Aspect-Level Cross-lingual Sentiment Classification with Constrained SMT

Patrik Lambert
Universitat Pompeu Fabra, Barcelona, Spain
`patrik.lambert@upf.edu`

Abstract

Most cross-lingual sentiment classification (CLSC) research so far has been performed at sentence or document level. Aspect-level CLSC, which is more appropriate for many applications, presents the additional difficulty that we consider subsentential opinionated units which have to be mapped across languages. In this paper, we extend the possible cross-lingual sentiment analysis settings to aspect-level specific use cases. We propose a method, based on constrained SMT, to transfer opinionated units across languages by preserving their boundaries. We show that cross-language sentiment classifiers built with this method achieve comparable results to monolingual ones, and we compare different cross-lingual settings.

1 Introduction

Sentiment analysis (SA) is the task of analysing opinions, sentiments or emotions expressed towards entities such as products, services, organisations, issues, and the various attributes of these entities (Liu, 2012). The analysis may be performed at the level of a document (blog post, review) or sentence. However, this is not appropriate for many applications because the same document or sentence can contain positive opinions towards specific aspects and negative ones towards other aspects. Thus a finer analysis can be conducted at the level of the aspects of the entities towards which opinions are expressed, identifying for each opinionated unit elements such as its target, polarity and the polar words used to qualify the target.

The two main SA approaches presented in the literature are (i) a machine learning approach, mostly supervised learning with features such as opinion words, dependency information, opinion

shifters and quantifiers and (ii) a lexicon-based approach, based on rules involving opinion words and phrases, opinion shifters, contrary clauses (but), etc. Thus in most SA systems we may distinguish three types of resources and text:

TRAIN Resources (collection of training examples, lexicons) used to train the classifier.

TEST Opinions to be analysed.

OUT Outcome of the analysis. It depends on the level of granularity. At the document or sentence level, it is the polarity of each document or sentence. At the aspect level, it may the set of opinion targets with their polarity.

The internet multilingualism and the globalisation of products and services create situations in which these three types of resources are not all in the same language. In these situations, a language transfer is needed at some point to perform the SA analysis or to understand its results, thus called cross-lingual sentiment analysis (CLSA).

Sentences or documents are handy granularity levels for CLSA because the labels are not related to specific tokens and thus are not affected by a language transfer. At the aspect level, labels are attached to a specific opinionated unit formed by a sequence of tokens. When transferring these annotations into another language, the opinionated units in the two languages have thus to be mapped.

This paper is one of the first ones to address CLSA *at aspect level* (see Section 3). It makes the following specific contributions:

(i) an extended definition of CLSA including use cases and settings specific to aspect-level analyses (Section 2);

(ii) a method to perform the language transfer preserving the opinionated unit boundaries. This avoids the need of mapping source and target opinionated units after the language transfer via methods such as word alignment (Section 4);

The paper also reports (in Section 5) experiments comparing different settings described in Section 2.

2 Use Cases and Settings

We can think of the following use cases for CLSA:

Use case I. There are opinions we want to analyse, but we do not avail of a SA system to perform this analysis. We thus want to predict the polarity of opinions expressed in a language L_{TEST} using a classifier in another language L_{TRAIN}. We can assume that the language L_{OUT} of the analysis outcome[1] is the same as the one of the opinions. In this case, equation 1 applies, yielding CLSA settings a and b as follows (see also Figure 1).

$$L_{TRAIN} \neq L_{TEST}; L_{OUT} = L_{TEST} \quad (1)$$

(a) available training resources are transferred into the test language to build a classifier in the test language.

