Development of Affective Lexicon for Spanish with Mexican Slang Expressions

Noé Alejandro Castro-Sánchez[1], Yolanda Raquel Baca-Gómez[2], and Alicia Martínez[1]

[1] Centro Nacional de Investigación y Desarrollo Tecnológico/Tecnológico Nacional de México,
Cuernavaca, Mexico

[2] INFOTEC Centro de Investigación e Innovación en Tecnologías
de la Información y Comunicación
Mexico

{ncastro, amartinez}@cenidet.edu.mx, yolanda.baca@infotec.mx

Abstract. Nowadays exists a growing interest in the automatic extraction of subjective expressions (opinions, emotions and feelings) in texts. To identify the semantic orientation of a text, it is assumed that the occurrence of expressions that belong to some emotional category can be regarded as evidence that there is an affective state. Based on this assumption, we create an affective lexicon, consisting in the translation from English to Spanish of various lexical resources, including works based on psychological theories to identify words associated with emotions. The lexicon was manually enriched through semantic relationships as inclusion and synonymy using explanatory dictionaries. Expressions used in Mexican slang were also included in the lexicon. Every word in the lexicon was labeled with its semantic orientation; these are: "very positive", "very negative", "positive" or "negative", for the translated words and "positive" or "negative" for the Mexican slang. The lexicon currently consists of 3550 words and 255 slang expressions.

Keywords: Affective lexicon, emotions, semantic orientation, Mexican slang

1   Introduction

Recently, emotions play an important role in intelligent behavior researching in Artificial Intelligence. The rapidly growing field of affective computing aims at developing systems and resources for predict, understand, and process emotions [1]. Defining what an emotion is, is a very difficult problem. Emotions are not linguistic things; however the most convenient access we have to them is through language, thus one reasonable way to separate emotions from non-emotions is to consider referents of emotion or opinion words [2, 3]. Opinion words are the most important indicators of sentiments; these words are commonly used to express positive or negative sentiments. A list of such words and phrases is called a *sentiment lexicon*, *opinion lexicon* or *affective lexicon*. This kind of lexicon is instrumental to sentiment analysis in the

lexicon-based method, which uses a dictionary of sentiment words and phrases with their associated orientations and strength, and incorporates intensification and negation to compute a sentiment score [4].

The purpose of this work is to present the development of a resource for sentiment analysis. This resource is an Affective lexicon composed by the translation from English to Spanish of sentiment words and also by Mexican slang expressions. The paper is organized as follows: Section 2 presents related works with the creation of affective and emotional lexicons. Section 3 describes the method followed for the creation of the affective lexicon. Section 4 details the results and lessons learned. Finally the Section 5 exposes the main conclusions and ideas for further works.

2   State of the Art

In this section we describe some works from both theoretical and computational approaches. These approaches are mainly useful for categorizing and classifying emotions, and also for identifying the intensity and the valence or semantic orientation of the emotions. From the psychological point of view we described some examples like the work presented in [2], where an Affective Lexicon was developed with a taxonomy of affective conditions using a list of 500 words used by other psychologists in their studies of emotion, including words from the work described in [5]. Also, in [5] a geometric representation was built, this representation consists of the relations among the 28 emotion words by placing them in a Euclidean space, where the 28 terms are definable in a two dimensional bipolar space pleasure-displeasure and degree of arousal.

Another work is the one referred in [6] where a corpus was built by collecting a representative sample of words denoting emotions by inspecting some lexical resources like [7], this corpus is composed by emotional words according to a communicative theory in which there should be a set of terms that refers to basic emotions, the theory implies that any emotional term should devolve upon one of the basic emotion modes, or some subset of them.

Now, we described some examples of the methods followed in the computational approach, like the work described in [8] where an affective resource called WordNet-Affect was created from WordNet, through the manually selection of a subset of words and by the labeling of every word of the subset with its affective category. Another example is the proposed method in [9] where a Spanish lexicon was built by the combination and translation from English to Spanish of resources like Opinion-Finder, WordNet and SentiWordNet.

