Emotion Cause Events: Corpus Construction and Analysis

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

[†]Department of Chinese and Bilingual Studies [‡]Institute of Linguistics
The Hong Kong Polytechnic University   Academia Sinica
{sophiaym, chenying3176, shoushan.li, churenhuang}@gmail.com

Abstract

Emotion processing has always been a great challenge. Given the fact that an emotion is triggered by cause events and that cause events are an integral part of emotion, this paper constructs a Chinese emotion cause corpus as a first step towards automatic inference of cause-emotion correlation. The corpus focuses on five primary emotions, namely *happiness*, *sadness*, *fear*, *anger*, and *surprise*. It is annotated with emotion cause events based on our proposed annotation scheme. Corpus data shows that most emotions are expressed with causes, and that causes mostly occur before the corresponding emotion verbs. We also examine the correlations between emotions and cause events in terms of linguistic cues: causative verbs, perception verbs, epistemic markers, conjunctions, prepositions, and others. Results show that each group of linguistic cues serves as an indicator marking the cause events in different structures of emotional constructions. We believe that the emotion cause corpus will be the useful resource for automatic emotion cause detection as well as emotion detection and classification.

1. Introduction

Emotion plays an important role in human communication. Its immediate responses are usually expressed via speaking, facial expression, or human behavior. With the advances of computer-mediated communication, emotional information is widely conveyed by means of written languages such as email, messenger, web blogs, forum, and chat rooms. Understanding emotions is becoming increasingly important not only for speech processing but also for written text. Hence, textual emotion analysis has begun to attract plenty of attention and has been a popular topic for research in Natural Language Processing (NLP). Most research has focused on emotion detection and classification where emotion types are identified, for instances *happiness* and *sadness*, for a given sentence or document (Mihalcea and Liu 2006, Tokuhisa et al. 2008). However, on top of this surface level information, deeper level information regarding emotions such as the experiencer, cause, and result of an emotion, needs to be extracted and analyzed for many real world applications. In this paper, we aim at mining one of the crucial deep level information, i.e. emotion cause, which provides useful information for applications ranging from economic forecast, public opinion mining, to product design.

Most theories of emotion treat recognition of a triggering cause event as an integral part of emotion (Descartes 1649, James 1884, Plutchik 1962, Wierzbicka 1996). In this study, cause events refer to the explicitly expressed arguments or events that trigger the presence of the corresponding emotions. For example, "they like it" is the cause event of the emotion *happiness* in the sentence "I was very happy that they like it". By detecting and analyzing the emotion cause events, we will have a clear picture of what induces emotions so that certain emotions can be inspired or avoided. Therefore, we attempt to build an emotion corpus annotated with cause events so as to help develop automatic detection of emotion cause events.

This study considers emotions that are explicitly expressed, which are usually presented by keywords, while implicit emotions that require inference or connotation are not discussed. First, we build a Chinese emotion corpus with five primary emotions, i.e. *happiness*, *sadness*, *anger*, *fear*, and *surprise*. Then, we annotate the cause event of each emotion verbs. Based on the annotated data, we examine the various linguistic cues which help detect emotion cause event: the position of cause event, causative verbs, perception verbs, epistemic markers, prepositions, conjunctions, and others. With the help of these linguistic cues, the causes of emotions can be effectively detected.

We believe that the corpus of emotion cause event is useful for many real world applications. Since manual detection on cause events is labour-intensive and time-consuming, the obtained corpus will be the useful resource for automatic systems detecting cause events for emotion processing in text. Furthermore, the corpus will contribute to emotion detection and classification as it provides important clues for the identification of implicit emotions.

The paper is organized as follows. Section 2 discusses the related work on various aspects of emotion analysis. Section 3 discusses the notion of cause events. Section 4 describes the construction of the emotion cause corpus. Section 5 presents our corpus analysis based on the various linguistic cues. Section 6 concludes the paper.

