

# Emotion in Code-switching Texts: Corpus Construction and Analysis

Sophia Yat Mei Lee<sup>†</sup>, and Zhongqing Wang<sup>†,‡</sup>

<sup>†</sup> Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

<sup>‡</sup> Natural Language Processing Lab, Soochow University, China

{sophiaym, wangzq.antony}@gmail.com

## Abstract

Previous researches have focused on analyzing emotion through monolingual text, when in fact bilingual or code-switching posts are also common in social media. Despite the important implications of code-switching for emotion analysis, existing automatic emotion extraction methods fail to accommodate for the code-switching content. In this paper, we propose a general framework to construct and analyze the code-switching emotional posts in social media. We first propose an annotation scheme to identify the emotions associated with the languages expressing them in a Chinese-English code-switching corpus. We then make some observations and generate statistics from the corpus to analyze the linguistic phenomena of code-switching texts in social media. Finally, we propose a multiple-classifier-based automatic detection approach to detect emotion in the code-switching corpus for evaluating the effectiveness of both Chinese and English texts.

## 1 Introduction

Due to the popularity of opinion-rich resources (e.g., online review sites, forums, and the microblog websites), emotion analysis in text is of great significance in obtaining useful information for studies on social media (Pang et al., 2002; Liu et al., 2013; Lee et al., 2014). Previous researches have mainly focused on analyzing emotion through monolingual text (Chen et al., 2010; Lee et al., 2013a). However, code-switching posts are also common in social media. Emotions can be expressed by either monolingual text or bilingual text in the code-switching posts. Code-

switching text is defined as text that contains more than one language ('code') (Adel et al., 2013; Auer, 1999). [E1-E3] are three examples of code-switching emotional posts on *Weibo.com* that contain both Chinese and English texts. [E1] expresses the *happiness* emotion through English, and the *sadness* emotion in [E2] is expressed through both Chinese and English, while the *sadness* emotion in [E3] is expressed through a mixed Chinese-English phrase (hold 不住 'cannot take it').

[E1] 玩了一下午轮滑 **so happy** !

(*I went rollerblading the whole afternoon, so happy!*)

[E2] 开学以来，浮躁的情绪。不安稳的心态。确实该自己检讨一下了。。。 **sigh**~~~

(*I have been grumpy and emotional since the first day of school, unstable mindset too. It's really time to self-evaluate...sigh~~~*)

[E3] 上了一天的课，嗓子 **hold** 不住了啊

(*I have been teaching the whole day, my throat can't take it anymore.*)

Despite the important implications of code-switching for emotion analysis, existing emotion analysis approaches fail to accommodate for the code-switching content. Thus, there is a crucial need for analyzing emotions in code-switching texts.

In this paper, we provide a well-defined and efficient method for constructing and analyzing a large-scale code-switching corpus from social media. We believe the annotated corpus provides a valuable resource for both linguistic analysis as well as natural language processing of emotion and code-switching texts. We construct and analyze the corpus using the below steps: First, we extract and filter the code-switching posts from the large-scale dataset by removing monolingual

and noise posts. Second, we propose an annotation scheme to annotate both emotions and the language(s) expressing the emotions (hereafter caused language(s)) in the data set. Third, we analyze the agreement of the corpus to verify the quality of the annotation and effectiveness of the scheme. We also show some observations and statistics on the corpus to analyze the linguistic phenomena of code-switching texts on social media. Finally, we propose a multiple-classifier-based automatic detection approach to detect emotion in the annotated code-switching corpus for indicating the effectiveness of both Chinese text and English text in code-switching posts in detecting emotions.

The rest of the paper is organized as follows. In Section 2, we give an overview on the related work. In Section 3, we introduce our data collection method and the annotation scheme. In Section 4, we report the analysis of the corpus including the inter-annotator agreement as well as other relevant statistics. In Section 5, we propose an automatic emotion detection framework on code-switching text. Finally, we conclude our work in Section 6.

## 2 Related Work

In this section, we discuss related works on emotion analysis and code-switching text analysis.

