	• Pro/contra la legislación	(Thomas, Pang and Lee 2006)
Comlex	• Adverbios de actitud	(Macleod, Grishman and Meyers 1994)
Customer review datasets	• Si se expresa una opinión – Características de un listado determinado	(Hu and Liu 2004)
Review-search results sets	• Subjetividad o no	(Pang, Lee and Vaithyanathan 2002)
Whissell's Dictionary of Affect in Language	Normas de subjetividad para el ingles y adjetivos de sentimiento	(Sweeney and Whissell 1984)
Computational semantic lexicon of French verbs	• Sensación • EmoEmocióntion • Estados psicológicos	(Mathieu 2006)
Economining	• Puntuación automatica de sentimientos	(Ghose, Ipeirotis and Sundararajan 2007)
Multiple-aspect restaurant reviews	– Escala de 1 a 5 para distintos aspectos	(Snyder and Barzilay. 2007)
Multi-Domain Sentiment Dataset	– Escala de 1 a 5 (estrellas)	(Blitzer, Dredze and Pereira 2007)

Tabla 2:

Creación de Recursos

NOMBRE	MANUAL	SEMI/AUTOM.	REFERENCIA
MPQA	X		(Wiebe, Wilson and Cardie 2005)
Cornell movie-review		X	(Pang and Lee 2004), (Pang, Lee and Vaithyanathan 2002)
SentimentWortschatz		X	(Remus, Quasthoff and Heyer 2010)
WordNet Affect		X	(Strapparava and Valitutti 2004)
Sentiwordnet		X	(Esuli and Sebastiani 2006)

NOMBRE	MANUAL	SEMI/AUTOM.	REFERENCIA
General Inquirer	X		(Stone, et al. 1966)
Opinion Finder	X		(Wilson, Hoffmann, et al. 2005)
Micro-WNOP	X		(Cerini, et al. 2007)
ISEAR corpus			(Scherer and Wallbott, The ISEAR Questionnaire and Codebook 1997)
CINEMO	X		(Brendel, Zaccarelli and Deuvillers 2010)
Gold Standard for Dutch	X		(Maks and Vossen 2010)
TREC test collection		X	http://ir.dcs.gla.ac.uk/test collections/ access to data.html
Gold standard for Dutch subjectivity words	X		(Jijkoun and Hofmann 2009)
		X	(Yu and Hatzivassiloglou 2003)
GermanPolarityClues	X		(Waltinger 2010)
Q-WordNet		X	(Agerri and García-Serrano 2010)
Congressional floor-debate transcripts	X		(Thomas, Pang and Lee 2006)
Comlex	X		(Macleod, Grishman and Meyers 1994)
Customer review datasets	X		(Hu and Liu 2004)
Review-search results sets	X		(Pang, Lee and Vaithyanathan 2002)
Whissell's Dictionary of Affect in Language	X		(Sweeney and Whissell 1984)
Computational semantic lexicon of French verbs	X		(Mathieu 2006)
Economining	X	X	http://economining.stern.nyu.edu/datasets.html
Multiple-aspect restaurant reviews		X	(Snyder and Barzilay 2007)

NOMBRE	MANUAL	SEMI/AUTOM.	REFERENCIA
Multi-Domain Sentiment Dataset			(Blitzer, Dredze and Pereira 2007)
Lexicon of appraisal terms	X		(Somasundaran, et al. 2006)

Tabla 3:
lengua, dominio y tamaño de los recursos

NAME	LANGUAGE	DOMAIN	SIZE	REFERENCE
MPQA	Inglés	General	10000 frases	(Wiebe, Wilson and Cardie 2005)
Cornell movie-review	Inglés	Películas	2000 reseñas 10662 frases 10000 frases	(Pang and Lee 2004), (Pang, Lee and Vaithyanathan 2002)
The NTCIR multilingual corpus	Inglés, Chino, japonés	20 temas	6000 frases	http://research.nii.ac.jp/ntcir/index-en.html
SemEval 2007 Task 18 —Affective Text	Inglés	General	1000 frases 1000 frases de prueba	(Strapparava and Mihalcea 2007)
ISEAR corpus	Inglés	Vida real	7000 frases	(Scherer and Wallbott, The ISEAR Questionnaire and Codebook 1997)
CINEMO	Francés	General (escénas de películas)	4k partes de conversaciones	(Brendel, Zaccarelli and Deuvillers 2010)
TREC test collection	Inglés	Temas diferentes	100649 blogs	http://ir.dcs.gla.ac.uk/test collections/ access to data.html
	Inglés	General	1336 palabras semilla	(Yu and Hatzivassiloglou 2003)

NAME	LANGUAGE	DOMAIN	SIZE	REFERENCE
Congressional floor-debate transcripts	Inglés	Política	38 debates (entren.), 10 (prueba), 5 (desarrollo)	(Thomas, Pang and Lee 2006)
Customer review datasets	Inglés	5 productos electronicos	n.d.	(Hu and Liu 2004)
Review-search results sets	Inglés	Reseñas	20 resultados (Yahoo!) de 69 preguntas son "reseña"	
Economining	Inglés	Precios de transacciones, opinion de comerciantes	n.d.	
Multiple-aspect restaurant reviews	Inglés	Restaurantes	4488 reseñas	(Snyder and Barzilay. 2007)
Multi-Multi MultiDomain	Inglés	Productos	n.d.	(Blitzer, Dredze and Pereira 2007)

Además, dichos recursos han sido aplicados a tareas de búsqueda de respuestas y generación de resúmenes automáticos. Las principales aportaciones las describimos en la siguiente tabla (Tabla 4).

Tabla 4:
Aplicaciones de los recursos a tareas de PLN

APPROACH	AUTHOR
OPINION QUESTION ANSWERING	
(Stoyanov, Cardie and Wiebe 2005) (Pustejovsky and Wiebe 2005)	Peculiaridades de preguntas de opinión
(Cardie, et al. 2004)	Resumens de opinions para apoyar el sistema de Multi-Perspective QA system, para identificar las respuestas a preguntas de opinion para un determinada collección de preguntas
(Yu and Hatzivassiloglou 2003)	Opiniones separadas de hechos y resumidas como respuesta a preguntas de opinion

APPROACH	AUTHOR
(Kim and Hovy 2006)	Fuentes de opinion identificadas, fundamentals para la recuperación de respuesta
TAC 2008	
The Alyssa system (Shen, et al. 2007)	Un clasificador SVM entrenado con el MPQA, NTCIR8 en Inglés y lexicones basados en reglas
(Varma, et al. 2008)	Análisis de la pregunta para detectar la polaridad de la pregunta usando reglas. Filtrado de opinions usando un clasificador basado en Naïve Bayes con características de unigramas, asignando a cada frase una puntuación (combinación linear entre opinion y polaridad)
PolyU system (Li, et al. 2008)	La orientación de sentimiento de la frase usando la medida de divergencia de Kullback-Leibler con los dos modelos de elenguage par alas categorías de pos y neg.
The QUANTA system (Fangtao, et al. 2008)	Detección de la fuente de opinion, objeto y polaridad. Usan un etiquetador semántico basado en PropBank y patrones definidos manualmente. Para la clasificación de la subjetividad extraen y clasifican las palabras de opinión. Para la recuperación de respuestas puntuan los extratos recuperados dependiendo de la presencia del tema y palabras de opinion y eligen larespuesta major puntuada.
NTCIR 7 MOAT	
	La mayoría de participantes emplea técnicas de aprendizaje automático usando patrones sintácticos aprendidos del corpus MPQA.
OPINION SUMMARISATION	
Granularidad fina, resúmenes basados en caracaterísticas de opinión	(Hu and Liu 2004)
	(Stoyanov and Cardie 2006)
	(Saggion and Funk, Interpreting SentiWordNet for opinion classification 2010)
	(Saggion, Lloret and Palomar 2010)
OPINION PILOT TRACK AT THE TEXT ANALYSIS CONFERENCE	
La mayoría de participantes añade nuevas caractarísticas (sentimiento, positivo/negativo, opinion positiva/negativa)para verificar la rpesencia de opinions positivas y/o negativas	CLASSY (Conroy and Schlesinger 2008); CCNU (He, et al. 2008); LIPN (Bossard, Généreux and Poibeau 2008); IIITSum08 (Varma, et al. 2008) Italic (Cruz, et al. 2008)

Como se puede deducir de todos los recursos existentes hay una la falta de recursos para lenguas distintas al inglés. Además, la mayoría de los trabajos realizados, se han centrado en tratar textos pertenecientes al dominio periodístico y se ha dado menos importancia a los nuevos géneros textuales. Por lo tanto, es necesario crear modelos de aprendizaje que sean capaces de entender el lenguaje humano y discriminar entre objetivo y subjetivo, así como también clasificar la información a distintos niveles.

A pesar de que se ha trabajado en este aspecto, nos encontramos con que la mayor parte de la investigación se ha centrado en un tipo de granularidad gruesa para el idioma inglés. De ahí que se identifique la necesidad de crear recursos y corpus en otros idiomas anotados con un nivel de granularidad más fina, con la finalidad de que permitan entender de una manera más profunda el uso del lenguaje subjetivo en los nuevos textos de la Web 2.0. En los próximos apartados nos centraremos en definir, proponer, analizar y evaluar nuestro esquema de anotación (EmotiBlog-Annotation-Model) y el recurso creado (EmotiBlog-Corpus).

Emotiblog

Teniendo en cuenta nuestras consideraciones previas, el recurso que presentamos, EmotiBlog-Annotation-Model, ha sido desarrollado para poder satisfacer las necesidades mencionadas y solventar las problemas existentes: falta de recursos en otras lenguas, anotación de granularidad fina y a múltiples niveles (documento, frase y elemento) y también corpus formados por extractos de nuevos géneros textuales.

Basándonos en el modelo de anotación que proponemos, hemos creado un corpus utilizando los nuevos géneros textuales: EmotiBlog-Corpus. Este corpus está compuesto por una colección de entradas de blog extraídas manualmente de la Web durante el 2009. Por lo tanto las características que distinguen nuestra colección de las otras creadas previamente son: multilingualidad, el uso de nuevos géneros textuales, multiplicidad de temas y anotación a granularidad fina.

Como podemos ver en la tabla 5, hemos considerado tres idiomas: inglés, español e italiano. El primero de ellos por su relevancia a nivel mundial, el segundo por ser la segunda lengua más hablada y el italiano dado que consideramos que está en un nivel de madurez suficiente en términos de recursos de PLN para poderse emplear. Además, la elección de estos tres idiomas no ha sido casual. Tener tres muestras distintas nos permite evitar coincidencias casuales, y por lo tanto ser más precisos y acertados en nuestra investigación. Todo ello tendrá como resultado la creación de un producto fiable.

