Cross-Lingual Projections vs. Corpora Extracted Subjectivity Lexicons for Less-Resourced Languages

Xabier Saralegi, Iñaki San Vicente, and Irati Ugarteburu

Elhuyar Foundation,
Osinalde Industrialdea 3,
20160 Usurbil, Spain
{x.saralegi,i.sanvicente,i.ugarteburu}@elhuyar.com

Abstract. Subjectivity tagging is a prior step for sentiment annotation. Both machine learning based approaches and linguistic knowledge based ones profit from using subjectivity lexicons. However, most of these kinds of resources are often available only for English or other major languages. This work analyses two strategies for building subjectivity lexicons in an automatic way: by projecting existing subjectivity lexicons from English to a new language, and building subjectivity lexicons from corpora. We evaluate which of the strategies performs best for the task of building a subjectivity lexicon for a less-resourced language (Basque). The lexicons are evaluated in an extrinsic manner by classifying subjective and objective text units belonging to various domains, at document- or sentence-level. A manual intrinsic evaluation is also provided which consists of evaluating the correctness of the words included in the created lexicons.

Keywords: Sentiment Analysis, Subjectivity Detection, Less Resourced Languages.

1 Introduction

Opinion mining or sentiment analysis are tasks involving subjectivity detection and polarity estimation. Both tasks are necessary in many sentiment analysis applications, including sentiment aggregation and summarization or product comparisons. Researchers into sentiment analysis have pointed out the frequent benefit of a two-stage approach, in which subjective instances are distinguished from objective ones, after which the subjective instances are further classified according to polarity ([1,2,3]). Pang and Lee [2] obtain an improvement from 82.8% to 86.4% for polarity classification by applying a subjectivity classifier in advance. So, developing a method for subjectivity detection seems an adequate first step for building an Opinion mining system for a certain language.

When dealing with subjectivity, some authors proposed rule-based methods [4] which use subjectivity lexicons. Other authors propose supervised methods based on machine learning techniques [1]. In both cases, subjectivity lexicons

A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 96–108, 2013.

are an important knowledge resource. So it is clear that subjectivity lexicons are a key resource for tackling this task. Nowadays, there are widely used lexicons, such as OpinionFinder [5], Sentiwordnet [6] and General Inquirer [7], but, as is the case with many NLP resources, those lexicons are geared towards major languages. This means that new subjectivity lexicons must be developed when dealing with many other languages.

As manual building is very costly and often uneconomic for most languages, especially less-resourced languages, machine building methods offer a viable alternative. In that sense, several methods [8,9,10,11,12] have been proposed for building subjectivity lexicons. The methods rely on two main strategies: building the lexicon from corpora or trying to project existing subjectivity resources to a new language. The first approach often produces domain specific results, and so, its performance in out-of-domain environments is expected to be poorer. Projecting a lexicon to another language would produce a resource that would *a priori* be more consistent in all environments. However, as the projection involves a translation process, the errors ocurring at that step could reduce the quality of the final lexicon as shown by Mihalcea et al. [10].

In our research we compared these two cost-effective strategies for building a subjectivity lexicon for a less-resourced language. We assumed that for languages of this type the availability of parallel corpora and MT systems is very limited, and that was why we avoided using such resources. Our contribution lies in a robust cross-domain evaluation of the two strategies. This experiment was carried out using Basque. First, we compared the correctness of the resulting lexicons at word level. Then, the lexicons were applied in a task to classify subjectivity and objectivity text units belonging to different domains: newspapers, blogs, reviews, tweets and subtitles.

The paper is organized as follows. The next chapter offers a brief review of the literature related to this research, and discusses the specific contributions of this work. The third section presents the resources we used for building the subjectivity lexicons, the experiments we designed and the methodology we followed. In the fourth chapter, we describe the different evaluations we carried out and the results obtained. Finally, some conclusions are drawn and we indicate some future research directions.

2 State of the Art

Wilson et al. [13] define a subjective expression as any word or phrase used to express an opinion, emotion, evaluation, stance, speculation, etc. A general covering term for such states is private state. Quirk et al. [14] define a private state as a state that is not open to objective observation or verification: "a person may be observed to assert that God exists, but not to believe that God exists". Belief is in this sense 'private'. So, subjectivity tagging or detection consists of distinguishing text units (words, phrases sentences...) used to present opinions and other forms of subjectivity from text units used to objectively present factual information. Detection is part of a more complex task which

Wilson [15] called subjectivity analysis, which consists of determining when a private state is being expressed and identifying the attributes of that private state. Identifying attributes such as the target of the opinion, the polarity of the subjective unit or its intensity, is outside the range of this work.