2. Related Theoretical Issues

2.1 Emotion Classes

Various approaches to emotion analysis were proposed in different fields, such as philosophy (Spinoza 1675, James 1884), biology (Darwin 1859), linguistics (Wierzbicka 1996, Kövecses 2000), neuropsychology (Plutchik 1962, Turner 1996), and computer science (Ortony et al. 1988, Picard 1995), as well as varying from language to language. Although there is no agreement among different theories on the emotion classes, a small number of primary emotions are commonly assumed whereas other emotions are secondary emotions which are the mixtures of the primary emotions.

Researchers have attempted to propose the list of primary emotions, varying from two to ten basic emotions (Ekman 1984, Plutchik 1980, Turner 2000). *Fear* and

anger appear on every list, whereas *happiness* and *sadness* appear on most of the lists. These four emotions, i.e. *fear*, *anger*, *happiness*, and *sadness*, are the most common primary emotions. Other less common primary emotions are *surprise*, *disgust*, *shame*, *distress*, *guilt*, *interest*, *pain*, and *acceptance*.

This study starts with a list of primary emotions agreed upon by most. It adopts Turner's emotion classification (2000), which identifies five primary emotions, namely *happiness*, *sadness*, *fear*, *anger*, and *surprise*.

2.2 Emotion Processing in Text

Research on textual emotion processing is still in its early stages in NLP. Most of the previous works focus on the emotion classification given a known emotion context such as a sentence or a document using either rule-based (Masum et al. 2007, Chaumartin 2007) or statistical approaches (Mihalcea and Liu 2005, Kozareva et al. 2007). However, the performance is far from satisfactory. What is more, many basic issues remain unresolved, for instances, the relationships among emotions, emotion class selection, etc. Tokuhisa et al. (2008) was the first to explore both the issues of emotion detection and classification. It created a Japanese emotion-provoking event corpus for emotion classification task using unsupervised approach. However, only 49.4% of cases are correctly labeled. Chen et al. (2009) developed two cognitive-based Chinese emotion corpora using semi-unsupervised approach, i.e. an emotion-sentence (sentences containing emotions) corpus and a neutral-sentence (sentences contain no emotion) corpus. They showed that studies based on the emotion-sentence corpus (~70%) outperform previous corpora.

Little research, if not none, has been done to examine the interactions between emotions and the corresponding cause events, which may make a great step towards an effective emotion classification model. The lack of research on cause events restricted current emotion analysis to simple classificatory work without exploring the potentials of the rich applications of putting emotion 'in context'. In fact, emotions are invoked by perceptions of external events and in turn trigger reactions. The ability to detect implicit invoking causes as well as predict actual reactions will add rich dimensions to emotion analysis and lead to further research on event computing.

3. Cause Events

In the context of emotion constructions, the causal relation is established by the link between a cause and the emotional state. Following Talmy (2000), the cause of an emotion should be an event itself. In this work, it is called a cause event. By cause event, we do not necessarily mean the actual cause of the emotion or what leads to the emotion. Rather, it refers to the immediate cause of the emotion, which can be the actual trigger event or the perception of the trigger event. The emotion *fear* is more likely to be linked to the perception of the trigger event, for instance a snake is not the direct cause of *fear* in a person, but through the cognitive awareness of the possible danger that the snake may bring to the person. In other words, any event that is highly associated with the presence of an emotion is a cause event.

Adapting TimeML annotation scheme (Sauri et al. 2004), events refer to situations that happen or occur. They can be punctual or last for a period of time, as well as predicates describing states or circumstances in which something obtains or holds true. In this study, cause events specifically refer to the explicitly expressed arguments or events that are highly linked with the presence of the corresponding emotions. They are usually expressed by means of verbs, nominalizations, and nominals. In Chinese, cause events are categorized into two types: verbal events and nominal events. Verbal events refer to events that involve verbs or nominalization as in (1) and (2), whereas nominal events are simply nouns as in (3).