TABLA 5:

Descripción de temas, tamaño e idiomas incluidos en EmotiBlog-Corpus

TOPIC	SIZE	LANGUAGE
El Protocolo de Kyoto	30,000 palabras	Inglés, Español, Italiano
Las elecciones en Zimbabwe	30,000 palabras	Inglés, Español, Italiano
Las últimas elecciones en EEUU	30,000 palabras	Inglés, Español, Italiano
TOTAL		270,000 palabras

Modelo

Después de haber analizado empíricamente las necesidades de nuestro corpus creamos la estructura del modelo de anotación EmotiBlog-Annotation-Model, que se presenta en las tablas a continuación:

TABLA 6:

Niveles de anotación contemplados por EmotiBlog-Annotation-Model

<table>
<tr><th>NIVEL</th><th>ELEMENTO</th></tr>
<tr><td>Discurso objetivo</td><td></td></tr>
<tr><td colspan="2"></td></tr>
<tr><td rowspan="9">Discurso subjetivo</td><td>Adjetivos</td></tr>
<tr><td>Advebios</td></tr>
<tr><td>Preposiciones</td></tr>
<tr><td>Verbos</td></tr>
<tr><td>Nombres</td></tr>
<tr><td>Mayúscolas</td></tr>
<tr><td>Onomatopeias</td></tr>
<tr><td>Puntuación</td></tr>
<tr><td>Frases hechas, colocaciones, dialecto, otros idiomas</td></tr>
<tr><td></td><td></td></tr>
<tr><td>Multi-post</td><td>Correferencia</td></tr>
</table>

TABLA 7:
Vista completa de EmotiBlog-Annotation-Model

NIVEL	ELEMENTO	ATRIBUTO											
		CONFIANZA	FUENTE	TEMA	POLARIDAD	NIVEL	DISC. SUBJ.	MODIFICADOR	TIPO	MODO	ANTECEDENTE	EMOCIÓN	COMENTARIO
Discurso objetivo	**Discurso objetivo**	**X**	**X**	**X**									
Discurso Subjetivo	Frase, título	X	X	X	X	X						X	X
	Frases hechas, colocaciones, dialecto, otros idiomas	X											X
	Adjetivos	X	X	X	X	X	X	X				X	X
	Adverbios	X	X	X	X	X	X	X				X	X
	Preposiciones	X	X	X	X	X	X	X				X	X
	Verbos	X	X	X	X	X	X	X		X		X	X
	Nombres	X	X	X	X	X	X					X	X
	Mayúscolas	X	X	X	X	X	X					X	X
	Puntuación	X	X	X	X	X	X					X	X
	Onomatopeias[85]	X	X	X	X	X	X					X	X
Multi-post	Correferencia	X	X						X		X		X

Evaluación intrínseca

Con este tipo de evaluación, nuestro propósito es comprobar que la anotación se puede aplicar sin problemas conceptuales que dificulten el proceso. Además vamos a determinar si la anotación producida es fiable y por lo tanto si se puede emplear por

86 This element has been removed alter the feature impact experiments for not being relevant

parte de sistemas de aprendizaje automático. El motivo de centrarnos en estos aspectos es verificar si el EmotiBlog-Annotation-Model es útil para los sistemas que necesitan discriminar entre datos objetivos y subjetivos y entrar en detalle en sus matices.

La anotación ha sido desarrollada en distintos pasos. El primer idioma que se anotó es el español, debido a que tiene una estructura sintáctica mucho más compleja comparado con el inglés. Por lo tanto queríamos estar seguros de que nuestro modelo contemplara tal complejidad. La anotación para los tres idiomas tratados no se ha realizado en paralelo, dado que nuestra idea era desarrollar distintas fases de anotación y pruebas, como se presenta a continuación:

- Anotamos la parte española de EmotiBlog-Corpus
- Calculamos el acuerdo entre anotadores
- Realizamos una selección de características general para el corpus anotado en español
- Realizamos una selección de características teniendo en cuenta la reducción de dimensionalidad para la parte del corpus anotada en español enfocada a la tarea de clasificación de la polaridad
- Anotamos la parte del corpus inglés
- Evaluamos la parte del inglés después de a reclasificación
- Mejoramos el EmotiBlog-Annotation-Model
- Anotamos la parte en italiano
- Llevamos a cabo la selección de características para la parte del corpus en italiano

Dos expertos anotan el corpus en español sobre el protocolo de Kyoto de forma independiente (30,000 palabras) y después medimos el acuerdo entre anotadores con la medida que se presenta abajo:

Cálculo de acuerdo entre anotadores

Los elementos de nuestro modelo que han sido evaluados se muestran en la siguiente tabla (Tabla 9).

Analizando los resultados obtenidos, podemos deducir que los elementos con mejor rendimiento son *mayúscula* y *puntuación* y los que obtienen resultados más bajos son *frase* e *inglés*. Deducimos que interpretar correctamente expresiones en otro idioma es extremadamente complejo y para ello se necesita un alto nivel de conocimiento del contexto que analizamos.

TABLA 9:
resultados del acuerdo entre anotadores

Anotación	**a**	**b**	**a||b**	**b||a**	**promedio**
Nombre	A	E	0,783	0,753	0,765
Adjetivo	A	E	0,782	0,613	0,681
Verbo	A	E	0,863	0,742	0,802
Adevrbio	A	E	0,831	0,764	0,794
Preposición	A	E	0,862	0,672	0,763
Puntuación	A	E	0,784	0,891	0,832
Mayúscolas	A	E	0,663	1	0,831
Otro idioma (English)	A	E	0,273	1	0,632
Otro idioma (Latin)	A	E	0,662	0,662	0,661
Frase	A	E	0,524	0,662	0,592
Objetivo	A	E	0,762	0,734	0,745
Promedio total					0,736

Un aspecto que cabe destacar es que el proceso de evaluación ha sido muy complejo debido a la fina granularidad del modelo. En varias ocasiones, cuando los anotadores detectaban la misma expresión, sus anotaciones podían diferenciarse por límites, elementos y sus atributos esto producirá un decremento de los resultados. Aparte de esto, no había garantía sobre el hecho de que los anotadores identificaran la misma expresión.

De todas formas, observando los resultados obtenidos, el porcentaje global de acuerdo es de 0,736. Este valor mejora con respecto al estado de la cuestión (0,71). De esto deducimos que el modelo está bien estructurado, no es ambiguo y por lo tanto la anotación es aplicable y puede llevarse a cabo sin mayores problemas ni dudas.

Para comprender mejor el impacto de cada elemento de nuestro modelo, presentamos a continuación los resultados de los experimentos de selección de características. Para ello, mediremos la importancia de cada elemento y verificaremos nuestra hipótesis de que una anotación considerando una granularidad fina es más conveniente para sistemas de aprendizaje automático. En otras palabras, lo que queremos hacer aquí es comprobar la validez y la granularidad fina de nuestro modelo de anotación para el propósito de clasificación. Empezaremos con el español debido a su complejidad sintáctica. En el caso del inglés, puesto que se compone de estructuras más sencillas, se obtiene un porcentaje de acierto más alto. Posteriormente, confirmamos los resultados obtenidos analizando el italiano, cuya estructura sintáctica es parecida al español.

Selección de características

Resultados para el Español

Teniendo en cuenta los elementos anotados en el corpus, los extraemos adoptando un modelo de bolsa de palabras. En algunas ocasiones eliminamos las palabras de parada , dado que no añaden información semántica al contenido del texto. De ese conjunto de palabras de parada, sí que tenemos en cuenta las partículas negativas para probar el efecto de la negación. Finalmente tenemos que tener en cuenta que hay palabras que adquieren un sentido u otro dependiendo del contexto. Estas formas incluyen coloquialismos y frases hechas y por lo tanto hay que realizar un tratamiento a nivel de todo el conjunto de la expresión y no para los términos individuales que la componen.

Continuando con la evaluación intrínseca que tiene como objetivo comprobar si nuestro modelo de anotación es consistente y si el listado de elementos con sus atributos tienen un efecto positivo para la tarea de clasificación automática, realizamos unos experimentos de selección de características iniciales en los tres idiomas para tener una idea de la validez del modelo en términos de utilidad a nivel multilingüe.

Para ello empleamos los algoritmos de aprendizaje automático más utilizados, en concreto: SVM y MNB. Estos experimentos nos permitirán ver la utilidad de EmotiBlog-Annotation-Model para el propósito de clasificación y medimos el impacto de cada elemento del modelo.

TABLA 10:
resultados en términos de F-measure y accuracy

Combinac.	ML caract.	MNB		SVM	
		Accuracy	F-measure	Accuracy	F-measure
BL	941	0,647	0,592	0,685	0,644
RS+RN	877	0,566	0,477	0,654	0,610
RS	878	0,532	0,420	0,625	0,572
SC	875	0,588	0,511	0,663	0,620
ST	**819**	**0,672**	**0,625**	**0,714**	**0,683**
RS+RN+ST	764	0,594	0,516	0,661	0,618
RS+ST	765	0,622	0,556	0,689	0,652
SC+ST	781	0,617	0,554	0,694	0,659

BL (baseline), RS (removing stopwords), RN (removing negation), SC (sayings and collocations as single features) and ST (stemming).

Resultados para el Inglés

Después de la reclasificación del corpus en términos de precisión cobertura y medida F, los resultados del corpus en lengua inglesa se presentan abajo (Tabla 11):

TABLA 11:
reclasificación en inglés

	Precision	Recall	F-Measure
Subj.	0,922	0,754	0,830
Obj.	0,756	0,721	0,738
Posit.	0,721	0,823	0,767
Neg.	0,924	0,924	0,924
Neut.	0,956	0,985	0,970

Resultados para el Italiano

Comparando el impacto obtenido por los idiomas español e inglés, refinamos el modelo y llevamos a cabo experimentos de selección de características en italiano. Empleamos un método de validación cruzada que divide el corpus en 10 partes (una para entrenamiento y las restantes para testeo) y las va combinando entre ellas. Con esto, comprobamos si podemos clasificar correctamente las frases en nuestro corpus italiano. Calculamos la medida de similitud de Lesk entre las frases y elementos anotados y representamos cada frase como un vector de características compuestas por la puntuación de similitud obtenida entre los elementos anotados:

TABLA 12:
resultados para el italiano

	Precision	Recall	F1
Subj.	0,731	0,563	0,636
Obj.	0,861	0,675	0,754
Posit.	0,712	0,732	0,722
Neg.	0,894	0,951	0,922

Después de realizar estos experimentos medimos el impacto de los elementos del modelo en inglés y español:

TABLA 13:
Impacto de los elementos en español

ELEMENTO	IMPACTO
verb	2,998%
phrase	2,664%
adjective	2,244%
noun	1,756%
preposition	0,338%
pronoun	–0,323%
onomatopoeic	–0,784%
adverb	–0,914%

TABLA 14:
Impacto de los elementos en inglés

ELEMENT	IMPACTO
phrase	2,951%
verb	0,560%
pronoun	0,337%
adjective	0,221%
noun	−0,177%
onomatopoeic	−0,278%
preposition	−0,283%
adverb	−0,525%

Teniendo en cuenta los resultados obtenidos refinamos el modelo. Aunque los adverbios obtienen resultados negativos los mantenemos porque creemos que son importantes para capturar la subjetividad en nuestros textos. Mantenemos también los elementos que nos dan resultados positivos y los que la suma entre los distintos idiomas tiene como resultado valores positivos. Sin embargo, eliminamos las onomatopeyas porque los resultados demuestran que no son útiles.