### 2.1 Emotion Analysis

The earliest research on emotion has focused on the representation and processing of emotion in facial expressions and body language (Andrew, 1963; Ekman and Friesen, 1978). More recently, there has been mounting research on the neurobiological basis of emotion (Olson et al., 2007; Hervé et al., 2012) and how emotion is linked with other aspects of human cognition (Smith and Lazarus, 1993; Smith and Kirby, 2001; Bridge et al., 2010).

Emotion has been well studied in natural language processing, while most previous researches focused on analyzing emotions in monolingual text. Some of these studies focus on lexicon building, for example, Rao et al. (2012) automatically building the word-emotion mapping dictionary for social emotion detection, and Yang et al., (2014) propose a novel emotion-aware topic model to build a domain specific lexicon. Moreover, emotion classification is one of the important tasks in emotion analysis. For example, Liu et al., (2013) used co-training framework to

infer the news reader’s and comment writer’s emotion collectively; Wen and Wan (2014) used class sequential rules for emotion classification of micro-blog texts by regarding each post as a data sequence.

The research of emotion has also been linked to the field of bilingualism. Previous studies have demonstrated that emotion is closely related to second language learning and use (Arnold, 1999; Schumann, 1999), as well as bilingual performance and language choice (Schrauf, 2000; Pavlenko, 2008). For example, there are a number of factors that may impact the use of emotion vocabulary, such as sociocultural competence, gender, and topic (Dewaele and Palvenko, 2002).

Despite a growing body of research on emotion, little has been done on the analysis of emotion in code-switching contexts due to the complications in processing two languages at the same time.

### 2.2 Analysis of Code-switching Texts

Research on code-switching can be traced back to the 1970s. Several theories have been proposed to account for the motivation behind code-switching such as diglossia (Blom and Gumperz, 1972), communication accommodation theory (Giles and Clair, 1979), the markedness model (Myers-Scotton, 1993), and the conversational analysis model (Auer, 1984).

Code-switched documents have also received considerable attention in the NLP community. Several studies have focused on identification and analysis, including mining translations in code-switched documents (Ling et al., 2013), predicting code-switched points (Solorio and Liu, 2008), identifying code-switched tokens (Lignos and Marcus, 2013), adding code-switched support to language models (Li and Fung, 2012), and learning poly-lingual topic models from code-switching text (Peng et al., 2014).

Another related research topic, multilingual natural language processing, has begun to attract attention in the computational linguistic community due to its broad real-world applications. Relevant studies have been reported in different natural language processing tasks, such as parsing (Burkett et al., 2010), information retrieval (Gao et al., 2009), text classification (Amini et al., 2010), and sentiment analysis (Lu et al., 2011).

However, none have studied the multilingual code-switching issues in the task of emotion detection and classification. This area of research is especially crucial when public emotions are mostly expressed on the Internet. Additionally,

the important implications of code-switching in emotion analysis serve as a first step towards an automatic multilingual classification system.

### 3 Data Collection and Annotation

In this section, we describe how to collect and filter code-switching posts on *Weibo.com*. We also discuss the annotation scheme and the annotation tool.

#### 3.1 Data Collection

We sourced our data set from *Weibo.com*, one of the famous SNS websites in China. We identified a post as code-switched if at least two predicted languages, i.e. Chinese and English, appeared in the text. As the encoding of Chinese and English characters is different (the maximum number of encoded English characters is less than 128), we thus utilized each character code to identify the language in a simple manner. We also remove the noise, and advertisement posts ([E4] and [E5] are the examples of noise and advertisement posts).

[E4] 分享 **Carpenters** 的歌曲《**Close To You**》

(*Share Carpenters' music <Close To You>*)

[E5] **the face shop** 提供新款化妆品  
(*the face shop provides new make-up*)

#### 3.2 Annotation Scheme

Five basic emotions were annotated, namely *happiness*, *sadness*, *fear*, *anger* and *surprise* (Lee et al., 2013b). Two languages, Chinese and English, were annotated as causal languages. Since emotion can be expressed through the two languages separately or collectively, and also could be expressed through mixed phrases e.g. “笑cry” (*very happy*), we thus need to annotate four kinds of causal situations, i.e. *English*, *Chinese*, *Both*, and *Mixed*. Following are descriptions of these situations:

➤ **Chinese (CN)** means the emotion of the post is individually expressed through the Chinese text. As *Weibo.com* is a Chinese SNS Website, Chinese is the dominant language on this website. Most of the posts express emotions through the Chinese text. [E6] is an example. The emotion of *surprise* is expressed through the Chinese text.

