

# ***Lost in Translation: Viability of Machine Translation for Cross Language Sentiment Analysis***

Balamurali A R<sup>1,2</sup>, Mitesh M. Khapra<sup>3</sup>, Pushpak Bhattacharyya<sup>1</sup>

<sup>1</sup>Indian Institute of Technology Bombay, India

<sup>2</sup>IITB-Monash Research Academy, India

<sup>3</sup>IBM Research, India

balamurali@cse.iitb.ac.in, mikhapra@in.ibm.com, pb@cse.iitb.ac.in

**Abstract.** Recently there has been a lot of interest in Cross Language Sentiment Analysis (CLSA) using Machine Translation (MT) to facilitate Sentiment Analysis in resource deprived languages. The idea is to use the annotated resources of one language (say,  $L_1$ ) for performing Sentiment Analysis in another language (say,  $L_2$ ) which does not have annotated resources. The success of such a scheme crucially depends on the availability of a MT system between  $L_1$  and  $L_2$ . We argue that such a strategy ignores the fact that a Machine Translation system is much more demanding in terms of resources than a Sentiment Analysis engine. Moreover, these approaches fail to take into account the divergence in the expression of sentiments across languages. We provide strong experimental evidence to prove that even the best of such systems do not outperform a system trained using only a few polarity annotated documents in the target language. Having a very large number of documents in  $L_1$  also does not help because most Machine Learning approaches converge (or reach a plateau) after a certain training size (as demonstrated by our results). Based on our study, we take the stand that languages which have a genuine need for a Sentiment Analysis engine should focus on collecting a few polarity annotated documents in their language instead of relying on CLSA.

## **1 Introduction**

In these times of multilingual information processing, there is a keen interest in bringing NLP capability to resource deprived languages by leveraging the resources of a rich language. This is true in the case of Sentiment Analysis (SA) also, where, polarity annotated documents in one language are used for building a SA engine for another language through the instrument of Machine Translation [1]. This task is known as Cross Language Sentiment Analysis (CLSA) wherein the following steps are commonly observed:

1. The polarity marked documents of a resource rich language  $L_1$  are translated to  $L_2$
2. An SA machine  $M$  is trained on these translated documents
3.  $M$  is then applied to a test document  $D$  of language  $L_2$  to detect its polarity

Another alternative is to (i) train a SA machine  $M$  for the resource rich language  $L_1$  (ii) given a document  $D$  in  $L_2$ , first translate it to  $L_1$  and (iii) apply  $M$  to this translated  $D$

to detect its polarity. However, the first alternative is better because it does not involve any translation at test time and hence has lesser test-time complexity and cost (it just has a fixed training time cost).

We claim with quantitative analysis that MT based CLSA at document level is fundamentally not a sound idea. One will instead do better by investing in creating *direct* resources for sentiment analysis. More explicitly, we say that "*if you want to do sentiment analysis in your language and have a limited amount of money, spend the money in creating polarity marked documents for your language, instead of using MT and then doing CLSA*".

Our focus is on document level SA wherein documents are classified into polarity classes (positive and negative) [2]. It is obvious that a case for developing sentiment analysis engine exists for a given language, if many polar documents (e.g., product or movie reviews) are available in electronic form in that language. Given such documents, the effort in annotating them with correct polarity is very little, especially compared to the effort in building an MT system needed for CLSA. For example, it is possible for a single lexicographer to annotate 500 reviews with correct polarity using minimal effort<sup>1</sup>. Our experiments suggest that 500 polarity annotated reviews are sufficient for building a good SA engine for a language (see section 5). Any additional document produces very marginal gain- *the proverbial case of saturation* (see Figure 12 which shows that this happens for three different languages).

Given that the effort involved in collecting polarity annotated documents is quite small, the next question is of performance. We define this performance of a SA engine in terms of its sentiment classification accuracy. Our experiments involving 4 languages suggest that the performance of a SA engine trained using *in-language* polarity annotated documents is better than that obtained using CLSA (see section 5). This is not contrary to intuition, and the reasons are not far to seek:

1. Training a sentiment analysis engine on the own-language corpus ensures that divergences due to cultural differences between two languages are minimal.
2. MT systems are not very accurate and as a result there is always noise in the polarity annotated documents translated from the source language.