Evaluación extrínseca

La siguiente fase de nuestro trabajo está centrada en llevar a cabo una evaluación extrínseca de nuestro recurso con la finalidad de comprobar si es útil para su aplicación a otras tareas de PLN que necesitan tratar la información empleando un nivel de granularidad fino. Elegimos minería de opiniones, dado que por su naturaleza implícita requiere el tratamiento de información subjetiva. Además seleccionamos la búsqueda de respuestas puesto que hasta ahora los investigadores se han centrado en desarrollar la tarea empleando exclusivamente información objetiva y por lo tanto creando sistemas que nos son efectivos para el tratamiento de la información subjetiva. Por último, seleccionamos la tarea de generación de resúmenes automáticos ya que es necesario el desarrollo de métodos que contemplen y traten los datos subjetivos, de manera que el contenido del resumen no se vea alterado cuando trabajamos en un contexto en el que la información subjetiva es la predominante.

En general podemos decir que para desarrollar nuestros experimentos hemos empleado:

- El EmotiBlog-Corpus anotado con nuestro modelo (con la lista de elementos total o parcial)
- El EmotiBlog-Corpus anotado y enriquecido con diferentes recursos léxicos, creando nuevos corpus
- Otros corpus disponibles para poder comparar nuestros resultados

Además, empleamos diversos algoritmos y técnicas de tratamiento de datos para entrenar nuestro sistema dependiendo de los aspectos de EmotiBlog que queramos explotar más o los aspectos que queramos comprobar.

- El EmotiBlog-Corpus anotado con nuestro modelo (su total o parcial lista de elementos)
- El EmotiBlog-Corpus anotado enriquecido con diferentes recursos léxicos, creando nuevos corpus
- Otros corpus disponibles para poder comparar nuestros resultados

Además empleamos diversos algoritmos y técnicas de tratamiento de datos para entrenar nuestro sistema dependiendo de cuales aspectos de EmotiBlog queremos explotar más o los aspectos que queríamos comprobar.

Experimentos de minería de opiniones

en los experimentos de minería de opiniones nuestro objetivo es ver si tanto el modelo como el corpus anotado son recursos útiles para este tipo de sistemas, teniendo en cuenta el hecho de que ofrecen un tipo de información analizada a una granularidad fina en comparación con los recursos citados previamente.

Siguiendo el mismo esquema que para la evaluación intrínseca, utilizamos el método de validación cruzada que divide el corpus en 10 partes y describimos un método para extraer opiniones a partir de la entrada del usuario en español usando n-gramas y el nivel de similitud entre los elementos anotados del corpus EmotiBlog.

Todo esto se realiza para poder explotar los distintos niveles de anotación que ofrece EmotiBlog –Annotation-Model, así como las relaciones entre los distintos niveles relacionados. De esta manera, potenciamos la anotación de nuestro modelo y probamos el rendimiento del corpus en inglés, además de comprobar si mejoramos el estado del arte comparamos nuestros resultados con los de la competición del SemEval[87].

Para llevar a cabo nuestros experimentos también utilizamos otros recursos que contemplan palabras subjetivas para así poder mejorar los resultados. Además, evaluamos un método para clasificar la polaridad basado en n-gramas y similitud de características de las frases usadas con aprendizaje automático extraídas de la anotación del EmotiBlog-Corpus-Annotated.

Por otro lado, evaluamos con validación cruzada al igual que antes, las frases objetivas y subjetivas del corpus anotado. Como recurso alternativo empleamos un corpus de opiniones positivas y negativas sobre el tema del reciclado que previamente se creó *ad hoc* y que se anotó con nuestro modelo.

Resultados para el Español

Procesamos la anotación y representamos cada frase sobre la que hemos llegado a un acuerdo como un vector de características. Cada frase se clasifica y compara con las anotaciones en el corpus EmotiBlog y su puntuación de similitud.

Después probamos nuestro método clasificando las frases como subjetivas u objetvias (subj/obj). Dicha clasificación se ha hecho en las 150 frases sobre el tema del reciclaje. Los resultados se muestran en la tabla 15.

TABLA 15:
clasificación empleando ten fold cross validation

	Precision	**Recall**	**F1**
Subj.	0,988	0,632	0,771
Obj.	0,682	0,892	0,773
Posit.	0,799	0,511	0,623
Neg.	0,892	0,969	0,929

87 http://semeval2.fbk.eu/semeval2.php

Dicha clasificación se ha hecho en las 150 frases sobre el tema del reciclaje.

Para la clasificación usamos n-grama de distinta longitud (1 a 4), dónde dichos n-gramas serán las palabras. Para determinar la importancia de las palabras realizamos el siguiente experimento: las quitamos del vector y reclasificamos en términos de obj/subj como pos/neg:

TABLA 16:

clasificación usando todos los n-gramas y los >2

	Precision	Recall	F1
Resultados de clasificación usando n-gramas, n>2			
Subj.	0,977	0,619	0,758
Obj.	0,442	0,954	0,604
Posit.	0,881	0,769	0,821
Negat.	0,923	0,962	0,942
Resultados de clasificación usando n-gramas, n>2			
Subj.	0,933	0,601	0,731
Obj.	0,432	0,743	0,546
Posit.	0,834	0,642	0,726
Negat.	0,902	0,910	0,906

Como podemos observar en los resultados, cuando usamos todos los n-gramas nos resulta más sencillo distinguir las frases subjetivas. Con respecto a la clasificación positiva y negativa los resultados son altos y balanceados, y por lo tanto demostramos la validez de nuestro método. Cuando eliminamos de los datos de entrenamiento palabras individuales junto con sus polaridades los resultados obtenidos no son tan buenos. Esto demuestra que la granularidad fina es un elemento clave para una buena clasificación

Resultados para el Inglés

Con estos experimentos y resultados demostramos como la anotación a nivel de elemento puede ser usada para la clasificación de la polaridad y la detección de emociones asociadas a su polaridad, pero esta vez centrándonos en el inglés.

Para ello creamos los modelos de entrenamiento tal y como hicimos para el español y los evaluamos. La evaluación de la polaridad e intensidad se lleva a cabo empleando EmotiBlog I y II , pero además se emplean recursos externos, en concreto el JRC quotes y SemEval. Dado que las citas del recurso JRC quotes contienen más de una frase consideramos la polaridad e intensidad del global como el resultado más frecuente en cada clase y que corresponde a las frases que los componen.

TABLA 17:
Resultados para la clasificación de polaridad e intensidad usando el modelo creado de las anotaciones de EmotiBlog

Test Corpus	Eval. type	Precision	Recall	F1
JRC quotes I	Polarity	32,131	54,09	40,314
	Intensity	36,002	53,21	42,943
JRC quotes II	Polarity	36,421	51,001	42,945
	Intensity	38,731	57,812	46,386
SemEval I	Polarity	38,572	51,323	44,043
	Intensity	37,394	50,941	43,129
SemEval II	Polarity	35,833	58,682	44,496
	Intensity	32,342	50,413	39,404

Comparando nuestros resultados con los obtenidos en la competición de la edición del Semeval 2007, podemos apreciar una mejora. En los casos en los que hemos tenido que evaluar la granularidad fina, hemos empleado la medida de correlación de Pearson entre los resultados del sistema y el gold standard.

Para poder unificar los resultados y obtener evaluaciones comparables, medimos el rendimiento del sistema usando la estructura dimensional alternativa.

TABLA 18:
Resultados de la clasificación de emoción usando el modelo creado de las anotaciones de EmotiBlog

Corpus de Test	Tipo de eval.	Precision	Recall
JRC quotes EmotiBlog Model I	Emotions	24,723	15,082
JRC quotes EmotiBlog Model II	Emotions	33,651	18,981
SemEval EmotiBlog Model I	Emotions	29,032	18,893
SemEval EmotiBlog Model II	Emotions	32,984	18,453
ISEAR EmotiBlog Model I	Emotions	22,312	15,012
ISEAR EmotiBlog Model II	Emotions	25,624	17,831

Analizando nuestros resultados, podemos ver que los mejores han sido obtenidos para las emociones que generalmente se expresan de una manera directa como por ejemplo “enfado”, mientras que emociones que generalmente se expresan de una manera más matizada, como por ejemplo “vergüenza” son más complejos de detectar con claridad. Además podemos deducir que los mejores resultados se obtienen en la detección de la subjetividad en el cuerpo de las noticias y no en los titulares. Esto se debe a que generalmente, los titulares emplean un estilo más complejo por el reducido espacio disponible o la necesidad de captar la atención del lector en pocas palabras.

Búsqueda de Respuestas

Como ya introdujimos cuando presentamos este tipo de evaluación, hemos considerado también la tarea de búsqueda de respuestas. Hasta ahora, los sistemas de búsqueda de respuestas que se han desarrollado sólo soncapaces de tratar la información objetiva, pero debido al contexto subjetivo en el que trabajamos, la creación de sistemas para la búsqueda de respuestas que permitan tratar con datos subjetivos es imprescindible.

En nuestros experimentos usamos la colección EmotiBlog y el corpus del TAC 2008 Opinion Pilot Test y comparamos los resultados con los obtenidos en (Balahur, Boldrini et al. 2009) pero también con el mejor y peor sistema del TAC 2008. Para ambos casos, creamos una colección de preguntas y empleamos roles semánticos para detectar el tópico esperado, fuente y polaridad. Posteriormente, para cada corpus realizamos dos pruebas y recuperamos 1 y 3 frases, respectivamente, usando técnicas de expansión de la pregunta utilizando técnicas de *Latent Semantic Analysis* y filtrado de acierto con el tópico esperado. Empleamos también análisis de sentimientos para poder seleccionar los fregmentos cuya polaridad es la misma que la polaridad esperada para la pregunta/respuesta. Tenemos en cuenta las primeras 5 respuestas hasta un máximo de 50.

TABLA 19:
Resultados para el inglés

Q	T	A	Número de Respuestas							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
1	F	5	0	0	0	2	0	3	4	4
2	O	5	0	0	0	1	0	1	0	3
3	F	2	1	1	2	1	2	1	2	1
4	F	10	1	1	2	1	6	2	10	4
5	O	11	0	0	0	0	0	0	0	0
6	O	2	0	0	0	0	0	1	0	2
7	O	5	0	0	0	0	0	1	0	3
8	F	5	1	0	3	1	3	1	5	1
9	F	5	0	1	0	2	0	2	1	3
10	F	2	1	0	1	0	1	1	2	1
11	O	2	0	1	0	1	0	1	0	1
12	O	3	0	0	0	1	0	1	0	1
13	F	1	0	0	0	0	0	0	0	1
14	F	7	1	0	1	1	1	2	1	2

Q	T	A	Número de Respuestas							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
15	F/O	1	0	0	0	0	0	1	0	1
16	F/O	6	0	1	0	4	0	4	0	4
17	F	10	0	1	0	1	4	1	0	2
18	F/O	1	0	0	0	0	0	0	0	0
19	F/O	27	0	1	0	5	0	6	0	18
20	F/O	4	0	0	0	0	0	0	0	0

TABLA 20:

Resultados para el español

Q	T	A	Número de Respuestas							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
1	F	9	1	0	0	1	1	1	1	3
2	F	13	0	1	2	3	0	6	11	7
3	F	2	0	1	0	2	0	2	2	2
4	F	1	0	0	0	0	0	0	1	0
5	F	3	0	0	0	0	0	0	1	0
6	F	2	0	0	0	1	0	1	2	1
7	F	4	0	0	0	0	1	0	4	0
8	F	1	0	0	0	0	0	0	1	0
9	O	5	0	1	0	2	0	2	0	4
10	O	2	0	0	0	0	0	0	0	0
11	O	5	0	0	0	1	0	2	0	3
12	O	2	0	0	0	1	0	1	0	1
13	O	8	0	1	0	2	0	2	0	4
14	O	25	0	1	0	2	0	4	0	8

Q	T	A	Número de Respuestas							
			@1		@5		@10		@ 50	
			TQA	OQA	TQA	OQA	TQA	OQA	TQA	OQA
15	0	36	0	1	0	2	0	6	0	15
16	0	23	0	0	0	0	0	0	0	0
17	0	50	0	1	0	5	0	6	0	10
18	0	10	0	1	0	1	0	2	0	2
19	0	4	0	1	0	1	0	1	0	1
20	0	4	0	1	0	1	0	1	0	1

Para recuperar las posibles frases candidatas (snippets) utilizamos la herramienta de recuperación de información JIRS. Después calculamos la polaridad de los posible candidatos con distintos métodos y usamos la similitud entre frases anotadas con el EmotiBlog corpus.

