

Annotating Events in an Emotion Corpus

Sophia Yat Mei Lee[†], Shoushan Li[‡], Chu-Ren Huang[†]

[†]The Hong Kong Polytechnic University [‡]Soochow University

Hong Kong

China

E-mail: ymlee@polyu.edu.hk, lishoushan@suda.edu.cn, churen.huang@polyu.edu.hk

Abstract

This paper presents the development of a Chinese event-based emotion corpus. It specifically describes the corpus design, collection and annotation. The proposed annotation scheme provides a consistent way of identifying some emotion-associated events (namely pre-events and post-events). Corpus data show that there are significant interactions between emotions and pre-events as well as that of between emotion and post-events. We believe that emotion as a pivot event underlies an innovative approach towards a linguistic model of emotion as well as automatic emotion detection and classification.

Keywords: Emotion corpus, pre-events, post-events

1. Introduction

Emotion is basic to human experience and communication, and at the same time very abstract and complex in nature. It is commonly defined as the bodily reaction to actual external stimuli which in turn leads to potential consequences. For example, the emotion of fear can be evoked by a threatening situation which in turn motivates potential actions to remove fear. For example,

- [1] 他看了這種現象，感到非常害怕，於是也趕快逃跑了。
‘[When] he saw such a thing, he was frightened. Then [he] ran away very quickly.’

Sentence [1] shows the context of an emotion ‘害怕’ (*frightened*) in which the event ‘他看了這種現象’ (*he saw such a thing*) induced the emotion, which in turn led to the another event ‘趕快逃跑’ (*[he] ran away very quickly*). Although the connection of the various events is logically assumed to be true, the linguistic analysis of emotion-event interaction is underexplored. In this paper, we take the position that emotions serves as a pivot event linking provoking events to potential actions by the experiencer (Lee et al. 2012).

This paper aims to describe the development of the Chinese event-based emotion corpus, with a focus on its design, collection, and annotation. It also accounts for the linguistic interactions between evoking events (i.e. pre-events) and caused emotions, and the correlations between emotions and events induced (i.e. post-events) in texts. Such a study will offer rich structured data allowing the development of a theory of emotion as events.

2. Related Work

The earliest research on emotion has focused on the representation and processing of emotion in facial expressions and body language (Andrew, 1963; Ekman & Friesen, 1978). More recently, there has been mounting research on the neurobiological basis of emotion (Olson et al., 2007; Craig, 2009; Hervé et al., 2012) and how emotion is linked with other aspects of human cognition (Smith & Lazarus, 1993; Smith and Kirby, 2001; Bridge et al., 2010).

Previous work has also reported on a close link between types of emotion and various prosodic features. Kehrein (2002) showed that the type of emotion resulted in different intensity and rate of speech. For example, *anger* is closely related to higher intensity and faster speech than *frustration*. Abelin and Allwood (2000), on the other hand, suggested that speech duration varies with emotion types, such as *fear* and *sadness* are more connected to speeches with shorter duration as compared to *happiness* and *surprise*.

This line of research has been extended to cover emotion in verbal language as well. It has been shown that when humans encounter emotional language, the processing of language and the processing of emotion are highly intertwined (Hervé et al., 2012). The language-processing part of the brain is responsible for extracting linguistic cues for emotion, while the emotion-processing part makes inferences about the emotional content based on linguistic meanings.

Emotion has also been well studied in natural language processing (Picard, 1995; Dellaert et al., 1996; Ortony et al. 1990). Recent research has begun to place more emphasis on automatic emotion detection and classification from textual input (Mihalcea and Liu, 2006; Ahmad, 2008). Most of the previous studies have focused on classifying descriptive emotions given a known emotion context, such as a sentence or a document, using either rule-based (Chaumartin, 2007) or statistical learning (Mihalcea and Liu, 2006) approaches. Some works began to explore both emotion detection and classification. Tokuhisa et al. (2008) created a Japanese emotion cause corpus using an unsupervised approach, whereas Chen et al. (2009) developed a Chinese emotion corpus and an English emotion corpus using a semi-supervised approach.

