On the Identification of Emotions and Authors' Gender in Facebook Comments on the Basis of their Writing Style

Francisco Rangel[1,2] and Paolo Rosso[2]

[1] Autoritas Consulting, C/ Lorenzo Solano Tendero 7, 28043 Madrid, Spain,
francisco.rangel@autoritas.es,
http://www.kicorangel.com

[2] Natural Language Engineering Lab, Universitat Politècnica de València,
Camino de Vera, S/N, 46011 Valencia, Spain
prosso@dsic.upv.es,
http://users.dsic.upv.es/~prosso

Abstract. In this paper, we propose a method for automatic identifying emotions in written texts in social media with high proliferation such as Facebook. For that task we try to model the way people use the language to express themselves, and also use this model for identifying the gender of the authors. We focused on Spanish due to the lack of studies and resources in that language.

Keywords: affective processing, emotion identification, gender identification, Spanish Facebook

1 Introduction

World is rapidly changing, social media are growing day by day and, in a sense, customers are becoming users looking for new experiences. The emotional aspect of the life is acquiring a growing importance and with it, the need of automatically processing the affective content of such social media, in order to know what users want and need.

The potentiality offered by social networking is undoubtful from lots of perspectives like marketing, security or health. But it is also undoubtful that the information users include about themselves, if they include it, may lack credibility. Age, gender, affiliation, likes... many users invent them, use linguistic devices such as sarcasm and irony, or simply, they have never reported them. Getting to know the demographic and psychosocial profile of such users is an opportunity for organizations and companies, and a challenge for natural language processing technologies, due to the fact that the unique certainty we can have is what we can obtain from what the users write and share in such social media.

Studies like [11] link the use of the language with some traits like the gender of the author, but the vast majority of such investigations are limited to English and traditional media, which should be extended to the (different?) use of language in the new technologies and media, and to other languages such as Spanish.

This investigation presents a method for automatically identifying emotions in Facebook and in Spanish language, taking into account another dimension of personality of growing interest in the scientific community: the author gender[3]. The main objective is to establish a common framework and a series of resources to investigate the relationship among demographics and emotions, and in the future with personality traits, in social media.

In Section 2 we describe the related work on resources and affective processing. In Section 3 we present our proposal for modeling the style of the language to automatically identify emotions and gender with a machine learning algorithm. In Section 4 the methodology is described and the results are presented in Section 5. In Section 6 we present the conclusions and future work to achieve our objective.

2 Related Work

Classification of related work can be done from two perspectives: the generation of affective resources and the affective processing methods.

2.1 Generation of affective resources

Dictionaries which include the affective dimension are the most common resources, being pioneers the Lasswell Value Dictionary [14], where each word is annotated with the existence of dimensions such as wealth, power, rectitude, respect, enlightenment, skill, affection or wellbeing, and the General Inquirer [29], where each word is annotated with the existence of dimensions such as active, passive, strong, weak, pleasure, pain, feeling, arousal, virtue, vice, overstated or understated. Both dictionaries use binary tags without considering the degree of occurrence.

Like the previous dictionaries, based on obtaining the existence of certain emotional dimensions, the Clairvoyance Affect Lexicon [9] labels categories such as anger, joy and fear, and also adds some dimensions as centrality and strength, in order to complete the relationship between the word and its affective class.

In the line of identifying emotional dimensions, the Dictionary of Affect in Language (DAL) [34] consists of a set of 8,842 words labeled by their activation and ability to imagine the emotion; or the Affective Norms for English Words (ANEW) [1] whose objective is to have measured the maximum number of English words in terms of activation, evaluation and control. On the other hand, Strapparava and Valitutti [30] developed WordNetAffect as a subdomain of Wordnet, where each word is labeled according to its emotional category, evaluation and activation.

In [16] the authors used Mechanical Turk[4] for creating a high-quality, moderate-size, emotion lexicon of about 2,000 terms. They showed how terms related to

[3] `http://www.uni-weimar.de/medien/webis/research/events/pan-13/pan13-web/author-profiling.html`

[4] https://www.mturk.com/mturk/welcome

emotions are among the most common unigrams and bigrams, and also identified which emotions tend to be evoked simultaneously by the same term. They used automatically generated word choices to detect and reject erroneous annotations.

