The Impact of Machine Translation on Sentiment Analysis

Gayane Shalunts*, Gerhard Backfried*, Nicolas Commeignes*

*SAIL LABS Technology GmbH
Vienna, Austria
Email: {gayane.shalunts, gerhard.backfried, nicolas.commeignes}@sail-labs.com

***Abstract*—The article explores the impact of Machine Translation on sentiment analysis, employing the combination of two state-of-the-art tools - the multilingual sentiment analysis tool SentiSAIL and the Machine Translation tool SDL Language Weaver. The original corpora are in German, Russian and Spanish in the domain of general news. The output language of translation is English. Firstly, the work presents the development and evaluation of SentiSAIL features in a newly supported language - Spanish. Further experimental setup reveals that the performance rates of sentiment analysis on the original and translated corpora are comparable. Thus a given tool, performing high quality Machine Translation from a target language to English, can eliminate the necessity to develop specific sentiment analysis resources for that language.**

Keywords–Sentiment analysis; machine translation.

I. Introduction

Sentiment analysis refers to a classification task in the Natural Language Processing (NLP) community, the goal of which is commonly to determine the polarity (positive/negative) of the input text. Whereas subjectivity analysis deals with the detection of private states (opinions, emotions, sentiments, beliefs, speculations) [1], classifying the textual input as objective/subjective. The main parameters defining the scope of a sentiment analysis approach are the target language, domain and media type (traditional or social media). Due to automation and the ability to process big amounts of data, sentiment analysis has found a broad range of applications in marketing, e.g., monitoring of public opinions of product reviews [2] [3] [4], political science, e.g., observation of public opinions during election campaigns, social science, economics, etc. Generally, sentiment analysis approaches may be divided into lexicon-based and machine-learning-based groups [5]. In machine learning approaches labeled data is employed to train classifiers [5] [6]. The demand of costly labeled data and the narrow context of applicability are the major drawbacks of these methods. Lexicon-based methods use a predefined list of words as features, also referred to as *sentiment dictionary* or *lexicon*, where each word is associated with a specific sentiment [5]. Here, the challenging task is to obtain a sentiment dictionary applicable in various contexts. Thus lexicon-based methods are tuned to cover specific target domains and media types, as traditional media exhibits formal language and social media - colloquial language, slang.

Whereas the research field is very active, the majority of publications are limited to the domains of movie and product reviews in English only. Here a straightforward question arises, if the performance of the state-of-the-art Machine Translation (MT) systems allows to translate an input text in an original language into English and to apply sentiment analysis in English afterwards. The objective of the current work is to evaluate the effect of MT on sentiment analysis. The goal of the evaluation is to compare the performance of the SentiSAIL tool on original German, Russian and Spanish corpora and on the corresponding corpora in English, translated employing the MT tool SDL Language Weaver [7]. The performance of SentiSAIL on the original self-compiled corpora in German and Russian is reported in [8]. The current paper also contributes an equivalent annotated traditional media corpus in Spanish and evaluates the classification of SentiSAIL on it. The comparison examines the impact of two factors on sentiment analysis - the translation noise and the difference of sentiment lexicons in English and original languages. Note that the English sentiment lexicon is well-tested and more extensive than those in other languages, which is likely to lead to better performance in English. The comparison reveals equivalent performance rates of sentiment analysis on original and translated data, leading to a conclusion that the state-of-the-art MT systems can provide an alternative to the costly development of language features to realize sentiment analysis in multiple languages.

SentiSAIL is a multilingual sentiment analysis system [8]. It employed the methodology of the lexicon-based system SentiStrength [9] and expanded it into the domains of general and disaster related news multilingually. The SentiStrength and SentiSAIL features in English, German and Russian are compared in [8] on a self-compiled traditional media corpus, reporting SentiSAIL performance improvement to be slight in English and considerable in German and Russian. SentiSAIL is integrated into the SAIL LABS Media Mining System (MMS) [10], which is a state-of-the-art Open-Source-Intelligence system, incorporating speech and text-processing technologies. Sentiment analysis forms a part of MMS automatic multifaceted processing of multilingual unstructured textual and speech data.