Also, in [10] a method for a dictionary creation was presented, in the dictionary created the words are labeled by multiple annotators with the six basic emotions, and the dictionary was evaluated with Kappa and PFA (Probability Factor of Affective Use). Finally in [11] an emotional lexicon called SentiSense was created, the creation of this lexicon is based in psychological theories with the purpose of obtain not only the semantic orientation but also the intensity of the emotion, and in this work it is also used WordNet as a reference.

3 Methodology

In this section we described the method for the creation of the affective lexicon for Spanish with Mexican slang expressions. According to the state of the art, we decided to start a translation with words already classified in emotional categories. Fig. 1 shows our solution methodology for developing the affective lexicon, which consists of three phases: (i) Translation of resources from English to Spanish, (ii) manual enrichment using semantic relationships and (iii) manual enrichment of Mexican slang.
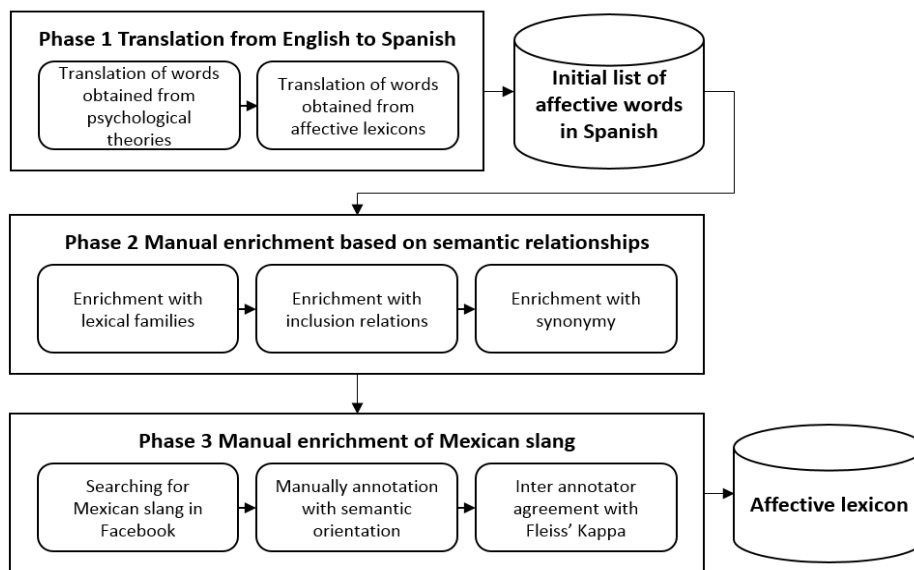
Phase 1 Translation from English to Spanish
- Translation of words obtained from psychological theories
- Translation of words obtained from affective lexicons
- Initial list of affective words in Spanish

Phase 2 Manual enrichment based on semantic relationships
- Enrichment with lexical families
- Enrichment with inclusion relations
- Enrichment with synonymy

Phase 3 Manual enrichment of Mexican slang
- Searching for Mexican slang in Facebook
- Manually annotation with semantic orientation
- Inter annotator agreement with Fleiss' Kappa
- Affective lexicon

Fig. 1. Solution Methodology for developing Affective Lexicon

3.1 Phase 1: Translation of Lexical Resources from English into Spanish

In this phase words obtained from psychological theories listed below were translated from English to Spanish. The translation is a problem in this kind of research, since many terms have somewhat different denotation and connotation meaning in different languages [12]. Thus, the meaning of the words is analyzed into its context where it can be used.

Translation of words obtained from psychological theories

1. A Circumplex Model of Affect [5].
2. Geneva Emotion Wheel Rating Study [12].
3. The GRID meets the Wheel: Assessing emotional feeling via self-report [13].
4. What are emotions? And how can they be measured? [14].
5. Structure of emotions [15].