Linguistic Inquiry and Word Count (LIWC) is a software for obtaining features from text. It was developed by Pennebaker et al. [22] Through using text analysis that provides up to 70 dimensions such as the degree of positive and negative emotions, self-references, causal words, and so on.

There are hardly any resources in Spanish language, highlighting the Spanish Adaptation of ANEW developed by [27]. With the help of 720 participants, they labeled the translation of 1,034 words of ANEW in the dimensions of polarity, activation and control. Moreover, the Spanish Emotion Lexicon (SEL) [28] consists on 2,036 words associated with the measure of "Probability Factor of Affective use" (PFA) related to one of the six basic emotions of Ekman[5]: joy, disgust, anger, fear, sadness, surprise. SEL defines four possible degrees of relationship with each emotion (null, low, medium, high). 19 annotators indicated these values for each word and the PFA was calculated as an average of the percentages assigned to each degree.

Although it is not a resource, we must cite the investigation carried out in [20]. They studied the necessity, or not, of using affective dictionaries in the emotion analysis, trying to answer questions as if they improve the identification or if they could be replaced by general purpose dictionaries.

2.2 Affective processing methods

Automatic processing of affectivity has been focused mainly on sentiment analysis, where one of the dimensions of the emotions is investigated: evaluation (or polarity). However, there are a series of methods oriented to classify documents in the corresponding emotional category, usually based on the six basic emotions of Ekman.

We highlight three methods presented in SemEval 2007, where the task of identifying emotions was included for 1,000 news headlines. UPAR7 [2] used the Stanford syntactic parser for identifying what the main topic was speaking about, estimating each word polarity with the help of Senti Wordnet and Wordnet Affect and obtaining incrementally the global classification. UA [12] utilized three search engines for searching all the words in the headline combined with each emotion, and then calculating the Pointwise Mutual Information according to the number of returned documents. SWAT [10] was a supervised system based on unigrams and trained with another 1,000 news manually annotated by their authors and which used the Roget thesaurus to expand synonyms and build the features.

In [32] results are presented and compared with five own proposals [31]. WN-AFFECT PRESENCE identified emotions based on the presence of words from WordnetAffect. LSA SINGLE WORD calculated the LSA similitude between each text and each emotion, taking some words like *joy* as representatives of the emotional class. LSA EMOTION SYNSET added Wordnet synonyms and LSA ALL EMOTIONS also included all the annotated words from WordnetAffect, as an emotion containers. NB TRAINED ON BLOGS was based on a Naive Bayes classifier trained on a corpus of blogs. Results are shown in Figure 1.

	Fine	Coarse		
	r	Prec.	Rec.	F1
WN-AFFECT PRESENCE	9.54	**38.28**	1.54	4.00
LSA SINGLE WORD	12.36	9.88	66.72	16.37
LSA EMOTION SYNSET	12.50	9.20	77.71	13.38
LSA ALL EMOTION WORDS	9.06	9.77	**90.22**	**17.57**
NB TRAINED ON BLOGS	10.81	12.04	18.01	13.22
SWAT	25.41	19.46	8.61	11.57
UA	14.15	17.94	11.26	9.51
UPAR7	**28.38**	27.60	5.68	8.71

Fig. 1. Results of SemEval 2007

The first global performance, measured by F1, was obtained by LSA ALL EMOTION WORDS with 17.57%. However, methods presented in SemEval performed better r measure[5].

Other investigations related to the identification of emotions are: [6] based on detecting keywords; [21] based on lexical affinity according to the probability of certain words to be related to certain emotions and [15] based on the OMCS2 knowledge base.

Previous methods obtain features and approaches by analyzing the semantic content of the texts. On the other hand, in [3] authors introduced style features as the identification of imperative sentences, exclamation signs, the use of capital letters or the use of present and future, in order to identify polarity and emotional category.

Trying to unlink the method from the language, English in all the cases seen so far, in [7] is described a modular architecture with semantic disambiguation per language or the use of affective dictionaries as ANEW. This system has also been applied to Spanish.

Following the line of style features and for a language different than English, in [33] authors used substantives, adjectives and verbs with the identification of keywords and types of sentences in Japanese in order to identify emotions.