The paper is organized as follows: Section II reviews the literature on the impact of MT on sentiment analysis, as well as sentiment analysis in German, Russian and Spanish. Section III clarifies SentiSAIL methodology and the development of Spanish resources. Section IV presents the experimental setup, performance evaluation and results. And finally, Section V draws conclusions from the work presented.

II. Literature review

The authors of [11] explore the impact of MT on sentiment analysis in French, German and Spanish. They employ three MT systems for comparison - Bing Translator [12], Google Translate [13] and Moses [14]. An original dataset in English was divided into training and testing sets. Afterwards corresponding training and testing datasets were generated by translating the original English data into French, German and Spanish by the three MT systems mentioned above. Classification models were trained and tested per language in two different experimental scenarios, using unigrams and bigrams as features. Firstly, training and testing datasets were created per language by each translator separately. The performances of the sentiment analysis on original English and translated corpora were comparable. The performance difference reached 8% in the worst case. Secondly, the corresponding training datasets generated by the three translators were combined together, resulting in increase of the noise level and performance drop. The paper concludes that the state-of-the-art MT systems are reliable enough for creating training data for languages other than English.

The approach in [2] experiments with polarity-annotated datasets in English and Turkish from the domain of movie and products reviews. The authors report that the polarity detection task is not affected considerably by the amount of the artificial noise introduced by MT. [15] proposes two approaches for mapping existing subjectivity resources in English to Romanian. The first approach builds the Romanian lexicon by translating the Opinion Finder lexicon [16] using a bilingual dictionary. The second approach generates a subjectivity-annotated corpus in Romanian by projecting annotations from an automatically annotated English corpus. The authors find out that the corpus projections preserve subjectivity more reliably than the lexicon translations. This observation was also made in their previous work, stating that subjectivity is a property associated not with words, but with word meanings [17].

Three further approaches for generating subjectivity resources in a target language from English are presented by [18]. The approaches on Romanian and Spanish show promising results, being comparable to those obtained using manually translated corpora. In the first approach the annotations of the Multi-Perspective Question Answering (MPQA) corpus are automatically translated, yielding subjectivity annotated sentences in Romanian. In the second one, they use the automatically translated entries in the Opinion Finder lexicon to annotate a set of sentences in Romanian. In the third experiment, the direction of translation is reversed to verify the assumption that subjective language can be translated and thus new subjectivity lexicons can be acquired for languages lacking such resources.

Another method to build lexicons for languages with scarce resources is presented by [19]. In this research, the authors apply bootstrapping to generate a subjectivity lexicon for Romanian, starting with a set of seed subjective entries, using electronic bilingual dictionaries and a training set of words.

The authors of [20] translate the MPQA corpus from English into 5 languages - Arabic, French, German, Romanian and Spanish. Their empirical results indicate that including multilingual information while modeling subjectivity is able not only to transfer English resources into other languages, but can also improve subjectivity classification in the source language itself. They showed that an English classifier was improved by using out-of-language features, achieving a 4.9% error reduction in accuracy with respect to using English alone. The work proposed by [3] constructs a polarity co-training system, using the multi-lingual views obtained through the automatic translation of English product-reviews into Chinese.

Further articles address sentiment analysis in languages, examined in the current work - German, Russian and Spanish. A German language sentiment analysis method, called SentimentWortschatz or SentiWS, is presented in [21]. The approach targets the domain of financial newspaper articles and respective blog posts on a German stock index [21]. The sentiment lexicon is developed from the General Inquirer (GI) lexicon [22] by semiautomatic translation into German using Google Translate and is manually revised afterwards. The lexicon post-translation revision included the removal of inappropriate words and addition of words from the finance domain [21]. The usage of the GI lexicon as a base is justified by the fact that it is widely accepted in the sentiment analysis community and has a broad coverage. Another method in German, introduced by [23], targets the domain concerning German celebrities. The approach utilizes the SentiStrength tool [9] and permits the classification of mixed sentiments. Here also the English opinion dictionary was automatically translated into German and manually improved afterwards by two German native speakers.

The publications [4] [24] illustrate the sentiment analysis research in Russian. Authors in [24] propose an approach for domain specific sentiment lexicon extraction in the meta-domain of products and services. [4] describes and evaluates the state-of-the-art sentiment analysis systems in Russian.

