SENTIMENT ANALYSIS IN SOCIAL MEDIA CONTENTS USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Author: Tamara Álvarez López
Advisors: Enrique Costa Montenegro, Milagros Fernández Gavilanes
Affiliation: University of Vigo

AtlantTIC

UniversidadeVigo

MOTIVATION

- Growth of social networks and their users.
- New way of expression.
- Need of extracting useful information from social networks and other online platforms.
- Very valuable information for companies, who need to analyse it.

- Great amount of information available online.
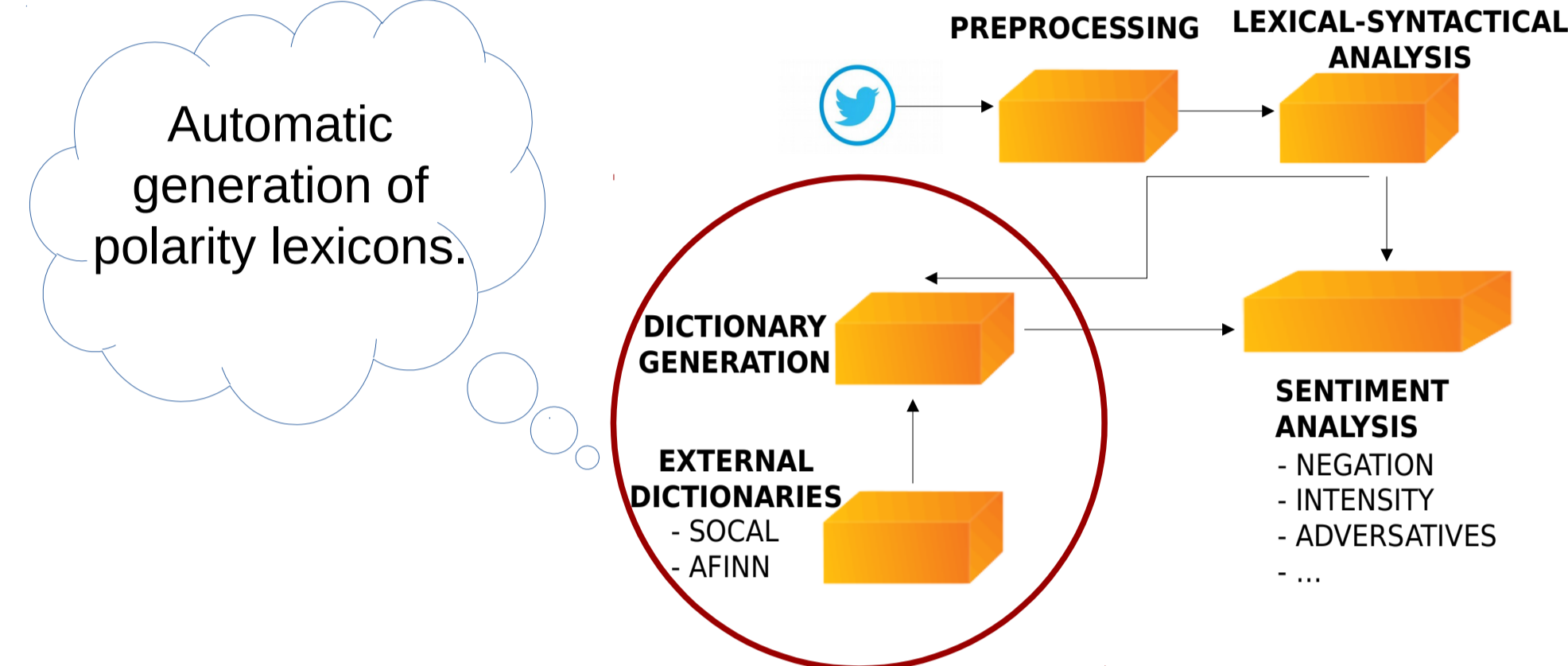- Difficult to analyse.

- Natural Language Processing (NLP) techniques.
- Applied in fields like: **Sentiment Analysis (SA)**, **Aspect Extraction**, Text Generation, Entity Recognition, etc.

THESIS OBJECTIVES

- Build an automatic sentence-level SA system competitive to other existing ones [1].
- Develop a fine-grained SA system, by means of aspect detection.
- Application and adaptation to different datasets and contexts → Analysis and comparison.
- Implementation of these systems for both English and Spanish languages.

PREVIOUS RESEARCH

- Study of state of the art in NLP techniques applied to SA.
- First system implementation of a complete SA system over Twitter platform.
- Participation in international competitions: SemEval-2015 (English) and TASS-2015 (Spanish).
- Application for detecting jihadist ideologies in Twitter.