2.1 Subjectivity Detection Methods

Methods for subjectivity detection can be divided into two main approaches. Rule-based methods which rely on subjectivity lexicons, and supervised methods based on classifiers trained from annotated corpora.

Wiebe et al. [16] use manually annotated sentences for training Naive Bayes classifiers. Pang and Lee [2] successfully apply Naive Bayes and SVMs for classifying sentences in movie reviews. Wang and Fu [17] present a sentiment density-based naive Bayesian classifier for Chinese subjectivity classification. Das and Bandyopadhyay [18] propose a Conditional Random Field (CRF)-based subjectivity detection approach tested on English and Bengali corpora belonging to multiple domains.

Lexicon-based systems are also proposed in the literature. Turney [8] computed the average semantic orientation of product reviews based on the orientation of phrases containing adjectives and adverbs. The classifier proposed by Riloff and Wiebe [4] uses lists of lexical items that are good subjectivity clues. It classifies a sentence as subjective if it contains two or more of the strongly subjective clues. Das and Bandyopadhyay [19] proposed a classifier which uses sentiment lexicons, theme clusters and POS tag labels.

A third alternative would be to combine both approaches. Yu and Hatzivassiloglou [1] obtain 97% precision and recall using a Bayesian classifier that uses lexical information. This proves that subjectivity lexicons are indeed important resources.

According to Yu and Kübler [20], opinion detection strategies designed for one data domain generally do not perform well in another domain, due to the variation of the lexicons across domains and different registers. They evaluated the subjectivity classification in news articles, semi-structured movie reviews and blog posts using Semi-Supervised Learning (SSL) methods, and obtained results that vary from domain to domain. Jijkoun and de Rijke [21] propose a method to automatically generate subjectivity clues for a specific topic by extending a general purpose subjectivity lexicon.

2.2 Methods for Subjectivity Lexicon Building

Text corpora are useful for obtaining subjectivity and polarity information associated with words and phrases. Riloff et al. [22] adopt a bootstrapping strategy based on patterns to extend a seed set of 20 terms classified as strongly subjective. Baroni and Vegnaduzzo [23] apply the PMI (Pointwise Mutual Information) method to determine term subjectivity. Subjectivity level is measured according to the association degree with respect to a seed set of 35 adjectives marked as subjective.

When tackling the problem of the lack of annotated corpora, many authors propose using MT techniques. Mihalcea and others [10] annotate an English corpus using OpinionFinder [5] and use cross-lingual projection across parallel corpora to obtain a Romanian corpus annotated for subjectivity. Following the same idea, Banea et al. [11] use machine translation to obtain the required parallel corpora. In this case they apply the method for Romanian and Spanish. Wan [12] also proposed the generation of Chinese reviews from English texts by Machine Translation.

Another approach to building a subjective word list in a language is the translation of an existing source language lexicon by using a bilingual dictionary. Mihalcea et al. [10] used a direct translation process to obtain a subjectivity lexicon in Romanian. Their experiments concluded that the Romanian subjectivity clues derived through translation are less reliable than the original set of English clues, due to ambiguity errors in the translation process. Das and Bandyopadhyay [18] proposed improving the translation of ambiguous words by using a stemming cluster technique followed by SentiWordNet validation. Jijkoun and Hofmann [24] apply a PageRank-like algorithm to expand the set of words obtained through machine translation.

Banea et al. [25] compare different methods of subjectivity classification for Romanian. Among subjectivity lexicon building methods, there are bootstrapping a lexicon by using corpus-based word similarity, and translating an existing lexicon. They conclude that the corpus-based bootstrapping approach provides better lexicons than projection.

In this work we wanted to analyse strategies for developing a subjectivity lexicon for a Less-Resourced Language. We assumed that such languages can only avail themselves of monolingual corpora and bilingual lexicons. So parallel corpora, MT system-based approaches and approaches based on large subjectivity annotated corpora are not contemplated. We focused on a corpus-based approach and projection onto the target language.