We substantiate our arguments by extensive evaluation of well-established CLSA techniques (described in section 3) for four languages, *viz., English, French, German and Russian*.

The remainder of this paper is organized as follows. In section 2, we discuss related work on CLSA. In section 3 we present the CLSA approaches employed in our work. Section 4 describes the experimental setup and datasets used for evaluation. In section 5 we present the results, followed by discussions in section 6. Section 7 concludes the paper.

---

<sup>1</sup> For instance, the authors of this paper were able to annotate 50 reviews with their correct polarity in 1 hour. It would thus take 10 hours to annotate 500 documents with their respective polarity labels. Compare this effort with the effort required to collect or generate parallel corpora for creating an SMT system, which is much larger

## 2 Related Work

To reduce the need of developing annotated resources for SA in multiple languages, cross-lingual approaches [3–6] have been proposed. To use the model trained on  $L_1$  on the test data from  $L_2$ , a Machine Translation (MT) system or a bilingual dictionary is used for transfer between the two languages.

In [6], a cross-lingual approach based on Structured Correspondence Learning (SCL) was proposed, which aims at eliminating the noise introduced due to faulty translations by finding a common low dimensional representation shared by the two languages. In [7], lexicon based and supervised approaches for cross language sentiment classification are compared. Their results show that lexicon based approaches perform better. In [3] and [4], cross-lingual methods which exploit existing tools and resources in English to perform subjectivity analysis in Romanian are proposed.

The state of the art in CLSA is an approach based on co-training. For example, in [5] labeled English data and unlabeled Chinese data was used to perform sentiment classification in Chinese. Here, the English features and the Chinese features are considered as two different views of the same document (one view is formed by English features and the other view is formed by Chinese features extracted after translating the document). Two classifiers are trained using these two views, and each classifier is then applied to the unlabeled Chinese data. The instances which get tagged with high confidence by both the classifiers are then added to the initial training data. Note that the approach requires two MT systems ( $L_1 \rightarrow L_2$  and  $L_2 \rightarrow L_1$ ).

Most, if not all, of the above methods advocate that even a low quality translation engine is adequate for performing CLSA. Our experiments involving 4 languages and 24 combinations of source-target pairs suggest that this argument is not correct. Further, we believe that it is hard to capture sentiment in a language using documents in another language, because of the disparate ways in which sentiments are expressed across languages, a result of cultural diversity amongst different languages. A good example, which we found in our data is that English users use the word ‘*suck*’ frequently to express negative opinion (as in ‘This X sucks’ where X could refer to a movie, actor, director, *etc*). However, the translation of ‘This X sucks’ (which contains the French word *suce/sucer/succion*) was never seen in the French corpus. This suggests that French speakers do not use the equivalent of ‘This X sucks’ to express negative sentiment. Hence, training an English SA by translating training documents from French would most likely not work on an English documents if the word ‘*sucks*’ is the only negative sentiment bearing word in the document.

## 3 CLSA Techniques we Use

Depending on the available tools and resources, (*viz.*, annotated corpus in  $L_1$ , MT between  $L_1$  and  $L_2$ , bilingual dictionary, unannotated corpus in  $L_2$ , we discuss four established methods [4, 3, 5] of performing document level CLSA.

**1. Resource rich  $L_1$  helps resource disadvantaged  $L_2$  using MT ( $MT-X$ ):** Build a Sentiment Analysis system for  $L_2$  by leveraging the annotated resources of  $L_1$  and a

Machine Translation (MT) system from  $L_1$  to  $L_2$ . The approach is outlined in Algorithm 1:

---

**Algorithm 1** *MT-X*

---

$LD_1 :=$  Polarity annotated data from  $L_1$   
 $LD'_2 := \text{translateUsingMT}_{L_1}^{L_2}(LD_1)$   
 $\phi :=$  model trained using  $LD'_2$   
 $\text{test}(\phi, \text{testDocument}_{L_2})$

---

*MT-X* stands for “a resource rich language  $X$  assists a target language using *MT*”.