A step forward to the link between emotions analysis and personality traits was given in [17], sentiment analysis by gender was incorporated. They analyzed three kind of emails: love letters, hate emails and suicide notes.

In Spanish, in [4] a method based on the SEL dictionary is presented, together with annotated short stories. Different machine learning methods are compared, demonstrating an improvement over baseline.

[5] Pearson's Kappa to measure the correlation between the obtained result and the random chance

3 Style-based Identification

The vast majority of investigations on emotions analysis are oriented to obtain representative characteristics of the semantics of the documents, that is, they are focused on the analysis of the content, what can imply overfitting and dependency on the domain, context or thematics.

On the basis of what was already studied for English by authors such as Pennebaker [23], we carried out some experiments to investigate the use of the different morphosyntactic categories for Spanish. The aim was verifying whether their use was different or not, depending on the channel [24] and its related language register. The final goal was using the morphosyntactic categories information for identifying emotions and subsequently age and gender [26].

With the aim of modeling the style of writing we considered readability features as well as the use of emoticons. We used also the Spanish Emotion Lexicon, an affective dictionary specially built for Spanish [28]. All these features are topic-independent. The complete set is described below. Each item is a list of individual features represented by frequencies and combined into a vector space model. We obtained the readability features (frequencies and punctuation marks) and emoticons using regular expressions, whereas the morphosyntactic categories where obtained with the Freeling library[6].

- (F)requencies: Ratio between number of unique words and total number of words; words starting with capital letter; words completely in capital letters; length of the words; number of capital letters and number of words with flooded characters (e.g. Heeeelloooo).
- (P)unctuation marks: Frequency of use of dots; commas; colon; semicolon; exclamations; question marks and quotes.
- Grammatical (C)ategories or Part-of-speech: Frequency of use of each grammatical category; number and person of verbs and pronouns; mode of verb; number of occurrences of proper nouns (NER) and non-dictionary words (words not found in dictionary).
- (E)moticons[7]: Ratio between the number of emoticons and the total number of words; number of the different types of emoticons representing emotions: joy, sadness, disgust, angry, surprise, derision and dumb.
- (SEL) Spanish Emotion Lexicon: We obtained the Probability Factor of Affective use value from the SEL dictionary for each lemma of each word. If the lemma does not have an entry in the dictionary, we look for its synonyms. We add all the values for each emotion, building one feature for each emotion.

We do not use any content/context dependent features in order to obtain total independence from the topics.

[6] http://nlp.lsi.upc.edu/freeling/

[7] http://es.wikipedia.org/wiki/Anexo:Lista_de_Emoticonos

4 Methodology

In the following sections, we describe the raw dataset, the labeling process and the machine learning approaches.

4.1 Dataset

We focused on social media since we are interested in everyday language and how it reflects basic social and personality processes. Due to that, we chose Facebook comments in Spanish language as the source of data for our experiments. Facebook comments have the freedom of expression (and style) without editorial guidelines unlike traditional media like newsletters and the spontaneity in the use of language unlike blogs. Facebook is massively used by people and the expected affectivity in such media is very high. Facebook also allows us to obtain demographics such as gender, unlike similar media like Twitter, so that we will be able to link this task with groundbreaking tasks such as Author Profiling at PAN 2013 [25].

We also chose Spanish because although its high penetration in Internet[8], the amount of available resources is still low especially if compared with English. We selected three thematics, with high volume of participation[9], and susceptible of emotional comments: politics, football and public figures. We balanced the data by theme and gender.

Neither selection nor cleaning has been done except for language filtering and for ensuring that comments have some text (not only links). Information about the dataset is shown in Table 4.1.