Translation of affective lexicons

1. WordNetAffect [8]
2. General Inquirer [16].
3. Opinion Finder [17].

Translation process

1. A word was taken from one of the lexical resources.
2. The word was translated using *Google*[1] and *Linguee*[2].
3. The context of the word was verified by searching their meaning in both English and Spanish. We used the *Oxford Dictionary* for English and *The Dictionary of Spanish Language of the Royal Spanish Academy* for Spanish.
4. Based on the meanings of the word, we choose the best translation.
5. The translated word was added to the Affective Lexicon and we labeled it with the semantic orientation specified in the lexical resources. This semantic orientation can be "very positive", "positive", "very negative" and "negative".
6. The resource where the word comes from is also specified.

Fig. 2 shows a brief content of the Affective lexicon. The first column corresponds to word translated, the second column is the polarity of the word, and in this case: "+" for *positive*, "++" for *very positive*, "-" for *negative* and "--" for *very negative*. The third column is the emotion associated to the translated word; the emotions were obtained from the psychological theories. From fourth column to nine are the lexical resources, where "GI" means General Inquirer, "WNA" means WordNetAffect and "OF" means Opinion Finder. In the final column appears the word in its original language (i. e., English). Columns from fourth to nine can contain one of the symbols "+", "++", "-" or "--" which means that the word was found in the lexical resource specified in the column header with the polarity represented by the symbol.

Spanish	Polarity	Emotion	Morgan-Heise	Russell	Scherer	GI	WNA	OF	English
Regocijo	+	Joy				+			Rejoice
Angustiado	--	Anxiety	--						Anguish
Tranquilo	+	Serenity		+					Calm
Deplorable	-	Sadness						-	Deplorable
Despojado	-	Sadness						-	Bereft
Abatido	--	Sadness			--				Dejected
Furioso	--	Anger	--						Furious
Furiosamente	--	Anger							Fiercely
Colérico	--	Anger							Choleric
Solidaridad	+	None						+	Solidarity
Placentero	++	Hapiness	++	++	++				Delighted

Fig. 2. Brief example of the content of the Affective lexicon

---

[1] Free online Translation service provided by Google.
[2] Both an editorial dictionary and a search engine for translations from the bilingual web.

3.2 Phase 2: Manual Enrichment based on Semantic Relationships

In this phase, the initial list (the one generated in the previous phase) was enriched with semantic relationships that are explained below. The types of enrichment were obtained from *The Dictionary of Spanish Language of the Royal Spanish Academy, The Reverse Dictionary, A printed Dictionary of Synonyms* and *WordReference*[3].

Enrichment with lexical families

A lexical family consists of a base word and all its derived and inflected forms. So, for the word *"pervertir"* (pervert), the words *"perverso"* (perverse), *"pervertido"* (perverted), *"perversidad"* (perversity), *"perversión"* (perversion), *"pervertidor"* (perverter), may be all members of the same lexical family [18]. So, for each translated word, members of its lexical family were also included keeping the same polarity.

Enrichment with inclusion relationships

The inclusion relationships describe situations where one entity type comprises or contains other entity types. Class inclusion is the standard subtype/super type relationship that frequently appears in data modeling. Examples include: *"coche"* (car) is a type of *"vehículo"* (vehicle), *"rosa"* (rose) is a type of *"flor"* (flower), and *"robo"* (robbery) is a kind of *"crimen"* (crime) [19].

Enrichment with synonymy

Synonyms are words that have the same or nearly the same meaning. For example, the word *"aprehender"* (apprehend) and *"detener"* (detain) are synonyms [19]. Dictionary of Synonyms were used for obtaining synonyms of the translated words.

3.3 Phase 3: Manual Enrichment of Mexican slang

In this phase, the lexicon was enriched with Mexican slang and other expressions like emoticons and interjections commonly used. Firstly, the vocabulary was obtained; then the expressions were annotated with the semantic orientation; and secondly, the annotator agreement was evaluated.

Searching for Mexican slang in Facebook

A Software System for automatic extraction of comments obtained from Facebook was developed in order to identify common expressions used in the Mexican slang. According to the context in which the word was used, the meaning of the expression was also added. Table 1 presents some Mexican Slang Expressions. First column shows the Mexican Slang Expressions. Second column describes the meaning of the expressions according to its context. Finally, third column presents an example, where the word is used.