The authors of [25] present a lexicon-based sentiment analysis system in Spanish, called Sentitext, which employs three major feature sets - the dictionary of individual words, the dictionary of multiword expressions and the set of context rules. They conclude that the proper management and extensive coverage of multiword expressions is critical for successful textual sentiment analysis. Sentitext is also used in [26] to detect sentiments on Twitter messages in Spanish. [27] describes machine learning technique for opinion mining in blogs. The experimental Spanish corpus was created by their Emotiblog system. The authors of [28] adapt an existing English semantic orientation system [29] to Spanish, comparing several alternative approaches. Their experiments prove that although language-independent methods show decent baseline performance, automation cost is considerable and the development of language-specific knowledge and resources provides the best long-term improvement. [30] introduces a framework, where the Spanish lexicons derived from manually and automatically annotated English lexicons yield an accuracy of 90% and 74% respectively.

III. SentiSAIL methodology and support of Spanish

The SentiSAIL sentiment analysis tool, introduced in [8], performs processing of both traditional and social media data. The target domains in traditional media are the general news and particularly the coverage of disasters/crises in general news. In addition to English, German and Russian the current version of SentiSAIL supports also Spanish, French and Arabic. In this work we introduce SentiSAIL in Spanish - the development of Spanish resources and the performance evaluation on a self-compiled traditional media corpus (Section IV).

SentiSAIL employs the algorithm of SentiStrength [9]. SentiStrength, like [29], is a lexicon-based approach, using as the main feature list a lexicon of sentiment patterns associated with scores of positive or negative orientation. The positive patterns are weighed in the range [1; 5], the negative ones - [-5; -1] in a step 1, e.g., ”charming” 4, ”cruel” -4. To account for the formation of diverse words from the same stem (inflection and declension), stemming of the lexicon words is implemented. E.g., ”sympath∗” will match all the words starting with ”sympath”, e.g., ”sympathize”, ”sympathizes”, ”sympathized”, ”sympathy”, ”sympathetic”, etc. The text to be processed is treated as a Bag of Words, and each word is compared to the predefined stemmed lexicon patterns for matching. Employing unigram sentiment terms as the main feature introduces less noise during translation compared to higher level n-grams. In order to model the structure and semantics of the language observed the following additional feature lists are employed:

Boosters. Sentiment words may be intensified or weakened by words referred to as *boosters*. E.g., ”less charming” will weigh 3 and ”very cruel” -5.

Negations. It is assumed that negating a positive word inverts the sentiment to negative and weakens it twice, whereas negating a negative word neutralizes the sentiment. E.g., ”not charming” scores -2, whereas ”isn’t cruel” equals 0. The boosters and negations affect up to 2 following words.

Phrases and idioms. We define a *phrase* as a combination of words, expressing sentiment only in the given sequence, e.g., ”high quality” 3. An *idiom* is also a combination of words, but unlike a phrase it expresses a figurative, not literal meaning, e.g., ”crocodile tears” -2. Idioms and phrases score as a whole, overriding the scores of their component words. The sentiment lexicon comprises the polarity of individual words (*prior polarity*) [31]. The polarity of a word in a sentence (*contextual polarity*) may be different from its prior polarity [31] and is determined in the context of negations, boosters, phrases and idioms.

SentiSAIL, like [32], solves a dual classification task by classifying a text into one of the following 4 classes: positive, negative, mixed (both positive and negative) or neutral (neither positive, nor negative). The dual classification scheme is motivated by the human ability to experience positive and negative emotions simultaneously [33].

The class of the input text is determined as a result of taking the following steps:

1) The sentiment on the granularity level of line is determined by computing a pair of positive/negative scores using a combining algorithm. Three combining algorithms were applied with no significant difference on the final classification accuracy [8]. The algorithms are listed below:

a) *Maximization*. The scores of the most positive and the most negative terms of the line are assigned to its positive and negative scores respectively.

b) *Averaging*. Positive and negative scores of each line are calculated respectively as the average of its all positive and negative word scores.

c) *Aggregation*. Positive and negative scores of each line are obtained from respective aggregation of the scores of all positive and negative words of the line, bounded by the maximum positive and negative values.