3 Experiments

Projection-based lexicon building requires a subjectivity lexicon L_S_s in a source language s and a bilingual dictionary $D_{s \to t}$ from s to the target language t. In our experiments we took the English subjectivity lexicon (L_S_{en}) introduced in [5] as a starting point. L_S_{en} contains 6,831 words (4,743 strong subjective and 2,188 weak subjective). According to the authors, those subjective words were collected from manually developed resources and also from corpora. Strong subjective clues have subjective meanings with high probability, and weak subjective clues have a lower probability of having subjective meanings. As for the bilingual dictionary, a bilingual English-Basque dictionary $D_{en \to eu}$ which includes 53,435 pairs and 17,146 headwords was used.

Corpora-based lexicon extraction requires subjective and objective corpora. Subjective and objective corpora can be built by using simple heuristics. News from newspapers or Wikipedia articles can be taken as objective documents.

Opinion articles from newspapers can be taken as subjective articles. Those heuristics are not trouble free, but then again, they allow us to create low-cost annotated corpora. Using news as an objective corpus can be a rough heuristic because, according to Wiebe et al. [26], many sentences (44%) included in news are subjective. On the other hand, as Wikipedia belongs to a different domain from that of newspaper opinion articles, some divergent words can be incorrectly identified as subjective if we compare a Wikipedia corpus with a subjective corpus comprising opinion articles, due to the fact that they are a feature in the journalism domain but not in Wikipedia texts.

We built a subjective corpus TC_S_{eu} by taking 10,661 opinion articles from the Basque newspaper Berria[1]. Two objective corpora were built: one by collecting 50,054 news items from the same newspaper TCN_O_{eu}, and the other by gathering all the articles (143,740) from the Basque Wikipedia TCW_O_{eu}. A subset of TCN_O_{eu} containing the same number of articles as TC_S_{eu} was also prepared for parameter tuning purposes which we will name $TCN\_O'_{eu}$.

3.1 Cross-Lingual Projection of the Subjectivity Lexicon

We translated the English subjectivity lexicon L_S_{en} by means of a bilingual dictionary $D_{en \rightarrow eu}$ to create a Basque subjectivity lexicon L_P_{eu}. Ambiguities are resolved by taking the first translation[2]. Using this method we obtained translations for 36.67% of the subjective English words: L_P_{eu} includes 1,402 strong and 1,169 weak subjective words. The number of translations obtained was low, especially for strong subjective words. Most of these words are inflected (e.g., *"terrified"*, *"winners"*, ...) forms or derived words where prefixes or suffixes have been added (e.g., *"inexact"*, *"afloat"*, ...).

According to Mihalcea et al. [10] translation ambiguity is another problem that distorts the projection process. In their experiments Romanian subjectivity clues derived through translation were less reliable than the original set of English clues. In order to measure to what extent that problem would affect our projection, we randomly selected 100 English words and their corresponding translations. Most of the translations (93%) were correct and subjective according to a manual annotation involving two annotators (97% inter-tagger agreement, Cohen's k=0.83). So we can say that the translation selection process is not critical. We annotated as correct translations those corresponding to the subjective sense of the English source word. Unlike Mihalcea et al. [10], we did not analyse whether the translated word had less subjective connotation than the source word.

3.2 Corpus-Based Lexicon Building

Our approach was based on inferring subjective words from a corpus which includes subjective and objective documents. So, we identified as subjective words

[1] `http://berria.info`

[2] The bilingual dictionary has its translations sorted according to their frequency of use, so the first translation method should provide us with the most common translations of the source words.

those whose relevance in subjective documents is significantly higher than in objective documents. We adopted a corpus-based strategy, because it is affordable and easily applicable to less-resourced languages. We extracted Basque subjectivity lexicons in accordance with various relevance measures and objective corpora. TC_S_{eu} was used as the subjective corpus, and TCW_O_{eu} (Wikipedia) or TCN_O_{eu} (News) as objective corpora. For each word w in the subjective corpus we measured its degree of relevance with respect to the subjective corpus as compared with the objective corpus. That way we obtained the most salient words in a certain corpus, the subjective corpus in this case. We took that degree of relevance as the subjectivity degree $bal(w)$. That degree was calculated by the Log Likelihood ratio (LLR) or by the percentage difference ($\%DIFF$). Maks and Vossen [27] compared LLR and $\%DIFF$ for that purpose, and obtained better results by using $\%DIFF$.