**2. Resource rich language helps a resource disadvantaged language using a bilingual dictionary (*BD-X*):** Here, the aim is same as above, but instead of using a MT system, a bilingual dictionary (BD)<sup>2</sup> is used for translating polarity annotated documents from  $L_1$  to  $L_2$ . This method thus caters to situations where a MT system is not available for a language pair. Every word in an  $L_1$  document is replaced by its translation in  $L_2$  as found in a bilingual dictionary. The approach is outlined in Algorithm 2:

---

**Algorithm 2** *BD-X*

---

$LD_1 :=$  Polarity annotated data from  $L_1$   
 $LD'_2 := \text{translateUsingBiDict}_{L_1}^{L_2}(LD_1)$   
 $\phi :=$  model trained using  $LD'_2$   
 $\text{test}(\phi, \text{testDocument}_{L_2})$

---

*BD-X* stands for “a resource rich language  $X$  assists a target language using a bilingual dictionary (*BD*)”.

**3. Multiple resource rich languages assist a resource deprived language using MT (*MMT-X*):**

Here, instead of using the labeled data available in one language, we use the labeled data available in multiple resource rich languages to help a resource deprived language. *MMT-XYZ* stands for “*M*ultiple resource rich languages  $X$ ,  $Y$  and  $Z$  assist a target language using Machine Translation (*MT*)”.

**4. Co-Training (*CoTr-X*):** Here, a co-training based approach is used which harnesses the unlabeled data in  $L_2$ . The steps involved in this algorithm are as follows:

### Training

---

<sup>2</sup> BD is created by taking all the unique words present in the resource disadvantaged language and translating them at word-level to resource rich language using Microsoft’s online translation services (<http://www.bing.com/translator>).

---

**Algorithm 3** *MMT-XYZ*

---

```
 $LD'_2 := \text{empty}$   
 $n := \text{number of assisting languages } (n > 1)$   
for  $i = 1 \rightarrow n$  do  
   $LD_i := \text{Polarity annotated data from } L_i$   
   $LD'_2 := LD'_2 + \text{translateUsingMT}_{L_i}^{L_2}(LD_i)$   
end for  
 $\phi := \text{model trained using } LD'_2$   
 $\text{test}(\phi, \text{testDocument}_{L_2})$ 
```

---

- **Step 1:** Translate annotated data ( $LD_1$ ) from  $L_1$  to  $L_2$  ( $LD'_2$ ) using an MT system.
- **Step 2:** Translate unannotated data ( $UD_2$ ) from  $L_2$  to  $L_1$  ( $UD'_1$ ) using an MT system.
- **Step 3:** Train models  $\theta_1$  and  $\theta_2$  using  $LD_1$  and  $LD'_2$  respectively.
- **Step 4:** Use  $\theta_1$  and  $\theta_2$  to label the reviews in  $UD'_1$  and  $UD_2$  respectively.
- **Step 5:** Find  $p$  positive and  $n$  negative reviews from  $UD'_1$  which were labeled with the highest confidence by  $\theta_1$ . Add these to  $LD_1$  and add their translations to  $LD'_2$ .
- **Step 5:** Find  $p$  positive and  $n$  negative reviews from  $UD_2$  which were labeled with the highest confidence by  $\theta_2$ . Add these to  $LD'_2$  and add their translations to  $LD_1$ .
- **Step 6:** Repeat Steps 1 to 5 for  $i$  iterations.

**Testing**

- **Step 7:** Test data from from  $L_2$  using  $\theta_2$ .