---

[3] Online Lenguage Dictionaries.

Table 1. Mexican Slang Expressions with meanings and context examples.

#	Mexican Slang Expression	Meaning	Context example
1	*Qué pedo!* – What the fuck!	*Enojo* (Angry)	*Qué pedo contigo!! Te estas pasando de ojete!!* (What the fuck with you!! You're doing wrong!!)
2	*Qué pedo!* – What's up	*Saludo* (Salute)	*Qué pedo wey, ya saliste de la uni, vamos por una frías!* (What's up buddy, are you out of school?, Do you wanna go for a beer?)
3	*Madreado*	*Golpeado* (Beaten)	*Lo dejaron bien madreado por andar en donde no debe.* (He was very beaten because he made something wrong)
4	*Madreado*	*Cansado* (Tired)	*Tuve un chingo de trabajo hoy, terminé bien madreado.* (I had a lot of work, and I'm so tired)
5	*Chingar*	*Robar* (Steal)	*Estoy que me lleva la... fui al centro y me chingaron mi celular.* (I can't believe it... I went to the downtown and someone stole my cell phone)
6	*Chingar*	*Molestar* (Annoy)	*Esos del banco siempre están chingando por teléfono.* (The bank cashiers always annoy me by phone)

Manual annotation

Every word in the list previously generated was labeled by five people as "positive" or "negative" and also if they agree with the meaning of the word. The purpose of this manual annotation is to validate the quality of the semantic orientation and the meaning of the Mexican slang expressions that we previously found in the Facebook comments. With the annotations of the five people we generated a table as shown in Table 2.

Table 2. Example of manual annotations of the Mexican slang.

#	Person 1				Person 2				Person 3				Person 4				Person 5			
	P	N	A	D	P	N	A	D	P	N	A	D	P	N	A	D	P	N	A	D
1		X	X			X	X			X		X		X	X			X	X	
2	X			X	X			X	X			X	X			X	X			X
3		X	X			X	X			X	X			X	X			X	X	
4		X		X		X	X			X	X			X	X			X	X	
5		X	X			X	X			X		X		X	X			X	X	
6		X	X			X	X			X	X			X	X			X	X	

The first column makes reference with Table 1, the next columns are the annotations of the five people, where the letters used mean: P = Positive, N = Negative, A = Agree and D = Disagree.

Other expressions

Some interjections, abbreviations and emoticons frequently used in the Facebook comments were also added. Table 3 shows some examples of these expressions.

Table 3. Examples of interjections, abbreviations and emoticons.

#	Exp.	Meaning	Context example
1	ash	*Desagrado* (Displeasure)	*Ash, no lo puede hacer alguien mas??? Porqué siempre yo???* (Ash, can someone else do this??? Why always me???)
2	nhp	*No hay pedo* (No problem)	*nhp!!! Yo pongo mi casa para la party!!!!!* (nhp!!! We can celebrate the party in my home!!!!!)
3	npi	*Ni puta idea* (No fucking idea)	*npi de cuando se publiquen los resultados.* (npi when the results will be published)
4	mms	*Mamadas* (Foolishness)	*Qué mms son esas de postear feliz viernes!!* (What a mms are those of post happy Friday)
5	chin	*Frustración* (Frustration)	*No llegué a tiempo chin!! No hice el examen.* (I'd not in time chin!!! I didn't resolved the exam)
6	mta	*Enojo* (Anger)	*Mta!!!! llevo mil horas esperando la ruta!!!* (Mta!!!! I've been waiting a thousand hours for the microbus arrive)
7	T_T	*Triste* (Sad)	*Alguien puede ayudarme!!! T_T No sé que hacer!!! :(* (Can anybody help me!!! T_T I don't know what to do!!!)
8	>.<	*Molesto* (Annoying)	*No es posible que existan personas así!!!! >.<* (How is possible that people like that could exist!!!!)
9	:D	*Feliz* (Happy)	*Wiiii El día tan esperado llegó, por fin de vacaciones!!! :D* (Wiiii The long awaited day is here, vacation!!!)
10	^.^	*Feliz* (Happy)	*Que tu novio te regale chocolates, no tiene precio! ^.^* (If your boyfriend gives you chocolates, is priceless!)

