Discovering Emotions through the Building of a Linguistic Resource

Manuela Angioni and Franco Tuveri

Center for Research and Scientific Studies in Sardinia, Bld. 1, Piscina Manna, Pula (CA), Italy

Keywords: Linguistic Resources, Natural Language Processing, Sentiment Analysis.

Abstract: Specific linguistic resources, syntactically annotated and distinctive for each language, related to the affective sphere are important in discovering terms or phrases associated with emotions in order to detect expressed emotions. The paper proposes the initial version of a linguistic resource for the Italian language, mapped on WordNet, where each concept, whose meaning falls into the sphere of emotions, is enriched by a category, allowing to better specify the type of emotion expressed by the term, and by a polarity value, whether the emotion is positive or negative. The resource is based on the model of emotions proposed by Robert Plutchik and has been developed, within a national project of Work-School Alternation, in collaboration with some high school students. The work has a twofold value. On one hand, the development of a linguistic resource, on the other the educational and didactic aspect of students' involvement. Working on the analysis of literary texts with the task of elaborating and defining the emotions described, the students, assisted by their teachers and two researchers, had to face with their feelings and talk more freely about their affective states, recognizing the emotions and giving them a name.

1 INTRODUCTION

Emotions constitute a fundamental aspect of people's lives. All the things people do or say reflect their emotions somehow and express it in all the communication ways. The automatic identification of emotions in a collection of textual data is becoming increasingly important, in order to be able to understand in more detail the mood and the sentiments and to better interpret the human experience and the type of communication expressed in the texts.

The building of linguistic resources is part of Computational Linguistics, the discipline that deals with the study of natural language by means of automatic tools, through the definition of algorithms that allow the automatic analysis of texts and the extraction of the meaning expressed in them

The analysis of the texts involves several phases, including syntactic and semantic analysis of the sentences in the text. A further fundamental step to detect emotions in a text is to identify terms or phrases dealing with affective and emotional sphere. To this end, the availability of linguistic resources, syntactically annotated and specific for the language used in the text, is necessary.

The paper proposes the creation of a linguistic resource for the Italian language named

PlutchikWN_ita, based on the model of emotions proposed by Robert Plutchik (Plutchik, 1980), whose terms are related to the emotions and mapped on WordNet (Miller, 1995). Each concept is enriched by a category, that allows to better specify the type of emotion expressed by the term, and by a polarity value. The resource is built on the basis of the experience of the daily discovery of the emotions of some adolescent students and the impact of emotions on their lives.

The work therefore has a twofold value.

On one hand, the development of a linguistic resource that allows a deeper analysis of textual resources, through the recognition of terms related to the emotional sphere and the type of emotion expressed.

On the other hand, the educational and didactic aspect of students' involvement. In the school environment it is in fact essential to promote a culture in which students can learn to discover their sentiments and moods, and teachers and students together can freely talk about feelings. Speaking about emotions, it is important to recognize the feelings that the students have, so that they are able to give them a name. It is thus important to learn students to identify their different moods and feelings educating them in recognize, manage, and express their emotions.

DOI: 10.5220/0008352803510357 In Proceedings of the 15th International Conference on Web Information Systems and Technologies (WEBIST 2019), pages 351-357 ISBN: 978-989-758-386-5

Copyright © 2019 by SCITEPRESS - Science and Technology Publications, Lda. All rights reserved

Discovering Emotions through the Building of a Linguistic Resource

To this aim, in agreement with the teachers, some students were involved in the analysis of some literary texts, forming part of their course of study, with the task of elaborating and defining the feelings and the emotions described and aroused by the reading of the texts. From the dialogue with the teachers and the class emerged in fact the common difficulty among students of similar age in recognizing and naming their emotions, as well as in speaking openly about them.

It was therefore decided to include the creation of the linguistic resource in a broader context, combining the strongly multidisciplinary aspect characteristic of this activity that requires a heterogeneous workgroup and a constant dialogue between professionals with very different skills.

The building of such linguistic resource combines in fact humanistic skills, to address aspects closer to Linguistics or Psychology, with others coming from Communication and Information Technologies (ICT). This made it possible to establish a dialogue on several fronts with the students, involving them in a multidisciplinary project.

The paper describes also the experience conducted in collaboration with the third year students of the "Liceo Classico, Linguistico e delle Scienze Umane B. R. Motzo" of Quartu Sant'Elena, Cagliari. The activities are part of the CRS4 initiatives within a national project of Work-School Alternation, which aims to promote the training of students in the field of digital technologies and scientific research.