Llevamos a cabo varias evaluaciones: a) seleccionamos los snippets con la misma polaridad que el tipo de polaridad esperada y tópico; b) misma polaridad pero no tópico; c) añadimos los resultados del LSA para filtrar los snippets de longitud 1 frase (usando Yahoo!); y por último d) filtramos los resultados usando la herramienta de roles semánticos Semrol.

TABLA 21:

Resultados con la colección de preguntas del TAC 2008

System	F-measure
Best TAC	0,534
Worst TAC	0,101
JIRS + SA+ET (1 phrase)	0,377
JIRS + SA+ET (3 phrases)	0,431
JIRS + SA+ET+LSA (1 phrase)	0,489
JIRS + SA+ET+LSA (3 phrases)	0,505
JIRS + SA+ET+LSA+SR (1 phrase)	0,533
JIRS + SA+ET+LSA+SR (3 phrases)	**0.571**

Los resultados obtenidos nos dan pie para reflexionar sobre varios problemas como por ejemplo la ambigüedad de las preguntas y del texto, pero también la necesidad de recuperar la respuesta que se encuentra en más de una frase. Para que sea efectivo, el procedimiento necesita un análisis y comprensión que trate el tema y la fuente de manera conjunta. Estos elementos están incluidos en nuestro modelo de anotación y por tanto puede ser de gran ayuda para solventar este problema.

Generación de Resúmenes Automáticos

La última fase de la evaluación extrínseca ha sido realizada aplicando y explotando nuestro recurso en la tarea de generación de resúmenes automáticos. Nuestro objetivo es comprobar si EmotiBlog es un recurso útil para poder desarrollar la tarea con la información subjetiva, es decir para poder generar de forma automática resúmenes subjetivos. Si en la minería de opiniones los experimentos que hicimos servían para mejorar la tarea y en búsqueda de respuestas mejoramos los sistemas creados para el tratamiento de la información subjetiva, ahora elegimos la tarea de generación de resúmenes automáticos con el objetivo de comprobar si EmotiBlog podría ser útil para la creación de sistemas enfocados a resumir datos subjetivos.

Con este propósito hemos recopilado una colección de entradas de blogs sobre distintos temas: economía, ciencia, tecnología, cocina, sociedad y deporte. Una vez recopilada la colección la anotamos con algunos elementos del modelo y clasificamos automáticamente la polaridad de las frases y del documento explotando el corpus EmotiBlog anotado. Finalmente produjimos resúmenes con distintos niveles de compresión, que sirvieran para representar de forma breve y concisa la información subjetiva más relevante de cada uno de los blogs recopilados.

Los resúmenes contienen opiniones positivas y negativas discriminadas dependiendo de su polaridad y para poder discriminar correctamente dichas polaridades a lo largo de los documentos, EmotiBlog ha tenido un papel fundamental en este proceso

Para evaluar los resúmenes utilizamos los siguientes criterios: redundancia, corrección gramatical, tópico, dificultad. Para cada uno de estos aspecto se establectión una escala cualitativa (no aceptable, legible y aceptable) para evaluar cómo de buenos eran los resúmenes generados en cada uno de los criterios establecidos. En todos estos aspectos se obtuvieron buenos resultados, a pesar de la dificultada de la tarea,, por lo que se puede deducir queEmotiBlog puede ser empleado también en este tipo de tareascomo elemento clave para un buena análisis del texto a resumir, permitiendo identificar y distinguir entre la información objetiva y subjetiva.

Los resultados obtenidos se presentan abajo (Tabla 22) y contemplan los distintos tamaños de resumen elegidos.

TABLA 22:

Resultados de la evaluación para ratio de compresión 10%, 15% AND 20%

RATIO DE COMPRESIÓN 10%			
	Non Accept.	**Understand**	**Accept**
Redun.	26%	45%	29%
Gram.	4%	22%	74%
Focus	33%	43%	24%
RATIO DE COMPRESIÓN 15%			
Redun.	0%	6%	94%
Gram.	2%	27%	71%
Focus	26%	29%	45%
RATIO DE COMPRESIÓN 20%			
Redun.	4%	10%	86%
Gram.	0%	55%	45%
Focus	14%	47%	39%

Como podemos apreciar en la tabla 22, la gramaticalidad disminuye conforme vamos aumentando el tamaño.

Además, debido al nivel de informalidad de los blogs, otra de nuestras hipótesis era que obtendríamos muchos errores gramaticales, cosa que no se verifica de los resultados obtenidos. Agrupando los porcentajes, obtenemos 65, 82 y 92 para 10, 15 y 20, respectivamente. El tamaño de resúmenes del 20% obtiene los mejores resultados.

Por lo tanto de aquí podemos ver que EmotiBlog es útil para el tratamiento de la información subjetiva que constituye una fase determinante del proceso de resúmenes, especialmente cuando la finalidad es producir resúmenes subjetivos. De hecho, sin la incorporación de una etapa intermedia de análisis de sentimientos, el sistema no sería capaz de interpretar correctamente la información y por lo tanto, producir un resumen fiel al texto de origen.

Conclusiones

Esta tesis doctoral está centrada en el área de análisis de sentimientos, subtarea del PLN. Concretamente, nos hemos centrado en el análisis, propuesta, desarrollo y evaluación de un recurso ,que comprende un modelo de anotación EmotiBlog-Annotation-Model y su corpus anotado, EmotiBlog-Corpus. Con el desarrollo de este recurso hemos contribuido a solventar problemas y desafíos que presentaban los recursos existentes para el análisis de sentimientos en la actualidad. Nuestro recurso es multilingüe y multidominio creado para detectar la subjetividad en los nuevos géneros textuales, y por tanto hemos contribuido a mejorar el estado del arte en análisis de sentimientos.

Nuestra motivación se basa principalmente en el hecho de que hoy en día con la web 2.0 tenemos a disposición una gran cantidad de información subjetiva sobre un amplia abanico de temas valiosos para explotar y crear numerosas aplicaciones de uso y utilidad real.

Esta información es muy importante dado que numerosas encuestas y estudios demuestran que pueden influenciar la actitud, y decisiones de numerosos usuarios.

El enfoque que hemos adoptado en el desarrollo de nuestro recurso sido a nivel multilingue (inglés, español e italiano) y centrado en el nuevo género textual de los blogs. Este enfoque ha sido evaluado tanto intrínseca como extrínsecamente, obteniendo buenos resultados en ambos tipos de evaluación. Por lo tanto, de los experimentos y evaluación realizados concluimos que Emotiblog (tanto el modelo de anotación como el corpus anotados con el modelo en los tres idiomas propuestos) es adecuado y útil para ser utilizado para anotar otros corpus, como para ser integrado y utilizado en otras tareas de PLN.

En resumen, nuestras contribuciones principales han sido:

- Presentar y analizar la importancia y utilidad de los nuevos géneros textuales nacidos con la web 2.0, especialmente los blogs.
- Describir en detalle las características y peculiaridades que presentan los blogs, y cómo se diferencias de otros géneros textuales más tradicionales. Realizar un estudio detallado del estado de la cuestión en análisis de sentimientos.
- Identificar los problemas y las limitaciones de los recursos existentes. En base a la revisión del estado de la cuestión, hemos detectado los siguientes problemas, aportando las correspondientes soluciones:
  - o Existe una escasez evidente de corpus en otras lenguas y compuestos por entradas de blogs, entonces:
    - Creamos un corpus multilingüe de entradas de blogs en 3 lenguas sobre 3 temas
  - o Hay escasez de modelos de anotación de granularidad fina para poder capturar las expresiones de subjetividad.
    - Después de analizar nuestro corpus, creamos nuestro modelo de anotación EmotiBlog un esquema de anotación de granularidad fina para detectar la subjetividad en los nuevos géneros textuales y anotamos el corpus.
    - Realizar una exhaustiva evaluación, que se compone de una evaluación intrínseca y extrínseca.
  - o En cuanto a la evaluación intrínseca, se concluye que:
    - El modelo es claro y permite unua anotación no compleja, factible y clara. El modelo permite una clasificación correcta.

  - El modelo se puede mejorar a partir del análisis del impacto de los elementos que lo componen, y por tanto lo mejoramos para obtener la versión definitiva.

- En relación a la evaluación extrínseca, se concluye que:

  - EmotiBlog corpus es un recurso útil y necesario para mejorar el rendimiento de tareas de PLN que manejan información subjetiva.

  - El esquema de anotación de EmotiBlog funciona con textos que no sean blogs y que pertenecen a otras temáticas.

  - Para la tarea de minería de opiniones, EmotiBlog es útil para entrenar un sistema de aprendizaje automático y ayudar en la clasificar correctamente las frases en subjetiva/objetiva positiva/negativa, así como también para determinar la intensidad de las emociones (alta/media/baja)

  - Para la búsqueda de respuestas, EmotiBlog ayuda al sistema a clasificar las respuestas en positivas/negativas. Además, EmotiBlog contempla los elementos necesarios (fuente–tema) imprescindibles para el buen desarrollo de la tarea y proporciona una mejora de la - misma.

  - Para la tarea de generación de resúmenes automáticos, EmotiBlog es de gran utilidad para clasificar entre polaridades necesarias, permitiendo que el sistema de generación de resúmenes disponga de información acerca de la subjetividad contenida en los documentos, posibilitando de esta manera, la generación de resúmenes subjetivos.

References

Agerri, R., and A. García-Serrano. “Q-WordNet: Extracting Polarity from WordNet Senses.” *In Proceedings of LREC, Malta 2010*.

Arnold, M. B. “Emotion and personality.” *New York: Columbia University Press*, 1960.

Artstein, R., and M. Poesio. “Inter-coder agreement for computational linguistics (survey article).” *Computational Linguistics 34(4): 555-596*, 2008.

Balahur, A., and A. Montoyo. “Applying a Culture Dependent Emotion Triggers Database for Text Valence and Emotion Classification.” *In Proceedings of the AISB 2008 Symposium on Affective Language in Human and Machine, Aberdeen, Scotland*, 2008.