[E6] 静静坐下来看别人 **show** 啦。刚刚在节目里看到妈咪和弟的视频真的很意外!

(*I set down quietly to watch someone else's show. To my surprise, both my mother and brother appeared on the programme.*)

➤ **English (EN)** means the emotion of the post is individually expressed through the English text. As English is the minority language, there are fewer English words in the posts to express emotions. [E1] is an examples expressing *happiness* emotion and expressed through English text.

➤ **Both (BOTH)** means the emotions of the post are expressed through both Chinese and English text. Note that the emotions expressed through the two languages would either be the same or different. [E2] and [E7] are two examples. The *anger* emotion of [E2] is expressed through both the Chinese and English text. However, the *happiness* emotion of [E7] is expressed through the Chinese text, while the *surprise* emotion is expressed through English.

[E7] 太感动这真是一个大 **surprise** 看的时候就鸡冻屎了

(*I was so touched and excited to see this great surprise.*)

➤ **Mixed (MIXED)** means the emotion of the post is expressed through a Chinese-English mixed phrase, such as the emotion being expressed through the mixed phrase “hold不住” in [E3]. Note that there are limited mixed patterns, and Table 1 illustrates the examples of mixed phrases in our dataset.

Moreover, the emotions of some posts are expressed implicitly, and do not contain explicit keywords to express emotions. [E8] and [E9] are examples of this, while these two posts both express a *sadness* emotion, [E8] is expressed through Chinese text, and [E9] is expressed through both Chinese and English text.

[E8] 英语的魅力在于，好不容易看懂每个word却看不懂组成的 **sentence**.

(*The charm of English is that you can't always understand the meaning of the sentence, even though you understand the meaning of each word in the sentence.*)

[E9] **stream flow, slowly away a few leaf, also taking the memory**.溪水缓慢地流动着，带走了几片落叶，也带走了记忆。

(The Chinese text is translated from English text)

Pattern	Examples
有 feel (sense)	-
hold 住	hold 住 ( <i>can take it</i> ) hold 不住 ( <i>cannot take it</i> )
XX cry	笑 cry ( <i>smile, very happy</i> ) 感动 cry ( <i>touched</i> ) 帅 cry ( <i>awesome</i> )
太 man 了 (handsome)	-

Table 1: Examples of mixed phrases

### 3.3 Annotation Tool and Format

An annotation tool is designed to facilitate the annotation process which allows better consistency.

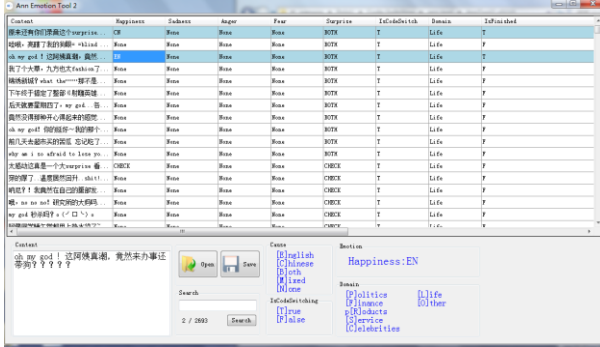


Figure 1: A sample of code-switching emotion annotation using the annotation tool

Figure 1 shows an example instance annotated with both emotion and caused languages using our annotation tool. For each emotion, annotators marked whether the post expresses emotion, together with the caused languages toward the emotion.