2) The sentiment on the granularity level of document is calculated likewise as a pair of positive/negative scores by averaging the pairs of the positive/negative scores of all lines respectively.

3) The final sentiment class of a text is produced by double thresholding of the pair of the positive/negative scores on the granularity level document. The classification of the positive and negative classes is straightforward. Documents passing both thresholds are classified into the mixed class, those failing both thresholds are classified as neutral.

Though SentiStrength comprises the mentioned feature lists in 14 languages, the lexicons in languages other than English are short and lack stemming. The development of the SentiSAIL lexicon in Spanish comprised four stages.

At the first stage the initial SentiStrength short lexicon in Spanish (286 words) is taken as a base and improved. We prefer to revise and expand the lexicon manually, since automation introduces also false hits [28]. A Spanish native speaker went through the SentiStrength lexicon, performing stemming and removing incorrect terms.

At the second stage the patterns from the parallel SentiStrength lexicon in English were translated into Spanish, stemmed and scored manually. At this step automation is not realizable, as the stemmed patterns may not be meaningful words (e.g., sympath*). Though the automatic translation of meaningful words may also be ambiguous due to multiple meanings. In addition weights of equivalent words in different languages may also vary due to cultural factors.

At the third stage additional sentiment words were manually selected and added from the sentiment dictionary generated by [30].

The fourth stage of the lexicon extension aims to cover the domains of general news and natural disasters. A database of 100 articles in Spanish from the target domains was collected from the web with that purpose. Half of the articles were chosen randomly as the training dataset, from which domain-specific sentiment terms were manually compiled and added to the lexicon together with their associated scores. To obtain a richer lexicon articles covering diverse topics were chosen.

As a result SentiSAIL’s sentiment lexicon in Spanish grew from the initial 286 to 2654 patterns. Note for comparison that SentiSAIL English lexicon comprises currently 2830 patterns. The development of Spanish resources was concluded by revising and expanding the lists of negations, boosters, phrases and idioms. The support of new languages in SentiSAIL can be achieved by taking steps equivalent to those taken for Spanish feature creation.

SentiSAIL is implemented in Perl. SentiSAIL performance speed is proportional to $log_2 N$, where N is the number of sentiment lexicon patterns in the language processed. Logarithmic performance speed is the result of running binary search on the sentiment lexicon. As SentiSAIL is typically deployed in a near real-time environment, high speed is a requirement.

IV. Evaluation and Results

Evaluating sentiment analysis systems is challenging, since there is no single ground truth. Each person classifies the observed text into one of the available sentiment classes depending on his/her cultural and educational background, age, political views, current mood and emotional state, etc. Thus the relation of the average inter-annotator agreement rate to

TABLE I. PERFORMANCE EVALUATION OF SENTISAIL IN SPANISH.

	Annotator 1	Annotator 2	Annotator 3	Average
Training set				
Annotator 1	-	76%	81%	**78%**
Annotator 2	-	-	77%	
SentiSAIL (Aggregation)	73%	65%	72%	70%
SentiSAIL (Averaging)	79%	71%	72%	74%
SentiSAIL (Maximization)	83%	73%	76%	**77.3%**
SentiSAIL (Maximization, SentiStrength features)	47%	41%	44%	44%
Testing set				
Annotator 1	-	77%	78%	**76%**
Annotator 2	-	-	73%	
SentiSAIL (Aggregation)	73%	68%	75%	72%
SentiSAIL (Averaging)	73%	66%	71%	70%
SentiSAIL (Maximization)	77%	74%	75%	**75.3%**
SentiSAIL (Maximization, SentiStrength features)	46%	43%	50%	46.3%

the average SentiSAIL-annotator agreement rate is chosen as an evaluation criterion for SentiSAIL. If the mentioned average rates are comparable, the performance of SentiSAIL system is considered as good as that of a human annotator.

The experimental setup comprises two stages. The first stage evaluates SentiSAIL's performance of the newly supported language - Spanish. The objective of the second stage is to compare SentiSAIL performance on the original German and Russian datasets, illustrated in [8], and on the Spanish dataset, introduced newly in this paper to the performance on the parallel corpora translated into English. The translations were performed automatically using the SDL Language Weaver (5.3.32 release), which is a statistical state-of-the-art MT tool [7]. The statistical translation models are generated automatically by applying machine learning technique on parallel collections of human translations.