In order to evaluate the adequacy of the measurements (LLR or $\%DIFF$) and the various corpus combinations (Wikipedia or News for the objective part), we analysed how subjective and objective words are distributed through the rankings corresponding to the different combinations (*LLR_News*, *DIFF_News*, *DIFF_Wiki* and *LLR_Wiki*). For that aim, two references were prepared. The first one includes only subjective words, while the second one includes both objective and subjective words. The first reference was built automatically by taking the strong subjective words of L_P_{eu}. For the second reference three annotators manually tagged subjective and objective words in a sample of 500 words selected randomly from the intersection of all candidate dictionaries (*DIFF_Wiki*, *DIFF_News*, *LLR_Wiki* and *LLR_News*). The overall inter-agreement between the annotators was 81.6% (Fleiss' k=0.63). Simple majority was used for resolving disagreements (27% of the words evaluated).

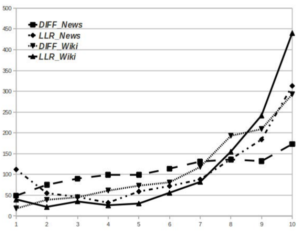

Fig. 1. Distribution of subjective words with various measure and corpus combinations

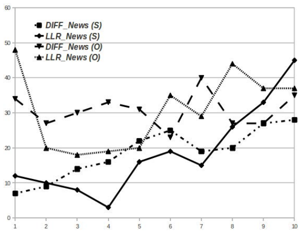

Fig. 2. Distribution of subjective and objective words using TCN_O_{eu} as objective corpus

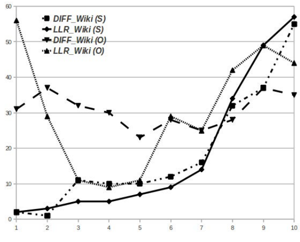

Fig. 3. Distribution of subjective and objective words using TCW_O_{eu} as objective corpus

According to the results shown in Figures 1, 2 and 3 Wikipedia seems to be a more adequate objective corpus. It provides a higher concentration of subjective words in the first positions of the rankings[3] (i.e. last intervals) than News when

[3] In Figures 1, 2, 3 and 4, higher intervals contain words scoring higher in the rankings.

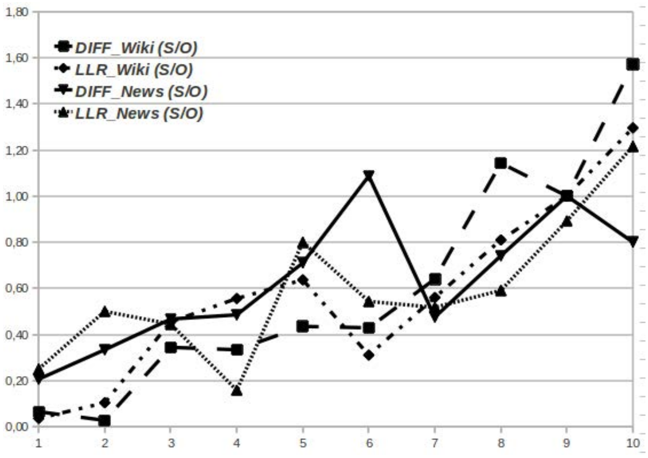

Fig. 4. Subjective/objective ratio with respect to ranking intervals

using both measurements and for both references. In addition, the concentration of objective words in the first positions is slightly lower when using TCW_O_{eu}, compared with using TCN_O_{eu} as the objective reference corpus.

Regarding the measurements, LLR provides better distributions of subjective words than $\%DIFF$ for both reference corpora. The highest concentration of the subjective words is in the first positions of the rankings. However $\%DIFF$ seems to be more efficient for removing objective ones from first ranking positions. Figure 4 plots the distribution of subjective/objective word rates across different ranking intervals. The best ratio distribution is achieved by the $\%DIFF$ measurement when used in combination with TCW_O_{eu}.

In terms of size, corpora-based lexicons are bigger than the projection-based one. For high confidence thresholds, $LLR > 3.84$, *p-value*< 0.05; and $\%DIFF > 100$ [27], corpora-based lexicons provide 9,761; 6,532; 8,346 and 6,748 words for *DIFF_Wiki*, *DIFF_News*, *LLR_Wiki* and *LLR_News*, respectively. These will be the dictionaries used in the evaluation presented in the next section. The sizes of these dictionaries are close to that of the source English lexicon L_S_{en} (6,831 words). However, after projecting it to Basque, this number goes down to 2,571. So it seems that the corpora-based strategy provides bigger subjectivity lexicons. Then again, we have to take into account that corpus-based lexicons include several objective words (See Figure 1.). In addition, corpus-based lexicons are biased towards the domain of journalism.