(b) we translate the test into the language of the classifier, classify the opinions in the test, and then transfer back the analysis outcome into the source language by projecting the labels or/and opinionated units onto the test set.
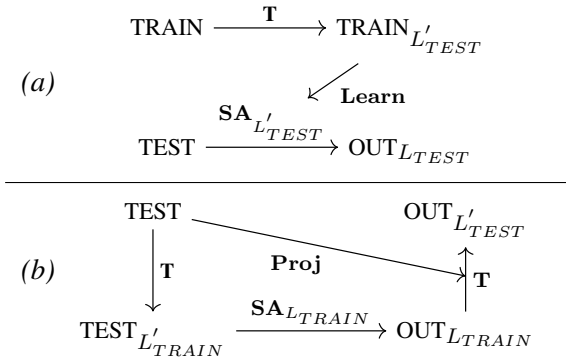

Figure 1: Use case I settings. **SA** refers to Sentiment Analisys, **T** to Translation, **Proj** to Projection and **Learn** to Learning, and the prime symbol designs a language into which a set has been automatically translated.

Use case II. We may have training resources in the language of the opinions, but we need the re-

sult of the analysis in a different language. Here, the inequality of Eq. 2 applies, yielding CLSA settings c and d as follows (see also Figure 2).

$$L_{OUT} \neq L_{TEST} \quad (2)$$

(c) $L_{TRAIN} = L_{TEST}$; the test opinions are first analysed in their language, then the analysis outcome is transferred into the desired language.

(d) $L_{TRAIN} = L_{OUT}$; the test set is first transferred into the desired outcome language, and the SA is performed in this language.

Figure 2: Use case II settings.

Use case II only makes sense for aspect-level analysis,[2] and to our knowledge, it was not addressed in the literature so far.

Use case III. We want to benefit from data available in several languages, either to have more examples and improve the classifier accuracy, or to have a broader view of the opinions under study.

In this paper we focus on use cases I and II.

3 Related Work

The main CLSC approaches described in the literature are via lexicon transfer, via corpus transfer, via test translation and via joint classification.

In the lexicon transfer approach, a source sentiment lexicon is transferred into the target language and a lexicon-based classifier is build in the target language. Approaches to transfer lexica include machine translation (MT) (Mihalcea et al., 2007), Wordnet (Banea et al., 2011; Hassan et al., 2011; Perez-Rosas et al., 2012), relations between dictionaries represented in graphs (Scheible et al., 2010), or triangulation (Steinberger et al., 2012).

The corpus transfer approach consists of transferring a source training corpus into the target language and building a corpus-based classifier in the target language. Banea et al. (2008) follow this approach, translating an annotated corpus via MT. Balamurali et al. (2012) use linked Wordnets to

---

[1] As mentioned above, at the aspect level, the outcome of the analysis may be a set of opinion targets with their polarity. It may also be more complex, such as a set of opinion expressions with their respective target, polarity, holder and time (Liu, 2012). The outcome may need to be in another language as the opinions themselves. For example, a company based in China may survey the opinions of their Spanish-speaking customers, and then transfer the SA outcome into Chinese so that their marketing department can understand it.

[2] For document and sentence-level classification, the outcome is a set of polarity labels independent on language.

replace words in training and test corpora by their (language-independent) synset identifiers. Gui et al. (2014) reduce negative transfer in the process of transfer learning. Popat et al. (2013) perform CLSA with clusters as features, bridging target and source language clusters with word alignment.

In the test translation approach, test sentences from the target language are translated into the source language and they are classified using a source language classifier (Bautin et al., 2008).

Work on joint classification includes training a classifier with features from multilingual views (Banea et al., 2010; Xiao and Guo, 2012), co-training (Wan, 2009; Demirtas and Pechenizkiy, 2013), joint learning (Lu et al., 2011), structural correspondence learning (Wei and Pal, 2010; Prettenhofer and Stein, 2010) or mixture models (Meng et al., 2012). Gui et al. (2013) compare several of these approaches.