The performance evaluation of SentiSAIL is reported in [8] on self-collected and labeled trilingual text corpus. The training dataset includes 32 news articles in English, 32 - in German and 48 - in Russian. The testing dataset comprises 50 news articles in each language. Since SentiSAIL is a lexicon-based method (as opposed to a machine learning based one), the training dataset was employed to extract additional domain-specific sentiment words manually, but not to train a classifier. We introduce an equivalent corpus in Spanish, comprising 100 traditional media articles, divided equally into training and testing datasets.

Table I details the experiments on the Spanish corpus. Identical experiments are conducted on training and testing datasets separately to show that the performance rates on the training dataset and unfamiliar data are comparable. Both training and testing datasets were labeled by 3 annotators by sentiment class labels (Positive, Negative, Neutral, Mixed). The average agreement rate among 3 annotators on training texts reached 78% (Table I). The following 3 lines in Table I present the agreement rates of SentiSAIL with the annotators, using the line scoring algorithms Aggregation, Averaging and Maximization in sequence. The best scoring algorithm is Maximization with 77.33% rate, which is competitive with the average human agreement rate of 78%. The next row in Table I shows that the improvement of the Spanish lexicon by SentiSAIL over the initial SentiStrength lexicon is considerable, having improved the performance rate from 44% to 77.33%. The main reason is that the initial SentiStrength lexicon is very short (286 words) and lacks stemming. The majority of sentiment terms are not detected and the classification is neutralized (42 out of 50 texts were classified as neutral). SentiSAIL achieves equivalent performance accuracy while running the same set of experiments on a previously unseen dataset (Testing set section in Table I).

The second stage of the experimental setup evaluates the impact of translation on the trilingual corpus. Since the Maximizaton algorithm scored the highest, it is chosen in the further experiments. Table II shows that the average inter-annotator agreement rate on the German training dataset scored 79.17% and SentiSAIL-annotators average agreement rate even outperforms it with 81.25% [8]. Running the equivalent experiment in English, i.e., performing sentiment analysis on the German into English translated corpus and using the English sentiment lexicon, yielded exactly the same average performance accuracy - 81.25% (Table II). Whereas the average performance rate on the original Russian corpus scored 82.99%, the equivalent rate on the English translated corpus decreased slightly to 80.9% (Table II). The third portion in Table II reports the empirical results on the newly supported language - Spanish. The average SentiSAIL-annotators agreement rate scored 77.33%, which is almost as high as the inter-annotator rate (78%). The average accuracy on the parallel English corpus recorded the highest decrease of 5% among 3 translated languages.

Table III presents the results of the identical experimental setup as Table II, but on testing datasets. Here the performance rate drop as an outcome of English translation of the trilingual corpora remains within negligible 1%. Table III also shows that SentiSAIL analysis accuracy on unfamiliar and training data are comparable.

V. CONCLUSION

Firstly, the work presented the development and evaluation of Spanish resources for the multilingual sentiment analysis tool SentiSAIL. Secondly, it explored empirically the impact of MT on sentiment analysis performance. The translation quality of the SDL Language Weaver allowed to achieve equivalent performance rates on original and translated parallel corpora while performing bipolar sentiment analysis by SentiSAIL. The original corpora were compiled in the traditional media domain in German, Russian and Spanish. The translation output language was English, supported by the majority of the state-of-the-art sentiment analysis systems. The performance decrease in the worst case remained within negligible 5%. The

TABLE II. PERFORMANCE EVALUATION ON THE ORIGINAL GERMAN, RUSSIAN, SPANISH TRAINING DATASETS AND THE EQUIVALENT ENGLISH TRANSLATIONS [8].