Theme	Gender	Comments
Politics	Male/Female	200/200
Football	Male/Female	200/200
Public People	Male/Female	200/200

Table 1. Dataset of Facebook comments in Spanish

4.2 Manual labeling

Three independent annotators labeled 1,200 documents with the six basic emotions of the Ekman's theory. Annotators were provided with the information of Figure 2 that was obtained by Greenberg [8] on the basis of psychological relationships of emotional states with the six basic emotions of Ekman. It is remarkable that some secondary emotions are shared by more than one primary

[8] http://eldiae.es/wp- content/uploads/2012/07/2012_el_espanol_en_el_mundo.pdf
[9] http://www.pewglobal.org/files/2012/12/Pew-Global-Attitudes-Project-Technology-Report-FINAL-December-12-2012.pdf

emotion; for example, *indignation* (indignación) is shared by *anger* and *disgust*, and *fascination* (fascinación) is shared by *joy* and *surprise*. This issue hinders the unique identification of such basic emotions, as it was evidenced in [19]. Besides, the identification of multiple emotions and the absence of any has been allowed.

ALEGRÍA	ENFADO	MIEDO	REPULSIÓN	SORPRESA	TRISTEZA
Agradecido	Agresivo	Acomplejado	Aborrecimiento	Extrañeza	Abatido
Alegre	Colérico	Alarmado	Desagrado	Sobresalto	Agobiado
Animado	Crispado	Angustiado	Grima	Susto	Apenado
Calmado	Descontento	Ansioso	Repulsión	Consternación	Confuso
Confiado	Enfadado	Atemorizado	Antipatía	Pasmo	Decepcionado
Contento	Enojado	Aterrado	Aversión	Desconcierto	Deprimido
Dichoso	Excitado	Avergonzado	Repugnancia	Estupor	Desalentado
Encantado	Fastidiado	Confuso	Disgusto	Asombro	Desanimado
Entusiasmado	Furioso	Desesperado	Repudia	Fascinación	Desdichado
Eufórica	Insatisfecho	Desorientado	Repulsa	Admiración	Desmoralizado
Esperanzado	irascible	Horrorizado	Odio	Confusión	Frustrado
Feliz	Malhumorado	Inquieto	Manía	Chasco	Nostálgico
Gozoso	Molesto	inseguro	Rabia	Impresión	Soledad
Satisfecho	Nervioso	Intranquilo	Animadversión	Exclamación	Triste
Tranquilo	Rabioso	Pánico	Nauseabundo	Conmoción	Infeliz
Complacido	Tenso	Preocupado	Indignación	Estupefacción	Desconsolado
Libre	Violento	Temeroso	Enfado		Afligido
Fascinado	Irritado	Tenso	Desprecio		Amargado
Seguro	Indignado	Indeciso	Distanciamiento		Impotente
		Impotencia			

Fig. 2. Secondary emotions related to the six basic emotions

We calculated the inter-annotator agreement with the Kappa_DS method [4], which allows multiple annotators (three in our case: A1, A2 and A3) and multinomial variables (six not mutually exclusive, the six basic emotions). We show results in Table 4.2.

	A1	A2	A3	REST
A1	-	0.0587	0.2738	0.1662
A2	0.0587	-	0.1042	0.0814
A3	0.2738	0.1042	-	0.1890
TOT		0.1455		

Table 2. Kappa DS: Inter-annotators agreement

The average value for Kappa, equal to 0.1455, shows a low index of agreement according to the recommendations of [13]. But, as it is shown by [4], we have to bear in mind the amount of variables intervening in the evaluation for the right interpretation of such index, that makes it not comparable to their original recommendation. We also grouped the nearest emotions, those which share secondary emotions, as we highlighted in figure 2: *joy / surprise* and *anger / disgust*. Results are shown in table 4.2. In this case, Kappa shows a higher value for the agreement, what suggests us that we have to bear in mind such discordance among annotators when assessing the results, especially with respect to *joy/surprise* and *anger/disgust*.

	A1	A2	A3	REST
A1	-	0.6618	0.5656	0.6137
A2	0.6618	-	0.5773	0.6196
A3	0.5656	0.5773	-	0.5715
TOT		0.6016		

Table 3. Kappa DS: Inter-annotators agreement with grouped emotions: joy/surprise and anger/disgust

The final selection of emotional tags for each document has been based on the concordance of at least two of three annotators. Figures are shown in table 4.2. The low number of documents labeled with the *fear* category did not allow us to perform experiments with this emotion.

