Computer Science and Information Systems 16(1):273–287 https://doi.org/10.2298/CSIS171001036J

An Experience in Automatically Building Lexicons for Affective Computing in Multiple Target Languages

Francisco Jurado and Pilar Rodriguez

Department of Computer Engineering
Universidad Autónoma de Madrid
Francisco Tomas & Valiente 11, Madrid
{Francisco.Jurado; Pilar.Rodriguez]@uam.es

Abstract. Affective Computing in text attempts to identify the emotional charge reflected in it, trying to analyse the moods transmitted while writing. There are several techniques and approaches to perform Affective Computing in texts, but lexicons are their common point. However, it is difficult to find solutions for specific languages different from English. Thus, this article presents an experience in automatically generating lexicons to perform Affective Computing following a multiple-target languages approach. The experience starts with some initial seeds of words in English that define the emotions we want to identify. It then expands them as much as possible with related words in a bootstrapping process and finally obtains a lexicon by processing the context sentences from parallel translated text where the terms have been used. We have checked the resulting lexicons by conducting an exploratory analysis of the affective fingerprint on a parallel corpus with books translated from and to different languages. The obtained results look promising, showing really similar affective fingerprints in different language translations for the same books.

Keywords. Sentiment Analysis, Affective Computing, Multiple Target Languages, Automatically Building Lexicons

1 Introduction

Sentiment Analysis and *Opinion Mining* can be defined as the *"computational treatment of opinion, sentiment, and subjectivity in text"* [1, 2]. In recent years, it has been applied to many different contexts, such as reviewing customers' products and services, monitoring reputations in social networks, tracking people's feelings about politicians, promoting marketing campaigns, etc. [3]. So it has become a growing research discipline [4]. Moreover, because currently social media texts are widely available and need to be analysed to create knowledge for decision-making processes, new approaches rise to perform sentiment analysis on code repositories like GitHub [5] and social networks like Twitter [6] or YouTube [7]. However, it looks as though *"affective computing and sentiment analysis are still finding their own voice as new interdisciplinary fields"* [8].

Given the variety of contexts and applications where these techniques can be applied, we can note some distinctions and nuances in terminology. Thus, Subjectivity Analysis is used to classify a given text into subjective or objective. Once a text has been classified

as subjective, a Sentiment Analysis (also called Polarity Analysis) can be performed by assigning a score (positive or negative) to the opinion underlying the text.

Regardless of the subjectivity and objectivity of the text, Affective Computing attempts to identify the emotional charge (happiness, sadness, fear, anger-passion, etc.) that is reflected and hence transmitted in the text, depending on the words used.

Accordingly, it is straightforward to see the difference between examining the opinion about something (a product or service) or somebody (an actor or politician) with a Polarity Analysis in the subjective text and analysing the moods reflected and transmitted in a text, either subjective or objective. We will focus our attention on this last kind of analysis, namely, Affective Computing.

Independently of the kind of analysis to be performed, there are several techniques and approaches to computing, including and Affective Computing in the text [1, 2, 9–11], among which we can mention Naive Bayes, Maximum Entropy, Support Vector Machines, Lexicon-based, etc. Nevertheless, the common point is a central piece known as lexicon [3].

Fortunately, there are several lexicons available. As examples, we can mention works like SentiWordNet [12, 13], which gives positivity, negativity and objectivity scores to each synset (synonym set) of WordNet [14]; WordNet-Affect [15], which assigns one or more affective labels (a-labels) for those synsets from WordNet that represent emotions (with emotional categories), emotional valence (positive, negative, ambiguous and neutral), moods, cognitive states, behaviour, etc.; the Affective Norms for English Words (ANEW) [16], which provides valence, arousal and dominance scores for the 1,034 most used words in English; and the NRC Word-Emotion Association Lexicon also known as EmoLex [17], which provides polarity and emotional annotations (anger, anticipation, disgust, fear, joy, sadness, surprise and trust) for 14,182 English words and their corresponding automatic translations to more than 40 languages.

However, the main drawback is that most of them have English as the goal language and it is difficult to find solutions for other languages [18].

While building a lexicon we must make some decisions. In particular, high precision lexicons can be obtained manually but with a low coverage, and in the opposite situation we find those automatically derived from pre-existing knowledge [19]. That is, the more human supervision devoted to classification, the more precision, but the more automatically derived, the more coverage. In particular, developing a supervised approach will be slower but more precise than an unsupervised one. However, because human supervision is not always available and we are focused on working with multiple target languages, we will centre our attention on this last kind of approach.

From this perspective, the main achievements of this paper can be summarised as follows:

- Facing the issue of automatically generating lexicons to perform Sentiment Analysis on texts following a target language independent approach and using automatic translation systems.
- Detailing an automatic approach that makes use of initial seeds of words in English that define the emotions we want to identify (anger, disgust, fear, joy, sadness and surprise). Then applying bootstrapping to expand them as much as possible with related words, and finally obtaining a target independent language lexicon to detect emotions in a text by processing the context sentences where the terms have been used in a parallel text.