3. Emotion Corpus

3.1 Emotion and the Associated Events

As mentioned in Section 1, emotion is regarded as a sub-type of event interacts with other associating events, namely pre-events and post-events (Lee et al. 2013). The

interaction between the emotion and the two types of events as presented in sentence [1] is shown in Figure 1.

- [1] 他看了這種現象，感到非常害怕，于是也趕快逃跑了。
 ‘[When] he saw such a thing, he was frightened. Then [he] ran away very quickly.’

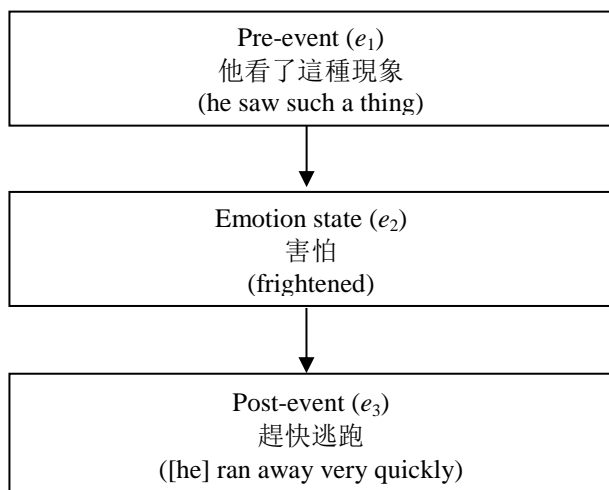


Figure 1: Emotion-event interaction

A pre-event (e_1) refers to the event that is triggered by or highly linked with the presence of the corresponding emotions (e_2). A post-event (e_3) is an event triggered by the emotion (e_2), which shows a clear cause-effect relation. Both types of events are usually linguistically represented in the form of verbs as in [2], nominal as in [3] and nominalizations as in [4].

- [2] 他非常傷心，於是哭了一場。(e₃: verb)
 ‘He was terribly heartbroken, and so cried many tears.’
- [3] 對於未來，老實說我很害怕。(e₁: nominal)
 ‘Regarding **the future**, I’m honestly quite scared.’
- [4] 剛剛你那副爬不起來的樣子，是有點叫我害怕。
 (e₁: nominalization)
 ‘That he couldn’t get up frightened me.’

Although we argue that emotion is treated as a pivot event which links the pre-events and post-events (as described in [1]), it is not assumed that the two events always exist in every instance of emotion. All the following combinations of the various events are found in the corpus:

- [5] a. pre-event(s) + emotion + post-event(s)
 b. pre-event(s) + emotion
 c. emotion + post-event(s)
 d. emotion (without any associating event)

3.2 Corpus Data

Based on our proposal that emotion is treated as a pivot event which connects the pre-events and post-events, we have constructed an event-based emotion corpus. Data are extracted from the Sinica Corpus using a pattern-based. Based on the list of 91 Chinese primary emotion keywords identified in Chen et al. (2009), we extracted 8,973 instances of sentences by keyword matching from the Sinica Corpus, which is a tagged balanced corpus of Mandarin Chinese containing ten million words¹. Each instance contains the focus sentence with the emotion keyword “<FocusSentence>,” plus the sentence before “<PrefixSentence>” and after “<SuffixSentence>” it. A sample instance is given in Figure 2. The emotion is indicated as <emo id=0>生氣(anger)</emo> whereas the pre-event is marked with “[...]” and the post-event “{...}”.

```

snc_11417 Y 0/生氣/Anger
<PrefixSentence> 過些時候，[[魯班的妻子懷孕了]]，肚子一天比一天大。 </PrefixSentence>

<FocusSentence>魯班的父親很<emo id=0>生氣</emo>，就{{罵她}}說：我兒子在涼州做事，離家那麼遠，很久沒有回來過，你竟然懷孕了，真是可耻！ </FocusSentence>

<SuffixSentence>魯班的妻子受了冤枉，很不甘心，就把魯班每晚乘木鸞回來的情形告訴他父親。 </SuffixSentence>

snc_11417 Y 0/生氣/Anger
<PrefixSentence>After some time, [[Luban's wife became with child]]. Her pregnant belly grew with each day. </PrefixSentence>

<FocusSentence>Luban's father was <emo id=0>furious</emo>, {{scolding her}}: "My son works in faraway Liangzhou and has not come back for so long. Yet you have conceived. What a shameful act!" </FocusSentence>

<SuffixSentence>Luban's wife could not bear the wrongful accusation, and told her father-in-law of how Luban had returned home each night by riding a wooden kite to be with her. </SuffixSentence>
  