The basic idea here is to treat  $LD_1$  and  $LD'_2$  as two different views of the same data. The unlabeled instances which are classified with a high confidence by a classifier trained on one view can then help to improve the classifier trained on the other view. Note that *CoTr-X* stands for “a resource rich language  $X$  assists a target language using *Co-Training*.” Two MT systems ( $L_1 \rightarrow L_2$  and  $L_2 \rightarrow L_1$ ) are needed for this approach thus making it heavily dependent on MT systems.

## 4 Experimental Setup

We performed an extensive evaluation using four languages, *viz.*, *English, French, German and Russian*. We downloaded movie reviews for English, French and German from IMDB<sup>3</sup>. The reviews for these languages were downloaded separately and randomly. Reviews with rating greater than 7 (out of 10) were labeled as positive. and those with the rating of less than 3 were labeled as negative. We ignored reviews having ratings between 3-7 as we found them to be ambiguous. For Russian, since we did not find enough movie review data, we focused on book reviews [8], a domain closely related to movie reviews<sup>4</sup>.

<sup>3</sup> <http://www.imdb.com>, <http://www.imdb.fr>, <http://www.imdb.de>

<sup>4</sup> This gave us chance to study cross domain CLSA

We collected 3000 positive and 3000 negative reviews for English, French and German and 500 positive and 500 negative reviews for Russian. The data in each language was translated to all of the other 3 languages using the Bing<sup>5</sup> translation service. We did not use Google translate because the APIs are no longer freely available. Even though we collected upto 3000 positive and 3000 negative reviews, we found that in almost all cases the performance showed saturation after 400 documents.

We report CLSA results by increasing the training documents in the *source* language  $L_1$  from 50 to 400 in steps of 50. The number of test documents in each language were 200 (*i.e.*, 100 positive and 100 negative reviews). Further, to ensure that our results are not biased to a particular training set and test set we created 10 different sets of 400 positive and negative reviews in each language as well as 10 different sets of 100 positive and negative reviews in each language. Training set 1 in  $L_1$  was then used to perform CLSA on test set 1 in  $L_2$ . We repeated this procedure with all the 10 sets and reported the average accuracy obtained over the 10 sets (similar to 10 fold cross validation albeit in a cross language setting).

We used SVM as the classifier because it is known to give the best results for sentiment classification [2]. Specifically, we used C-SVM (linear kernel with parameters optimized over training set using 5 fold cross validation) available as a part of the LibSVM<sup>6</sup> package. The feature set comprises of unigrams extracted from the seed labeled data. We also experimented with bigram features but did not find much difference in the performance. Further, using higher n-grams features would be unfair to the CLSA systems because most existing MT systems do not produce translations having a good syntactic structure. Hence, we stick to unigram features in this work.

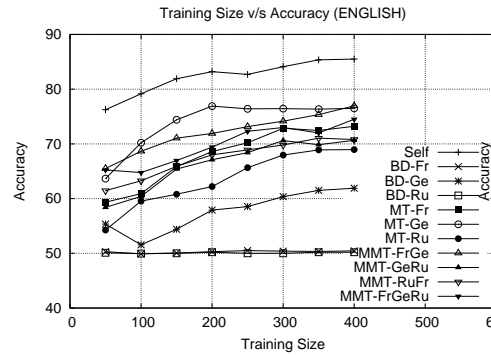
## 5 Results

The results of our experiments are presented in Figures 1 to 12. Figure 1 compares the performance of *MT-X*, *BD-X* and *MMT-XYZ* using different source languages and English as the target language. Figures 2, 3 and 4 present the same comparison with French, German and Russian as the target language. Next, we also wanted to see if one or more resource rich languages can help in improving the performance of another resource rich language (as opposed to assisting a resource poor language). To test this we used  $k$  polarity annotated documents from the target language and added  $k$  polarity annotated documents each translated from one or more source languages. These results are presented in Figures 5 to 8. For ease of understanding and representation, we report the overall accuracy over both positive and negative test documents. In all the graphs, we use the following language codes for representing languages: **En**→*English*, **Fr**→*French*, **Ge**→*German* and **Ru**→*Russian*. Along the X-axis, we represent the number of documents used for training and along the Y-axis we represent the accuracy. To help the reader in interpreting the graphs we explain the different curves in Figure 1 and Figure 5 with English as the target language. The curves in the other graphs can be interpreted similarly.