- Checking the validity of the approach by conducting an exploratory analysis on a parallel corpus with books translated into different target languages to analyse the affective fingerprint identified in the texts and comparing the results between two different target languages.

Thus, the rest of the article is structured as follows: section 2 shows some related works; section 3 outlines the approach we followed to build the lexicons; section 4 shows some implementation issues; section 5 explains some inspections we carried out to the lexicons; section 6 exposes the conducted exploratory analysis we performed on a parallel corpus; section 7 provides a discussion and some further works; and finally, section 8 ends with the conclusions.

2 Related Works

Although most of the research in Sentiment Analysis has been done for the English language, the literature offers a few examples of multilingual automatic translation approaches to build resources and tools for Sentiment Analysis.

Thus, [18, 20, 21] use bilingual dictionaries and parallel corpora to rapidly create tools for Subjectivity Analysis in the new language. The authors employ resources available for English and apply machine translation to generate resources for subjectivity analysis in other languages different from English. Comparing evaluations of Romanian and Spanish, they conclude that automatic translation is a viable alternative for the construction of resources and tools for Subjectivity Analysis in a new target language.

In the same way, [22, 23] face the issue of sentiment detection in different languages using distinct machine translation systems (Bing Translator, Google Translate and the Moses statistical machine translation system), concluding that those systems can be used to obtain training data and to build Sentiment Analysis tools for languages other than English with comparable performance to the one obtained for English.

All these previous approaches make use of existing English resources and perform the corresponding translation to a target language; however, none of them builds the original resources on their own.

For their part, [24] try to produce a set of lexicons for 136 major languages via graph propagation. They use the Lius' lexicon [2] as the initial seed of words in English and make use of several resources like Wiktionary [25], Google Translate and WordNet [14]. As a result, they are able to build polarity-annotated lexicons for the major languages.

As far as we know, although all these researchers have proved the validity of using automatic machine translations and propagation to build resources for Sentiment Analysis, the above approaches have not been developed with an aim to build an emotional lexicon to assemble tools that allow analysis of the moods reflected in a text, but also the polarity.

The only exception we have found is the NRC Word-Emotion Association Lexicon (also named *EmoLex* by their authors) developed by [17]. These authors used Mechanical Turk (https://www.mturk.com) to crowdsource a lexicon that provides polarity (positive and negative) and emotional annotations (anger, anticipation, disgust, fear, joy, sadness, surprise and trust) for 14,182 English words. Starting from these English words, they provide the corresponding automatic translations for more than 40 languages. In a very

similar way, [26] have used crowdsourcing for harvesting emotional annotations of news articles rather than the words. After processing these annotated articles, they were able to build an English lexicon for Sentiment Analysis. However, they do not provide translation into any other languages.

In these last-mentioned works, although the translation to the non-English languages can be automatically performed, the translation is not validated, and the time and effort consumed by the crowdsourcing process cannot always be tackled depending on the application domain.

Although not designed for multilingual purposes, [6] presents SentiCircles, an interesting lexicon-based approach that takes into account the co-occurrence of words to provide a context-specific sentiment orientation. Their authors argue that *"words that co-occur in a given context tend to have certain relation or semantic influence"*. Thus, the SentiCircles approach modifies the pre-established sentiment value by taking into consideration the context in which the term was used and building lexicons that use a contextual representation of words.

Aside from the fact that SentiClircles was designed to perform polarity analysis on Twitter (not tested in general texts) and only in the English language, what is the reason for limiting this idea of co-occurrence of words to single words but not the associated emotion? If words are used that appear near emotional and affective terms, they are somehow related to that emotion and must be incorporated in the emotional lexicon. That is, we will identify those words that co-occur to emotional related terms, and then we will label them with the emotion with which they co-occur.

From this perspective, we will show an approach that makes use of the co-occurrence of words for emotional terms in sentences retrieved from parallel texts to build lexicons for affective computing in multiple target languages.

3 An Approach to Automatically Building Lexicons for Affective Analysis

To automatically generate lexicons to perform Affective Analysis that annotates words with emotions in different target languages, our approach will start from the initial seeds of words in English and then use linguistic tools to perform expansion and translation. The outline of the approach can be summarized as follows:

1. Build sets of emotional words in English starting from some initial seeds for each emotion.
2. Retrieve parallel sentences in English and in the target language, where the words from the initial sets in English have been used.
3. Compute the most frequently used words in the parallel context sentences obtained for the target language.
4. Build the lexicon with the most frequently used words for each emotion.

Thus, we will first build a set of words in English for each emotion we want to identify. To do this, the set associated with each emotion initially contains only two seeds. These seeds correspond to the emotion noun and its related adjective in English.

In order to identify emotions to work with, we must consider disjoined categories. Although initially conceived to identify facial expressions within the Computer Image Analysis field, one of the most common models is the one proposed by Ekman [27], who defined a six basic emotions model that takes into account these emotions: anger, disgust, fear, happiness, sadness, and surprise.