```

Figure 2: A sample of an event-annotated instance

4. Event Annotation

This section describes the annotation procedure and guidelines. It begins with a description of the annotation tool as well as annotators, followed by a set of guidelines to explain exactly how the task should be carried out.

¹ An earlier version of the Chinese event-based Emotion Corpus contained 3,000 instances (Lee et al. 2013).

4.1 Annotation Tool

An annotation tool is designed to facilitate the annotation process which allows for better consistency. Four annotators were recruited for the annotation task. Two annotators annotated the pre-events and post-events of the same set of emotion instance. Figure 3 shows an example instance annotated with the corresponding pre-event and post-event using our annotation tool. For each identified event, annotators marked whether it is a pre-event or a post-event, together with other information including event type and event subject. Event type refers to verbal or nominal event. A verbal event is a linguistic expression denoting an event that involves a verb or nominalization (indicated as ‘event’ in the annotation tool), whereas a nominal event is simply a noun (indicated as ‘nominal’ in the annotation tool). Event subject suggests whether the subject of pre- or post-event is the experiencer of the emotion.

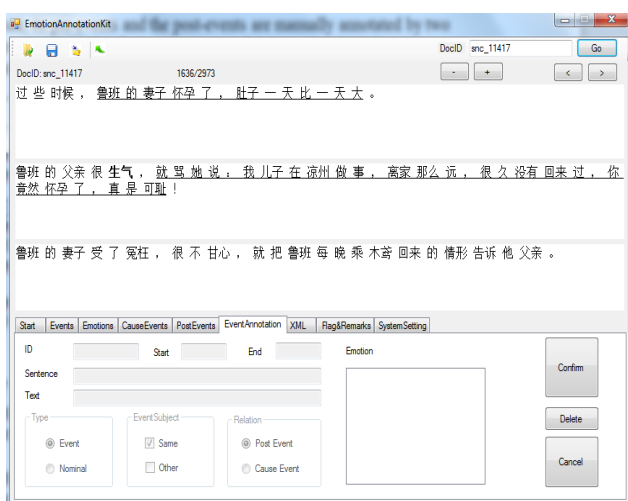


Figure 3: A sample of event annotation using the annotation tool

4.2 Annotation Guidelines

We mark the shortest meaningful pre-events and post-events that are closest to the emotion keywords. Some guidelines for marking the events are given below:

Determining the event boundaries

- Only the immediate pre-events or post-events are annotated. For example, in [6], ‘身體僵硬’ (*body stiff*) is considered the direct reaction to fear, i.e. the post event. The follow-up actions such as ‘痛哭求饒’ (*broke down in tears and begged for mercy*) not marked as the post-event.

[6] ... 一動也不動的站著，因恐懼而身體僵硬。最後甚至痛哭求饒。
‘..(he) stood still, unmoved, body stiff from fear. At last, (he) broke down in tears and begged for mercy.’

- When the two events are closely tied, both events are marked as the pre-events or post-events. For example, in (7), ‘he fell ill’ and ‘[he] passed away’ are closely tied both syntactically and semantically.