	TOTAL	%
Joy	338	28.17
Anger	151	12.58
Fear	3	0.25
Disgust	129	10.75
Surprise	390	32.50
Sadness	76	6.33
Neutral	262	21.83

Table 4. Documents per emotion

4.3 Learning and evaluation

A binary classifier has been proposed for each emotion with the aim to determine whether a given text contains such emotion. Each classifier was trained with the labeled examples for its emotion as a positive samples, and with the rest as negative samples. The evaluation method was 10-fold cross validation.

We carried out two different evaluations, as in SemEval 2007, the first one based on Pearson's Kappa to measure the correlation between the obtained result and the random chance, and the second one based on precision, recall and F1.

We tested four learning algorithms implemented in Weka[10] with their default parameters: J48 trees, Naive Bayes, Bayes Net and Support Vector Machines.

5 Experimental Results

The objective was to obtain an automatic method for identifying emotions in Facebook comments in Spanish language, attempting the maximum independence from the thematics and trying to link the emotional information with other personal dimensions such as gender. Our starting hypothesis was the style features described in Section 3.

5.1 Emotions identification

Style features were enriched with the information of the SEL affective dictionary, allowing the construction of an adequate and competitive model for identifying emotions. We retrieved the described features in Section 3 for training each classifier and results are shown in Table 5.1. The best results obtained according to each individual metric are marked in bold.

We can appreciate that different methods have different strengths. J48 has the highest precision in most of the cases at the cost of low recall. In similar way, BayesNet obtains better recall but reducing precision, although it is the best method in terms of F1. In terms of r, values seems to be less correlated with the method. However in most cases the best methods are the statistical ones (Naive Bayes and BayesNet).

With respect to emotions, results for *joy* and *surprise* are the highest, mainly for F1 measure, which correlates with the size of the training dataset. Results for *sadness* are lower than the rest, probably due to the fact that the total number of documents labeled for this emotion is much lower than for the rest (see Table 4.2). This fact implies some dependency of the machine learning approach with the number of samples used in the training and it must be studied further by the parametrization of the methods.

It is necessary to remark the lower results of the SVM method in some experiments, due to the imbalance of the class and the small amount of training data, being this method more sensible to both factors. This could be improved by tunning its configuration parameters.

The proposed features, all independent from thematics and mainly based on the style of writing, achieved competitive results compared to the state-of-the-art approaches in social media in the Spanish language.

5.2 Gender Identification

In order to link emotions with demographics, we carried out an experiment consisting in using the features we used for identifying emotions, to learn a new

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

Emotion	Algorithm	r	Prec.	Rec.	F1
Joy	J48	27.1	49.7	43.2	46.2
	NB	**27.9**	45.4	56.8	50.5
	BN	25.6	40.9	**73.7**	**52.6**
	SVM	24.9	**56.9**	30.5	39.7
Anger	J48	16.6	**32.3**	19.9	24.6
	NB	**22.6**	25.9	60.3	**36.3**
	BN	22.2	25.6	**60.9**	36.0
	SVM	10.8	25.8	15.2	19.2
Disgust	J48	21.7	**36.1**	23.3	28.3
	NB	15.7	19.7	55.8	29.1
	BN	**24.9**	25.5	**64.3**	**36.5**
	SVM	6.2	11.7	5.4	7.4
Surprise	J48	**25.8**	**50.4**	48.7	49.5
	NB	20.6	42.7	**67.2**	**52.2**
	BN	20.7	43.0	64.6	51.6
	SVM	17.2	49.4	30.5	37.7
Sadness	J48	12.1	**20.0**	14.5	16.8
	NB	6.1	9.8	35.5	15.4
	BN	**16.7**	16.3	**51.3**	**24.7**
	SVM	8.2	17.9	0.92	12.2
Average results					
	J48	20.7	**37.7**	29.9	33.1
	NB	18.6	28.7	55.1	36.7
	BN	**22.0**	30.3	**63.0**	**40.3**
	SVM	13.5	32.3	16.5	23.2

Table 5. Results of identification of basic emotions

model to identify gender of the authors of the Facebook comments. The hypothesis was that proposed features, which describe the authors' style of writing, could be useful for identifying personal dimensions such as gender.

We trained the Support Vector Machine method implemented in Weka. We experimented with different parameters and finally used a Gaussian kernel with g=0.01 and c=3,500. Results for gender identification are shown in Table 5.2.