Within the Natural Language Processing field, inter-annotator agreement studies for each of these six emotions have been performed in order to measure how well the annotators can make the same annotation decision for a certain category [15, 20, 28, 29], showing that the six emotions from Ekman's model are applicable. Thus, Table 1 shows the initial seeds associated with each emotion. As we can see, the seeds contain the emotion noun and its related adjective in the English language.

Table 1. Seeds associated with each emotion (noun and adjective) in English.

Emotion	Seeds (noun + adjective)
Joy	joy, joyful
Sadness	sadness, sad
Anger	anger, angry
Fear	fear, afraid
Disgust	disgust, displeased
Surprise	surprise, surprised

Then, to build the sets of emotional words in the English language, each word from these seeds is expanded with its synonyms from WordNet [14] and terms with similar meaning taking into account the target language from bab.la [30]. So far, we have obtained a set of words for each emotion in the English language. Afterwards, we collect context sentences in English where the emotional words have been used, and the corresponding translation of these sentences into the target language. That is, we get a corpus of parallel sentences where the emotional words are used. To build it, we have made use of the bab.la service [30]. Finally, we annotate the most frequently used words from the context sentences in the target languages with the corresponding emotion from which they come from, and then create the lexicon with those words.

4 Details of the Experience on Applying the Approach

In this section, we will explain how we applied the steps outlined in the previous section in the way detailed in Algorithm 1. Thus, as introduced in the outline section, we started with a seed ω of words in English for each emotion, particularly, the emotion noun and its related adjective in English. Then, we expanded these initial seeds with their related words and synonyms in a bootstrapping process. As a result, we obtained a set of words related to each emotion in English. To perform these tasks, we made use of bab.la [30], a language portal that provides multilanguage dictionaries. These dictionaries provide not only translations but also a full list of terms with a similar meaning for each term used, a list of synonyms based on different thesauri depending on the target language, such as WordNet [14] for English or OpenThesaurus [31] for German. Because bab.la does not provide an API interface, we have developed the necessary routines which request specific terms and scrapes the retrieved HTML in order to extract the relevant information.

1. Bootstrap the initial seeds of emotional words in English
1.1 For each *Emotion*
1.1.1 $\omega_{\{Emotion\}}$ = set {$noun_{\{Emotion\}}$, $adjective_{\{Emotion\}}$}
1.1.2 For each word in $\omega_{\{Emotion\}}$
1.1.2.1 $\omega_{\{Emotion\}} = \omega_{\{Emotion\}} \cup$ Synonyms(word)
2. Retrieve the context sentences from parallel texts
2.1 For each *Emotion*
2.1.1 For each *word* in the $\omega_{\{Emotion\}}$
2.1.1.1 $\Omega_{\{Emotion\}}$= Context_sentences(*word*, lang)
3. Obtain the sets of words for the target language
3.1 π = {entities} $\cup$ {abbreviations} $\cup$ {acronyms}
3.2 For each *Emotion*
3.2.1 For each *cs* in $\Omega_{\{Emotion\}}$
3.2.1.1 $\omega'_{\{Emotion\}}$ = tokenize(*cs*) - π
3.3 $\Upsilon = \varnothing$
3.4 For each *E1, E2* in combination(*Emotions*)
3.4.1 $\Upsilon = \Upsilon \cup (\omega'_{\{E1\}} \cap \omega'_{\{E2\}})$
3.5 For each *Emotion*
3.5.1 $\omega'_{\{Emotion\}} = \omega'_{\{Emotion\}} - \Upsilon$
4. Build the lexicon with the words for each emotion
4.1 $\varphi = \varnothing$
4.2 For each *Emotion*
4.2.1 For each *word* in $\omega'_{\{Emotion\}}$
4.2.1.1 $\varphi = \varphi \cup$ { (word, Emotion) }

Algorithm 1. Steps to implement the approach.

Following Algorithm 1, the next step was to retrieve the context sentences in a target language. To perform this step, we again made use of bab.la, which provides a large list of parallel context sentences to exemplify the usage of the words for both the source and the target languages.

Thus, for each word from the corresponding set ω, we fetched as many context sentences as possible. Bab.la reports these sentences as parallel examples where each word is used in both the original language (English) and its corresponding translation into a target language. Therefore, we filter only the translated sentences. The result is the context sentences Ω in the target language for each emotion.

Because entities, abbreviations and acronyms do not provide affective information, we needed to remove them. To perform this task in an unsupervised way, we simply used regular expressions to identify them. Thus, to find entities we used a regular expression looking for those words that start with only one uppercase letter, are followed by several lowercase letters and are not at the beginning of a sentence. In the same way, to detect abbreviations we used a regular expression looking for those words whose length is not greater than three letters, start with an uppercase letter and are not at the beginning of a sentence. Finally, to identify acronyms, we used a regular expression looking for combinations of uppercase letters, all of them joined by dots. Obviously, the best way to perform this task is to count with a set of well-known entities, abbreviations and acronyms, but because of the automatic nature of our approach, we were looking for a language-

independent approach, and this is an exploratory study, we considered this method a good starting point.