Figure 2 is a sample of an annotated instance. Each instance contains the caused language with the emotion tag, e.g., “<Happiness>CN </Happiness>”, while the example tag means the post expresses the *happiness* emotion through Chinese text.

```

<Post id="1">
  <Happiness>
  CN
  </Happiness>
  <Sadness>
  None
  </Sadness>
  <Anger>
  None
  </Anger>
  <Fear>
  None
  </Fear>
  <Surprise>
  None
  </Surprise>
  <Content>
  baby 生日快乐! 附加订婚: 此女贤良
 淑德 拥有现代女性智慧和古典女性的温婉 诚
 征凹凸曼 非诚勿扰
  </Content>
</Post>

```

Figure 2: A sample of an annotated instance

## 4 Statistics and Analysis

In this section, we analyze the agreement of the corpus, and present some observations and statistics.

### 4.1 Agreement Analysis

To verify the quality of the annotation, two human annotators were asked to annotate 1,000 posts. We then calculated the inter-annotator agreement between them using Cohen’s Kappa coefficient. Table 2 shows the results of agreement analysis. We find that the agreement is high, indicating that the quality of the annotation and scheme is effective. In addition, the agreement of emotion annotation is lower than that of caused language, which probably due to the fact that some posts express more than one emotion, and some emotions are expressed implicitly.

	Kappa score
Emotion	0.692
Caused Language	0.767

Table 2: Results of agreement analysis

### 4.2 Statistics and Observations

In this subsection, we discuss some statistics from the dataset.

### General Distribution of Data

Out of 4,195 annotated posts, 2,312 posts are found to express emotions. Moreover, 81.4% of emotional posts are expressed through Chinese. Although English contains relatively fewer words in each post, there are still 43.5% of emotional posts are expressed through English. This indicates that English is of vital importance to emotion expression even in code-switching contexts dominated by Chinese. Note that, there are overlaps between Chinese and English emotional posts, since some emotional posts are conducted in both Chinese and English. Besides, although some posts express the same emotion through both Chinese and English text ([E2]), there are still some posts expressed different emotions through different languages. For example, the *happiness* emotion in [E7] is expressed through Chinese, while the *surprise* emotion is expressed through English.

Moreover, as shown in Figure 3, we find that most posts describe people’s daily lives, since people like to discuss their life on their micro-blogs, and posts from financial and political domains were limited.

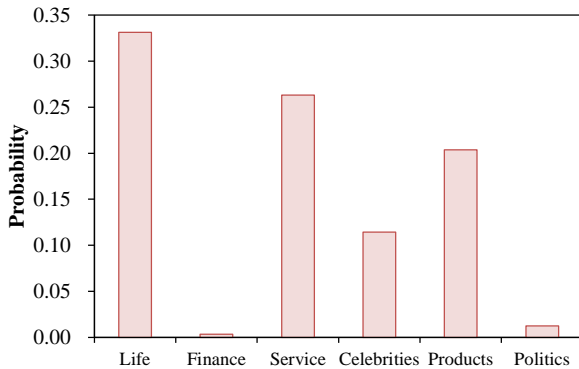


Figure 3: Domain statistics from the data set

### Joint Distribution of Emotions and Caused Languages

For the purpose of analyzing the distribution of emotions and the caused languages, we first calculate the joint distribution between emotions and caused languages as in Figure 4. The Y-axis of the figure presents the conditional probability of a post expressing the emotion  $e_i$  given that  $l_j$  is the caused language,  $p(e_i | l_j)$ .

It is suggested in Figure 4 that: 1) *happiness* occurs more frequently than other emotions; 2) people prefer to use English text to express *happiness* more than *sadness*; 3) the distribution of emotions expressed through Chinese and English

text are similar; and 4) *fear* and *surprise* occur less frequently in English text.

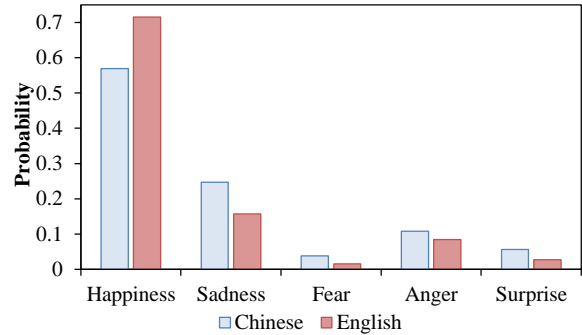


Figure 4: Joint Distribution of Emotions and Caused Languages

### Transfer Probability between Emotions

We then examine the conditional probabilities of a post expressing emotion  $e_i$  given that the post contains emotion  $e_j$ . The conditional probabilities are shown as in Table 3.