Automatic generation of polarity lexicons.

PREPROCESSING — LEXICAL-SYNTACTICAL ANALYSIS

DICTIONARY GENERATION

EXTERNAL DICTIONARIES
- SOCAL
- AFINN

SENTIMENT ANALYSIS
- NEGATION
- INTENSITY
- ADVERSATIVES
- ...

RESEARCH PLAN

- State of the art for Aspect Based Sentiment Analysis (ABSA) → Fine-grained SA [2].
- ABSA system implementation:

2015-2016
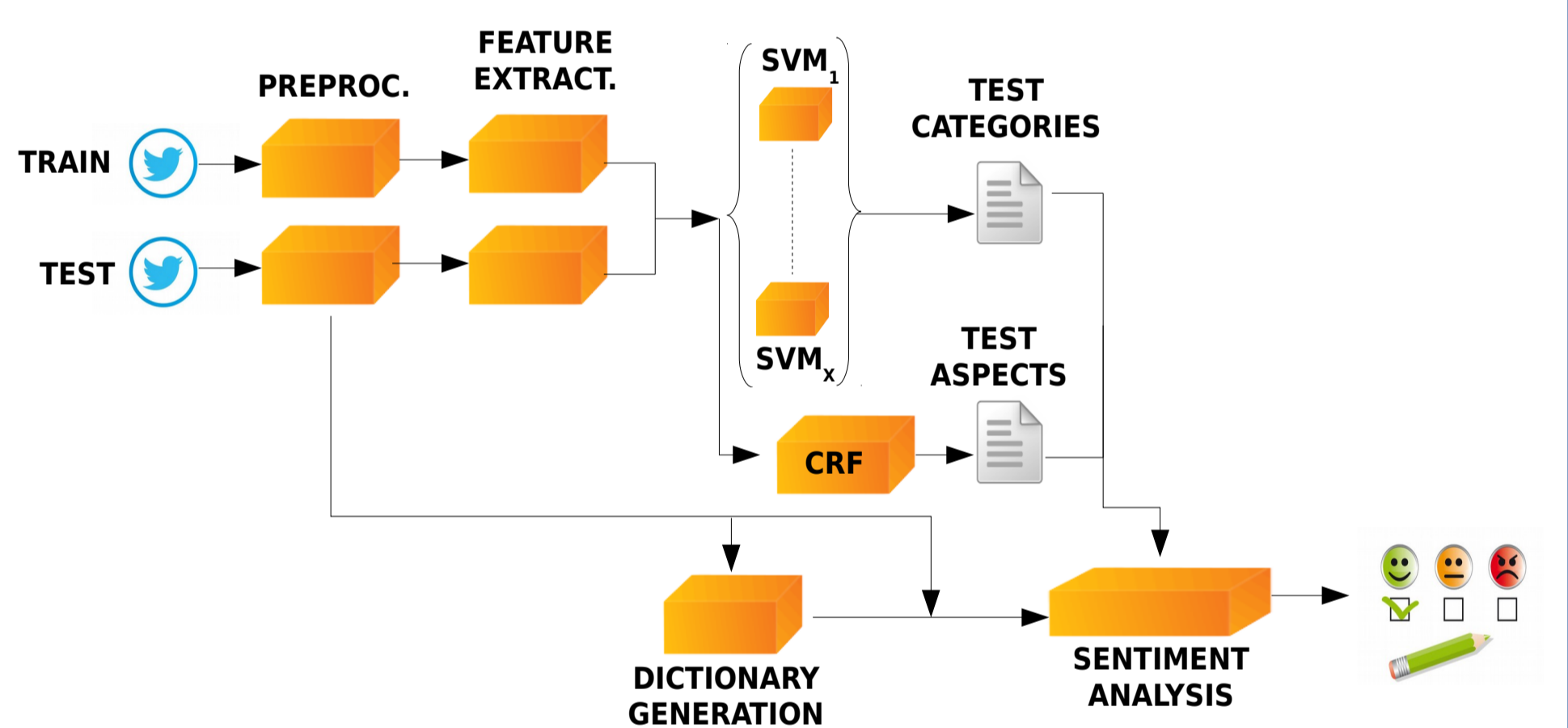
- Article publication in a scientific journal [3].
- Application of the ABSA system to different datasets and both English and Spanish languages.
- Attendance and presentation of the work in international conferences – SemEval 2016 [4, 5] and TASS 2016 [6].