Following those steps, after removing the entities, abbreviations and acronyms that appear in the context sentences, we created a set ω' of words for each emotion in the target language by including the most frequently used words in the context sentence for the corresponding emotion. Because our aim was to obtain exclusive sets of words for each emotion, we removed the intersection $\curlyvee$ between sets of words so that we eliminated stop-words, domain-specific words and ambiguous words that could be associated with more than one emotion. Moreover, we did not include in the sets those words that have appeared only in one context sentence.

Finally, the lexicon φ is composed by the ω' sets containing the words in the target language and their corresponding emotion. To measure the precision of these obtained sets, we use EmoLex [17] as labelled data.

To summarize and help make the rest of the article easier to follow, we can provide these definitions:

- $\omega_{\{Emotion\}}$ is the initial set of emotion-related words in English, including the seeds we use to build the lexicon (i.e., the emotion nouns and their related adjectives), as well as their synonyms.
- $\Omega_{\{Emotion\}}$ contains the context sentences in the target language for the specific emotion, where the words from $\omega_{\{Emotion\}}$ have been used in parallel translated texts.
- $\omega'_{\{Emotion\}}$ is the set of words in the target language for the specific emotion.
- φ is the resulting lexicon containing the aggregation of $\omega'_{\{Emotion\}}$ with all the emotions.

5 Checking the Obtained Lexicons

To validate the set of English words ($\omega_{\{Emotion\}}$ from step one of Algorithm 1, we checked their precision against a gold-labelled lexicon for two different languages with very different origins and roots, namely, Spanish and German. As far as we know, currently the only manually labelled emotional English lexicon is *EmoLex* [17] with 14,182 words. Thus, we will use *EmoLex* as a reference dataset.

Table 2 shows the number of words labelled under certain emotions in *EmoLex* and how many of them can be found in the initial sets of emotional words we obtained in English. As expected, the initial sets contain a reduced number of words compared with those contained in *EmoLex*. For their part, Table 3 shows the precision computed for each emotion after using bab.la for the Spanish and German languages to obtain the initial sets in English. To compute the precision we followed Equation (1).

$$\text{precision}_{\{\text{emotion}\}} = \text{card}(\text{EmoLex}_{\{\text{emotion}\}} \cap \omega_{\{Emotion\}})/\text{card}(\omega_{\{Emotion\}}), \quad (1)$$

where $\text{EmoLex}_{\{\text{emotion}\}}$ is the subset from the reference dataset containing the words for the specific emotion, and $\omega_{\{\text{emotion}\}}$ is the set of words we have obtained for the same emotion.

Table 2. Number of words for each emotion in *EmoLex* and in the emotional words in English for Spanish and German.

		Spanish	German
Emotion	Emolex	Emolex$_{\{emotion\}} \cap \omega_{\{Emotion\})}$	Emolex$_{\{emotion\}} \cap \omega_{\{Emotion\}}$
Anger	1247	54	108
Disgust	1058	29	65
Fear	1476	60	191
Joy	689	57	83
Sadness	1191	68	140
Surprise	534	20	39

Table 3. Precision for each emotion for the initial sets of emotional words in English.

	Precision for Spanish	Precision for German
Anger	0.333	0.315
Disgust	0.379	0.246
Fear	0.317	0.204
Joy	0.404	0.337
Sadness	0.529	0.393
Surprise	0.350	0.231

Table 4. Number of missing words we can find in the obtained initial seeds of emotional words but which do not appear in *EmoLex*.

	Spanish	German
Anger	17	52
Disgust	9	22
Fear	19	68
Joy	28	44
Sadness	20	64
Surprise	5	15

Table 5. Number of retrieved context sentences and tokens per sentence scrapped from bab.la

			Tokens per sentence		
Emotion	Lang.	Nr. of sentences	mean	std	Var
Fear	German	38,303	13.831	4.954	24.541
	Spanish	11,440	15.981	5.599	31.344
Anger	German	6,695	14.563	7.082	50.149
	Spanish	5,326	18.326	8.249	68.047
Surprise	German	5,441	14.947	5.889	34.675
	Spanish	3,396	15.078	4.959	24.593
Disgust	German	5,394	15.561	5.930	35.167
	Spanish	3,134	18.668	7.256	52.654
Sadness	German	12,134	15.230	6.643	44.132
	Spanish	9,875	16.751	6.081	36.974
Joy	German	9,039	15.286	5.475	29.972
	Spanish	4,837	15.822	7.721	59.619

As we can observe, the obtained precision is a bit low. The reason for this can be found by looking inside the lexicon: although the initial sets clearly contain emotional words, those words are not contained in the *EmoLex* lexicon. Table 4 shows the number of seeds per emotion we have obtained but are missed in the *EmoLex* dataset. Details about the whole list of missing seeds can be found in Appendix A.

According to these results, it looks obvious that the initial sets contain a great percentage of emotional words related to their corresponding emotion, but currently there are is not a gold-labelled lexicon that we can use to validate them.