(1) 沒想到**她說的都是真話**，讓他震驚不已。
Mei2-not *xiang3dao4*-think **ta1**-3.SG.F **shuo1**-say **de**-POSS **dou1**-all **shi4**-is **zhen1**-true **hua4**-word, *rang4*-lead *ta1*-3.SG.M *zhen4jing1*-shocked *bu4yi3*-very
"He was very shocked that what she said was the truth."

(2) 他對**這個充滿濃厚愛意的想法**高興的手舞足蹈。
Ta1-3.SG.M *dui4*-for *zhe4*-DET *ge4*-CL **chong1man3**-full of **nong2hou4**-dense **ai4yi4**-love **de**-DE **xiang3fa3**-idea *gao1xing4*-happy *de*-DE *shou3wu3zu2dao3*-flourish
"He was very happy about this loving idea."

(3) **尼奧的話**很令凱洛琳傷心。
Ni2ao4-Leo **de**-POSS **hua4**-word *hen3*-very *ling4*-make *kai3luo4lin2*-Helena *shang1xin1*-sad
"Helena was very sad by Leo's words."

Sentences (1)-(3) show examples of cause events which do not require coreference. Yet, in the corpus, there are many cases requiring coreference. For instance, in (4), the determiner *ci3/this* is considered the cause event of *anger*. It is an anaphoric expression that refers to the previous clause "*The Taiwan authorities refused the person involved returning to Taiwan on political grounds*". In cases such as (4), both are marked as cause events in that the determiner marks as a nominal cause event while the antecedent a verbal cause event.

(4) **台灣當局以政治理由拒絕當事人返台**。當事人對**此**憤懑。
Tai3wan1-Taiwan **dang1ju2**-authority **yi3**-as **zheng4zhi4**-political **li3you4**-reason **ju4jue2**-refuse **dang1shi4ren2**-the person involved **fan3**-return **tai2**-Taiwan 。 *dang1shi4ren2*-the person involved *dui4*-for **ci3**-this *fen4men4*-angry
"**The Taiwan authorities refused the person involved returning to Taiwan on political grounds**. Because of **this**, the person involved was furious."

When the antecedent is out of context, only the anaphoric expression is marked, as in (5):

(5) 夫妻倆 昨天都對**此喜訊**興奮不已。
Fu1qi1liang3-husband and wife *zuo2tian1*-yesterday *dou1*-both *dui4*-for **ci3**-DET *xi3xun4*-good news *xing4fen4*-happy *bu4yi3*-very。
"Yesterday, the couple were very excited about **this good news**."

Since the existence of covert subjects and covert objects is a common phenomenon in Chinese syntactic structure, it is not surprising that a single verb alone is the cause event. Consider the cause event in (6), 我聽了-wo3ting1le-after I heard, the object is dropped, which is actually referring to the previous sentence, i.e. what mom said. There are also cases where the actual cause events are not expressed in the context as in (7) and (8). For consistency, instead of marking the previous sentence as the cause event, we mark the trigger event that is closest to the emotion keyword as the cause. Besides, the actual cause event is usually too long and complicated, which involves several events as we can see in (6).

(6) 媽媽為了鼓勵我考試得滿分，便告訴我說：如果你考一張一百分，媽媽就給你五元做獎勵。**我聽了**好高興。
Ma1ma3-mom *wei4le*-for *gu3li4*-encourage *wo3*-1.SG *kao3shi4*-exam *de2*-get *man3*-full *fen1*-mark，*bian4*-then *gao4su4*-tell *wo3*-1.SG *shuo1*-say: *ru2guo3*-if *ni3*-2.SG *kao3*-exam *yi1*-one *zhang1*-CL *yi1bai3*-one hundred *fen1*-mark ，*ma1ma3*-mom *jiu4*-then *gei3*-give *ni3*-2.SG *wu3*-five *yuan2*-dollar *zuo4*-as *jiang3li4*-reward 。*wo3*-1.SG *ting1*-hear *le*-ASP *hao3*-very *gao1xing4*-happy
"In order to encourage me to get full marks in the exam, Mom told me, 'if you get 100 points, Mom will give you five dollars as an award'. I was very happy when I **heard [this]**."