2016-2017

Aix-Marseille Univ. (Oct'16-Dec'16)

- Analysis of book reviews (OpenEdition/Amazon) structure and distinctive features.
- ABSA system adaptation to French book reviews.
- Development of a new dataset for book reviews, annotated with aspects, categories and sentiment [7]:

April'17–Sept'17

```
-<sentence id="000_0007175000_79:6">
  -<text>
    The characters are likeable, the plot is complicated yet compelling and the writing superb
  </text>
  -<Opinions>
    <Opinion category="CONTENT#CHARACTERS" occurrence="1" polarity="positive" target="characters"/>
    <Opinion category="CONTENT#PLOT" occurrence="1" polarity="positive" target="plot"/>
    <Opinion category="BOOK#QUALITY" occurrence="1" polarity="positive" target="writing"/>
  </Opinions>
</sentence>
```

Aix-Marseille Univ. (Sept'17-Jun'18)

- Evaluation of the new dataset.
- ABSA over domains: evaluation and analysis [8].
- Informal reviews about fiction books vs formal reviews about scholarly books [9]
  - Lexical and semantic analysis
  - Review classification system
  - Designing new schemas

Oct'17-March'18

- Apply unsupervised approaches for ABSA in book reviews.
- Integration of ABSA in recommendation systems.
- Wrapping up, conclusions and writing.

April'18-Dec'18

RESULTS AND DISCUSSIONS

- **New book reviews dataset for ABSA** [7]:

Category	P	N	NEU	Total	%
C#CHARACTERS	167	155	1022	1344	35,06
C#PLOT	150	135	782	1067	27,84
B#GENERAL	373	73	13	459	11,97
B#AUTHOR	143	21	211	375	9,78
C#GENRE	65	8	82	155	4,04
B#TITLE	7	1	140	148	3,86
B#AUDIENCE	95	21	20	136	3,55
B#QUALITY	59	7	4	70	1,83
B#STRUCTURE	22	14	8	44	1,15
B#PERIOD	-	1	10	11	0,29
C#PERIOD	-	-	11	11	0,29
B#LENGTH	-	4	5	9	0,23
B#PRICE	2	-	2	4	0,10

*Available online: http://www.gti.uvigo.es/index.php/en/book-reviews-annotated-dataset-for-aspect-based-sentiment-analysis

- **ABSA across domains** [8]:

Dataset	Aspect Extraction			Category Detection		
	P	R	F1	P	R	F1
Restaurants	0,69	0,64	0,66	0,72	0,64	0,68
Electronics	0,64	0,31	0,42	-	-	-
Laptops	-	-	-	0,66	0,40	0,50
Books	0,59	0,14	0,22	0,53	0,30	0,38

- **Informal (Amazon) vs. Expert book reviews (OpenEdition)** [9]:
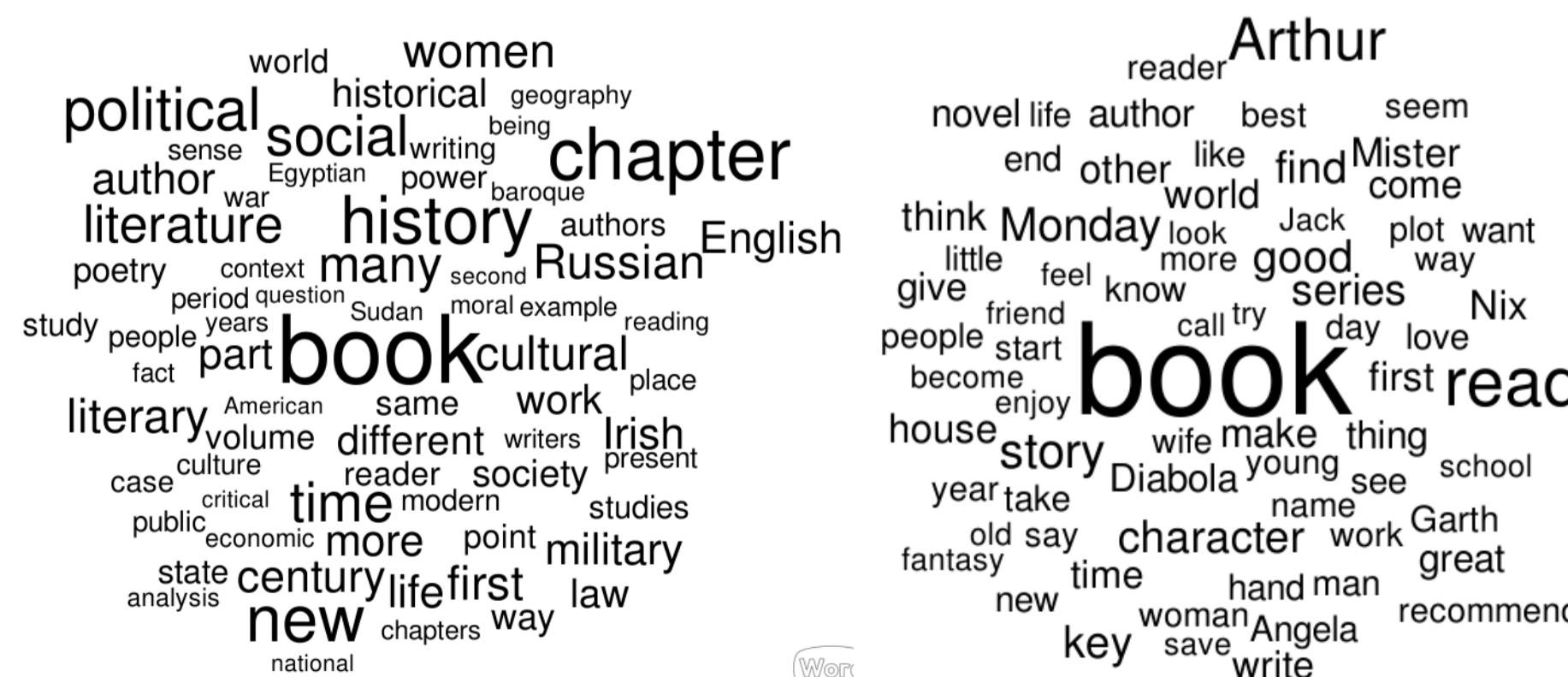
- Review classification system (SVM):