Balahur, A., and A. Montoyo. “Applying a culture dependent emotion triggers database for text valence and emotion classification.” *Procesamiento del Lenguaje Natural, 40(40)*, 2008.

Balahur, A., E. Lloret, O. Ferrández, A. Montoyo, M. Palomar, and R. Muñoz. “The DLSIUAES Team’s Participation in the TAC 2008 Tracks.” *In Proceedings of the Text Analysis Conference 2008 Workshop*, 2008.

Balahur, A., and A. Montoyo. “Semantic Approecahes to fine and coarse grained feature based opinion mining.” *In Proceedings of the International Coneference on Application of Natural Language to Information Systems*, 2009.

Balahur, A., Boldrini, E. Boldrini, A. Montoyo, and P. Martínez-Barco. “Fact versus Opinion Questions Classification and Answering: Challenges and Keys.” *In ICAI’09 - The 2009 International Conference on Artificial Intelligence. Las Vegas, Nevada, USA*, 2009.

Balahur, A., E. Boldrini, A. Montoyo, and P. Martínez-Barco. “Cross-topic opinion mining for real-time human-computer interaction.” *In Proceedings of the Workshop on Natural Language and Cognitive Science , NLPCS 2009: 13-22*, 2009.

Balahur, A., E. Boldrini, A. Montoyo, and P. Martínez-Barco. "Opinion and Generic Question Answering Systems: a Performance Analysis." *In Proceedings of ACL. Singapur*, 2009.

Balahur, A., R. Steinberger, E. van der Goot, and B. Pouliquen. "Opinion Mining from Newspaper Quotations." *In Proceedings of the Workshop on Intelligent Analysis and Processing of Web News Content, 2009 IEEE/WIC/ACM International Conference on Web Intelligence held in conjunction with IAT'09, September 2009, Milan, Italy*, 2009.

Balahur, A., et al. "Sentiment Analysis in the News." *In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010), pp. 2216-2220. Valletta, Malta*, 2010.

Balahur, A., and A. Montoyo. "OpAL: Applying Opinion Mining Techniques for the Disambiguation of Sentiment Ambiguous Adjectives in SemEval-2 Task 18." *In Proceedings of SemEval-2, the 5th International Workshop on Semantic Evaluation, satellite workshop to ACL 2010*.

Banea, C., R. Mihalcea, and J. Wiebe. "A bootstrapping method for building subjectivity lexicons for languages with scarce resources." *In Proceedings of the Conference on Language Resources and Evaluations (LREC 2008)*.

Banea, C., R. Mihalcea, J. Wiebe, and S. Hassan. "Multilingual subjectivity analysis using machine translation." *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*.

Banea, C., R. Mihalcea, and J. Wiebe. "Multilingual subjectivity: are more languages better?" *In Proceedings of the International Conference on Computational Linguistics (COLING 2010), Beijing, China*, 2010.

Banfield, A. *Unspeakable sentences: Narration and Representation in the Language of Fiction. Routledge and Kegan Paul.* Edited by Routledge and Kegan Paul. 1982.

Blitzer, J., M. Dredze, and F. Pereira. "Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification." *Association of Computational Linguistics (ACL), 2007*.

Boldrini, E., A. Balahur, P. Martínez-Barco, and A. Montoyo. "EmotiBlog: an Annotation Scheme for Emotion Detection and Analysis in Non-traditional Textual Genres." *In Proceedings of DMIN 2009, Las Vegas. Nevada*, 2009.

Bossard, A., M. Généreux, and T. Poibeau. "Description of the LIPN systems at TAC 2008: Summarizing information and opinions." *In Proceedings of the Text Analysis Conference of the National Institute for Standards and Technology*, 2008.

Bradley, M. M., and P. J. Lang. "International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2)." *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*, 1999.

Brendel, M., R. Zaccarelli, and L. Deuvillers. "CINEMO — A French Spoken Language Resource for Complex Emotions: Facts and Baselines." *In Proceedings of LREC. Malta 2010*.

Breugelmans, S. M., et al. "Body sensations associated with emotions in Raramuri Indians, rural Javanese, and three student samples." *Emotion, 5, 166–174*, 2005.

Cardie, C., J. Wiebe, T. Wilson, and D. Litman. "Low-Level Annotations and Summary Representations of Opinions for Multiperspective QA." *In Mark Maybury (ed), New Directions in Question Answering , AAAI Press/MIT Press*, 2004.

Carletta, J. "Assessing agreement on classification task: the kappa statistic." *Computational Linguistics, 22(2):249–254*, 1996.

Cerini, S., V. Compagnoni, A. Demontis, M. Formentelli, and G. Gandini. "Micro-wnop: A gold standard for the evaluation of auto-matically compiled lexical resources for opinion mining." *Franco Angeli Editore, Milano, IT*, 2007.

Chaovalit, P., and L. Zhou. "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches." *In Proceedings of HICSS-05, the 38th Hawaii International Conference on System Sciences*, 2005.

Choi, Y., C. Cardie, E. Riloff, and S. Patwardhan. "Identifying sources of opinions with conditional random fields and extraction patterns." *In Proceeding of the conference on empirical methods in natural language processing (EMNLP 2005), October 6–8, 2005 (pp. 355–362). Vancouver, B.C., Canada*, 2005.

Cohen, J. "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement, 20:37–46*, 1960.

Conroy, J., and S. Schlesinger. "Classy at TAC 2008 metrics." *In Proceedings of the Text Analysis Conference of the National Institute for Standards and Technology*, 2008.

Copeck, T., D. Inkpen, A. Kazantseva, A. Kennedy, D. Kipp, and S. Szpakow. *In Proceedings of DUC 2007* (Catch What You Can), 2007.

Cruz, F., J. Troyano, J. Ortega, and F. Enríquez. "The Italica system at TAC 2008 opinion summarization task." *In Proceedings of the Text Analysis Conference of the National Institute for Standards and Technology*, 2008.

Cui, H., V. Mittal, and M. Datar. "Comparative Experiments on Sentiment Classification for Online Product Reviews." Edited by In Proceedings of the 21st National Conference on Artificial Intelligence AAAI 2006. *. In Proceedings of the 21st National Conference on Artificial Intelligence AAAI 2006.*.

Damasio, A. R. "Descartes' error: emotion, reason, and the human brain." *New York: G. P. Putnam; 1994*.

Das, S., and M. Chen. "Yahoo! for Amazon: Extracting market sentiment from stock message boards." *In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).*, 2001.

Dave, K., S. Lawrence, and D. Pennock. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews." *In Proceedings of WWW-03*, 2003.

Davidson, R., K. Sherer, and H. Goldsmith. "Handbook of Affective Sciences." Edited by Boldrini. *Oxford University Press*, 2003.

De Rivera, J. "A structural theory of the emotions." *New York: International Universities Press*, 1977.

Deerwester, S., S. Dumais, G. Furnas, T. Landauer, and R. Harshman. "Indexing by latent semantic analysis." *Journal of the American Society for Information Science, 41(16):391–407*, 1990.

Donaway, R., K. W. Drummey, and L. A. Mather. "A comparison of rankings produced by summarization evaluation measures." *In Proceedings of NAACL-ANLP 2000 Workshop on Automatic Summarization*, 2008.

Dyer, M. "Emotions and their computations: three computer models." *Cognition and Emotion, 1, 323-347*, 1987.

Elliot, C.D. "The Affective Reasoner: a process model of emotions in a multi-agent system." *Ph.D. thesis, Northwestern University, Evanston, Illinois*, 1992.

Ekman, P. "Basic Emotions." (In T. Dalgleish and M. Power (Eds.). Handbook of Cognition and Emotion. Sussex, U.K.: John Wiley & Sons, Ltd., 1999.) 1999.

Ekman, P. "Universals and cultural differences in facial expressions of emotion." *In J. Cole (Ed.), Nebraska symposium on motivation, 1971 (pp. 207-283). Lincoln: University of Nebraska Press*, 1972.

Ekman, P., and W. V. Friesen. "The repertoire of nonverbal behavior: Categories, origins, usage, and coding." *Semiotica, 1, 49–98*, 1969.

Ekman, P., W. V. Friesen, and P. Ellsworth. "What emotion categories or dimensions can observers judge from facial behavior?" *In P. Ekman (Ed.), Emotion in the human face (pp. 39-55). Cambridge, England: Cambridge University Press*, 1982.

Erkan, G., and D.R. Radev. "LexRank: Graph-based centrality as salience in text summarization." *Journal of Artificial Intelligence Research (JAIR)*, 2004.

Esuli, A., and F. Sebastiani. "Determining the semtnaic orientation of terms through gloss analysis." *In Proceedings of CIKM 2005*.

Esuli, A., and F. Sebastiani. "Sentiwordnet: A publicly available resource for opinion mining." *In Proceedings of the 6th International Conference on Language Resources and Evaluation*, 2006.

Flew, T. "New Media: An Introduction." (Oxford University Press, UK, pg. 13) 2002.

Fangtao, L., et al. "THU QUANTA at TAC 2008 QA and RTE track." *In Proceedings of Human Language Technologies Conference/Conference on Empirical methods in Natural Language Processing (HLT/EMNLP), Vancouver, BC, Canada, 2008*.

Fontaine, J., Y. Poortinga, B. Setiadi, and S. Markam. "Cognitive structure of emotion terms in Indonesia and the Netherlands." *Cognition & Emotion, 16, 61-86*, 2002.

Frijda, N.H. "The emotions." (Cambridge Cambridge University Press.) 1986.

Gamon, M. "Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis." *In Proceedings of COLING-04, the 20th International Conference on Computational Linguistics (pp. 841–847). Geneva, CH*, 2004.

Ghose, A., P. G. Ipeirotis, and A. Sundararajan. "Opinion Mining using Econometrics: A Case Study on Reputation Systems." *In Proceedings of the ACL 2007*.

Goldberg, A. B., and J. Zhu. "Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization." *In HLT-NAACL*

2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing, 2006.

Gray, J. A. "The whole and its parts: Behaviour, the brain, cognition and emotion." *Bulletin of the British Psychological Society. 38, 99-112*, 1985.

Gómez, J. M., P. Rosso, and E. Sanchis. "JIRS Language-Independent Passage Retrieval System: A Comparative Study." *5th International Conference on Natural Language Proceeding (ICON 2007)*, 2007.

Harrigan, J. A., R. Rosenthal, and K. Scherer. "The new handbook of methods in nonverbal behavior research." *Series in Affective Science. Oxford University Press, Oxford*, 2005.

Hatzivassiloglou, V., and J. Wiebe. "Effects of adjective orientation and gradability on sentence subjectivity." *In Proceedings of COLING 2000*.

Hatzivassiloglou, V., and K. McKeown. "Predicting the semantic orientation of adjectives." *In Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL*, 1997.

He, T., J. Chen, Z. Gui, and F. Li. "CCNU at TAC 2008." *In Proceedings of the Text Analysis Conference of the National Institute for Standards and Technology*, 2008.

Hovy, E. "Automated text summarization." *Oxford University Press, Oxford, UK. In Mitkov, R., editor, The Oxford Handbook of Computational Linguistics, pages 583–598.*, 2005.