Brooke et al. (2009) and Balamurali et al. (2013) conclude that at document level, it is cheaper to annotate resources in the target language than building CLSA systems. This may not be true at aspect level, in which the annotation cost is much higher. In any case, when the skills to build such annotated resources are lacking, CLSA may be the only option. In language pairs in which no high-quality MT systems are available, MT may not be an appropriate transfer method (Popat et al., 2013; Balamurali et al., 2012). However, Balahur and Turchi (2014) conclude that MT systems can be used to build sentiment analysis systems that can obtain comparable performances to the one obtained for English.

All this work was performed at sentence or document level. Zhou et al. (2012) and Lin et al. (2014) work at the aspect level, but they focus on cross-lingual aspect extraction. Haas and Versley (2015) use CLSA for individual syntactic nodes, however they need to map target-language and source-language nodes with word alignment.

4 Language Transfer

In aspect-level SA, there may be several opinionated segments in each sentence. When performing a language transfer, each segment in the target language has to be mapped to its corresponding segment in the source language. This may not be an obvious task at all. For example, if a standard MT system is used for language translation, the source opinionated segment may be reordered and split in several parts in the target language. Then the different parts have to be mapped to the original segment with a method such as word alignment, which may introduce errors and may leave some parts without a corresponding segment in the source language. To avoid these problems, we could translate only the opinionated segments, independently of each other. However, the context of these segments, which may be useful for some applications, would then be lost. Furthermore, the translation quality would be worse than when the segments are translated within the whole sentence context.

To solve these problems, we translate the whole sentences but with reordering constraints ensuring that the opinionated segments are preserved during translation. That is, the text between the relevant segment boundaries is not reordered nor mixed with the text outside these boundaries.[3] Thus the text in the target language segment comes only from the corresponding source language segment. We use the Moses statistical MT (SMT) toolkit (Koehn et al., 2007) to perform the translation. In Moses, these reordering constraints are implemented with the `zone` and `wall` tags, as indicated in Figure 3. Moses also allows mark-up to be directly passed to the translation, via the `x` tag. We use this functionality to keep track, via the tags `<ou[id][-label]>` and `</ou[id]>`, of the segment boundaries (`ou` stands for Opinionated Unit), of the opinionated segment identifier (`[id]`) and, for training and evaluation purposes, of the polarity label (`[-label]`). In the example of Figure 3, the id is 1 and the label is P.

5 CLSA experiments

In order to compare CLSA settings a and b (of use case I), we needed data with opinion annotations at the aspect level, in two different languages and in the same domain. We used the OpeNER[4] opinion corpus,[5] and more specifically the opinion expression and polarity label annotations of the hotel review component, in Spanish and English. We split the data in training (train) and evaluation (test) sets as indicated in Table 1.

The SMT system was trained on freely avail-

---

[3]However, reordering within the segment text is allowed.
[4]http://www.opener-project.eu/
[5]Described in deliverable D5.42 (page 6) at: http://www.opener-project.eu/project/publications.html. This corpus will be freely available from June 2016 on, and until then can be used for research purposes.

Source: On the other hand `<zone>` `<x translation="ou1-P">x</x>` `<wall/>` a big advantage `<wall/>` `<x translation="/ou1">x</x>` `</zone>` of the hostel is its placement
Translation: por otra parte `<ou1-P>una gran ventaja</ou1>` del hostal es su colocación

Figure 3: Source text with reordering constraint mark-up as well as code to pass tags, and its translation.

	Lang	Docs	Words	Op. Units
Train	EN	346	32149	3643
	ES	359	31511	3905
Test	EN	49	4256	496
	ES	50	3733	484

Table 1: Number of documents (Docs), words and opinionated units (Op. Units) in the OpeNER annotated data for English (EN) and Spanish (ES).