[7] 那少年很傷心，生了一場病便死了。

‘The youth was deeply heartbroken; he fell ill and passed away.’

- A pre-event can trigger different emotions. Similarly, a post-event can be triggered by different emotion keywords. In [7], the underlined part is the post-event of the two events of 害怕 ‘fear’ as shown in text.

[7] 娼妓會成為「人類最古老的行業」，主要原因就是社會為了保全家庭，害怕亂倫及不正常的婚外性關係，害怕性苦悶沒有合法的解決管道，所以藉開放周邊價值來保全中心價值。

‘Prostitution became “the world’s oldest profession” primarily because of society’s need to protect the family. As [it] feared the existence of incest and abnormal extramarital relationships, and [it] feared that sexual boredom could not find release through appropriate channels, [it] conceded on marginal values to preserve its core ones.’

- Some peripheral information is not marked as part of the pre-events or post-events which includes:

- Reported verbs, e.g. ‘談到’，‘說到’(speaking of) [8] 談到「寶來國際金融機場」，寶來證券集團董事長白文正難掩興奮之情。

(Speaking of ‘The Polaris International Financial Airport’, the chairman of the board of Polaris Securities, Bai Wenzheng, could barely contain his excitement.)

- Prepositions, e.g. ‘對於’，‘關於’(regarding)

[9] 對於未來，老實說我很害怕。
(Regarding the future, I’m honestly quite scared.)

- Conjunctions, e.g. ‘於是’，‘所以’(and so)

[10] 他非常傷心，於是哭了一場。
(He was terribly heartbroken, and so cried many tears.)

- Adverbs, e.g. ‘也’(also), ‘就’(then) unless they are in the middle of the pre-event or post-event [11] 傷心的醜小鴨，也就離開美麗的大池塘了
(The heartbroken ugly duckling, then left the beautiful lake.)

- sentence final particles, e.g. 了 as in [11]

- unnecessary punctuation marks, e.g. commas and full stops; necessary ones are question marks, exclamation marks, open and close quotation marks

- The subject of the pre-event or post-event should be marked when it is present in the context. For example, ‘辛巴’(Xinba) in [12] should be marked as part of the post-event.

[12] 當辛巴聽到父親死去的消息後，傷心的不得了，認為是自己害死了父親，再加上刀疤的慫恿，辛巴便離開家園。

(When Xinba heard the news of his father’s passing, he became terribly heartbroken. Under the impression he

had caused his father's death, and the influence of the scar, Shenba left his home.)

- Saying action is considered part of a pre-event or post-event. The ‘say’ verb together with the content should both be marked as in [13].

[13] 蘭妮公主聽到自己永遠不能恢復人形，傷心極了。只好對王子說：「忘了我吧！」

(On hearing that she would never be able to resume her human form again, Princess Lanny was overcome with grief. She could only tell the prince: “Forget me!”)

Determining the Appropriate Events

In Chinese, the usage of the three particles ‘得’，‘的’，and ‘地’ can be confusing. It is found that the three particles are used interchangeably in the corpus data, which is rather misleading. To determine the appropriate events to annotate, we have set the following guidelines:

- The structure ‘emotion word + 得 + verb’ denotes a cause-effect relation. For instance, in [14], ‘說不出話來’ (*[he] could not bear to speak*) is the effect of ‘傷心’ (*fear*). In this case, ‘說不出話來’ is considered a post-event.

[14] 尼奧傷心得說不出話來。

(Leo felt such pain [he] could not bear to speak.)

- The structure ‘emotion word + 的/地 + verb’ mostly denotes two simultaneous actions as shown in [15] and [16]. For example, in [15], the state of sadness and the action of bringing back the body of the deceased are not in a cause-effect relation.

[15] 一行人哀傷的將屍體運回。

(The grieving party brought back the body of the deceased.)

[16] 談到保養心得，黃嘉千心虛地說：「其實我很愛漂亮，可是又有點懶...