Hovy, E. H., and C. Y. Lin. "Automated Text Summarization in SUMMARIST." *In I. Mani and M. Maybury (eds), Advances in Automatic Text Summarization. MIT Press, 81-94. 1999*.

Hu, M., and B. Liu. "Mining and summarizing customer reviews." *In Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining. Aug. 22–25, 2004, Seattle, WA, USA.*, 2004.

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

Izard, C.E. "The face of emotion." *New York: Apple- ton-Century-Crofts*, 1971.

James, W. "What is an emotion?" *Mind, 9, 188-205*, 1884.

Jijkoun, V., and K. Hofmann. "Generating a Non-English Subjectivity Lexicon: Relations That Matter." *In Proceedings of EACL-2009, Athens, Greece*, 2009.

Kabadjov, M, A. Balahur, and E. Boldrini. “Sentiment Intensity: Is It a Good Summary Indicator?” *In Proceedings of the 4th Language Technology Conference LTC, pp. 380-384*, 2009.

Kilgarriff, A. and Grefenstette, G.. “Introduction to the special issue on the web as corpus.” *In computational Linguistics, 29(3): 333-347*, 2003.

Kim, S. M., and E. Hovy. “Automatic identification of pro and con reasons in online reviews.” *In Proceedings of the COLING/ACL Main Conference Poster Sessions, pp. 483–490*, 2006.

Kim, S. M., and E. Hovy. “Determining the Sentiment of Opinions.” *In Proceedings of COLING 2004*, 2004.

Knebel, T., S. Hochreiter, and K. Obermayer. “An SMO Algorithm for the Potential Support Vector Machine.” *Neural Computation(2008) 271-287*, 2008.

Koppel, M., and I. Shtrimberg. “Good or bad news? Let the market decide.” *In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, 2004.

LeDoux, J. E. “Cognitive-emotional interactions in the brain.” *Cognition and Emotion, 3, 267–89*, 1989.

Levi, A. W. “Nature and art.” *Journal of Aesthetic Education 18(3):5-21*, 1984.

Lewis, D., and W. Gale. “A sequential algorithm for training text classifiers.” *In Proceedings of the 17th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval. Springer-Verlag New York*, 1994.

Li, W., Y. Ouyang, Y. Hu, and F. Wei. “PolyU at TAC 2008.” *In Proceedings of Human Language Technologies Conference/Conference on Empirical methods in Natural Language Processing (HLT/EMNLP), Vancouver, BC, Canada, 2008*.

Lin, D., and P. Pantel. “Discovery of Inference Rules for Question Answering.” *Natural Language Engineering 7(4):343-360*, 2001.

Lindquist, K. A. “Language is Powerful.” *Emotion Review Journal*, 2009.

Liu, B. “Sentiment Analysis and Subjectivity.” *Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau)*, 2010.

Lloret, E., and M. Palomar. “Lecture Notes in Computer Science. 12th International Conference on Text, Speech and Dialogue.” *A Gradual Combination of features for Building Automatic Summarisation Systems*, 2009.

Elena Lloret. Text Summarisation based on Human Language Technologies and its Applications. PhD Thesis. University of Alicante, 2011.

Lüdeling, A., S. Evert, and M. Baroni. "Using Web Data for Linguistic Purposes." *Corpus Linguistics and the Web. Edited by Marianne Hundt, Nadja Nesselhauf and Carolin Biewer , pp. 7-24(18)*, 2006.

Lutz, C. A. "Unnatural emotions." *Chicago, IL: University of Chicago Press*, 1988.

McDonald, R., H. Kerry, N. Tyler, M. Wells, and J. C. Reynar. "Structured Models for Fine-to-Coarse Sentiment Analysis." *In Proc.eedings of the ACL*, 2007.

McDougall, W. "An introduction to social psychology." *Boston: Luce*, 1926.

Macleod, C., R. Grishman, and A. Meyers. "Creating a common syntactic dictionary of English." *In Proceedings of the International Workshop on Sharable Natural Language Resources*, 1994.

Maks, I., and P. Vossen. "Annotation Scheme and Gold Standard for Dutch Subjective Adjectives." *In Proceedings of LREC, Malta 2010*, 2010.

Mathieu, Y. Y. "A computational semantic lexicon of french verbs of emotion." *In Shanahan J. G., Qu Y. and Wiebe J., Eds., Computing attitude and affect in text: Theorie and applications, 109-124. Springer*, 2006.

Mihalcea, R., C. Banea, and J. Wiebe. "Learning multilingual subjective language via cross-lingual projections." *In Proceedings of the Conference of the Annual Meeting of the Association for Computational Linguistics 2007.*, 2007.

Moreda, P. "Los Roles Semánticos en la Tecnología del Lengauje Humano: Anotación y Aplicación." *Doctoral Thesis. University of Alicante*, 2008.

Morris, A. H., G. M. Kasper, and D. A. Adams. "Information Systems Research, Vol. 3 (1) 17-35." *The effect and limitation of automated text condensing on reading comprehension performance*, 1992.

Mowrer, O. H. "Learning theory and behavior." *New York: Wiley*, 1960.

Mullen, T., and M. Collier. "Sentiment Analysis Using Support Vector Machines with Diverse Information Sources." *In Proceedings of EMNLP 2004*, 2004.

Nasukawa, T., and J. Yi. "In Sentiment analysis: Capturing favorability using natural language processing." *In Proceedings of the Conference on Knowledge Capture (K-CAP)*, 2003.

Oatley, K., and P. N. Johnson-Laird. "Towards a cognitive theory of emotions." *Cognition & Emotion, 1, 29-50*, 1987.

Orasan, C. “Comparative evaluation of modular automatic summarisation systems using CAST.” *PhD Thesis*, 2006.

Ortony, A., and T. J. Turner. “What’s basic about basic emotions?” *Psychological Review, 97, 315-331*, 1990.

Ortony, A., G. L. Clore, and A. Collins. “The cognitive structure of emotions.” (Cambridge: Cambridge University Press) 1988.

Ounis, I., M. de Rijke, G. Mishne, and I. Soboroff. “Overview of the TREC 2006 blog track.” *In The Fifteenth Text Retrieval Conference (TREC 2006), National Institute for Science and Technogies.*, 2006.

Plutchik, R. “A general psychoevolutionary theory of emotion.” *In R. Plutchik & H. Kellerman (Eds.), Emotion: Theory, research, and experience: Vol. 1. Theories of emotion (pp. 3-33). New York: Academic*, 1980.

Pang, B., L. Lee, and S. Vaithyanathan. “Thumbs up? Sentiment classification using machine learning techniques.” *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002*, 2002.

Pang, B., and L. Lee. “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts.” *In Proceedings of the ACL 2004*, 2004.

Pang, B., and L. Lee. “Opinion mining and sentiment analysis.” *Foundations and Trends in Information Retrieval, Vol 2, Nr. 1-2, 2008*, 2008.

Pang, B., and L. Lee. “Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales.” *In Proceedings of the 43rd Annual Meeting of the ACL, pages 115–124*, 2003.

Panksepp, J. “Toward a general psychobiological theory of emotions.” *The Behavioral and Brain Sciences, 5, 407-467*, 1982.

Paquet, S. “Personal knowledge publishing and its uses in research.” *National Research Council of Canada*, 2002.

Pariser, E. “The Filter Bubble: What the Internet Is Hiding from You.” 2010.

Parrott, W. “Emotions in Social Psychology.” *Psychology Press, Philadelphia*, 2001.

Picard, R. “Affective Computing.” *MIT Press*, 1997.

Pustejovsky, J., and J. Wiebe. “Introduction to special issue on advances in question answering.” *Language Resources and Evaluation (2005), (39)*, 2005.

Quarteroni, S., A. Moschitti, S. Manandhar, and R. Basili. "Advanced structural representations for question classification and answer re-ranking." *In Proceedings of ECIR 2007, pages 234–245, Rome*, 2007.

Quirk, R. "A Comprehensive Grammar of the English Language." *Longman*, 1985.

Reisenzein, R. "Pleasure-arousal theory and the intensity of emotions." *Journal of Personality and Social Psychology, 67, 525-539*, 1994.

Remus, R., U. Quasthoff, and G. Heyer. "SentiWS - A Publicly Available German-language Resource for Sentiment Analysis." *In Proceedings of LREC 2010. Malta*, 2010.

Riloff, E, J. Wiebe, and W. Phillips. "Exploiting subjectivity classification to improve information extraction." *In Proceedings of the AAAI 2005*.

Riloff, E., and J. Wiebe. "Learning extraction patterns for subjective expressions." *In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP 2003*.

Ruimy, N., et al. "A computational semantic lexicon of italian: SIMPLE." *In: Linguistica Computazionale XVIII-XIX, Pisa, pp. 821–64*, 2003.

Russell, J. A. "Culture and the categorization of emotion." *Psychological Bulletin, 110, 426-450*, 1991.

Russell, J. A. "Pancultural aspects of the human conceptual organization of emotions." *Journal of Personality and Social Psychology, 45, 1281-1288*, 1983.

Russell, T. A., A. Qualter, and L. McGuigan. "Reflections on the implementation of National Curriculum Science Policy for the 5-14 age range: findings and interpretations from a national evaluation study in England." *International Journal of Science Education, 17, (4), pp. 481-492*, 1995.

Scherer, K. R. "Appraisal Considered as a Process of Multi-Level Sequential Checking." *in K.R. Scherer, A. Schorr and T. Johnstone (eds) Appraisal Processes in Emotion: Theory, Methods, Research, pp. 92–120. . New York and Oxford: Oxford University Press*, 2001.

Scherer, K. R. "Appraisal theories." *In T. Dalgleish, & M. Power (Eds.). Handbook of Cognition and Emotion (pp. 637–663). Chichester: Wiley*, 1999.

Scherer, K. R. "Emotion as a Multicomponent Process: A Model and Some Cross-Cultural Data." *In P. Shaver (ed.) Review of Personality and Social Psychology, Vol. 5, pp. 37–63. Beverly Hills, CA: Sage*, 1984.

Scherer, K. R. “Feelings Integrate the Central Representation of Appraisal- Driven Response Organization in Emotion’’,.” *in A.S.R. Manstead, N.H. Frijda and A.H. Fischer (eds) Feelings and Emotions: The Amsterdam Symposium, pp. 136–57. Cambridge: Cambridge University Press*, 2004.

Scherer, K. R. “Studying the Emotion-Antecedent Appraisal Process: An Expert System Approach.” *Cognition and Emotion 7: 325–55*, 1993.

Scherer, K. R. “What are emotions? And how can they be measured?” *Social Science Information., 44(4), 693–727*, 2005.

Scherer, K. R., and H. Wallbott. “The ISEAR Questionnaire and Codebook.” *Geneva Emotion Research Group*, 1997.

Scherer, K.R. “Criteria for Emotion-Antecedent Appraisal: A Review’.” *In V. Hamilton, G.H. Bower and N.H. Frijda (eds) Cognitive Perspectives on Emotion and Motivation, pp. 89–126. Dordrecht: Kluwer.*, 1988.

Scherer, K.R. “Toward a Dynamic Theory of Emotion: The Component Process Model of Affective States.” *Geneva Studies in Emotion and Communication 1: 1–98*, 1987.