	Available		Selected	
Corpus	EN	ES	EN	ES
Common Crawl	46.7	49.5	6.7	7.0
Europarl v7	54.6	57.1	1.7	1.7
News Commentary	4.5	5.1	4.5	5.1
UN	321.7	368.6	3.4	3.5
Booking	1.7	2.6	1.7	2.6
Trip Advisor	23.4	4.4	23.4	4.4

Table 2: Size of the available and selected corpora (in million words) in English (EN) and Spanish (ES) used to train the SMT system.
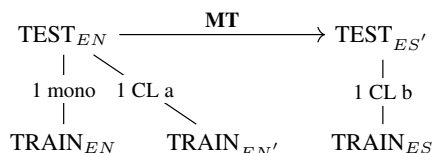

Figure 4: Experiments corresponding to group of rows 1 of Table 3. "mono" refers to monolingual and "CL a" and "CL b" refer to settings a and b of use case I (Sec. 2).

able data from the 2013 workshop on Statistical Machine Translation[6] (WMT 2013). We also crawled monolingual data in the hotel booking domain, from booking.com and TripAdvisor.com. From these in-domain data we extracted 100k and 50k word corpora, respectively for data selection and language model (LM) interpolation tuning. We selected the data closest to the domain in the English-Spanish parallel corpora via a cross-entropy-based method (Moore and Lewis, 2010), using the open source XenC tool (Rousseau, 2013). The size of available and selected corpora are indicated in the first 4 rows of Table 2. The LM was an interpolation of LMs trained with the target part of the parallel corpora and with the rest of the Booking and Trip Advisor data (last 2 rows of Table 2). We used Moses Experiment Management System (Koehn, 2010) with all default options to build the SMT system.[7]

Because the common crawl corpus contained English sentences in the Spanish side, we applied an LM-based filter to select only sentence pairs in which the Spanish side was better scored by the Spanish LM than with the English LM, and conversely for the English side.

We conducted supervised sentiment classification experiments for settings a and b of use case I (see Section 2). We trained and evaluated classifiers on the annotated data (Table 1), using as features the tokens (unigrams) within opinion expressions, and SP (Strong Positive), P (Positive), N (Negative) and SN (Strong Negative) as labels. We performed the experiments with the weka toolkit (Hall et al., 2009), using a filter to convert strings into word vectors, and two learning algorithms: SVMs and bagging with Fast Decision Tree Learner as base algorithm.

Figure 4 represents the experiments conducted with the EN test set. A monolingual classifier in English is trained with the EN training set, and evaluated with the EN test set (1 mono). The re-

Config	Train	Test	LM Filter		No Fil
			Bag.	SVM	SVM
1 mono	EN	EN	77.2	83.4	83.4
1 CL a	EN$'$	EN	70.3	75.4	75.8
1 CL b	ES	ES$'$	73.0	75.8	73.6
2 mono	ES	ES	76.8	81.1	81.1
2 CL a	ES$'$	ES	66.2	72.5	73.0
2 CL b	EN	EN$'$	74.5	77.6	76.8

Table 3: Accuracy (in %) achieved by the different systems. LM Filter and No Fil(ter) refer to the presence or not of the LM filter for the common crawl parallel corpus. "Bag." refers to bagging.

---

[6] http://www.statmt.org/wmt13/translation-task.html
[7] We kept selected parallel data of the common crawl corpus for tuning and test. We obtained BLEU scores of 42 and 45 in the English–Spanish and Spanish–English directions.

sults are reported in the first row of Table 3. To evaluate cross-lingual setting a, the ES training set is translated into English (see Section 4), and an English classifier is trained on the translated data and evaluated on the EN test set (1 CL a). To evaluate setting b, the EN test set is translated into Spanish, and this translated test is used to evaluate a classifier trained on the ES training set (1 CL b). With this very simple classifier, we achieve up to 83.4% accuracy in the monolingual case. With cross-lingual settings, we loose from about 4% to 8% accuracy, and with the higher quality SMT system (LM filter), CL-b setting is slightly better than CL-a.

The same three experiments were conducted for the ES test set (last three rows of Table 3). We achieved an accuracy of 81.1% in the monolingual case. Here the CL-b setting achieved a clearly better accuracy than the CL-a setting (at least 5% more), and only from 2.3% to 3.5% below the monolingual one. Thus with the higher quality SMT system, it is always better to translate the test data (CL-b setting) than the training corpus.