4 Evaluation

4.1 Classifier

In this work, we adopted a simple lexicon-based classifier similar to the one proposed in [28]. We propose the following ratio for measuring the subjectivity of a text unit tu:

$$subrat(tu) = \sum_{w \in tu} bal(w)/|tu| \qquad (1)$$

where $bal(w)$ is 1 if w is included in the subjectivity lexicon[4].

Those units that reach a threshold are classified as subjective. Otherwise, the units are taken as objective. Thresholds are tuned by maximising accuracy when classifying the training data at document level. Even if most of the evaluation data collections are tagged at sentence level, the lack of a sentence level annotated training corpus led us to choose this parameter optimisation method. In order to tune the threshold with respect to a balanced accuracy for subjective and objective classification, tuning is done with respect to a balanced training corpus comprising TC_S_{eu} and $TCN\_O'_{eu}$, which we will call $Train\_D$.

4.2 Annotation Scheme

We evaluated the subjectivity lexicons obtained by the different methods in an extrinsic manner by applying them within the framework of a classification task. That way we measured the adequacy of each lexicon in a real task. The gold-standard used for measuring the performance comprises subjective and objective text units that belong to different domains. As we mentioned in section 2.1, the performance of subjectivity classification systems is very sensitive to the application domain. In order to analyse that aspect, we prepared the following test collections:

- Journalism documents ($Jour\_D$) and sentences ($Jour\_S$): texts collected from the Basque newspaper Gara[5].
- Blog sentences ($Blog\_S$): texts collected from Basque blogs included in the website of Berria.
- Twitter sentences ($Tweet\_S$): tweets collected from the aggregator of Basque tweets Umap[6]. Only tweets written in standard Basque are accepted.
- Sentences of music reviews (Rev_S): reviews collected from the Gaztezulo[7] review site.
- Sentences of subtitles (Sub_S): subtitles of different films are collected from the azpitituluak.com site.

In the case of documents, no manual annotation was done. Following the method explained in section 3, we regarded all opinion articles as subjective, and all news articles as objective. The sentences were manually annotated. Our annotation scheme is simple compared to that used in MPQA [5] which represents private states and attributions. In contrast, our annotation is limited to tagging a sentence as subjective if it contains one or more private state expression; otherwise, the sentence is objective. A private state covers opinions, beliefs, thoughts, feelings, emotions, goals, evaluations, and judgements.

[4] We experimented using weights based on the strength of subjectivity but no improvement was achieved, and so, these results are not reported.
[5] http://www.gara.net
[6] http://umap.eu/
[7] http://www.gaztezulo.com/

Table 1. Statistics and class distribution of the reference collections

Source	Unit	Domain	# units	# sub+	# sub	# obj	# obj+
Train_D	document	Journalism	21,320	10,660		10,660	
Jour_D	document	Journalism	9,338	4,669		4,669	
Jour_S	sentence	Journalism	192	60	46	35	51
Blog_S	sentence	Blog	206	94	50	20	42
Tweet_S	sentence	Twitter	200	69	40	21	70
Rev_S	sentence	Music Reviews	138	54	36	24	24
Sub_S	sentence	Subtitles	200	98	31	20	51

We classified sentences according to four categories, depending on aspects such as the number of private state expressions, their intensity, etc.: completely subjective (sub+); subjective but containing some objective element (sub); mostly objective but containing some subjective element (obj); and completely objective (obj+). In order to obtain a robust annotation, three references per annotation were done by three different annotators. Disagreement cases were solved in two different ways. Firstly, annotators discussed all sentences including three different annotations or two equal annotations and a third that was to a distance of more than one category, until consensus was achieved. For dealing with the rest of the disagreement cases, majority voting was used. Table 1 shows the statistics for the test collections and the results of our annotation work.

4.3 Results

By means of our average ratio classifier, we classified the text units in the seven collections presented in the previous section. As mentioned in section 4.1, the units in the test collections were classified according to the subjectivity threshold tuned over the documents in *Train_D*. The optimum subjectivity threshold is computed for each lexicon we evaluated (L_P_{eu}, *DIFF_News*, *LLR_News*, *DIFF_Wiki* and *LLR_Wiki*).

Table 2 and 3 present overall accuracy results and F-score results of the subjective units achieved by the different lexicons in the various test collections.