The activities took place with a presentation of the work in front of the whole class through the involvement of all the students and the introduction of the topics concerned. A smaller group of three students worked on the realization of the linguistic resource through the skills of two CRS4 researchers, who provided the assistance and the training necessary to carry out the assigned task.

The remainder of the paper is organized as follows: Section 1 introduces the resource and the motivation behind. Section 2 describes the related works. Section 3 presents the building of the resource, with an overview of the Plutchik's Wheel of Emotions, and describes the interface used. Section 4 describes the current state of PlutchikWN_ita, while Section 5 presents the format of the resource. Finally, in Section 6 final remarks and future works directions are presented.

2 RELATED WORK

Several resources about emotions exist for the English language, such as NRC Emotion Lexicon (Mohammad and Turney, 2010), composed by frequent English terms annotated through Amazon Mechanical Turk with respect to eight emotions (e.g. joy, sadness, trust) and positive or negative sentiment. Another resource, WordNet-Affect (Valitutti et al., 2004), has been developed starting from WordNet, assigning one or more affective labels (a-labels) to a subset of synsets representing affective concepts that contribute to precise the affective meaning. In WordNet-Affect, Ekman's six basic emotions (anger, disgust, fear, joy, sadness, surprise) have been used.

Other affective lexicons are available but they often give a polarity evaluation of terms without a reference to an emotional category, such as SentiWordNet (Baccianella et al., 2010) devised for supporting sentiment classification and opinion mining. SentiWordNet is one of the publicly available lexical resources, that extends WordNet thanks to a semi-automatic acquisition of the polarity of WordNet terms, evaluating each synset according to positive, negative and objective values.

In this scenario we choose to develop a new resource based on the model of emotions proposed by Robert Plutchik (Plutchik and Conte, 1997), (Plutchik 2002), an American psychologist, Professor Emeritus at the Albert Einstein College of Medicine and Adjunct Professor at the University of South Florida. He developed a psychoevolutionary theory about the emotions and created a model based on a distinction between basic and complex ones (Kołakowska et al., 2015), introducing an emotion classification system known as the Wheel of Emotions.

3 BUILDING THE RESOURCE

The idea of the resource, named PlutchikWN_ita, arises from the experience gained in the field of Opinion Mining and of the automatic analysis of texts through a linguistic approach in which WordNet has been chosen as the main linguistic resource (Tuveri and Angioni, 2012a).

We realized at that time the need to enrich WordNet through some additional properties, initially linked to adjectives and adverbs and later extended to verbs, to get a deeper analysis of the sentences and improve the semantic disambiguation phase. A first resource has therefore developed (Tuveri and Angioni, 2012b) in which the synsets related to adjectives and adverbs are enriched with a group of properties and with a positive, negative and objective polarity value. The building of this first linguistic resource, related to adjectives and adverbs, and the analysis of the properties of the terms already included in WordNet and identified from the lexicon files, revealed the need to better specify the properties of the terms inherent to sentimental and emotional states. In fact, in both the resources there are properties related to emotions, such as *noun.feeling* and *verb.emotion* in WordNet or *emotion* and *morals/ethics* in the resource we developed.

For this reason we have built a linguistic resource that allows to better specify the type of emotion expressed by the term extending the WordNet properties with the categories arising from the Plutchik model.

Despite the fact that the resource is very limited in number of terms and concepts, its construction allowed to face some of the problems related to a resource dealt with the emotional sphere and to lay the foundations for the development of a more complete resource.

Although there is no fully exhaustive and universally accepted model of emotions, and looking for a substantially simple model for the students, it was decided to consider the model of emotions proposed by Robert Plutchik. In fact, the model defines a few basic emotions, making the problem of the development of the resource manageable, and having as strength point its capacity to simplify very complex concepts.

Moreover, the graphical representation of the emotions proposed by Plutchik enable to immediately visualize the emotions, and thus easily understand which combinations of emotions created a resulting emotion.

3.1 Plutchik's Wheel of Emotions

Plutchik proposes a graphic representation of the wheel of emotions that, starting from a circle of eight primary and fundamental emotions, develops in the structure shown in Figure 1.