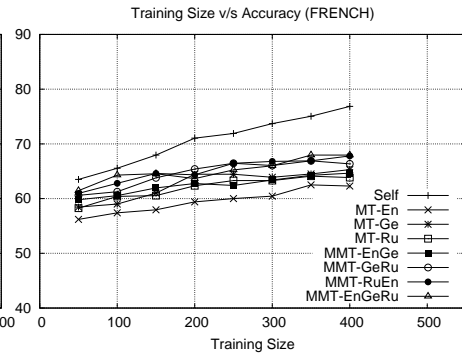
---

<sup>5</sup> <http://www.microsofttranslator.com/>

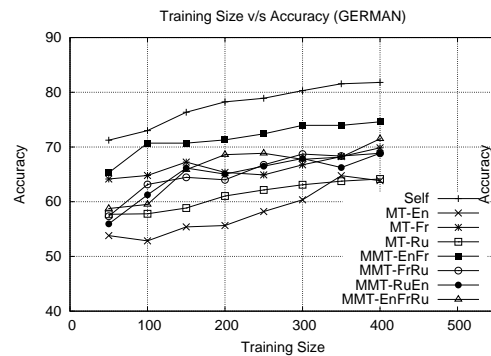
<sup>6</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvm>



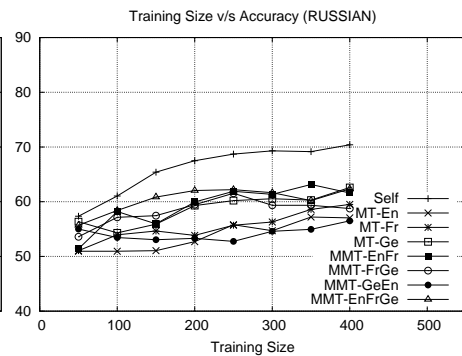
**Fig. 1.** Comparing the performance of different algorithms with English as the target language



**Fig. 2.** Comparing the performance of different algorithms with French as the target language



**Fig. 3.** Comparing the performance of different algorithms with German as the target language



**Fig. 4.** Comparing the performance of different algorithms with Russian as the target language

- *Self* : The accuracy obtained by training a sentiment analysis engine using polarity annotated documents in the target language itself.
- *MT-Fr* : The accuracy obtained by training a sentiment analysis engine using the polarity annotated documents translated from French (Fr) to English using MT.
- *BD-Fr* : The accuracy obtained by training a sentiment analysis engine using the polarity annotated documents translated from French (Fr) to English using a bilingual dictionary.
- *MMT-FrGe* : The accuracy obtained by training a sentiment analysis engine using the polarity annotated documents translated from French (Fr) and German (Ge) to English using MT.
- *MMT-FrGeRu* : The accuracy obtained by training a sentiment analysis engine using the polarity annotated documents translated from French (Fr), German (Ge) and Russian (Ru) to English using MT.



- *Self + MMT-FrGeRu* : This curve in Figure 5-8 plots the accuracy obtained by training a sentiment analysis engine using the polarity annotated documents in English plus the polarity annotated documents translated from French (Fr), German (Ge) and Russian (Ru) to English using a MT system.
- *CoTr-Fr* : This curve in Figure 9 plots the accuracy obtained by training a sentiment analysis engine using the Co-Training approach which uses the polarity annotated documents in French plus the unannotated documents in English.

## 6 Discussions

In this section, we discuss some important observations made from our evaluation.

