



## PROJECT VISION AND OBJECTIVES

EuroSentiment aims at creating a **shared pool of shared language resources for fostering sentiment analysis**, accessible by means of well-defined models and frameworks that leverage the promotion of SMEs in the emerging market of Sentiment Analysis products and services. The data pool will cover 6 languages -English, Catalan, German, Italian, Portuguese and Spanish- and will be validated through opinion mining demonstrators in two different domains (i.e. hotel and electronic). The targeted users are B2B including service developers, content providers and language resource owners.

The **specific objectives** of the project are:

- Provide **semantic interoperability** and connectivity between several multilingual sentiment analysis resources available online for the first time. Semantic interoperability is based on domain ontologies linked to a domain labelled WordNet and compatible with existing Linked Data initiative and EmotionML.
- Reduce the cost of aggregating new language resources to the shared resource pool by providing best-practice guidelines and QA procedures based on a **publicly available multilingual sentiment analysis corpus** on two different domains.
- Provide a **self-sustainable and profitable framework** for language resource sharing based on a **community governance model**, which offers contributors unwilling to grant free access the possibility of exploiting commercially the resources they provide.
- Demonstrate the impact of the developed pool by providing public access to a **multilingual demonstrator in the media domain**, which will show how the different resources can provide high quality results working with specialised language resources, integrate semantically their results and exploit these multilingual results with a **semantic front-end**.

## KEY INNOVATION

Sentiment analysis can foster the development of new products and services. Nevertheless, the main obstacle to develop these services is the **difficulty in accessing to multilingual language resources for sentiment analysis**. The main barriers we have identified are:

- The developed **language resources remain scattered** and restricted to their customers.
- **Lack of agreed language resource schemas** and available multilingual language resources for sentiment analysis.
- **Atomised sentiment analysis projects** resulting in reduced language resources visibility, accessibility and interoperability.

In this respect, **EuroSentiment will innovate providing a domain-oriented shared language resource based on WordNetDomains and aligned with WordNet Affect**. The pool will be multi-lingual and based on linked data, providing a self-sustainable and profitable framework for language resource sharing.

### *Project title*

Language Resource Pool for Sentiment Analysis in European Languages

*Grant Agreement no:* 296277

*Call:* FP7-ICT-2011-SME-DCL

*Starting date:* 01/09/2012

*Duration:* 24 months

### *Coordinator*

Paradigma Tecnológico (Spain)

### *Contact Person*

Daniel Molina

daniel.molina@paradigmatecnologico.com

### *Consortium*

Fondazione Bruno Kessler (Italy)

Expert System (Italy)

Sindice (Ireland)

Universidad Politécnica de Madrid (Spain)

National University of Ireland, Galway (Ireland)

*Total budget:* 2.560.179 Euros

*Total EC contribution:* 1.930.000 Euros

### *Project Website*

[www.eurosentiment.eu](http://www.eurosentiment.eu)



## TECHNICAL APPROACH

EuroSentiment proposes a twofold approach. First, since Opinion mining is strongly domain dependent, a semantic network relating entities and features will provide a **uniform semantic interface for users**, and entities will have uniform URI naming according to Linked Data conventions. In this way, different sentiment analysers (even in different languages) will return semantic triples that can be easily combined and even more, queried in a distributed fashion. Second, since WordNet is the most commonly used language resource, **WordNet-Domains will be extended for Sentiment analysis**, overcoming the limitations of SentiWordNet which is not domain driven. This domain orientation comes also from the analysis of the language resources provided by the consortium which are already domain-oriented. Once the resources are normalised, semantic and service access will be provided following a licence model. In order to ensure not only the sustainability of the language resource pool but its business orientation, a community governance model will be defined and applied, following the successful community approach of (profitable) open source communities.

**The Language Resource Pool will be built upon a Semantic and Service layer**, allowing business users and content providers to access the shared pool through **semantic Linked-data-based and REST-based requests** following a license model. Service layer will provide **LRP protocol level access mechanisms** in a seamless way, so third parties will not have to deal with issues such as interoperability, multilingualism, internal LRP transactions and secure and IPR functionalities by:

- Processing of datasets with arbitrary transformations in order to derive from source data many possible forms which are useful e.g. for specific reuses.
- Advanced search both at dataset and at specific record level including advanced operators and interactive functionalities  
Full set of APIs to programmatically interact with the repository
- Social aspects to foster crowdsourcing of comments and feedbacks. This will allow users to comment at high granularity resources and foster their uses e.g. allowing users to leave examples.

Recommendation aspects, algorithms that will suggest resources given other resources or specific user input datasets.

## IMPACT

- Improved **European competitive position** in a **multilingual digital market** through the provision of better products and services to citizens and businesses.
- **Novel forms of partnership** between new programme entrants and established players, reduced development costs and shorter time-to-market, thus stimulating innovation and expanding markets.
- **Result-driven knowledge transfer** between research centres (and their spin-offs) and progressive technology providers (especially SMEs), data brokers/aggregators and content providers.

## TARGET GROUPS AND EXPECTED OUTCOMES

- **Service developers** being SME, large company or end users will be able to use EuroSentiment dataset in order to develop new services by integrating the provided interfaces, based on a commercial license
- **Content providers** can use EuroSentiment dataset in order to increase the value of their contents by aggregating consumers' trends and customers' perceptions
- **Language resource owners** will benefit from having a shared pool where their language resources are valorised thanks to the integration with the rest of resources, resulting in higher visibility and return of investment.

On the supplier side, this language resource pool is targeted at research centres willing to unveil their language resources, while keeping visibility, and SMEs willing to provide complementary language resources to the shared pool and exploiting these resources thanks to its increased visibility which will increment its business opportunities.