Salton, G., and M. E. Lesk. “Computer evaluation of index- ing and text processing,.” *Pages 143–180. Prentice Hall, Ing. Englewood Cliffs, New Jersey*, 1971.

Saggion, H., and A. Funk. “Interpreting SentiWordNet for opinion classification.” *In Proceedings of LREC 2010*.

Saggion, H., E. Lloret, and M. Palomar. “Using text summaries for predicting rating scales.” *In Proceedings of the 1st Workshop on Subjectivity and Sentiment Analysis WASSA 2010*.

Sebastiani, F. “Machine Learning in automated text categorization.” *ACM Computing Surveys (CSUR), 34, 1-47*, 2002.

Shen, D., et al. “The alyssa system at trec qa 2007: Do we need blog06?” *In Proceedings of TREC 2007*.

Snyder, B., and R. Barzilay. “Multiple Aspect Ranking using the Good Grief Algorithm.” *In Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL). pp. 300–307*, 2007.

Somasundaran, S., J. Wiebe, P. Hoffmann, and D. Litman. “Manual Annotation of Opinion Categories in Meetings.” *COLING-ACL Workshop: Frontiers in Linguistically Annotated Corpora*, 2006.

Steinberger, J., M. Poesio, M. Kabadjov, and K. Jezek. "Two uses of anaphora resolution in Summmarization." *Information Processing and Management, 43(6):1663–1680. Special Issue on Text Summarisation (Donna Harman, ed.)*, 2007.

Stemmler, G. "Methodological considerations in the psychophysiological study of emotion." *In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer (Eds.), Handbook of affective sciences (pp. 225–255). New York: Oxford University Press*, 2003.

Stone, P., D.G. Dumphy, M. S. Smith, and D. M. Ogilvie. "The General Inquirer: A Computer Approach to Content Analysis." *The MIT Press*, 1966.

Stoyanov, V., C. Cardie, and J. Wiebe. "Multiperspective question answering using the opqa corpus." *In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, 2005.

Stoyanov, V., C. Cardie, D. Litman, and J. Wiebe. "Evaluating an Opinion Annotation Scheme Using a New Multi-Perspective Question and Answer Corpus." *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, 2004.

Stoyanov, V., and C. Cardie. "Toward opinion summarization: Linking the sources." *In Proceedingns of the COLING-ACL 2006 Workshop on Sentiment and Subjectivity in Text*, 2006.

Strapparava, C., and A. Valitutti. "Wordnet-affect: an affective extension of wordnet." *In Proceedings of the 4th International Conference on Language Resources and Evaluation LREC 2004*.

Strapparava, C., and R. Mihalcea. "Semeval 2007 task 14: Affective text." *In Proceedings of ACL 2007*.

Subasic, P., and A. Huettner. "Affect Analysis of text using fuzzy semantic typing." *IEEE Trasactions on Fuzzy System, 9, 483-496*, 2000.

Sweeney, K., and C. Whissell. "A dictionary of affect in language: I. Establishment and preliminary validation." *Perceptual and Motor Skills, 59: 695-698*, 1984.

Taylor, S. "Emotional Labour and Organizational Restructuring within the Service Sector: Some implications for working life and work culture." *in Kristensen, C. (ed) The Meeting of the Waters: Individuality and Community in Late Modernity Oslo: Scandinavian University Press.*, 1997.

Thomas, M., B. Pang, and L. Lee. "Get out the vote: Determining support or opposition from congressional floor-debate transcripts." *In Proceedings of EMNLP 2006*.

Tomkins, S. S. "Affect theory." *In K. R. Scherer & P. Ekman (Eds.), Approaches to emotion (pp. 163-195). Hillsdale, NJ: Erlbaum*, 1984.

Tomkins, S. S., and R. McCarter. "What and where are the primary affects: Some evidence for a theory." *Perceptual and Motor Skills, 18, 119–158*, 1964.

Tong, R. M. "An operational system for detecting and tracking opinions in on-line discussion." *In Proceedings of the Workshop on Operational Text Classification (OTC), 2001*, 2001.

Turney, P. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews." *In Proceedings of the Association for Computational Linguistics (ACL), pages 417–424, 2002.*, 2002.

Vapnik, V. "The nature of statistical learning theory." *Springer-Verlag New York*, 1995.

Varma, V., et al. "Iit hyderabad at tac 2008." *In Proceedings of the Text Analysis Conference (TAC) 2008*, 2008.

Vicedo, J. L., H. Rodríguez, A. Peñas, and M. Massot. "Los sistemas de Búsqueda de Respuestas desde una perspectiva actual." *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural. Num.31*, 2003.

Waltinger, U. "GermanPolarityClues: A Lexical Resource for German Sentiment Analysis." *In Proceedings of LREC, Malta 2010*, 2010.

Watson, J. B. "Behaviorism." *Chicago: University of Chicago Press*, 1930.

Weiner, B., and S. Graham. "An attributional approach to emotional development." *In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.)*, 1984.

Whissell, C. M., and M. J. Dewson. "A dictionary of affect in language: III. Analysis of two biblical and two secular passages." *Perceptual and Motor Skills, 62: 127-133*, 1986.

Whissell, W., and K. Charuk. "A Dictionary of Affect in Language: II." *Word Inclusion and Additional Validation, Perceptual and Motor Skills, 61 (1985): 65-66*, 1985.

Wilson, T. "Fine-Grained Subjectivity Analysis." *PhD Dissertation, Intelligent Systems Program, University of Pittsburg*, 2008.

Wilson, T., and J. Wiebe. "Annotating opinions in the world press." *In Proceedings of SIGdial 2003*.

Wilson, T., et al. "OpinionFinder: A system for subjectivity analysis." *In Proc. Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005) Companion Volume (software demonstration), 2005*.

Wilson, T., J. Wiebe, and P. Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." *In Proceedings of HLT-EMNLP 2005*.

Wilson, T., J. Wiebe, and R. Hwa. "Just how mad are you? Finding strong and weak opinion clauses." *In Proceedings of AAAI 2004*, 2004.

Wiebe, J. "Tracking point of view in narrative." (Computational Linguistics, 20) 1994.

Wiebe, J., and E. Riloff. "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts." *In Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing-06)*, 2006.

Wiebe, J., and T. Wilson. "Annotating attribution and private states." *In Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*, 2005.

Wiebe, J., T. Wilson, and C. Cardie. "Annotating expressions of opinions and emotions in language." *In Language Resources and Evaluation, volume 39*, 2005.

Wierzbicka, A. "Emotions across languages and cultures: Diversity and universals." *Cambridge: Cambridge University Press*, 1999.

Yang, J., and J. Pedersen. "A comparative study on feature selection in text categorization." *Proceedings of ICML-97, 14th International Conference on Machine Learning*, 1997.

Yi, J., Nasukawa, T., Bunescu, R. and Niblack, W. "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques." *In Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 2003.

Yu, D., and V. Hatzivassiloglou. "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 (EMNLP)*, 2003.

Appendix

-EMOTIBLOG-
-Guía de anotación esencial-

INTRODUCCIÓN

Esta guía tiene el objetivo de explicar:

- Qué es EmotiBlog y para qué sirve
- En qué tipos de textos se puede emplear
- Para qué lenguas
- Cómo se anota
- Presentar algunos ejemplos de anotación
- Resolver algunos casos especialmente ambiguos de anotación

EMOTIBLOG

EmotiBlog es un esquema de anotación que sirve para anotar la subjetividad en los textos.

Por subjetividad entendemos todas aquellas frases que o presentan la simple información objetiva, los hechos.

TIPOS DE TEXTOS

EmotiBlog está pensado para trabajar con los nuevos géneros textuales nacidos con la Web 2.0 y sobre todo con los blogs.

LENGUAS

Este modelo está pensado de momento para inglés, castellano e italiano.

CÓMO SE ANOTA

Los textos tienen que estar anotados en su totalidad, es decir cada frase tiene que estar anotada. No podemos dejar ninguna frase sin etiquetas (<xxx> </xxx>). No tenemos que preocuparnos por ellas, dado que el editor de anotación las genera automáticamente.

La primera discriminación que tenemos que hacer es entre discurso **objetivo** o **sujetivo**.

> *Ej. El protocolo de Kyoto se firmó en 1997*
>
> *Ej: Creo que Estados Unidos es una nación sin escrúpulos que no se plantea ni en reducir sus emisiones*

En primer ejemplo es una frase objetiva, mientras que el segundo es evidentemente una opinión del escritor.

En la frase objetiva tendremos que poner varios elementos:

> **Source:** quién habla, quien presenta los hechos
>
> **Topic:** el objeto del discurso

En este caso sería Source: writer y Topic: The Kyoto Protocol

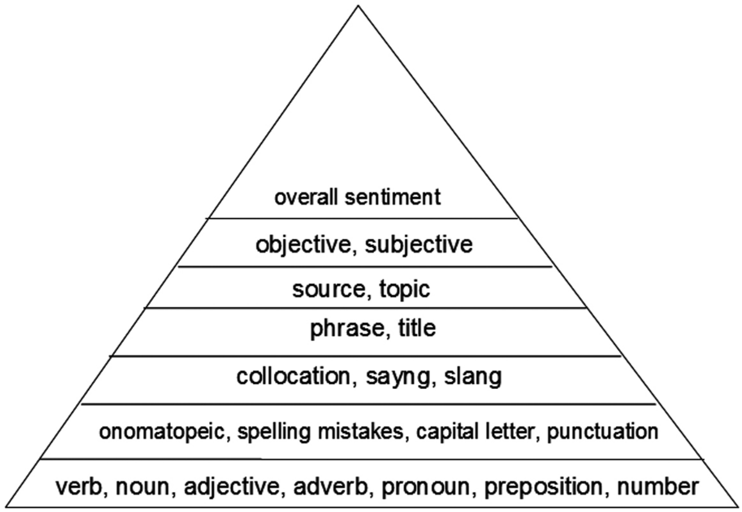
Veamos un caso de anotación concreto:

> <objective-speech-event target=”Kyoto protocol” source=”w”>Kyoto expires in 2010</objective-speech-event>

Como podemos ver, es un discurso objetivo (objective-speech-event), el tema es el Protocolo de Kyoto (target=”Kyoto protocol” y la fuente del discurso es el mismo escritor del post del blog (source=”w”).

En el caso de las frases sujetivas la anotación es un poco más compleja.

De hecho tenemos primero que anotar la frase en su totalidad como sujetiva y tendríamos que añadir como en el caso de las objetivas la Source y el Topic, pero a continuación deberemos proceder a anotar todos aquellos elementos que dan la matiz de subjetividad a la frase. Estos elementos de una manera esquemática son:



LOS ELEMENTOS Y SUS ATRIBUTOS

Ahora a continuación vamos a ver cada elemento y los distintos atributos que les pertenecen.

Elements	Description
Obj. speech	Confidence, comment, source, target.
Subj. speech	Confidence, comment, level, emotion, phenomenon, polarity, source and target.
Adjectives	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target.
Adverbs	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target.
Verbs	Confidence, comment, level, emotion, phenomenon, polarity, mode, source and target.