Table 2. Accuracy results for subjectivity and objectivity classification

	L_P_{eu}	*DIFF_Wiki*	*DIFF_News*	*LLR_Wiki*	*LLR_News*
Train_D	0.63	0.66	**0.90**	0.64	0.87
Jour_D	0.74	0.76	0.80	0.74	**0.87**
Jour_S	0.63	0.59	0.57	0.58	**0.64**
Blog_S	0.65	**0.73**	0.66	**0.73**	0.72
Tweet_S	**0.68**	0.58	0.62	0.59	0.60
Rev_S	**0.70**	**0.70**	0.67	0.67	0.67
Sub_S	0.67	**0.71**	0.70	0.67	0.67

Table 3. F-score results for subjectivity classification

	L_P_{eu}	*DIFF_Wiki*	*DIFF_News*	*LLR_Wiki*	*LLR_News*
Train_D	0.65	0.68	**0.90**	0.68	0.87
Jour_D	0.75	0.77	0.82	0.75	**0.86**
Jour_S	0.73	0.71	0.58	0.72	**0.74**
Blog_S	0.76	0.82	0.77	**0.83**	**0.83**
Tweet_S	**0.73**	0.69	0.70	0.70	0.71
Rev_S	0.79	0.77	0.78	0.75	**0.80**
Sub_S	0.78	**0.81**	0.79	0.78	0.79

In this evaluation, only a binary classification was performed, text units belonging to **obj** and **obj+** classes were grouped into a single category, and the same was done for **sub** and **sub+**. Firstly, according to those results, corpus-based lexicons compiled using TCN_O_{eu} (News) as objective reference (columns 3 and 5) are very effective for document classification. The projected lexicon L_P_{eu} performs significantly worse. Those results were expected, since the corpora-based lexicons have the domain advantage. However, L_P_{eu}'s performance is comparable to corpus-based lexicons' on non-journalistic domains. Moreover, it is better than the corpus-based lexicons in the Twitter domain, both in terms of accuracy and F-score of subjective units. Taking all the results into account, we can see that despite the better performance of corpus-based lexicons in most the domains, the performance of the projected lexicon is more stable across domains than the performance of corpus-based lexicons.

With regard to the corpus used as objective reference (columns 2 and 4 versus columns 3 and 5), the use of the wikipedia corpus TCW_O_{eu} improves the results of the News corpus only in non-journalistic domains and in terms of accuracy. Furthermore, Table 3 shows that if we only take into account the classification of subjective text units, TCN_O_{eu} performs better in all cases except for the subtitle domain collection.

Differences between LLR and $\%DIFF$ vary across the domains. In terms of accuracy, $\%DIFF$ provides better performance when dealing with tweets, reviews, and subtitles. On the contrary, in terms of F-score of subjective units, $\%DIFF$ is only better over subtitles.

We used 4 categories to annotate the references with different degrees of subjectivity. It is interesting how the performance of subjectivity detection changes depending on the required subjectivity degree. In some scenarios only the detection of highly subjective expressions is demanded. In order to adapt the system to those scenarios, we optimised the subjectivity threshold by maximising the $F_{0.5}$-score against training data. Table 4 shows precision and recall results for subjectivity detection if we only accept the ones that belong to the class **sub+** as subjective sentences. According to those results, with the new optimisation of the threshold, the system's performance for classifying **sub+** is similar to that of the initial system.

Table 4. Precision, recall and F-score results for detecting clearly subjective sentences

	L_P_{eu} sub+			LLR_News sub+		
	P	R	F	P	R	F
Jour_S	0.61	0.90	0.73	0.65	0.84	0.73
Blog_S	0.73	0.80	0.76	0.74	0.96	0.83
Tweet_S	0.67	0.82	0.73	0.64	0.83	0.72
Rev_S	0.73	0.86	0.79	0.65	0.99	0.79
Sub_S	0.69	0.88	0.78	0.68	0.99	0.80

5 Conclusions and Future Work

This paper has presented the comparison between two techniques to automatically build subjectivity lexicons. Both techniques only rely on easily obtainable resources, and are adequate for less-resourced languages.