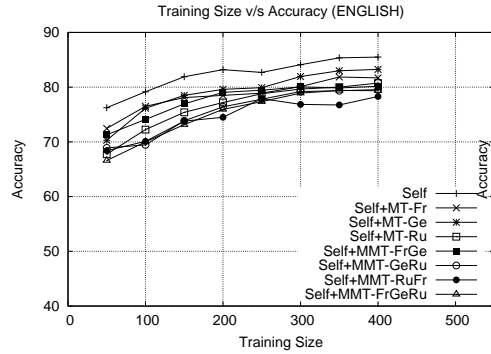
**1. In-language sentiment analysis clearly outperforms cross language sentiment analysis:** We first compare the performance of *MT-X* and *BD-X* with *Self*. In all the graphs (see Figures 1 to 4), the curve of *MT-X* and *BD-X* is much below the curve of *Self*. Specifically, if we compare the performance obtained by using 400 (positive and negative) in-language documents (*i.e.*, *Self*) with that obtained using 400 (positive and negative) cross-language documents, the performance of *Self* is better than *MT-X* by 8-10%. The same difference between *Self* and *BD-X* is much higher. The poor results for *BD-X* suggest that a strategy that simply uses word based translations and ignores the syntactic and semantic structure performs poorly. Thus, the argument that even a very low quality translation engine which ignores syntactic and semantic structure suffices for cross language sentiment analysis does not seem to hold true.

Next, we wanted to see if using data from multiple assisting languages as opposed to a single assisting language can help. The intuition was that taking training examples from multiple languages would increase the diversity in the collection and perhaps be a better strategy for cross language sentiment analysis. However, the results here are not consistent. In some cases, using cross-language data from multiple assisting languages, performs better than taking data from a single assisting language while in other cases it does not. For example, in Figure 1 taking a total of 400 documents from French, German and Russian (*MMT-FrGeRu*) performs better than individually using 400 documents from French or Russian (*MT-Fr*, *MT-Ru*). On the other hand, *MT-Ru* performs better than *MMT-FrGeRu*. However, for all the target languages, the results are in agreement with the stand taken in this paper, *i.e.*, the performance of cross language sentiment analysis using single/multiple assisting language/languages is lower when compared to in-language sentiment analysis.

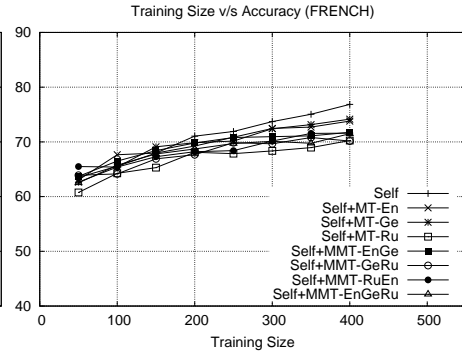
### 2. Does having unannotated data in the target language help?

We wanted to check the importance of unannotated data in the target language. Over all Co-Training seems to be the best CLSA technique, but, in general, it still does not outperform in-language sentiment analysis (Figure 9-11). Specifically, at small training sizes (50, 100), Co-Training does better than in language sentiment analysis but as the training size increases in-language Sentiment Analysis performs better than CLSA. These results contradict previously made claims that CLSA using Co-Training clearly outperforms in-language SA. Further, it should be noted that Co-Training requires (1)

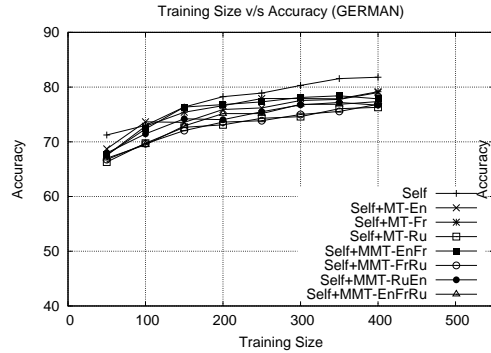




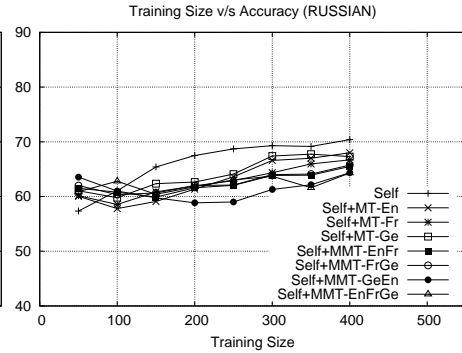
**Fig. 5.** Comparing the performance of different algorithms with English as the target language when self training data in English is also available