This model shows how different emotions can be combined or mixed together, as could do a painter with colors. According to his theory, other emotions are a combination of the basic emotions. Plutchik's studies have therefore highlighted 8 primary basic emotions: joy and sadness, trust and disgust, anger and fear, surprise and anticipation. Primary emotions are represented through a circle, in which the closest emotions are also the most similar and opposite emotions are located on opposite spokes of the wheel. So joy is opposed to sadness, anger is opposed to fear, acceptance is opposed to boredom, and surprise is in opposition to anticipation.



Figure 1: Plutchik's Wheel of Emotions.

Plutchik places the emotions similar to the primary ones, but of lower intensity, on the external part of the wheel, putting them all the more distant, the less their intensity. Following the same principle, similar but more intense emotions are placed within the wheel of primary emotions. In this way, a kind of flower is obtained whose petals can be joined towards the bottom creating a three-dimensional representation in a cone-shaped model.



Figure 2: Primary emotions combination or dyads.

Also the variations in the intensity of the color on each petal correspond to the variations in



Figure 3: Insertion of Italian terms and Pluchik's emotions.

intensity of the emotions. The eight primary emotions occupy the central part of the flower, with gradually less intense emotions by proceeding from the center towards the end of the petal. For example, the most intense form of fear is terror, the least intense is apprehension.

Figure 2 shows how primary emotions work in combination. The compositions of two emotions are called dyads. All not primary emotions are mixed or derived, that is, they appear as combinations or compositions of primary emotions and are identified as primary, secondary or tertiary dyads depending on whether adjacent emotions are combined, shifted by one or two petals.

Primary dyads made from adjacent primary emotions are considered "often felt". The emotions can then be combined in a variety of ways. Dyads including emotions further away on the wheel, express less felt emotions, until opposites create conflict. For example, joy and trust could combine to create love, joy with fear could give rise to guilt, anticipation combined with fear could give rise to apprehension (Taylor, 2017).

3.2 The Identification of Emotional Terms

The building of the resource PlutchikWN_ita has been performed by means of an ad-hoc developed web-based application, that allows to insert one or more terms in Italian and to specify both a polarity value and an emotion from the list of Plutchik's categories, as shown in Figure 3.

The search for Italian terms, which fell within the affective and emotional sphere, was conducted by involving the students.

They examine some of the most famous *carmina* written by Catullo and the Italian version of the "La notte d'amore tra Ginevra e Lancillotto", from the 12th-century Old French poem "Lancelot ou le Chevalier à la charrette" written by Chrétien de Troyes, highlighting the terms of interest through reading.

The use of these terms in the texts allows the author to communicate his mood to the reader. In them, therefore, the expressive function prevails and the attention focuses on emotions, thoughts and personal experiences.

Once identified an Italian term falling within the emotional sphere, the term in English with the corresponding meaning is detected, making use of several online dictionaries when necessary.

The application shows all the possible meanings included in WordNet pertains to the emotional sphere or to morality, starting from the identified English term and from the relative Part of Speech (POS). At this point, it is possible to insert in the resource PlutchikWN_ita the Italian term and its synonyms, corresponding to the English term and to the definition, and to specify the Plutchik's category that best reflects the type of emotion expressed by the term.

It was decided to exploit the terms in English because the English version of WordNet is the most

complete both as a number of concepts and definitions.

Each of the students had the task of independently inserting each term in Italian and the relative Plutchik's emotion. Once the insertion was completed, the terms and the categories were validated.

When disagreement occurred in the identification of the emotion, the students discussed the motivations that led to the specific assignment. In some cases a convergence of opinions has been found, otherwise the term was not included in the resource.

Through the analysis of the texts the students have identified the words expressing the emotions, relating them to the context and using them as a starting point to investigate the real meaning of the emotion by means of discussion and argumentation. The recognition of the emotions of the others refined their ability to identify themselves with the moods of others. At the end of the work the students also showed a greater awareness towards their emotions and a greater attention to their inner states.

4 THE CURRENT VERSION

The concepts included in the resource are currently very limited, 214 in total, and are divided between names (N), adjectives (A), adverbs (R) and verbs (V), as shown in Table 1.

It is possible to notice that the majority of the terms are related to adjectives, followed by nouns and verbs, while terms related to adverbs are irrelevant, but they could be easily extracted from the corresponding adjectives and nouns.

The resource is based on WordNet 3.1 version.

Table 1: The distribution of concepts for POS.