From the table, we find that the probability that a post contains more than one emotion is small. Moreover, the probability of polarity shifting between emotions (*happiness* vs. *sadness*, *fear*, *anger*) is limited.

	Happiness	Sadness	Fear	Anger	Surprise
Happiness	-	0.060	0.016	0.025	0.019
Sadness	0.088	-	0.023	0.033	0.023
Fear	0.114	0.114	-	0.068	0.023
Anger	0.090	0.079	0.034	-	0.011
Surprise	0.086	0.071	0.014	0.043	-

Table 3: The transfer probability between emotions

### Transfer Probability between Caused Languages

We also examine the conditional probabilities of the emotion(s) expressed in one language  $l_i$  given that the emotion is expressed in another language  $l_j$  simultaneously in a post. The conditional probabilities are shown as in Table 4.

	Chinese	English
Chinese	-	0.236
English	0.614	-

Table 4: Transfer probability between caused languages

From the table, we find that there is a high probability that the two languages both express emotions, especially when given that the emotion

is expressed in English. It is also highly likely that the emotion would be expressed in Chinese.

### Sentence Length Distribution of Each Language

Table 5 shows the statistics on the average sentence length of each language. We notice, as our data are always written by Chinese individuals, the length of Chinese words is longer than English words. Besides, the emotions expressed through English text are mostly single words, e.g., *happy*, *high*, and *surprise*. Note that, as mentioned above, although the length of Chinese words is longer than English words, English is of vital importance to emotion expressions even in code-switching context dominated by Chinese.

	#avg. word
Chinese	19.8
English	2.9

Table 5: Statistics on average word length

### Distribution of Cue Words

In addition, we count the top-10 frequency emotion cue words of both English and Chinese text as given in Table 6. We find that the most frequent cue words express *happiness* emotions, for example, *happy*, *nice*, and *喜欢 (like)*. What is more, there are several negative expressions in the top-10 English cue words, e.g. *sorry* and *shit*, while the top-10 Chinese cue words are all positive. This may be due to the fact that expressing the negative emotion through native language (Chinese) would be too explicit for Chinese individuals, while most of them tend to express their negative emotions implicitly.

English	Chinese
Happy	喜欢 ( <i>like</i> )
Love	快乐 ( <i>happy</i> )
Good	希望 ( <i>hope</i> )
Nice	开心 ( <i>joyful</i> )
Sorry	哈哈 ( <i>haha</i> )
Shit	幸福 ( <i>happiness</i> )
Luck	真心 ( <i>heartfelt</i> )
Thank	可爱 ( <i>cute</i> )
Perfect	感谢 ( <i>thank</i> )
Sweet	成功 ( <i>success</i> )

Table 6: Statistics of emotional cue words

## 5 Automatic Emotion Detection in Code-switching Texts

Based on the annotated corpus data, we attempt to detect emotion in code-switching text automatically. Results show both Chinese and English texts are effective, and the classifier combination approach which incorporates both Chinese and English text achieves the best performance.

### 5.1 Overview of Detection Approach

A straightforward approach to detect emotion in code-switching text is using a supervised learning approach to classify the mixed text without any processing. Besides, we extract unigrams as a feature for each post. As emotions could be expressed in either Chinese or English text, we also adopt two classification approaches which consider Chinese or English texts individually.