**Fig. 6.** Comparing the performance of different algorithms with Russian as the target language when self training data in Russian is also available



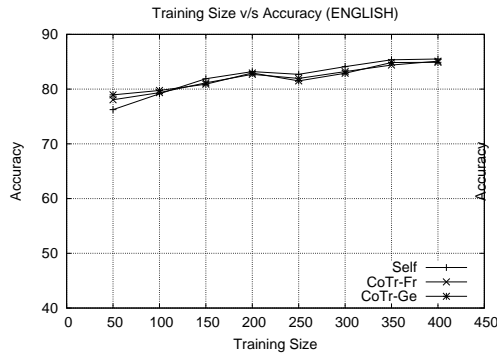
**Fig. 7.** Comparing the performance of different algorithms with German as the target language when self training data in German is also available



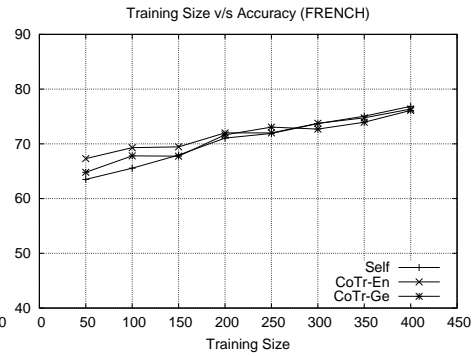
**Fig. 8.** Comparing the performance of different algorithms with Russian as the target language when self training data in Russian is also available

two MT systems and (2) untagged corpus in  $L_2$ . As mentioned earlier, if untagged documents are already available in  $L_2$  then the effort involved in annotating them is much less than the effort involved in building two MT systems.

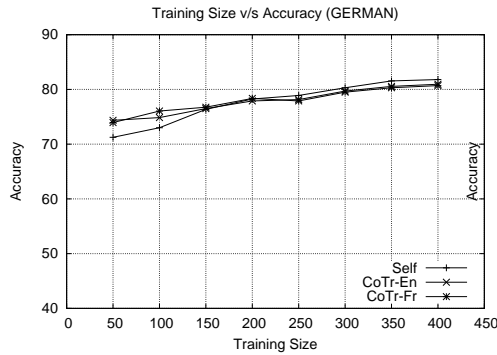
**3. Additional data from other languages does not improve the performance of in-language sentiment analysis:** Figures 5 to 8 suggest that in the presence of annotated data in the target language, adding additional data from other languages harms the performance. For all the target languages, the performance of *Self* is always better than *Self + MT-X* or *Self + MMT-XYZ*. There could be two possible reasons why the additional training data from other languages harms the performance. Firstly, the translations obtained using the MT system maybe erroneous and thereby add noise to the training



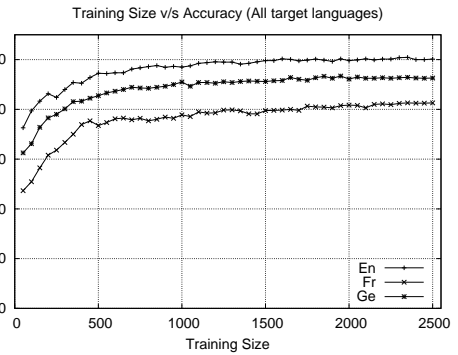
**Fig. 9.** Comparing the performance of *CoTr-X* and *Self* with English as the target language



**Fig. 10.** Comparing the performance of *CoTr-X* and *Self* with French as the target language



**Fig. 11.** Comparing the performance of *CoTr-X* and *Self* with German as the target language



**Fig. 12.** Effect of training size on accuracy (beyond 500 documents there is very little improvement in the accuracy)

process. One reason of for this is the incorrect spellings present in the reviews which can affect the translation quality but probably may not affect the self training because the same incorrect spellings may be present in the test set. Secondly, there might be cultural differences in the manner in which sentiment is expressed in different languages. For example, in some languages double negation is a common phenomenon. A unigram feature based Cross Language Sentiment Analysis trained in a language where such a phenomenon is rare may harm the classification accuracy. These differences again make the training data noisy leading to poor learning and consequently poor performance.