After performing step 2 of Algorithm 1 to obtain the context sentences for each emotion $\Omega_{\{Emotion\}}$ for German and Spanish as target languages, Table 5 summarizes the results. The table shows the number of context sentences we gathered from bab.la in December 2016 grouped by emotion and language. In addition, it presents the mean, standard deviation and variance for the number of tokens (words) per sentence, once we have removed acronyms, entities and abbreviations. As can be seen, there are differences in the number of gathered sentences per sentiment depending on the language. This is really significant because the resulting lexicon quality is directly related to the number of processed sentences.

Continuing applying step 3 of the algorithm to obtain the $\omega'_{\{Emotion\}}$, we retrieved the words for each language from the context sentences after removing the entities, abbreviations, acronyms, stop-words and domain-specific terms. As a result, we obtained the $\omega'_{\{Emotion\}}$ containing the words related to each one of the six emotions for German and Spanish.

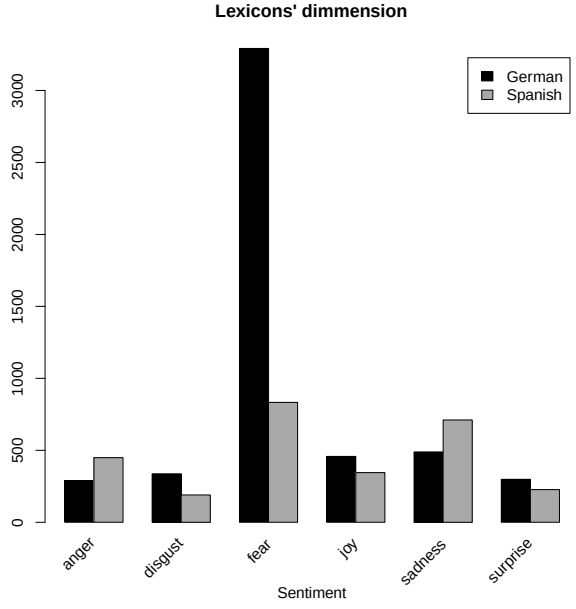

Fig. 1. Number of words per sentiment for the obtained lexicons.

By performing step 4 of the algorithm and obtaining the final lexicons, we got those words that appear at least twice, that is, in two sentences. As a result, Figure 1 shows the number of words per sentiment for the obtained φ lexicons for German and Spanish. As you can see, the number of words we obtained for each emotion may significantly differ among emotions and between languages.

6 Exploratory Analysis in a Parallel Corpora for German and Spanish

As previously stated, there is currently no gold-labelled lexicon that we can use to validate the obtained lexicons. Therefore, taking into account that most of the studies on Sentiment Analysis focused on documents [11], we conducted an exploratory Affective Analysis on a parallel corpus composed of well-known books in order to check the validity of the approach.

Thus, by checking the emotions identified in the texts as well as comparing the affective fingerprints between two different languages for the same books, we will see the viability of the proposal. In the next subsections, we will give the details and the initial results obtained.

6.1 Choosing the Corpus and Target Languages

The corpus we've used was developed in the OPUS project (Open Parallel corpUS) [32]. It constitutes a collection of translated aligned free texts from the web. In particular, from the corpora available in OPUS, we selected *"books"*, a collection of multilingually aligned copyright-free well-known books.

As target languages, we again selected German and Spanish. Accordingly, the books that we can find multilingually aligned in OPUS for those two languages are *Jane Eyre* by C. Bronte, *Alice in Wonderland* by L. Carroll, *Die Verwandlung* by F. Kafka, *The Fall of the House of Usher* by E.A. Poe, and *Anna Karenina* by L. Tolstoy. All of them are well-known titles.

6.2 Performing a Simple Emotional Identification to the Parallel Corpus

In order to check only the computed lexicon but not any classification algorithm, we simply performed a wordspotting approach to perform the Affective Analysis.

Accordingly, to avoid the effect of morphological change of the words, we repeated the algorithm to obtain a lexicon not of words but of stems, including those that appear at least ten times. Thus, we obtained a lexicon with stems categorized under a certain emotion. Later, to compute the affective fingerprint of a book, we simply increased the counter every time a stem categorized under a certain emotion appeared in the corresponding book. Lastly, we computed the percentage the stems with emotional charge represented in the whole book.

The obtained results are summarized in Figure 2. This figure shows a radar chart for each book with the six emotions in both target languages, namely, German (in black) and Spanish (in grey).

Recalling Figure 1, *fear* has over 3,000 words in the generated lexicon for the German language, but fewer than 1,000 for the Spanish language. Therefore, to avoid possible bias introduced by the size of the lexicon, in order to compare the emotions, values have been normalized for the German language following Equations (2), (3) and (4):

$$\%_{\{emotion\}} = 100 \text{ x } (\text{card}(Book_{\{emotion\}})/\text{card}(Book)) \quad (2)$$

$$factor_{\{emotion\}} = \text{card}(\varphi\text{-Spanish}_{\{emotion\}})/\text{card}(\varphi\text{-German}_{\{emotion\}}) \quad (3)$$

$$Normalized\ \%_{\{emotion\}} = \%_{\{emotion\}} \text{ x factor}_{\{emotion\}} \quad (4)$$

where $\%_{\{emotion\}}$ is the computed percentage of words from each book that are identified under a certain *emotion*, and $factor_{\{emotion\}}$ is the scaling factor in order to avoid lexicon size bias for the specific emotion.