(7) … 救了我的性命。**人魚 公主聽** 了很傷心。
…jiu4-save *le*-ASP *wo3*-1.SG *de*-POSS *xing4ming4*-life 。 *ren2yu2*- mermaid *gong1zhu3*-princess *ting1*-hear *le*-ASP *hen3*-very *shang1xin1*-sad
"… saved my life. The mermaid was very sad after hearing [this]."

(8) 他越**想**越生氣。
Ta1-3.SG.M *yue4*-more *xiang3*-think *yue4*-more *sheng1qi4*-angry
"The more he thought, the more angry he became."

In general, in terms of event annotation, emotion causes are loosely defined. We mark the shortest meaningful cause events that are closest to the emotion keywords.

4. Corpus Data and Annotation Scheme

Based on the list of 91 primary emotion keywords identified in Chen et al. (2009), we extract 6,058 entries of sentences by keyword matching from the Sinica data in our emotion corpus. Each entry contains the focus sentence with the emotion keyword "<FocusSentence>" plus the sentence before "<PrefixSentence>" and after "<SuffixSentence>" it. The extracted entries include all primary emotion keywords occurred in the Sinica Corpus except for the emotion class *happiness*, as the keywords of *happiness* exceptionally outnumbers other emotion classes (i.e. 2,544). In order to balance the number of each emotion class, we set the upper limit as about 1,600 entries for each primary emotion.

Note that the presence of emotion keywords does not necessarily convey emotional information due to different possible reasons such as negative polarity and sense ambiguity. Hence, we remove entries that 1) are non-emotional; 2) contain highly-ambiguous emotion

keywords, such as 如意 -ru2yi4-to be happy, 害羞 -hai4xiu1-to be shy, 為難 -wei2nan2-to feel awkward, from the corpus. After the removal of eight ambiguous emotion keywords, the number of the remaining entries in the emotion cause corpus is 5,629. Among the remaining entries, we also remove the emotion keywords in which the entries do not express that particular emotion and yet are emotional. The total number of emotion keywords in the corpus is 5,964.

For each emotional entry, we manually annotate cause events of each keyword. Since more than one emotion can be present in an entry, the emotion keywords are tagged as <emotionword id=0>, <emotionword id=1>, and so on. Two examples are given in Figure 1.

---

573 Y 0/傷心/Sadness <PrefixSentence>媽媽 也 問 了 鄰居 ， 但是 沒有 人 見到 小白 。 </PrefixSentence> <FocusSentence>為 了[*01n] 這 件 事[*02n] ， 我 曾 <emotionword id=0>傷心</emotionword> 了 好久 ， 但 也 無濟於事 。 </FocusSentence> <SuffixSentence>每當 再 看 到 白色 的 棄狗 ， 不禁 會 想起 小白 。 </SuffixSentence>

573 Y 0/*shang1xin1* /Sadness <PrefixSentence> Mom also asked the neighbors, but no one ever saw Little White. </PrefixSentence> <FocusSentence> Because of [*01n] this [*02n] , I has been feeling very <emotionword id=0> sad </emotionword> for a long time, but this did not help. </FocusSentence> <SuffixSentence> Whenever [I] see a white stray dog, [I] cannot help thinking of Little White. </SuffixSentence>

3016 Y 0/害怕/Fear, 1/高興/Happiness <PrefixSentence> 崔 西 及 約翰 • </PrefixSentence> <FocusSentence>麥卡錫 在 黎巴嫩 被 釋放 後， 立即 被 送 往 德國 的 美軍 醫院 做 檢查 時， 就 顯示出， 他們 雖然 身體 情況 良好， 但 仍 顯示出 許多 失調 的 微兆， 最 明 顯 的 就是， 崔西 的 混合 感覺 既 <emotionword id=1>高 興</emotionword>[*11e] 獲救[*12e] ， 又 <emotionword id=0>害怕</emotionword> [*01e]這 是 真 的[*02e] 。 </FocusSentence> <SuffixSentence>因為 被 囚禁 了 四年， 這 四 年 的 日子 及 環境 與 外界 的 多 不同 啊 ！ </SuffixSentence>