POS	Ν	А	R	V
Synsets/Terms	63	81	2	68

From the analysis of the resource emerges that the richest emotions as number of synsets are serenity (30), grief (20) and admiration (22), as shown in the following Table 2.

As an example, in PlutchikWN_ita the Italian terms like *tranquillo*, *sereno* and *calmo* have been classified as positive terms under the serenity Plutchik's emotion (*serenità* in Italian) and correspond to the English words calm, serene, tranquil, unagitated. For the sake of completeness, in WordNet v.3.1 the synsetID is 300531862 and the definition: is "not agitated; without losing self-

possession; "spoke in a calm voice"; "remained calm throughout the uproar" ".

Plutchik's	Syne	Plutchik's	Syns
Emotions	Syns	Emotions	
serenity	30	aggressiveness	8
grief	24	apprehension	7
admiration	22	low_love	6
joy	13	pensiveness	5
interest	12	high_aggressiveness	5
sadness	11	remorse	4
trust	9	submission	4
fear	9	high_remorse	4
high_love	9	boredom	3
love	8	annoyance	3

Table 2: Plutchik's emotions and number of synsets.

Other examples of terms included in the resource are the Italian adjectives *sofferente*, *penoso* e *doloroso*, corresponding to the English terms afflictive, painful, sore, that have been classified under the Plutchik's emotion grief (*dolore* in Italian) with a negative polarity. The term is related to the synsetID 301809309 in WordNet v.3.1 and is related to the definition: "causing misery or pain or distress; "it was a sore trial to him"; "the painful process of growing up" ".

5 THE RESOURCE FORMAT

PlutchikWN_ita is currently being developed in the WordNet LMF format (Soria, 2009), an extension of the ISO LMF format defined for the representation of lexical resources. LMF allows the coding of lexicons in natural language, and provides a common model for the creation and use of lexical.

The WordNet LMF has been introduced in the EU KYOTO project (Vossen et al., 2008) with the specific purpose of providing the WordNet, set in different languages, with a standardized interoperability format that would allow the exchange of coded lexico-semantic information in each of them. The use of this format enables WordNet to provide a format representation that allows easier integration between resources that share the same structure and, more importantly, between resources with different theoretical and implementation approaches. This fact introduced a new kind of problems especially with the enrichment of new synsets, not included in the English version of WordNet but specific of the language. Furthermore, the different wordnets can be built using different methods and starting from different

points: with the expand or the merge approach, manually or semi-automatically, using pre-existing resources or original bilingual resources available for the translation of the original English words in the target language.

All this led to the definition of CILI, the Collaborative Inter-Lingual Index, (Bond et al., 2016), based on the interlingual index ILI (Interlingual Index) and proposed for the first time in the EuroWordNet project (Vossen, 1998).

CILI is a single shared repository of concepts, defined as an inter-collaborative index designed to make possible to coordinate multiple generic wordnet projects in the Global WordNet Grid (GWG).



Figure 4: An example of PlutchikWN_ita in LMF format.

Figure 4 shows an example of the PlutchikWN_ita in WordNet LMF format. Each new wordnet has a Lexicon for each resource, which contains a name, an ID and a language. The language is specified by the Internet Engineering Task Force (IETF) Best Current Practice (BCP) 47 language tag, the code to identify human languages.

Each term inserted in the linguistic resource is defined as a Lexical Entry, and is characterized by a lemma, by at least a meaning and possibly by any number of Syntactic Behavior.

The Lemma is defined through its written form and the Part Of Speech (POS), which can be a name (n), a verb (v), an adjective (a), an adverb (r), a phrase (like phrase, p, o sentence, s) or an unknown pos (u). The Sense has a variable number of sense relations and a synset.

The Synset can have an optional definition and a variable number of sense relations.

The Definition is given both in the WordNet language from which it comes but also in the ILI definition language (in English). A definition can also have a sentence that provides an example of using of the term.

Synset Relation specifies the type of relationship, such as synonimy, hyperonymy, antonymy, taken from the list of reports used by the Global WordNet Grid and documented on the Global Wordnet Association website¹.

A Syntactic Behavior (verb frame) provides a sub-category frame in plain text.

Metadata from the Dublin Core system can be added to the Lexicon, Lexical Entry, Sense and Synset tags.

6 CONCLUSIONS

The paper describes the creation of a new linguistic resource, built on a set of terms mapped on emotions, by means of the everyday life of the discovery of emotions and their impact on the lives of students in adolescent age. In agreement with the teachers, some students were involved in the analysis of some literary texts, forming part of their course of study, with the task of elaborating and defining the feelings and the emotions described and aroused by the texts.