(On the topic of anti-aging secrets, Huang Jiaqian said diffidently, “I actually love beauty, but I’m also a bit lazy...”)

Since the three particles are sometimes used interchangeably in the corpus data, it is suggested that when the event can be rewritten as the double-conjunction structure ‘因為... 所以...’ (*because...therefore...*) which denotes a clear causal relation. For example,

因為傷心所以哭

(because [he was] heartbroken, [therefore] he cried)

*因為心虛所以說...

(because [she felt] diffident, [therefore] (she) said)

*因為哀傷所以把屍體運回

(because [they were] sad, therefore [they] brought back the deceased body).

- When the event involves bodily reaction, e.g., *cry* and *tremble*, it is considered a pre-event or post-event.

[17] 小寶一邊傷心的哭著

(Xiao Bao cried sadly)

5. Corpus Analysis

Table 1 shows that, out of the 8,973 instances of emotion, 73.9% contain a pre-event and 15.3% contain a post-event in the context. We also notice that pre-events tend to occur before the emotion keyword (64.1%), while post-events mostly occur after the emotion keyword (94.5%). This is in line with the assumption that there is a sequential ordering among the various events. Of the 35.9% of the pre-events that appear after the emotion keyword, most are represented in the form of ‘...emotion word + 的是’ ‘is’ + pre-event’ in that the pre-events are explicitly expressed as the reason for the emergence of that particular emotion.

	Pre-event	Post-event
Total	73.9%	15.3%
Before emo	64.1%	94.5%
After emo	35.9%	5.5%
Verb	83.6%	97.9%
Nominal	16.4%	2.1%

Table 1: Analysis of pre-events and post-events found in the corpus

In terms of linguistic representation, both pre-event and post-event are mostly expressed as verbs, which are 83.6% and 97.9% respectively. Nominal pre-events are more likely to appear as the topic of the sentence. In addition, the pre-events and post-events tend to be introduced by a list of linguistic cues, as shown in Table 2 and Table 3 respectively.

Type	Cue words
Causative verbs	讓，令，使 (to cause)
Reported verbs	說到，談到 (speaking of)
Epistemic markers	看到 (to see)，聽到 (to hear)
Say verbs	的說 (to say)
Prepositions	為了，對於 (for)
Conjunctions	因為，由於 (because)
Others	的是 (is)

Table 2: Linguistic cues associated with pre-events²

Type	Cue words
Causative verbs	讓，令 (to cause)
Particles	得，的 (to the extent that)
Adverbs	也 (also)，就 (then)，起來 (start to)，之後，(afterwards)，不禁 (can’t help)
Conjunctions	於是，因此，而 (so)，結果 (as a result)

Table 3: Linguistic cues associated with post-events

² With a larger set of data in this paper, we have also found similar linguistic cues identified in Lee et al. (2012). For a full version of the list of linguistic cues, please refer to Lee et al. (2012).

As for post-events, it is also observed that there is a close association between the emotion and the event type. For instance, the emotion of anger often triggers shouting events which are expressed as 罵 ‘to scold’, 大吼 ‘to yell’, 咆哮 ‘to roar’, etc. More examples of emotion-post-event association are shown in Table 4.

Emotions	Event types
Happiness	笑 (to laugh) 擁抱 (to hug) 跑 (to run) 大叫 (to shout)
Sadness	哭 (to cry) 死 (to die) 離開 (to leave) 下定決心 (to determine)
Fear	不敢 (not dare) 躲 (to hide) 逃 (to flee) 身體 (body)
Anger	罵 (to scold) 大吼 (to yell) 咆哮 (to roar) 破壞 (to destroy) 殺 (to kill)

Table 4: Emotion-Event Association

6. Conclusion

As part of a project, this paper presents the development of a Chinese event-based emotion corpus based on our proposal that emotion is treated as a pivot event which connects the pre-events and post-events. It specifically describes the corpus data and the annotation scheme. The proposed annotation scheme provides a consistent way of identifying some emotion-associated events. It also points out some challenges faced and the respective solutions in the process of event annotation. Corpus data indicate that emotions interact with pre-events and post-events in various ways, including syntactic representations, linguistic markers, and event types. We believe that emotion as a pivot event underlies an innovative approach towards a linguistic model of emotion/event as well as automatic emotion detection and classification.