Inter annotator agreement with Fleiss' Kappa

Fleiss' Kappa metric [20, 21] was applied in order to evaluate the agreement between the annotators. We conducted two assessments, the first one for the polarity annotation, and the second one for the annotation of the meaning. The results were the following: for polarity annotation a value of 0.82 was obtained and for the annotation

of the meaning a value of 0.79 was obtained. According to the Fleiss' Kappa metric the firs result means "very good" agreement and the second one means "good" agreement [21]. In both cases, the subjectivity involved in the interpretation and anno-tation of the affective words has an impact into the results.

4 Results and Lessons Learned

In this section the size of the Affective lexicon and some lessons learned in the crea-tion of the Affective lexicon are presented.

4.1 Size of the Affective Lexicon

Table 4 shows some examples of the elements of the categories that are part of the Affective Lexicon and the total number of expressions included in every category. In the last category, the emotion associated to the affective word was obtained from the lexical resources translated. First column presents the Affective Lexicon categories; Second column shows the total number of expressions in every category. Finally, last column details some examples of the expressions contained in the categories. In the Affective Words category, the emotion associated to the affective word is also pre-sented.

Table 4. Number of elements of the Affective lexicon

Category	Total expressions	Expressions	
		Semantic Orientation	**Example**
Emoticons	131	*Positive*	:D :-) n.n ^.^
		Negative	:-(T.T u.u </3
Interjections and abbreviations	60	*Negative*	snif, ogt, ash
		Positive	wi, aw, yeah
Mexican slang expressions	255	*Negative*	*Desmadrar* (Break)
			Chingar (Bothersome)
			Cagetiza (Scolding)
		Positive	*Poca madre* (Spectacular)
			Chingón (Able)
			Chulo (Nice-looking)
Affective words	3550	*Very Negative Emotion*	*Asesinar* (kill)
			Ira (rage)
		Negative Emotion	*Llorar* (cry)
			Tristeza (sadness)
		Very positive Emotion	*Entusiasta* (enthusiastic)
			Alegría (hapiness)
		Positive Emotion	*Plácido* (placid)
			Serenidad (serenity)

4.2   Lessons Learned

The translation took a lot of time because we analyzed the meaning and the context of every word, and the final translation is based on what we considered as the best option. We found some words and phrases that can be used in both positive and negative way, and also words with more than one meaning. In the lexicon we kept both positive and negative context and also all the possible meanings of the words. But, in the practice it is necessary to implement an algorithm to identify the context in which the word is used.

5   Conclusions and Future Works

The main motivation for the development of this research arises from the need to generate lexical resources for Sentiment Analysis in Spanish. In the literature we observed that lexical resources for Spanish are limited, hence we developed an Affective Lexicon.

A common expressions used in Mexican slang were included. These expressions cannot be obtained from other resources because they are designed for Castilian or English. Thus we can say that this project is pioneer in including Mexican slang expressions in an affective lexicon.

As future work we propose:

- The manual enrichment of the Affective lexicon takes a lot of time and it is susceptible to mistakes. So, the best option is to develop a tool for the automatic enrichment through semantic relationships.
- It is necessary to carry out a deeper analysis of comments posted in Social Networks for identifying more Mexican slang expressions to enrich the Affective lexicon with this kind of expressions.
- It would be interesting to identify and classify affective expressions by domain, for example to identify the most common of them used in football, politics, tourism, etc.