	Annotator 1	Annotator 2	Annotator 3	Average
German training set				
Annotator 1	-	78.1%	79.7%	**79.2%**
Annotator 2	-	-	79.7%	
German original	92.2%	73.4%	78.2%	81.3%
English translation	92.2%	73.4%	78.2%	81.3%
Russian training set				
Annotator 1	-	84.4%	79.2%	**82%**
Annotator 2	-	-	82.3%	
Russian original	86.5%	84.4%	78.1%	83%
English translation	85.4%	82.3%	75%	80.9%
Spanish training set				
Annotator 1	-	76%	81%	**78%**
Annotator 2	-	-	77%	
Spanish original	83%	73%	76%	77.3%
English translation	78%	70%	69%	72.3%

TABLE III. PERFORMANCE EVALUATION ON THE ORIGINAL GERMAN, RUSSIAN, SPANISH TESTING DATASETS AND THE EQUIVALENT ENGLISH TRANSLATIONS [8].

	Annotator 1	Annotator 2	Annotator 3	Average
German testing set				
Annotator 1	-	85%	76%	**76.7%**
Annotator 2	-	-	69%	
German original	81%	80%	77%	79.3%
English translation	80%	83%	72%	78.3%
Russian testing set				
Annotator 1	-	93%	93%	**92.7%**
Annotator 2	-	-	92%	
Russian original	92%	88%	90%	90%
English translation	90%	89%	89%	89.3%
Spanish testing set				
Annotator 1	-	77%	78%	**76%**
Annotator 2	-	-	73%	
Spanish original	77%	74%	75%	75.3%
English translation	76%	71%	80%	75.7%

conclusion drawn as an outcome of the extensive experimental setup is that substituting multilingual sentiment analysis by English sentiment analysis via MT may be an acceptable alternative, leading to inconsiderable performance drop. Such a setup may be advantageous when lacking the appropriate resources for a particular language and when fast deployment is crucial. In practical terms, the trade-off between the cost of the MT system and the effort for the development of language specific resources needs to be taken into consideration.

Future work will be in the direction of extending the list of languages further and evaluating the performance on data from multilingual social media platforms.

REFERENCES

[1] A. Montoyo, P. Martnez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," Decision Support Systems, vol. 53, no. 4, 2012, pp. 675 – 679.

[2] E. Demirtas and M. Pechenizkiy, "Cross-lingual Polarity Detection with Machine Translation," in Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, ser. WISDOM '13. New York, NY, USA: ACM, 2013, pp. 9:1–9:8.

[3] X. Wan, "Co-training for cross-lingual sentiment classification," in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (ACL-IJCNLP). Suntec, Singapore: ACL, 2009, pp. 235–243.

[4] I. Chetviorkin and N. Loukachevitch, "Evaluating Sentiment Analysis Systems in Russian," in Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, Sofia, Bulgaria, 2013, pp. 12–17.

[5] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and combining sentiment analysis methods," in Proceedings of the 1st ACM Conference on Online Social Networks (COSN 2013). Boston, USA: ACM, 2013, pp. 27–38.

[6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques." in Proceedings of the ACL conference on Empirical methods in natural language processing (EMNLP '02), Philadelphia, PA, USA, 2002, pp. 79–86.

[7] R. Soricu, N. Bach, and Z. Wang, "The SDL Language Weaver Systems in the WMT12 Quality Estimation Shared Task," in Proceedings of the 7th Workshop on Statistical Machine Translation. Montreal, Canada: 2012 Association for Computational Linguistics, 2012, pp. 145–151.

[8] G. Shalunts and G. Backfried, "SentiSAIL: Sentiment Analysis in English, German and Russian," in Proceedings of the 11th International Conference on Machine Learning and Data Mining, ser. MLDM '15, Hamburg, Germany, 2015, pp. 87–97.

[9] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment Strength Detection in Short Informal Text," J. American Society for Information Science and Technology, vol. 61, no. 12, Dec. 2010, pp. 2544–2558.

[10] G. Backfried et al., "Open source intelligence in disaster management," in Proceedings of the European Intelligence and Security Informatics Conference (EISIC). Odense, Denmark: IEEE Computer Society, 2012, pp. 254–258.

[11] A. Balahur and M. Turchi, "Multilingual Sentiment Analysis Using Machine Translation?" in Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, ser. WASSA '12. Stroudsburg, PA, USA: Association for Computational Linguistics,

2012, pp. 52–60.