The reading of literary texts and the identification in them of terms relating to the emotional and affective sphere, allowed the students to define their emotions by identifying themselves in the lives of others and recognizing themselves in other people. With this work, an alternative teaching method was proposed that allowed students to put the emotional intelligence at the center and stimulate it, as a primary element of the evolution of their personality.

The building of the linguistic resource, in a broader context, combines the strongly multidisciplinary aspect characteristic of this activity that requires a heterogeneous workgroup and a constant dialogue between professionals with very different skills, with the didactic-educational aspect that allowed the students to give voice to their emotions.

¹ Global WordNet Association, http://globalwordnet.org/

Currently the resource is only at the initial stage, but in the future we intend to extend the resource through the involvement and the systematic contribution of different classes of students supported by their teachers. The resource has not still been used in any practical application, but we are planning to realize a comparative work using similar resources, with the intent to provide its validation results.

ACKNOWLEDGEMENTS

We wish to thank the three high school students Katia Pedditzi, Ludovica Medved and Alessandra Pani, their whole class, and their teachers and tutors Cinzia Quattrocchi and Antonella Puggioni, from the "Liceo Classico, Linguistico e delle Scienze Umane B. R. Motzo" of Quartu Sant'Elena, Cagliari, to their contribution to the development of the resource.

REFERENCES

- Baccianella, S., Esuli, A., Sebastiani, F., 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In LREC 2010, 7th International Conference on Language Resources and Evaluation, Malta, pp. 2200-2204.
- Bond, F., Vossen, P., McCrae, J., Fellbaum, C., 2016. CILI: the Collaborative Interlingual Index. *Proceedings of the 8th Global WordNet Conference* 2016, in Bucharest, Romania, January 27-30.
- Global WordNet Association Homepage, http://globalwordnet.org/.
- Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wrobel, M.R., 2015. Modeling emotions for affect-aware applications. In book: *Information Systems Development and Application*, Chapter: 4. Publisher, Faculty of Management University of Gdańsk. Editors: Stanislaw Wrzycza.
- Miller, G. A., 1998. WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
- Mohammad, S.M., Turney, P.D., 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: *Proceedings of NAACL 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text.* ACL.
- Plutchik, R., 1980. Emotion: Theory, research, and experience: Vol. 1. Theories of emotion, 1, New York: Academic.
- Plutchik, R., H. R. Conte, 1997. Circumplex Models of Personality and Emotions. Washington, DC: American Psychological Association.

- Plutchik, R., 2002. *Emotions and Life: Perspectives from Psychology, Biology, and Evolution*. Washington, DC: American Psychological Association.
- Taylor, M. J., 2017. Classroom Emotions. https://www.tetraanalytix.com/classroom-emotions/.
- Tuveri, F., Angioni, M., 2012a. A Linguistic Approach to Feature Extraction Based on a Lexical Database of the Properties of Adjectives and Adverbs. *Proceedings of the 6th International Global Wordnet Conference* Tribun EU, pages 365-370, Global WordNet Association. ISBN: 978-80-263-0244-5.
- Tuveri, F., Angioni, M., 2012a. Definition of a Linguistic Resource for Opinion Mining. Proceedings of the 9th International Workshop on Natural Language Processing and Cognitive Science –NLPCS -SciTePress -june 2012. ISBN: 978-989-8565-16-7.
- Soria, C., Monachini, M., Vossen, P., 2009. Wordnet-LMF: Fleshing out a Standardized Format for Wordnet Interoperability. *Proceedings of the 2009 international* workshop on Intercultural collaboration. ACM.
- Valitutti, A., Strapparava, C., Stock, O., 2004. Developing affective lexical ressources. *Psychnology*: 2-2004.
- Vossen, P. (ed), 1998. EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht.
- Vossen, P., E. Agirre, N. Calzolari, C. Fellbaum, S. Hsieh, C. Huang, H. Isahara, K. Kanzaki, A. Marchetti, M. Monachini, F. Neri, R. Raffaelli, G. Rigau, M. Tescon, J. van Gent, 2008. KYOTO: A System for Mining, Structuring, and Distributing Knowledge Across Languages and Cultures. *Proceedings of the Fourth International Global WordNet Conference - GWC* 2008. Szeged, Hungary, January 22-25.