Our results show that subjectivity lexicons extracted from corpora provide a higher performance than the projected lexicon over most of the domains. Accuracies obtained with this method range from 87%, in case of the document classification, to 60-67%, in case of sentences. Projection provides a slight better performance only when dealing with non-journalistic domains. So, it could be an alternative for those domains. If we are interested in identifying only very subjective sentences, both methods offer a good performance (0.72-0.83 in terms of F-score), in particular, the corpora extracted subjectivity lexicons. Hence, the resources obtained with our methods could be applied in social-media analysis tasks where precision is the priority.

Regarding to ongoing and future work, as we have already mentioned, the methods we have researched in this paper are applicable to less-resourced languages because they only require widely available resources. At the moment, we are analysing the effect the characteristics (size, domain,...) of the resources used have on the quality of the final subjectivity lexicon. In the future, we plan to evaluate the Bootstrapping method proposed by Banea et al. [11], which also relies on corpora.

Acknowledgements. This work has been partially funded by the Industry Department of the Basque Government under grants IE12-333 (Ber2tek project) and SA-2012/00180 (BOM2 project).

References

1. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP 2003, pp. 129–136 (2003)

2. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, ACL 2004 (2004)
3. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT/EMNLP 2005, pp. 347–354 (2005)
4. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pp. 105–112 (2003)
5. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 39(2), 165–210 (2005)
6. Esuli, A., Sebastiani, F.: SENTIWORDNET: a publicly available lexical resource for opinion mining. In: Proceedings of LREC 2006, pp. 417–422 (2006)
7. Stone, P., Dunphy, D., Smith, M.: The general inquirer: A computer approach to content analysis (1966)
8. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL 2002, Philadelphia, Pennsylvania, p. 417 (2002)
9. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of HTML documents. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2007, Prague, Czech Republic, pp. 1075–1083 (2007)
10. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, vol. 45, p. 976 (2007)
11. 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, EMNLP 2008, pp. 127–135 (2008)
12. Wan, X.: Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, pp. 553–561 (2008)
13. Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: OpinionFinder. In: Proceedings of HLT/EMNLP on Interactive Demonstrations, Vancouver, British Columbia, Canada, pp. 34–35 (2005)
14. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: A comprehensive grammar of the English language. Pearson Education India (1985)
15. Wilson, T.A.: Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States. ProQuest (2008)
16. Wiebe, J.M., Bruce, R.F., O'Hara, T.P.: Development and use of a gold-standard data set for subjectivity classifications. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, ACL 1999, pp. 246–253 (1999)
17. Wang, X., Fu, G.H.: Chinese subjectivity detection using a sentiment density-based naive bayesian classifier. In: 2010 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 6, pp. 3299–3304 (2010)
18. Das, A., Bandyopadhyay, S.: Subjectivity detection in english and bengali: A CRF-based approach. In: Proceeding of ICON (2009)

19. Das, A., Bandyopadhyay, S.: Theme detection an exploration of opinion subjectivity. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, pp. 1–6 (September 2009)
20. Yu, N., Kübler, S.: Filling the gap: Semi-supervised learning for opinion detection across domains. In: Proceedings of the Fifteenth Conference on Computational Natural Language Learning, pp. 200–209 (2011)
21. Jijkoun, V., de Rijke, M.: Bootstrapping subjectivity detection. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, New York, NY, USA, pp. 1125–1126 (2011)
22. Riloff, E., Wiebe, J., Wilson, T.: Learning subjective nouns using extraction pattern bootstrapping. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, Edmonton, Canada, pp. 25–32 (2003)
23. Baroni, M., Vegnaduzzo, S.: Identifying subjective adjectives through web-based mutual information. In: Proceedings of the 7th Konferenz zur Verarbeitung Natürlicher Sprache, KONVENS 2004, pp. 613–619 (2004)
24. Jijkoun, V., Hofmann, K.: Generating a non-english subjectivity lexicon: relations that matter. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2009, pp. 398–405 (2009)
25. Banea, C., Mihalcea, R., Wiebe, J.: Multilingual sentiment and subjectivity analysis. In: Multilingual Natural Language Processing (2011)
26. Wiebe, J., Wilson, T., Bell, M.: Identifying collocations for recognizing opinions. In: Proceedings of the ACL 2001 Workshop on Collocation: Computational Extraction, Analysis and Exploitation, pp. 24–31 (2001)
27. Maks, I., Vossen, P.: Building a fine-grained subjectivity lexicon from a web corpus. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey (May 2012)
28. Wang, D., Liu, Y.: A cross-corpus study of unsupervised subjectivity identification based on calibrated em. In: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, WASSA 2011, pp. 161–167 (2011)