Comparing the SVM classification accuracy in the "LM Filter" and "No Fil" columns, we can see the effect of introducing noise in the MT system. We observe that the results were more affected by the translation of the test (-2.2% and -0.8% accuracy) than the training set (+0.5% accuracy in both cases). This agrees with the intuition than errors in the test directly affect the results and thus may be more harmful than in the training set, where they may hardly affect the results if they represent infrequent examples.

Regarding use case II, setting c implies a translation of the analysis outcome. We can use our method to translate the relevant opinionated units with their predicted label in their test sentence context, and extract the relevant information in the outcome language. In setting d, the test is translated in the same way as in setting b.

6 Conclusions and Perspectives

We extended the possible CLSA settings to aspect-level specific use cases. We proposed a method, based on constrained SMT, to transfer opinionated units across languages by preserving their boundaries. With this method, we built cross-language sentiment classifiers achieving comparable results to monolingual ones (from about 4 to 8% and 2.3 to 3.5% loss in accuracy depending on the lan-

guage and machine learning algorithm). We observed that improving the MT quality had more impact in settings using a translated test than a translated training corpus. With the higher MT quality system, we achieved better accuracy by translating the test than the training corpus.

As future work, we plan to investigate the exact effect of the reordering constraints in terms of possible translation model phrase pairs and target language model n-grams which may not be used depending on the constraint parameters, in order to find the best configuration.

References

Alexandra Balahur and Marco Turchi. 2014. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech & Language*, 28(1):56–75.

A.R. Balamurali, Aditya Joshi, and Pushpak Bhattacharyya. 2012. Cross-lingual sentiment analysis for Indian languages using linked wordnets. In *Proc. of the International Conference on Computational Linguistics (COLING)*, pages 73–82, Mumbai, India.

A. R. Balamurali, Mitesh M Khapra, and Pushpak Battacharyya. 2013. Lost in Translation: Viability of Machine Translation for Cross Language Sentiment Analysis. In *Proc. of International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, pages 38–49, Samos, Greece.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2008. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *Proc. of the International Conference on Linguistic Resources and Evaluation (LREC)*, pages 2764–2767, Marrakech, Morocco, May.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2010. Multilingual subjectivity: Are more languages better? In *Proc. of the International Conference on Computational Linguistics (COLING)*, pages 28–36, Beijing, China.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2011. Multilingual sentiment and subjectivity analysis. In D. M. Bikel and I. Zitouni, editors, *Multilingual Natural Language Applications: From Theory to Practice*. Prentice-Hall.

Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. 2008. International sentiment analysis for news and blogs. In *Proc. of the International Conference on Weblogs and Social Media*, pages 19–26, Seattle, U.S.A.

Julian Brooke, Milan Tofiloski, and Maite Taboada. 2009. Cross-Linguistic Sentiment Analysis: From English to Spanish. In *Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 50–54, Borovets, Bulgaria.

Erkin Demirtas and Mykola Pechenizkiy. 2013. Cross-lingual Polarity Detection with Machine Translation. In *Proc. of the International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '13*, pages 9:1–9:8, Chicago, Illinois, USA. ACM Press.

Lin Gui, Ruifeng Xu, Jun Xu, Li Yuan, Yuanlin Yao, Jiyun Zhou, Qiaoyun Qiu, Shuwei Wang, Kam-fai Wong, and Ricky Cheung. 2013. A Mixed Model for Cross Lingual Opinion Analysis. In *Second CCF Conference, Natural Language Processing and Chinese Computing*, pages 93–104.

Lin Gui, Ruifeng Xu, Qin Lu, Jun Xu, Jian Xu, Bin Liu, and Xiaolong Wang. 2014. Cross-lingual opinion analysis via negative transfer detection. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pages 860–865, Baltimore, Maryland.