#### 4. How much in-language data does one really need?

The answer to this question is important for making an informed choice regarding the number of documents needed to get a reasonably good accuracy in a language. Specif-

ically, we are interested in the number of documents beyond which the marginal gain in accuracy is negligible. To do so, we plotted the accuracies obtained using increasing amounts of data in the target language. We varied the training data size from 50 to 2500 in steps of 50 and observed that for all the three languages the knee of the curve is obtained at a training size of around 500 documents (we could not run this experiment for Russian as we had only 400 documents in Russian). Beyond this training size the marginal gain in accuracy is very small.

**5. A note on truly resource scarce scenarios:** Our experiments on CLSA were done using European languages which are politically and commercially important. As a result, the SMT systems available for these languages are of comparatively higher quality than those available for many other widely used languages. For example, consider some widely spoken languages like Hindi, Pashto, Punjabi, Sundanese, Hausa, Marathi, Gujarati, *etc*, which have a native speaker base of more than 25 million people<sup>7</sup>. Good quality translation engines are not available for these languages. The results obtained for European languages which have good MT systems suggests that such CLSA systems have very less hope in truly resource deprived scenarios. Further for many languages MT systems are not available at all. For example, to the best of our knowledge, *no* translation engines are publicly available for Pashto, Sundanese, Hausa, Marathi and Punjabi.

## 7 Conclusions and Future Work

We performed an exhaustive evaluation using four languages and different configurations centered around harnessing MT for Cross Language Sentiment Analysis. Our experimental results show that a system developed using *in-language* data performs much better than one developed on cross-language data. Two main reasons for the better performance are (i) CLSA fails to capture the cultural divergence between languages with respect to expression of sentiments and (ii) MT systems are not very accurate and hence introduce noise in the training data. Further, our study falsifies the claim that a crude translation using bilingual dictionary suffices to perform SA in the target language. We also observed that in the presence of training data in a language, adding additional data from other languages actually harms the performance. We would like to emphasize that our experiments were performed on languages which are commercially dominant and hence have much better MT systems than a host of other languages. The poor performance of CLSA in the presence of such better quality MT systems gives rise to the following question: *if there is a genuine interest in developing sentiment analysis engines for these languages then isn't it wiser to invest in collecting polarity annotated documents than to rely on a MT system which is much more complex and hard to obtain?*

---

<sup>7</sup> [http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_number\\_of\\_native\\_speakers](http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers)

## References

1. Brown, P.F., Pietra, V.J.D., Pietra, S.A.D., Mercer, R.L.: The mathematics of statistical machine translation: parameter estimation. *Comput. Linguist.* **19** (1993) 263–311
2. Pang, B., Lee, L.: Thumbs up? sentiment classification using machine learning techniques. In: *Proceedings of EMNLP-02.* (2002) 79–86
3. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: *ACL.* (2007)
4. Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual subjectivity analysis using machine translation. In: *EMNLP.* (2008) 127–135
5. Wan, X.: Co-training for cross-lingual sentiment classification. In: *ACL/AFNLP.* (2009) 235–243
6. Wei, B., Pal, C.: Cross lingual adaptation: an experiment on sentiment classifications. In: *Proceedings of the ACL 2010 Conference Short Papers.* ACLShort '10, Stroudsburg, PA, USA, Association for Computational Linguistics (2010) 258–262
7. Brooke, J., Tofiloski, M., Taboada, M.: Cross-linguistic sentiment analysis: From english to spanish. In: *International Conference on Recent Advances in NLP.* (2009)
8. Zagibalov, T.: Unsupervised and knowledge-poor approaches to sentiment analysis. PhD thesis, University of Sussex (2010)