Gender	Acc	r
Male / Female	59.0	18.0

Table 6. Results for gender identification in accuracy, Pearson's coefficient, Precision, Recall and F1

An r value equal to 18.0 means that the classifier works over the random chance and suggests that style features provide some kind of information about the gender, as [11] showed for English. An accuracy value of 59.0 allows us to think that our method is competitive for such task in comparison with approaches presented in the Author Profiling task at PAN 2013. We plan to perform further experiments with the dataset provided for this task.

The fact that features used for identifying emotions allowed us to identify gender with a good accuracy, suggests that there is a certain correlation between the use of emotions and the gender of the authors.

6 Conclusions and Future Work

We have built a dataset of Facebook comments for Spanish, manually labeled it with six basic emotions from Ekman's theory and carried out a Kappa-DS analysis of concordance.

We have proposed a method for automatically identifying emotions based on a combination of stylistic features with the use of the SEL affective dictionary, obtaining competitive results. We have also verified the difficulty to label, even for a human, among primary emotions which share secondary emotions, as is the case of *joy* and *surprise*, or *anger* and *disgust*.

Finally, we have employed the proposed approach for identifying authors' gender, showing that style features provide certain information valuable for such task.

As a future work we plan to investigate further what are the most relevant features for identifying emotions and gender, and their possible relationship. We will investigate the identification of combined emotions (*joy* and *anger* will be joined respectively with *surprise* and *disgust*), in order to verify if the current results are due to the difficulty of discriminating such emotions. We also plan to carry out with the PAN-AP13 dataset for the identification of gender and age. For that, we will include the detected emotions as new features in order to investigate the relationship between emotions and demographics. Finally, we aim at introducing some more features trying to obtain a better description of the way people use language (e.g. collocations) and, therefore, analyze discourse in depth).

ACKNOWLEDGEMENTS

The work of the first author was partially funded by Autoritas Consulting SA and by Ministerio de Economía de España under grant ECOPORTUNITY IPT-2012-1220-430000. The work of the second author was carried out in the framework of the WIQ-EI IRSES project (Grant No. 269180) within the FP 7 Marie Curie, the DIANA APPLICATIONS Finding Hidden Knowledge in Texts: Applications (TIN2012-38603-C02-01) project and the VLC/CAMPUS Microcluster on Multimodal Interaction in Intelligent Systems.

References

1. Bradley, M., Lang, P.: Affective norms for English words (ANEW): Instruction manual and affective ratings. Gainesville: Center for Research in Psychophysiology, University of Florida (1999)
2. Chaumartin, F. Upar7: A knowledge-based system for headline sentiment tagging. In Proceedings of SemEval2007, Prague, Czech Republic (2007)
3. Dhaliwal, K., Gillies, M., O'connor, J., Oldroyd, A., Robertson, D., Zhang, L.: Facilitating online role-play using emotionally expressive characters. Artificial and Ambient Intelligence, Proceedings of the AISB Annual Convention, 179-186 (2007)
4. Díaz Rangel, I.: Detección de afectividad en texto en español basada en el contexto lingístico para síntesis de voz. Tesis Doctoral. Instituto Politécnico Nacional. México (2013)
5. Ekman, P.: Universals and cultural differences in facial expressions of emotion. Symposium on Motivation, Nebraska, 207-283 (1972)
6. Elliot, C.: The affective reasoner: A process model of emotions in a multi-agent system. Northwestern University: Tesis doctoral, The Institute for Learning Sciences, Northwestern University (1992)
7. García, D., Alías, F.: Emotion identification from text using semantic disambiguation. Procesamiento del Lenguaje Natural. (50), 75-82 (2008)
8. Greenberg, L. Emociones: Una guía interna. Bilbao: Descleé De Brouwer (2000)
9. Huettner, A., Subasic, P.: Fuzzy Typing for Document Management. ACL 2000. Hong Kong (2000)
10. Katz, P., Singleton, M., Wicentowski, R.: Swat-mp:the semeval-2007 systems for task 5 and task 14. In Proceedings of SemEval-2007, Prague, Czech Republic (2007)
11. Koppel, M., Argamon, S., Shimoni, A.: Automatically categorizing written texts by author gender. Literay and Linguistic Computing 17 (4), 401-412 (2003)
12. Kozareva, Z., Navarro, B., Vazquez, S., Montoyo., A.: Ua-zbsa: A headline emotion classification through web information. In Proceedings of SemEval-2007, Prague, Czech Republic (2007)
13. Landis, R., Koch, G.: The measurement of observer agreement for categorical data. Biometrics(35), 159-174 (1977)
14. Lasswell, H., Namenwirth, J.: The Laswell Value Dictionary. Yale University Press. New Haven (1969)
15. Liu, H., Lieberman, H., Selker, T.: Automatic affective feedback in an email browser. MIT Media Lab Software Agents Group Technical Report (2002)
16. Mohammad, S., Turney, P.: Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Association for Computational Linguistics, Los Angeles, CA June (2010)