Michael Haas and Yannick Versley. 2015. Subsentential sentiment on a shoestring: A crosslingual analysis of compositional classification. In *Proc. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 694–704, Denver, Colorado.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update. *SIGKDD Explorations*, 11(1).

Ahmed Hassan, Amjad AbuJbara, Rahul Jha, and Dragomir Radev. 2011. Identifying the semantic orientation of foreign words. In *Proc. of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 592–597, Portland, Oregon, USA, June.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proc. of the 45th Annual Meeting of the Association for Computational Linguistics (Demo and Poster Sessions)*, pages 177–180, Prague, Czech Republic, June. Association for Computational Linguistics.

Philipp Koehn. 2010. An experimental management system. *Prague Bulletin of Mathematical Linguistics (PBML)*, (94):87–96.

Zheng Lin, Xiaolong Jin, Xueke Xu, Yuanzhuo Wang, Weiping Wang, and Xueqi Cheng. 2014. A cross-lingual joint aspect/sentiment model for sentiment analysis. In *Proc. of the ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 1089–1098, Shanghai, China.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Bin Lu, Chenhao Tan, Claire Cardie, and Benjamin K. Tsou. 2011. Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proc. of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 320–330, Portland, Oregon, USA.

Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, and Houfeng Wang. 2012. Cross-lingual mixture model for sentiment classification. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pages 572–581, Jeju Island, Korea.

Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pages 976–983, Prague, Czech Republic, June.

Robert C. Moore and William Lewis. 2010. Intelligent selection of language model training data. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 220–224, Uppsala, Sweden.

Veronica Perez-Rosas, Carmen Banea, and Rada Mihalcea. 2012. Learning sentiment lexicons in spanish. In *Proc. of the International Conference on Linguistic Resources and Evaluation (LREC)*, pages 3077–3081, Istanbul, Turkey, may.

Kashyap Popat, Balamurali A.R, Pushpak Bhattacharyya, and Gholamreza Haffari. 2013. The haves and the have-nots: Leveraging unlabelled corpora for sentiment analysis. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pages 412–422, Sofia, Bulgaria.

Peter Prettenhofer and Benno Stein. 2010. Cross-language text classification using structural correspondence learning. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pages 1118–1127, Uppsala, Sweden. Association for Computational Linguistics.

A Rousseau. 2013. XenC: An Open-Source Tool for Data Selection in Natural Language Processing. *Prague Bulletin of Mathematical Linguistics (PBML)*, (100):73–82.

Christian Scheible, Florian Laws, Lukas Michelbacher, and Hinrich Schütze. 2010. Sentiment translation through multi-edge graphs. In *Proc. of the International Conference on Computational Linguistics (COLING)*, pages 1104–1112, Beijing, China, August.

Josef Steinberger, Mohamed Ebrahim, Maud Ehrmann, Ali Hurriyetoglu, Mijail Kabadjov, Polina Lenkova, Ralf Steinberger, Hristo Tanev, Silvia Vázquez, and Vanni Zavarella. 2012. Creating sentiment dictionaries via triangulation. *Decision Support Systems*, 53(4):689 – 694.

Xiaojun Wan. 2009. 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*, pages 235–243, Suntec, Singapore.

Bin Wei and Christopher Pal. 2010. Cross lingual adaptation: An experiment on sentiment classifications. In *Proc. of the ACL 2010 Conference Short Papers*, pages 258–262, Uppsala, Sweden. Proc. of the Annual Meeting of the Association for Computational Linguistics.

Min Xiao and Yuhong Guo. 2012. Multi-view adaboost for multilingual subjectivity analysis. In *Proc. of the International Conference on Computational Linguistics (COLING)*, pages 2851–2866, Mumbai, India.

Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2012. Cross-Language Opinion Target Extraction in Review Texts. In *IEEE 12th International Conference on Data Mining*, pages 1200–1205, Brussels, Belgium.