3016 Y 0/*hai4pa4*/Fear, 1/*gao1xing4*/Happiness <PrefixSentence> Tracy andJohn </PrefixSentence> <FocusSentence> After McCarthy were released, [they] were sent to the U.S. military hospital in Germany for inspection, [the report] shows that, despite the fact that their physical condition is good, but [they] still shows signs of many disorders, the most obvious being Tracy's mixed feeling of being <emotionword id=1> pleased </emotionword>that [*11e] [she] was rescued [*12e] , and also <emotionword id=0>fear</emotionword> that [*01e] it is true[*02e]. </FocusSentence><SuffixSentence> because [they] were imprisoned for four years, the life and the environment in the past four years is very different from the outside world! </SuffixSentence>

Figure 6.1: Two Examples of Cause Event Annotation

For an emotion keyword tagged as <id=0>, [*01n] marks the beginning of its cause event while [*02n] marks

the end. The "0" shows which index of emotion keyword it refers to, "1" marks the beginning of the cause event, "2" marks the end, and "n" indicates that the cause is a nominal event. For an emotion keyword tagged as <id=1>, [*11e] marks the beginning of the cause event while [*12e] marks the end, in which "e" refers to a verbal event, i.e. either an event or a nominalization. An emotion keyword can sometimes be associated with more than one cause, in which case both causes are marked. The emotional sentences containing no explicitly expressed cause event remain as they are. Entries without causes explicitly expressed are mainly due to the following reasons: i) there is not enough contextual information, for instance the previous or the suffix sentence are interjections, e.g. 嗯哼! "aha"; ii) when the focus sentence is the beginning or the ending of a paragraph, no prefix sentence or suffix sentence can be extracted as the context. In this case, the cause may be beyond the context; iii) when the cause is obscure, which can be something very abstract or even with unknown reasons. This is especially true for heterogeneous emotion verbs.

The actual number of extracted entries of each emotion class to be analyzed, the emotional entries, and the entries with cause events are presented in Table 1.

Table 1: Summary of Cause Corpus Data

Emotions	No. of Entries		
	Extracted	Emotional	With Causes
Happiness	1,646	1,327 (81%)	1,132 (82%)
Sadness	902	616 (68%)	468 (75%)
Fear	898	689 (77%)	567 (81%)
Anger	1,177	847 (72%)	629 (73%)
Surprise	1,341	781 (58%)	664 (85%)
Total	5,964	4,305 (72%)	3,460 (80%)

Total number of extracted keywords is 5,964. We can see that 72% of the extracted entries express emotions, and 80% of the emotional entries have a cause. The corpus contains *happiness* (1,327) keywords the most and *sadness* (616) the least. For each emotion type, about 80% of the emotional sentences, on average, are considered containing a cause event, with *surprise* the highest (85%) and *anger* the lowest (73%). This indicates that an emotion mostly occurs with the cause event explicitly expressed in the text regardless of emotion classes, which actually implies the prominent role of cause events in expressing an emotion. Hence, an extensive investigation of cause events will have great implications to emotion analysis and processing.

5.   Corpus Analysis

Anaylzing the corpus data allows better prediction to the correlation between cause events and emotions as well as the linguistic cues faciliating cause event detection. We examine the position of case events relative to emotion verbs, as well as linguistic cues including causative verbs, perception verbs, epistemic markers, conjunctions, and prepositions, and a group of cue words called others. Each of these cues is shown to be collocated with cause events in certain ways.

5.1   Position of Cause Events

We calculate the distribution of cause event types of each emotion and the position of cause events relative to emotion keywords, as shown in Table 2.