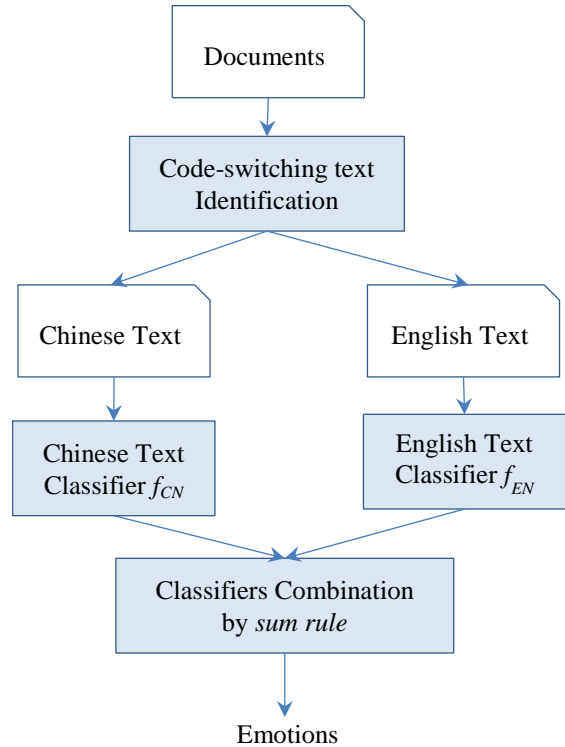


Figure 5: Overview of the multiple-classifiers-based detection framework

However, a more effective way to detect emotion in code-switching posts is incorporating both Chinese and English text through a Multiple Classifier System (MCS). The key issue in constructing a multiple classifier system is to find a suitable way to combine the outputs of the base classifiers. In MCS literature, various methods are available for combining the outputs, such as fixed rules including the voting rule, the product

rule and the sum rule (Kittler et al., 1998; Li et al., 2010). In this study, we adopt the sum rule, a popular fixed rule to combining the outputs of both Chinese and English text classifiers.

For utilizing MCS to detect emotion in code-switching texts, we first define the base classifiers. In this paper, we use the Chinese text classifier  $f_{CN}$  and English text classifier  $f_{EN}$  which only considers Chinese text or English text individually as two base classifiers. Each base classifier provides a kind of confidence measurement, e.g., posterior probabilities of the test sample belonging to each class. Formally, each base classifier  $f_i$  assigns a test sample (denoted as  $x_i$ )

a posterior probability vector  $\vec{P}(x_i)$  :

$$\vec{P}(x_i) = (p(c_1 | x_i), \dots, p(c_j | x_i), \dots, p(c_n | x_i))^t \quad (1)$$

Where  $p(c_j | x_i)$  denotes the probability that the  $i$ -th base classifier considers the sample belonging  $c_j$ .

After we define the two base classifiers, we can use a sum rule to combine the base classifiers by summing the posterior possibilities and using the sum possibility for decision, i.e.

$$\text{assign } y \rightarrow c_j \text{ where } k = \underset{j}{\operatorname{argmax}} \sum_i p(c_j | x_i)$$

Figure 5 illustrates the process of the multiple classifier system for emotion detection in code-switching texts.

## 5.2 Experiments

As described in Section 3, the data are collected from *Weibo.com*. We randomly select half of the posts as the training data and another half as the test data. We use *FudanNLP*<sup>1</sup> for Chinese word segmentation and Maximum Entropy (ME) as the basic supervised classification model, while the ME algorithm is implemented with the *MALLET Toolkit*<sup>2</sup>. Note that, as the number of posts which express *fear* and *surprise* are limited, we only detect the other three kinds of emotions, i.e. *happiness*, *sadness*, and *anger*.

As discussed in the above subsection, we use the following approaches for automatic emotion detection in code-switching text:

- $f_{ALL}$ : which uses all the words of each post as a feature to train a Maximum Entropy (ME) classification model.

- $f_{CN}$ : which only uses the Chinese text of each post as a feature to train a Maximum Entropy (ME) classification model.
- $f_{EN}$ : which only uses the English text of each post as a feature to train a Maximum Entropy (ME) classification model.
- $f_{comb}$ : which combines the results of the Chinese text classifier  $f_{CN}$  and English text classifier  $f_{EN}$  using the sum rule.

The results of emotion detection are shown in Table 7. The performance indicates the accuracy of detecting emotions in code-switching text.