References

1. Bonnefon Jean François, Longin Dominique, Nguyen Manh-Hung: A Logical Framework for Trust-Related Emotions. Proceedings of the Third International Workshop on Formal Methods for Interactive Systems (FMIS), Electronic Communications of the EASST, Vol. 22, 1–15 (2009)
2. Ortony Andrew, Clore Gerald L., Foss Mark A.: The referential structure of the Affective Lexicon. Cognitive Science 11, 341–364, University of Illinois at Urbana-Champaign (1987)
3. Ortony Andrew, Clore Gerald L., Collins Allan: The Cognitive Structure of Emotions. Cambridge University Press (1990)
4. Liu Bing: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2012)

5. Russell James A.: A Circumplex Model of Affect, Journal of Personality and Social Psychology. Vol. 39, No. 6, 1161–1178, University of British Columbia, Vancouver, Canada, (1980)

6. Johnson-Laird, Oatley Keith: The Language of Emotions: An Analysis of a Semantic Field. Cognition and Emotion, Vol. 3, No. 2, 81–123, United Kingdom (1989)

7. Clore Gerald L., Ortony Andrew, Foss Mark A.: The Psychological Foundations of the Affective Lexicon. Journal of Personality and Social Psychology, Vol. 53, No. 4, 751–766 (1987)

8. Strapparava Carlo, Valitutti Alessandro: WordNet-Affect: An Affective Extension of WordNet. Proceedings of the 4th International Conference on Language Resources and Evaluation, 1083–1086, ITC-irst, Instituto per la Ricerca Scientifica e Tecnologica, Povo Trento Italy (2004)

9. Pérez-Rosas Verónica, Banea Carmen, Mihalcea Rada: Learning Sentiment Lexicons in Spanish. Proceedings of the Eight International Conference on Language Resources and Evaluation, 3077–3081, Istanbul (2012)

10. Díaz Rangel Ismael, Sidorov Grigori, Suárez-Guerra Sergio: Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. Onomazein, No. 29, 1–23 (2014)

11. Plaza Laura, Carrillo de Albornoz Jorge, Gervás Pablo: SentiSense: An easily scalable concept-based affective lexicon. Proceedings of the Eight International Conference on Language Resources and Evaluation, 3562–3567, Istanbul (2012)

12. Sacharin V., Schlegel K., Scherer K. R.: Geneva Emotion Wheel rating study (Report). Geneva, Switzerland, University of Geneva, Swiss Center for Affective Sciences (2012)

13. Scherer K. R., Shuman V., Fontaine J. R. J., Soriano C.: The GRID meets the Wheel: Assessing emotional feeling via self-report. Components of emotional meaning: A sourcebook. Oxford, Oxford University Press (2013)

14. Scherer Klaus R.: What are emotions? And how can they be measured? Trends and developments: research on emotions. Social Science Information, SAGE Publications, Vol. 4, No. 4, 695–729 (2005)

15. Morgan Rick L., Heise David: Structure of Emotions. Social Psychology Quarterly, Vol. 5 No. 1, 19–31 (1988)

16. Stone Philip J., Dunphy Dexter C., Smith Marshall S., Ogilvie Daniel M.: The General Inquirer: A Computer Approach to Content Analysis. MIT Press, Oxford, England (1966)

17. Wilson Theresa, Hoffman Paul, Somasundaran Swapna, Kessler Jason, Wiebe Janyce, Choi Yejin, Cardie Claire, Riloff Ellen, Patwardhan Siddharth: OpinionFinder: A system for subjectivity analysis. Proceedings of HLT/EMNLP Interactive Demonstrations, 34–35, Vancouver (2005)

18. Bauer Laurie, Nation Paul: Word Families. International Journal of Lexicography, Vol. 6, No. 4, Oxford University Press, Victoria University of Wellington, New Zealand (1993)

19. Storey Veda C.: Understanding Semantic Relationships. VLDB Journal, Vol. 2, No. 4, 455–488 (1992)

20. Fleiss Joseph L., Levin Bruce, Paik Myunghee: Statistical Methods for Rates and Proportions. Third Edition, John Wiley & Sons Co. (2003)

21. Díaz Rangel Ismael, Suárez-Guerra Sergio, Sidorov Grigori: Detección de afectividad en texto en español basada en el contexto lingüístico para síntesis de voz. Instituto Politécnico Nacional, Centro de Investigación en Computación, Tesis Doctoral (2013)