[12] "Bing Translator," http://www.bing.com/translator, accessed: 2016-07-31.

[13] "Google Translate," https://translate.google.com, accessed: 2016-07-31.

[14] P. Koehn et al., Moses: Open Source Toolkit for Statistical Machine Translation. Association for Computational Linguistics, 6 2007, pp. 177–180.

[15] R. Mihalcea, C. Banea, and J. Wiebe, "Learning Multilingual Subjective Language via Cross-Lingual Projections," in Proceedings of the Association for Computational Linguistics (ACL), Prague, Czech Republic, 2007, pp. 976–983.

[16] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," in Proceedings of HLT/EMNLP, Vancouver, Canada, 2005, pp. 347–354.

[17] J. Wiebe and R. Mihalcea, "Word sense and subjectivity," in Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. Sydney, Australia: ACL, 2006, pp. 1065–1072.

[18] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, "Multilingual subjectivity analysis using machine translation," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '08). Honolulu, Hawaii: ACL, 2008, pp. 127–135.

[19] C. Banea, J. M. Wiebe, and R. Mihalcea, "A bootstrapping method for building subjectivity lexicons for languages with scarce resources," 2008.

[20] C. Banea, R. Mihalcea, and J. Wiebe, "Multilingual subjectivity: Are more languages better?" in Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10). Beijing, China: ACL, 2010, pp. 28–36.

[21] R. Remus, U. Quasthoff, and G. Heyer, "Sentiws - a german-language resource for sentiment analysis," in Proceedings of the 7th conference on International Language Resources and Evaluation (LREC), Valletta, Malta, 2010, pp. 1168–1171.

[22] P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie, The General Inquirer: A Computer Approach to Content Analysis. Cambridge, MA: MIT Press, 1966.

[23] S. Momtazi, "Fine-grained german sentiment analysis on social media," in Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC). Istanbul, Turkey: European Language Resources Association (ELRA), 2012, pp. 1215–1220.

[24] I. Chetviorkin and N. Loukachevitch, "Extraction of Russian Sentiment Lexicon for Product Meta-Domain," in Proceedings of the 24th International Conference on Computational Linguistics (COLING), Bombay, India, 2012, pp. 593–610.

[25] A. Moreno-Ortiz, C. Perez-Hernandez, and M. A. Del-Olmo, "Managing multiword expressions in a lexicon-based sentiment analysis system for spanish," in Proceedings of the 9th Workshop on Multi-word Expressions, Atlanta, Georgia, USA, 2013, pp. 1–10.

[26] A. Moreno-Ortiz and C. P. Hernández, "Lexicon-based sentiment analysis of twitter messages in spanish," Procesamiento del Lenguaje Natural, vol. 50, 2013, pp. 93–100.

[27] E. Boldrini, A. Balahur, P. Martnez-Barco, and A. Montoyo, "Emotiblog: an annotation scheme for emotion detection and analysis in non-traditional textual genres," in Proceedings of the 5th International Conference on Data Mining (DMIN). Las-Vegas, USA: CSREA Press, 2009, pp. 491–497.

[28] J. Brooke, M. Tofiloski, and M. Taboada, "Cross-linguistic sentiment analysis: From English to Spanish," in Proceedings of Recent Advances in Natural Language Processing (RANLP). Borovets, Bulgaria: RANLP 2009 Organising Committee / ACL, 2009, pp. 50–54.

[29] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational Linguistics, vol. 37, no. 2, 2011, pp. 267–307.

[30] R. M. Veronica Perez Rosas, Carmen Banea, "Learning Sentiment Lexicons in Spanish," in Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC). Istanbul, Turkey: European Language Resources Association (ELRA), 2012, pp. 3077–3081.

[31] M. M. S. Missen, M. Boughanem, and G. Cabanac, "Opinion mining: Reviewed from word to document level," Social Network Analysis and Mining, vol. 3, no. 1, 2013, pp. 107–125.

[32] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," Computational Linguistics, 2009, pp. 399–433.

[33] G. J. Norman et al., "Current emotion research in psychophysiology: The neurobiology of evaluative bivalence," Emotion Review, vol. 3, no. 3, 2011, pp. 349–359.