17. Mohammad, S.M., Yang, T.: Tracking sentiment in mail: how gender differ on emotional axes. Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. Portland, Oregon, 70-79 (2011)
18. Oberlander, J., Gill, A. J.: Language with character: A stratified corpus comparison of individual differences in e-mail communication. Discourse Processes (42), 239-270 (2006)
19. Ortony, A., Turner, T.: What's basic about basic emotions? Psychological Review (97), 315-331 (1990)
20. Osherenko, A., Andr, E.: Lexical Affect Sensing: Are Affect Dictionaries Necessary to Analyze Affect? Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, ACII '07, Pages 230 - 241 (2007)
21. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs Up? Sentiment Classification Using Machine Learning Techniques. Conference on Emprirical Methods in Natural Language Processing (2002)
22. Pennebaker, J. W., Booth, R. E., Francis, M. E.: Linguistic inquiry and word count: LIWC2007 - Operator's manual. Austin, TX: LIWC.net (2007)
23. Pennebaker, J. W., Mehl, M. R., Niederhoffer, K.: Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology, (54), 547-577 (2003)
24. Rangel, F., Rosso, P. El Uso del Lenguaje en los Diferentes Canales de Internet. In: Proceedings Comunica 2.0. Gandia, Spain, February 21-22. (2013)
25. Rangel, F., Rosso, P., Koppel, M., Stamatatos, E., Inches, G.: Overview of the Author Profiling Task at PAN 2013. In: Forner P., Navigli R., Tufis D.(Eds.), Notebook Papers of CLEF 2013 LABs and Workshops, CLEF-2013, Valencia, Spain, September 23- 26. (2013)
26. Rangel, F., Rosso, F.: Use of Language and Author Profiling: Identification of Gender and Age. In: 10th International Workshop on Natural Language Processing and Cognitive Sciences NLPCS 2013 CIRM, Marseille, France, October 13-17 (2013)
27. Redondo, J., Fraga, I., Padrón, I., Comesaña, M.: The Spanish adaptation of ANEW (Affective Norms for English Words). Behavior Research Methods (39), 600-605 (2007)
28. Sidorov, G., Miranda, S., Viveros, F., Gelbukh, A., Castro, N., Velásquez, F., Díaz, I., Suárez, S., Treviño, A., Gordon, J.: Empirical Study of Opinion Mining in Spanish Tweets. LNAI 7629-7630 (2012)
29. Stone, P., Dunphy, D., Smith, M.: General Inquirer: A Computer Approach to Content Analysis. M.I.T. Press. Oxford, England (1966)
30. Strapparava, C., Valitutti, A.: Wordnetaffect: an affective extension of wordnet. Proceedings of the 4th International Conference on Language Resources and Evaluation. Lisboa, 1083-1086 (2004)
31. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In Proceedings of the 2008 ACM Symposium on Applied Computing (SAC 2008), 1556-1560 (2008)
32. Strapparava, C. Mihalcea, R.: SemEval- 2007 Task 14: Affective Text. Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) (2008)
33. Sugimoto, F., Yoneyama, M.: A method for classifying emotion of text based on emotional dictionaries for emotional reading. Proceedings of the 24th IASTE D International Multi-Conference Artificial Intelligence and Applications. Innsbruck. 91-96 (2006)
34. Whissell, C. The dictionary of affect in language. Emotion: Theory, research and experience. The measurement emotions, 113- 131 (1989)