Elements	Description
Anaphora	Confidence, comment, type, source and target.
Capital letter	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target.
Punctuation	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target.
Names	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, and source.
Phenomenon	Confidence, comment, type: collocation, saying, slang, title, and rhetoric.
Reader Interpretation	Confidence, comment, level, emotion, phenomenon, polarity, source and target.
Author Interpretation	Confidence, comment, level, emotion, phenomenon, polarity, source and target.
Emotions	Confidence, comment, accept, anger, anticipation, anxiety, appreciation, bad, bewilderment, comfort, compassion, confidence, consternation, correct, criticism, disappointment discomfort, disgust, despondency, depression, envy, enmity, excuse, force, fear, fright, good, grief, guilt, greed, hatred, hope, irony, interesting, important, incorrect…

Ahora pasamos a tratar cada elemento.

Subj. speech	Level, emotion, phenomenon, polarity, source and target.

Frase sujetiva, que expresa una opinión y por lo tanto no se limita a describir unos hechos.

Level: es el nivel de la opinión. Puedes elegir entre High/Medium/Low

Emotion: tienes que seleccionar el tipo de emoción que se expresa. Algunas veces encontrarás que tienes que poner tres emociones. Son las emociones que expresa el escritor. Empieza por la fundamental y añade dos secundarias.

Phenomenon: puedes elegir entre Phrase/Slang/Title, Saying/Collocation

Polarity: Es la polaridad de la opinión que puede ser Positive/Negative

Source: Como explicábamos anteriormente se trata de quién expresa dicha opinión

Target: Es el objeto del discurso, pero hay que tener cuidado:

Ej: Odio a Bush que no es ambientalista

Ej: Bush dijo: “Creo que El Protocolo de Kyoto es un peligro para la industria de mi país y por esto estoy preocupado”

En el primer ejemplo la Source es el escritor y el objeto del discurso es el Protocolo de Kyopo, mientras que en el segundo se trata de una frase objetiva cuya Source es el Escritor y el Target es Bush, luego en la frase entre comillas la Source es Bush y el Target será el protocolo de Kyoto.

Veamos un ejemplo concreto de frase subjetiva:

<phenomenon target=”Kyoto Protocol” category=”phrase” degree=”medium” source=”w” polarity=”positive” emotion=”good”>The Onion has a great story today titled, “Bush Told to Sign Birthday Treaty for Someone Named Kyoto.”</phenomenon>

Podemos ver que se trata de una frase sujetiva, cuyo tema es el Protocolo de Kyoto. Es una frase normal con un nivel medio de polaridad positiva. La fuente es el propio escritor del post y el sentimiento que expresa es lo de algo bueno.

Adjectives	Level, emotion, phenomenon, modifier/not, polarity, source and target.

Adverbs	Level, emotion, phenomenon, modifier/not, polarity, source and target.

Los adjetivos y adverbios en muchos casos sirven para modificar el sustantivo u otros elementos y por lo tanto en muchos casos hay que anotarlos. Los anotaremos y añadiremos sus atributos correspondientes. Un nuevo atributo es Modifier/Not, que sirve para especificar si es un modificador o no.

Veamos un ejemplo concreto, por ejemplo para el caso de los adjetivos:

<phenomenon target=”Kyoto Protocol” category=”phrase” degree=”medium” source=”w” polarity=”positive” emotion=”good”>The Onion has a <adjective target=”Kyoto Protocol” phenomenon=”phrase” degree=”medium” polarity=”positive” emotion=”good” source=”w” ismodifier=”yes”>great</adjective> story today titled “Bush Told to Sign Birthday Treaty for Someone Named Kyoto.”</phenomenon>

Como podemos ver, el adjetivo está enmarcado en una frase sujetiva (ver ejemplo anterior) y "great" está marcado como un adjetivo que expresa subjetividad. Este adjetivo de hecho, no es objetivo, sino algo que depende del hablante. Entonces lo anotamos. Las etiquetas resultantes nos dicen que se trata de un adjetivo que hace referencia al Protocolo de Kyoto. Está enmarcado en una frase normal cuya fuente es el mismo escritor del post del blog. Tiene un nivel medio de polaridad positiva que expresa algo positivo, bueno.

Prepositions	Level, emotion, phenomenon, modifier/not, polarity, source and target.

Si es necesario y nos damos cuenta de que la preposición añade un matiz de subjetividad hay que anotarla. Esto pasa en casos raros.

<phenomenon target="Bush" category="phrase" degree="medium" source="w" polarity="negative" emotion="bad">I'm <preposition target="Bush" phenomenon="phrase" degree="medium" polarity="negative" emotion="bad" source="w" ismodifier="yes">against</adjective> Bush's view on the Kyoto Protocol</phenomenon>

Verbs	Level, emotion, phenomenon, polarity, mode, source and target.

En muchos casos el escritor decide emplear un verbo en vez de otro para y esta también es una estrategia que da subjetividad al texto. En todas aquellas veces que el verbo no es objetivo deberemos anotarlo y añadir todos sus atributos con los valores arriba explicados.

A continuación pasamos a analizar un ejemplo concreto.

<phenomenon target="Bush" category="title" degree="high" source="w" polarity="negative" emotion="criticism">Bush <verb target="Bush" degree="" source="w" tense="indicative" degree="medium" polarity="negative" emotion="bad" >Pulls</verb> U.S Out of Kyoto Treaty</phenomenon>

Como podemos apreciar en el ejemplo, el verbo está en una frase que esta vez es un título. El autor habla de Bush se trata de un indicativo con grado medio de un sentimiento negativo y entonces la polaridad será negativa.

Anaphora	Type, source and target.

Por anaphora entendemos la correferencia, pero sólo a nivel cross-document, es decirlas referencias entre blos posts.

Por correferencia entendemos *"un tipo de deixis que desempeñan ciertas palabras para recoger el significado de una parte del discurso ya emitida; p. ej., lo en dijo que había estado, pero no me lo creí".*

Nosotros anotamos los siguientes tipos:

Definite Description (descripción definida) Ej. La estudiante más alta de la clase

Pronominal (Pronominal): Él lo sabe

Adverbial (Adverbial): Todos estaban fuera. Allí contemplaban el paisaje

Ellipsis (elipsis que puede ser de sujeto o de objeto): Lo sé. / Ven a recogerlo.

Ejemplo de correferencia a nivel de cross-document

> *Ej.* ***Tu intervención*** *me ha gustado mucho. Estoy totalmente de acuerdo* ***contigo***. ***Eres*** *un buen escritor.*
>
> *Tu intervención se refiere al último post que ha escrito Paco.*
>
> *Contigo se refiere al escritor del post precedente*
>
> *Eres: es el escritor del post anterior.*
>
> *Veamos un ejemplo concreto en nuestro corpus:*
>
> *But what <anaphora type="pronominal" target="previous writer" source="w">you</anaphora> don't observe is that the EU, like the US, has basically followed a business-as-usual path on emissions*

Se trata de un fenómeno anafórico de tipo pronominal. La anáfora se refiere al autor del post anterior y la persona que habla es el escritor de este post.

Capital letter	Level, emotion, phenomenon, modifier/not, polarity, source and target.

Sirve para anotar aquellos casos en los cuales el autor decide emplear la mayúscula para expresar una opinión o estado de ánimo que no son los normales.

Ej. ESTOY HARTO DE LA POLITÍCA DE BUSH

En este caso es como si el autor quisiera expresar disgusto hacía la política de Bush y es casi como si estuviera gritando. Se nota la frase entera como capital letter y luego si hay algún elemento particular que expresa opinión se nota, sino se deja la frase como tal, pero normalmente sí que hay algún elemento para anotar dentro.

Punctuation	Level, emotion, phenomenon, modifier/not, polarity, source and target.

El elemento Punctuation sirve para anotar aquellos casos de uso anómalo de la puntuación que. Generalmente, sirve para enfatizar lo que se dice. Ten en cuenta que es imposible que aparezca una puntuación anómala como parte de frases objetivas.

Un ejemplo en el cual tendríamos que anotar el elemento Punctuation podría ser:

Ej. Odio la política de Bush en tema de medioambiente!!!

En este caso la frase sería sujetiva con todos sus atributos, anotaríamos el verbo odio y la puntuación con sus elementos.

Names	Level, emotion, phenomenon, modifier/not, polarity, and source.

En muchos casos el uso de un sustantivo más que otro para un concepto sirve para expresar subjetividad y una determinada inclinación de nuestra fuente. Por lo tanto este elemento lo consideramos relevante sólo si vemos que efectivamente tiene un matiz de subjetividad.

Ej. <phenomenon target="Bush" category="title" degree="high" source="w" polarity="negative" emotion="bad">Bush is a <name target="Bush" degree="" source="w" degree="high" polarity="negative" emotion="bad" >dictator</name></phenomenon>

Se trata de una frase sujetiva y "dictador" es un nombre que hace referencia a Bush, lo expresa el autor del post, tiene una polaridad de grado alto y expresa un sentimiento malo, de algo malo.

Phenomenon	Type: collocation, saying, slang, title, and rhetoric.

Con este elemento se anota la frase entera o parte de ella.

Entera cuando encontramos Title or Rethoric. El primero se refiere a un título y rethoric a una frase simple.

Collocation se refiere a una expresión fija, privativa de una lengua, cuyo significado no se deduce de las palabras que la forman *Ej. Troche y moche.*

Slang: son todas aquellas expresiones vulgares, extremadamente coloquiales.

Saying: son las frases hechas cuyo significado no depende del significado literal de cada una de las palabras que lo componen *Ej. Coger el toro por los cuernos.*

Emotions	Accept, anger, anticipation, anxiety, appreciation, bad, bewilderment, comfort, compassion, confidence, consternation, correct, criticism, disappointment discomfort, disgust, despondency, depression, envy, enmity, excuse, force, fear, fright, good, grief, guilt, greed, hatred, hope, irony, interesting, important, incorrect…

Reader Interpretation	Confidence, comment, level, emotion, phenomenon, polarity, source and target.

Con este elemento queremos anotar casos en los cuales lo que leo me suscita algo como lector.

Author Interpretation	Confidence, comment, level, emotion, phenomenon, polarity, source and target.

El objetivo de este elemento es lo de evidenciar algo que nos da información sobre la esfera personal del autor (religión, orientación política, etc).

Las emociones quizás sean la cosa más compleja de anotar debido a varias cosas como el estado de ánimo del anotador, etc.

También por esta razón ponemos una elección de tres.

Intenta que tu estado de ánimo no te afecte mucho y ponte en lugar de quién escribe y ten en cuenta el tema del que estamos hablando.

Durante los últimos años la **Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)**, en su constante interés en la actualización y divulgación de la investigación en este campo, ha convocado la Edición Anual de los premios SEPLN a la Investigación en Procesamiento del Lenguaje Natural.

Este concurso admite trabajos monográficos de investigación de extensión variable que no hayan sido publicados con anterioridad, y escritos por un miembro de la SEPLN.

Siguiendo con esta dinámica, esta publicación recoge la decimocuarta monografía correspondiente a la XIV Edición de los mencionados premios SEPLN a la Investigación en Procesamiento del Lenguaje Natural.

sociedad española para el Procesamiento del Lenguaje Natural

ISBN 978-84-608-1977-6

www.sepln.org