	Acc.
$f_{ALL}$	0.509
$f_{CN}$	0.521
$f_{EN}$	0.409
$f_{comb}$	0.539

Table 7: Results of emotion detection in code-switching text

From the table, we find that:

- 1) The performance of basic approach  $f_{ALL}$  which uses mixed text directly is inferior.
- 2) As Chinese is the dominant language, and the English text is loosely distributed, using Chinese text ( $f_{CN}$ ) outperforms both using all text ( $f_{ALL}$ ) and English text ( $f_{EN}$ ). Besides, as the English texts in the posts are always composed of single words,  $f_{EN}$  is much lower than the other two approaches.
- 3) As incorporating both Chinese classifiers and English classifiers to a multiple classifier system,  $f_{comb}$  achieves a better performance than the other approaches. It also indicates that both Chinese text and English text in code-switching posts are effective for detecting emotions.

## 6 Conclusion

This paper presents the development of a code-switching emotion corpus in which the emotion is expressed through either Chinese or English. We first collect and filter the data from *Weibo.com*, which is annotated with both emotion and caused language; we then analyze the inter-annotator agreement on the dataset, and present our findings and analysis. Finally, we propose a multiple-classifiers-based approach to detect emotion in the annotated code-switching corpus. Results show that both Chinese text and English text in code-switching posts are effective

<sup>1</sup> <https://code.google.com/p/fudanntp/>

<sup>2</sup> <http://mallet.cs.umass.edu>

in detecting emotions. We believe that emotions analysis in code-switching text underlies an innovative approach towards a linguistic model of emotion as well as automatic emotion detection and classification.

## Acknowledgments

The work is funded by an Early Career Scheme (ECS) sponsored by the Research Grants Council of Hong Kong (No. PolyU 5593/13H), and supported by the National Natural Science Foundation of China (No. 61273320, and No. 61375073) and the Key Project of the National Natural Science Foundation of China (No. 61331011).

Firstly, we need to thank the hard works of the annotators. We thank Prof. Shoushan Li for his useful discussion. We acknowledge Helena Yan Ping Lau for corpus analysis and insightful comments. We also thank anonymous reviewers for their valuable suggestions and comments.

## References

- Adel H., N. Vu, and T. Schultz. 2013. Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling. In *Proceedings of ACL-13*.
- Andrew R. 1963. Evolution of Facial Expressions. *Science*, 142, 1034-1041.
- Amini, M., C. Goutte, and N. Usunier. 2010. Combining Coregularization and Consensusbased Self-training for Multilingual Text Categorization. In *Proceeding of SIGIR-10*.
- Auer P. 1999. *Code-Switching in Conversation*. Routledge.
- Arnold J. 1999. *Affect in Language Learning*. Cambridge, MA: CUP.
- Auer P. 1984. *Bilingual Conversation*. Amsterdam: John Benjamins.
- Blom J., and J. Gumperz. 1972. Social Meaning in Linguistic Structures: Code Switching in Northern Norway. *Directions in Sociolinguistics*. New York: Winston.
- Bridge D., J. Chiao, and K. Paller. 2010. Emotional Context at Learning Systematically Biases Memory for Facial Information. *Memory & Cognition*, 38, 125-133.
- Burkett, D., and D. Klein. 2008. Two Languages are Better than One (for Syntactic Parsing). In *Proceedings of EMNLP-08*.
- Chen Y., S. Lee, S. Li, and C. Huang. 2010. Emotion Cause Detection with Linguistic Constructions. In *Proceeding of COLING-10*.
- Dasgupta S., and V. Ng. 2009. Mine the Easy, Classify the Hard: A Semi-Supervised Approach to Automatic Sentiment Classification. In *Proceedings of ACL-IJCNLP-09*.
- Dewaele J., and A. Pavlenko. 2002. Emotion Vocabulary in Interlanguage. *Language Learning*, 52 (2), 265-324.
- Ekman, P., and W.V. Friesen. 1978. *Facial Action Coding System*. California: Consulting Psychology Press.
- Gao W., J. Blitzer, M. Zhou, and K. Wong. 2009. Exploiting Bilingual Information to Improve Web Search. In *Proceedings of ACL/IJCNLP-09*.
- Giles, H., and R. Clair. 1979. *Language and Social Psychology*. London: Basil Blackwell.
- Hervé P., A. Razafimandimby, M. Vigneau, B. Mazoyer, and N. Tzourio-Mazoyer. 2012. Disentangling the Brain Networks Supporting Affective Speech Comprehension. *NeuroImage*, 61(4), 1255-1267.
- Kittler J., M. Hatef, R. Duin, and J. Matas. 1998. On Combining Classifiers. *IEEE Trans. PAMI*. 20.226-239.
- Lee S., H. Zhang, and C. Huang. 2013a. An Event-Based Emotion Corpus. In *Proceedings of CLSW 2013*.
- Lee S., Y. Chen, C. Huang, and S. Li. 2013b. Detecting Emotion Causes with a Linguistic Rule-Based Approach. *Computational Intelligence*, 29(3), 390-416.
- Lee S., S. Li, and C. Huang. 2014. Annotating Events in an Emotion Corpus. In *Proceedings of LREC-14*.
- Li S., S. Lee, Y. Chen, C. Huang, and G. Zhou. 2010. Sentiment Classification and Polarity Shifting. In *Proceeding of COLING-10*.
- Li Y., and P. Fung. 2012. Code-switch Language Model with Inversion Constraints for Mixed Language Speech Recognition. In *Proceedings of COLING-12*.
- Ling W., G. Xiang, C. Dyer, A. Black, and I. Trancoso. 2013. Microblogs as Parallel Corpora. In *Proceedings of ACL-13*.
- Liu H., S. Li, G. Zhou, C. Huang, and P. Li. 2013. Joint Modeling of News Reader's and Comment Writer's Emotions. In *Proceedings of ACL-13, shorter*.
- Lignos C., and M. Marcus. 2013. Toward Web-scale Analysis of Codeswitching. In *Proceedings of Annual Meeting of the Linguistic Society of America*.
- Lu B., C. Tan, C. Cardie, and B. Tsou. 2011. Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora. In *Proceedings of ACL-11*.