Reviews	Sentence-level			Review-level		
	P	R	F1	P	R	F1
Amazon	0,93	0,88	0,90	1,0	0,99	0,99
OpenEdition	0,73	0,83	0,77	0,97	1,0	0,98
Avg.	0,83	0,86	0,84	0,99	1,0	0,99

- Different annotation schemas.
  - Categories for expert reviews: judgement, organization, argumentation, writing style, technical features, etc.

NEXT YEAR PLANNING

- Unsupervised approaches for ABSA – Try to improve the aspect extraction and category detection tasks by means of unsupervised approaches.
- Integration of ABSA with recommender systems: Do the aspect detection improve the recommendation??
- Wrapping up, conclusions and writing.

REFERENCES

[1] Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. Annual Review of Linguistics, 2, 325-347.

[2] Marcheggiani D.,Täckström O., Esuli A., & Fabrizio Sebastiani. 2014. Hierarchical multilabel conditional random fields for aspect-oriented opinion mining. In Advances in Information Retrieval, pages 273–285. Springer.

[3] Fernández-Gavilanes, M., **Álvarez-López, T.**, Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. Expert Systems with Applications, 58, 57-75.

[4] Juncal-Martínez, J., **Álvarez-López, T**., Fernández-Gavilanes, M., Costa-Montenegro, E., González-Castaño, F. J. GTI: Training a Naive Bayes Classifier using Features of an Unsupervised System. SemEval-2016.

[5] **Álvarez-López, T.**, Juncal-Martínez, J., Fernández-Gavilanes, M., Costa-Montenegro, E., González-Castaño, F. J. GTI: SVM and CRF for Aspect Detection and Unsupervised Aspect-Based SentimentAnalysis. SemEval-2016.

[6] **Álvarez-López, T.**, Fernández-Gavilanes, M., García-Méndez, S., Juncal-Martínez, J., & González-Castaño, F. J. (2016). GTI at TASS 2016: Supervised Approach for Aspect Based Sentiment Analysis in Twitter. Comité organizador, 53.

[7] **Álvarez-López, T.**, Fernández-Gavilanes, M., Costa-Montenegro, E., Juncal-Martínez, J., García-Méndez, S., Bellot, P.: A Book Reviews Dataset for Aspect Based Sentiment Analysis. In: 8th Language & Technology Conference. pp.49–53. (2017)

[8] **Álvarez-López, T.**, Fernández-Gavilanes, M., Costa-Montenegro, E., Bellot, P. A Proposal for Book Oriented Aspect Based Sentiment Analysis: Comparison Over Domains. In: 23rd International Conference on Natural Language & Information Systems. (2018)

[9] **Álvarez-López, T.**, Bellot, P., Fernández-Gavilanes, M., Costa-Montenegro, E. From Genre Classification to Aspect Extraction: New Annotation Schemas for Book Reviews. Submitted to: IEEE/WIC/ACM International Conference on Web Intelligence, 2018.