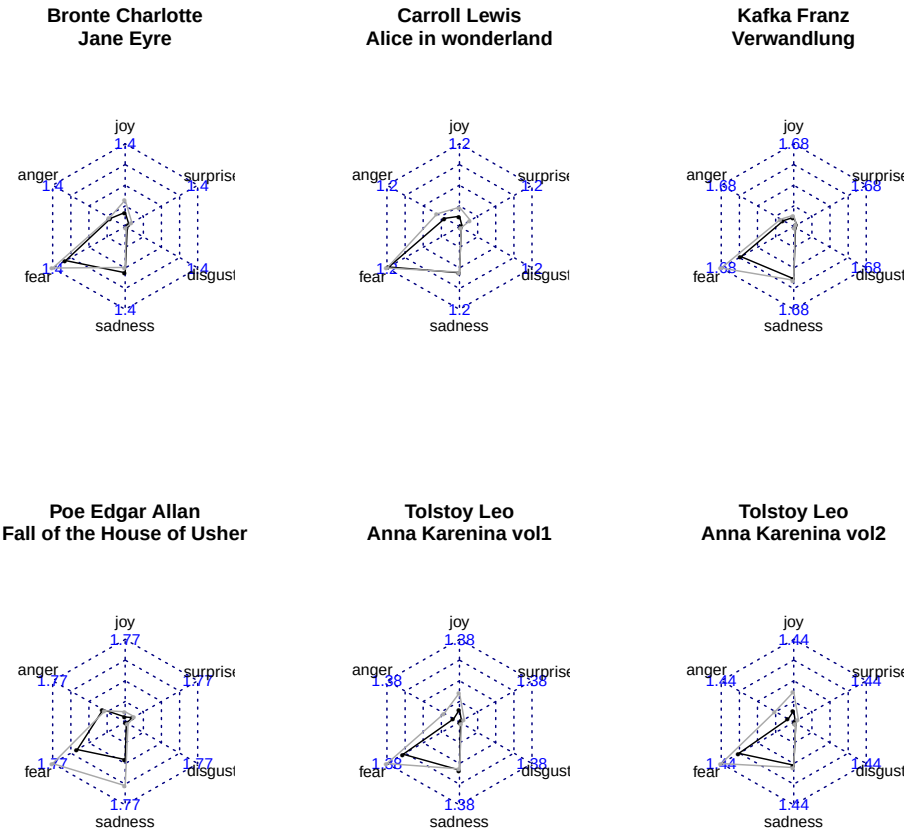

Fig. 2. Radar chart for the six emotions processed for each book in the German (black) and Spanish (grey) languages.

Because they are all well-known books, it is easy to check their emotional sign. Moreover, that affective fingerprint is really similar in both languages, that is, they present the same emotional tendency in both languages. It is interesting to notice that in fact *disgust* and *surprise* have almost no emotional load in the text. This agrees with the work performed by [33], who advise that the basic emotions are only four: *anger, fear, joy* and *sadness*.

6.3 Discussion and Further Work

Along with this article, we have detailed an approach for automatically generating target language independent lexicons to perform Affective Analysis. The approach starts with some initial seeds of nouns and adjectives for each emotion in English; these seeds are expanded by a bootstrapping process; and finally, the affective lexicon for the target language is obtained by processing the context sentence from the parallel text where the words for each emotion have been used.

We implemented the necessary routines that make use of bab.la, a language portal that provides multilanguage dictionaries, translations of common phrases and expressions, etc. in order to check our approach, to automatize the creation of sets of words with those words with similar meanings, to retrieve the context sentences and to perform the translation process.

Later, we conducted an exploratory analysis by applying the approach to a parallel corpus, checking the emotions identified in the texts and comparing the results between two different target languages. This exploratory analysis has shown that the approach provides coherent results and can be used to automatically generate target language independent lexicons to perform Affective Analysis.

However, although the conducted study has shown evidence about the benefits of the approach by providing us with an affective fingerprint in whole texts, it is necessary to improve its accuracy and enhance several critical points. The precision of the resulting lexicons depends on the quality of the context sentences as well as on the domain from which they are extracted. Validating the proposal using a set of context sentences for different specific domains will give us a better idea of the accuracy of our approach. In addition, it is necessary to test the proposed approach by using different bilingual dictionaries and translation tools such as Bing Translator, Google Translate, the Moses statistical machine translation system, and others.

7 Conclusions

In this article, we have presented an experience in automatically building lexicons for affective computing in multiple target languages. Starting with just a few seeds of words in English for each emotion we are interested in, we then bootstrapped them and retrieved context sentences from a parallel translated text where they had been used. By processing the translated sentences corresponding to the target language, we generated the affective lexicon for the target language. Although it is necessary to perform more work to get better precision, the exploratory analysis performed against parallel books in different languages has demonstrated its validity.