- Myers-Scotton C. 1997. *Duelling Language: Grammatical Structure in Code-switching*. Oxford: Clarendon.
- Olson I., A. Plotzker, and Y. Ezzyat. 2007. The Enigmatic Temporal Poles: A Review of Findings on Social and Emotional Processing. *Brain*.
- Pavlenko A. 2008. Structural and Conceptual Equivalence in Acquisition and Use of Emotion Words in a Second Language. *Mental Lexicon*, 3(1): 91-120.
- Peng N., Y. Wang, and M. Dredze. 2014. Learning Polylingual Topic Models from Code-Switched Social Media Documents. In *Proceedings of ACL-14*.
- Quan C., and F. Ren. 2009. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. In *Proceedings of EMNLP-09*.
- Rao Y., X. Quan, W. Liu, Q. Li, and M. Chen. 2012. Building Word-emotion Mapping Dictionary for Online News. In *Proceedings of SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data*.
- Schrauf R. 2000. Bilingual Autobiographical Memory: Experimental Studies and Clinical Cases. *Culture and Psychology*. 6 (4), 387-417.
- Schumann J. 1999. A Neurobiological Perspective on Affect and Methodology in Second Language Learning. *Affect in Language Learning*. Cambridge: CUP, 28-42.
- Smith C., and L. Kirby. 2001. Toward Delivering on the Promise of Appraisal Theory. *Appraisal processes in emotion: Theory, methods, research*. Oxford, UK: Oxford University Press.
- Smith C., and R. Lazarus. 1993. Appraisal Components, Core Relational Themes, and the Emotions. *Cognition and Emotion*, 7, 233-269.
- Solorio T., and Y. Liu. 2008. Learning to Predict Code-Switching Points. In *Proceedings of EMNLP-08*.
- Volkova S., W. Dolan, and T. Wilson. 2012. CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language. In *Proceedings of EACL-12*.
- Wen S. and X. Wan. 2014. Emotion Classification in Microblog Texts Using Class Sequential Rules. In *Proceedings of AAAI-14*.
- Xu G., X. Meng, and H. Wang. 2010. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources. In *Proceeding of COLING-10*.
- Yang M., B. Peng, Z. Chen, D. Zhu, and K. Chow. 2014. A Topic Model for Building Fine-grained Domain-specific Emotion Lexicon. In *Proceedings of ACL-14*.