7. Acknowledgements

This work is supported by a General Research Fund (GRF) sponsored by the Research Grants Council, Hong Kong (Project No. 543810) and a Faculty Research Grant sponsored by the Hong Kong Polytechnic University (Project no. 1-ZVB2).

8. References

Abelin, A. & Allwood, J. (2000). Cross linguistic interpretation of emotional prosody. *Proceedings of the ISCA ITRW on Speech and Emotion*, 110–113.

Ahmad, K. (Ed.). (2008). *Proceedings of the LREC Workshop on Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology*.

Andrew, R. J. (1963). Evolution of facial expressions. *Science*, 142, 1034-1041.

Bridge, D.J., Chiao, J.Y. and Paller, K.A. (2010). Emotional context at learning systematically biases memory for facial information. *Memory & Cognition*, 38, 125-133.

Chaumartin, F. R. (2007). A Knowledge-based System for Headline Sentiment Tagging. *Proceedings of the 4th International Workshop on Semantic Evaluations*.

Chen, Y., Lee, S. Y. M., and Huang, C.-R. (2009). A Cognitive-based Annotation System for Emotion Computing. *Proceedings of the Third Linguistic Annotation Workshop (The LAW III)*.

Craig, A.D. (2009). How do you feel now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10(1):59-70.

Dellaert, F., Polzin, T., & A. Waibel. (1996). Recognizing Emotion in Speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP-96)*.

Ekman, P. and W.V. Friesen. (1978). *Facial Action Coding System*. California: Consulting Psychology Press.

Hervé, P.-Y., Razafimandimby, A., Vigneau, M., Mazoyer, B. and Tzourio-Mazoyer, N. (2012). Disentangling the brain networks supporting affective speech comprehension. *NeuroImage*, 61(4), 1255-67.

Kehrein, R. (2002). The Prosody of Authentic Emotions. *Proceedings of Speech Prosody 2002*.

Lee, S. Y. M., Zhang H., and Huang, C. -R. (2013). An Event-based Emotion Corpus. In *Proceedings of Chinese Lexical Semantics Workshop*. LNCS Springer.

Lee, S. Y. M., Chen, Y., Huang, C.-R., and Li, S. (2012). Detecting Emotion Causes with a Linguistic Rule-based Approach. *Computational Intelligence, Special Issues on Computational Approaches to Analysis of Emotion in Text*. Wiley-Blackwell.

Mihalcea, R., and Liu, H. (2006). A Corpus-based Approach to Finding Happiness. *Proceedings of the AAAI Spring Symposium on Computational Approaches to Weblogs*.

Olson, I.R. Plotzker, A., Ezzyat, Y. (2007). The Enigmatic Temporal Poles: A Review of Findings on Social and Emotional Processing. *Brain*.

Ortony, A., G. L. Clore, and A. Collins. (1990). *The Cognitive Structure of Emotions*. Cambridge University Press.

Picard, R.W. (1995). *Affective Computing*. Cambridge: The MIT Press.

Smith, C.A., and Kirby, L.D. (2001). Toward delivering on the promise of appraisal theory. In K.R. Scherer, A. Schor, and T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research*. Oxford, UK: Oxford University Press.

Smith, C.A., and Lazarus, R.S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, 7, 233–269.

Tokuhisa, R., Inui, K., and Matsumoto, Y. (2008).
Emotion Classification Using Massive Examples
Extracted from the Web. In *Proceedings of
International Conference on Computational
Linguistics (COLING-08)*.