Acknowledgement. This research has been partially supported by the Ministry of Economy and Competitiveness (in Spanish *Ministerio de Economía y Competitividad*) through the projects REF: TIN2013-44586-R, TIN2014-52129-R and by the Ministry of Education, Youth and Sports, Community of Madrid (in Spanish *Consejería de Educación, Juventud y Deporte, Comunidad de Madrid*) through the project S2013/ICE-2715.

1. References

1. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Found. Trends Inf. Retr. 2, 1–135 (2008).
2. Liu, B.: Sentiment Analysis and Subjectivity. Presented at the (2010).
3. Feldman, R.: Techniques and Applications for Sentiment Analysis. Commun. ACM. 56, 82–89 (2013).
4. Piryani, R., Madhavi, D., Singh, V.K.: Analytical mapping of opinion mining and sentiment analysis research during 2000-2015. Inf. Process. Manag. 53, 122–150 (2017).
5. Jurado, F., Rodriguez, P.: Sentiment Analysis in monitoring software development processes: An exploratory case study on GitHub's project issues. J. Syst. Softw. 104, 82–89 (2015).
6. Saif, H., He, Y., Fernandez, M., Alani, H.: Contextual semantics for sentiment analysis of Twitter. Inf. Process. Manag. 52, 5–19 (2016).
7. Severyn, A., Moschitti, A., Uryupina, O., Plank, B., Filippova, K.: Multi-lingual opinion mining on YouTube. Inf. Process. Manag. 52, 46–60 (2016).
8. Cambria, E.: Affective Computing and Sentiment Analysis. IEEE Intell. Syst. 31, 102–107 (2016).
9. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. Presented at the (2002).
10. Liu, B.: Sentiment Analysis and Opinion Mining. Presented at the (2012).
11. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Syst. 89, 14–46 (2015).
12. Esuli, A., Sebastiani, F.: SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. Presented at the (2006).
13. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Presented at the (2010).
14. Wordnet: Wordnet, https://wordnet.princeton.edu/wordnet.
15. Strapparava, C., Valitutti, A.: WordNet-Affect: an affective extension of WordNet. Presented at the May (2004).
16. Bradley, M.M., Lang, P.J.: Affective Norms for English Words (ANEW): Instruction manual and affective ratings. (1999).
17. Mohammad, S.M., Turney, P.D.: Crowdsourcing a Word-Emotion Association Lexicon. Comput. Intell. 29, 436–465 (2013).
18. Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual Subjectivity Analysis Using Machine Translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 127–135. Association for Computational Linguistics, Stroudsburg, PA, USA (2008).
19. Gatti, L., Guerini, M., Turchi, M.: SentiWords: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis. IEEE Trans. Affect. Comput. 7, 409–421 (2016).
20. Strapparava, C., Mihalcea, R.: SemEval-2007 Task 14: Affective Text. Presented at the (2007).
21. Banea, C., Mihalcea, R., Wiebe, J.: Porting multilingual subjectivity resources across languages. IEEE Trans. Affect. Comput. 4, 211–225 (2013).
22. Balahur, A., Turchi, M.: Multilingual Sentiment Analysis Using Machine Translation? In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis. pp. 52–60. CEUR Workshop Proceedings, Stroudsburg, PA, USA (2012).

23. Balahur, A., Turchi, M.: Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Comput. Speech Lang. 28, 56–75 (2014).
24. Chen, Y., Skiena, S.: Building Sentiment Lexicons for All Major Languages. In: for Computational Linguistics, A. (ed.) Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. pp. 383–389. , Baltimore, Maryland, USA (2014).
25. Wiktionary, the free dictionary, https://www.wiktionary.org/.
26. Staiano, J., Guerini, M.: DepecheMood: a Lexicon for Emotion Analysis from Crowd-Annotated News. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. pp. 427–433. , Baltimore, Maryland, USA (2014).
27. Ekman, P.: Handbook of Cognition and Emotion. Presented at the (1999).
28. Strapparava, C., Mihalcea, R.: Learning to Identify Emotions in Text. Presented at the (2008).
29. Bhowmick, P.K., Mitra, P., Basu, A.: An Agreement Measure for Determining Inter-annotator Reliability of Human Judgements on Affective Text. In: Proceedings of the Workshop on Human Judgements in Computational Linguistics. pp. 58–65. Association for Computational Linguistics, Stroudsburg, PA, USA (2008).
30. Babla: Babla, http://bab.la.
31. TheOpenThesaurus: TheOpenThesaurus, https://www.openthesaurus.de.
32. Tiedemann, J.: Parallel Data, Tools and Interfaces in OPUS. In: Chair), N.C. (Conference, Choukri, K., Declerck, T., Dogan, M.U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S. (eds.) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). European Language Resources Association (ELRA), Istanbul, Turkey (2012).
33. Zinck, A., Newen, A.: Classifying emotion: a developmental account. Synthese. 161, 1–25 (2008).

Francisco Jurado is Lecturer in the Computer Engineering Department of the Universidad Autónoma de Madrid, Spain. He received his Ph.D. degree with honours in Computer Science from the University of Castilla-La Mancha in 2010. His research areas include Natural Language Processing, Intelligent Tutoring Systems, Heterogeneous Distributed eLearning Systems and Computer Supported Collaborative Environments.

Pilar Rodriguez is Associate Professor in the Computer Engineering Department of the Universidad Autónoma de Madrid. She received her Ph.D. degree in 1990, with a thesis on Computational Linguistics. She joined IBM in 1985, working at the IBM-UAM Scientific Center until 1989, when she was assigned to the Instituto de Ingeniería del Conocimiento, IIC. In November 1996 she joined UAM. At present, she is a member of the GHIA group at the UAM. Main research focuses on adaptive systems, especially for learning purposes, both in individual and collaborative environments.

Received: October 1, 2017; Accepted: September 30, 2018

Appendix A. Missing words in *Emolex* from the initial sets of emotional words in English

	Spanish	German
Anger	abuzz, angriness, annoyed, apotheosize, braveness, choler, courageousness, enraged, feral, heated-up, impassionated, involving, lawless, ricochet, tempestuous, thundery, vexed	aggravate, angrily, angriness, annoyed, blowy, blustering, bolshy, bombing, brawling, choler, cock-a-hoop, coltish, dissolute, ebullient, enrage, exasperated, extravagant, fantastic, feral, ferociously, great, groovy, hilariously, incensed, infuriated, infuriates, intemperate, irately, mean, miff, mindless, mustered, obstreperous, peeve, rabidly, rampant, random, rapturous, rioting, riotously, roaring, skittish, speeding, squally, sulfurous, sulphurous, super, tempestuous, vex, wrathful, wrathfully, wroth
Disgust	abhorrence, creeps, displease, nauseate, reluctancy, repel, repulse, sicken, uspet	abhorrence, abominate, appal, appall, cloy, disgruntle, displease, dissatisfy, exasperate, insubordination, insurgency, nauseate, odium, putsch, rebelliousness, repel, repugnance, repulse, scuff, sicken, spurn, vex
Fear	agonized, anguished, attend, belong, bow, brokenhearted, browbeat, bussy, carefulness, enshrine, fearfulness, heedfulness, incumbency, trouble, troubling, tyrannize, venerate, worrier, worrisome	accurateness, affect, agitate, angst, appertain, apprehensibility, apprehensiveness, attentiveness, biz, browbeat, carefulness, cause, cautiousness, chariness, consideration, cure, daunt, devotion, devotions, dignify, disconcertment, disquiet, disturb, edited, egis, elaborateness, enshrine, enshrinement, fearfulness, fosterage, funky, godawful, handling, harassment, heedfulness, honour, idolise, idolize, insurgency, involve, ministration, misgiving, mousey, mousy, neatness, nursing, obeisance, overawe, painstakingness, regard, regardfulness, revers, safe-keeping, scariness, shy, solicitousness, solicitude, therapy, timidly, trouble, turnout, unease, value, venerate, vex, wariness, wish, wonder
Joy	God, buoyant, cock-a-hoop, deity, delectation, elation, enjoyment, enthuse, euphoric, exhilarate, exhilarated, exult, gaiety, gladden, gleeful, gloat, godhead, jolly, joyfulness, joyousness, jubilate, jubilation, lively, luxuriate, mirthful, perky, playfulness, pleasure	blissfulness, bright, buoyant, chirpy, complacence, contentment, convivial, delectation, elate, elates, emboldened, encouraged, encourages, enjoyable, enthral, enthrall, exhilarated, exultant, exulting, gaiety, gay, gayety, gladden, gleeful, gloating, gloatingly, gratification, gratifying, gratifyingly, joyfully, joyfulness, joyousness, jubilantly, jubilate, jubilating, lively, nerved, please, pleasure, prosperousness, regale, sneering, zestful, zestfulness
Sadness	aggrieved, anguished, distressful, disturbing, gloominess, grief-stricken, grieved, heartbroken, lamentable, lugubriousness, paretic, pitiable, pitiful, regretful, sad, shamed, sorrowfulness, sorry, troubling, wistfulness	abominably, afflicting, afflictive, afflictively, afflicts, awkward, bewail, bitchiness, bitter, broody, caitiff, deplorably, depressingness, dim, disconcerting, dolefully, dolorousness, down-at-heel, dreariness, funereal, gloominess, grief-stricken, grilling, heartrending, heavy-hearted, hoodoo, lamentable, lamentably, lamentation, lugubrious, lugubriousness, meagre, mercifully, misadventure, niggling, penitence, piteous, pitiable, pitiably, pitiful, pitifully, pitying, plangent, regretfulness, regrettably, ruefulness, sad, saddening, serious, shattering, smart, sombreness, sorrowfulness, sorry, tearfulness, teariness, tormenting, trouble, troubled, troubling, upsetting, verminous, vexed, woebegone
Surprise	rainstorm, startled, surprisal, wonder, wonderment	amazed, amazes, astonish, astonished, besiege, dumbfound, dumfound, eye-opener, pelt, storminess, surprisal, welter, wonder, wondering, wonderment