Table 2: Cause Event Position of Emotion Verbs

Emotions	Cause Type (%)		Cause Position (%)	
	Event	Nominal	Left	Right
Happiness	93	7	74	26
Sadness	89	11	80	20
Fear	84	16	52	48
Anger	76	24	71	29
Surprise	86	14	59	41

Table 2 suggests that emotion cause events tend to be expressed by verbal events (~85% on average) than nominal events and that cause events tend to occur at the position to the left of the emotion keyword, with *fear* (52%) being no preference. This may be attributed to the fact that *fear* can be triggered by either factive or potential causes, which is rare for other primary emotions. For *fear*, factive causes tend to take the left position whereas potential causes tend to take the right position. With these observations, it is assumed that most of the undetected causes should come before the emotion keyword.

5.2   Linguistic Cues

We also examine the correlations between emotions and cause events in terms of other linguistic cues: causative verbs, perception verbs, epistemic markers, conjunctions, and prepositions, and others. We hypothesize that these cues mark cause events in some ways which help cause detection. First, lists of potential linguistic cues that are likely to be collocated with cause events are identified, as presented in Table 3.

Next, the frequencies of these linguistic cues collocating with cause events of a given emotion type are calculated. We set different window sizes for statistics: [ce-0, ce+0], [ce-1, ce+1], [ce-2, ce+1], [ce-2, ce+2], in which [ce-0, ce+0] stands for the clause where the cause event (ce) exists; [ce-1, ce+1] for the window size between the clause before the beginning of the cause event and the clause after the end of cause event, and so on. Within the window size, we calculate the position of the six groups of linguistic cue words relative to the cause event, i.e. to the left or right of the cause event.

Statistics indicates that the window size [ce-2, ce+1], i.e. between two clauses before the beginning of the cause event and one clause after the end of the cause event, shows a clear pattern marking the cause event. Note that the statistics do not indicate the accurate frequencies of the keywords functioning as a cause event cue since there are certainly occurrences of keywords in the context which do not serve as a cue. Even so, it provides a tendency indicating the reliability of the cue words as well as the position of the cue words relative to cause events. In the following sub-sections, each group of linguistic cues is discussed and evaluated based on the statistics.

Table 3: Lists of Potential Linguistic Cues

Group	Keywords
Causative Verbs	"to cause": 讓-*rang4*, 令-*ling4*, 使- *shi3*
Perception Verbs	'to think about': 想到-*xiang3dao4*, 想起-*xiang3qi3*, 一想-*yi1xiang3*, 想來-*xiang3 lai2* 'to talk about': 說到-*shuo1dao4*, 說起-*shuo1qi3*, 一說-*yi1shuo1*, 講到-*jiang3dao4*, 講起-*jiang3qi3*, 一講-*yi1jiang3*, 談到-*tan2dao4*, 談起-*tan2qi3*, 一談-*yi1tan2*, 提到-*ti2dao4*, 提起-*ti2qi3*, 一提-*yi1ti2*
Epistemic Markers	'to hear': 聽-*ting1*, 聽到-*ting1dao4*, 聽說-*ting1shuo1* 'to see': 看-*kan4*, 看到-*kan4dao4*, 看見-*kan4jian4*, 見到-*jian4dao4*, 見-*jian4*, 眼看-*yan3kan4*, 瞧見-*qiao2jian4* 'to know': 知道-*zhi1dao4*, 得知-*de2zhi1*, 得悉-*de2xi1*, 獲知-*huo4zhi1*, 獲悉-*huo4xi1*, 發現-*fa1xian4*, 發覺-*fa1jue2* 'to exist': 有-*you3*
Prepositions	'for' as in 'I will do this for you': 為-*wei4*, 為了-*wei4le* 'for' as in 'He is too old for the job': 對-*dui4*, 對於-*dui4yu2* 'as': 以-*yi3*
Conjunctions	'because': 因-*yin1*, 因為-*yin1wei4*, 由於-*you2yu2* 'so': 於是-*yu1shi4*, 所以-*suo3yi3*, 因而-*yin1er2* 'but': 可是-*ke3shi4*
Others	'is': 的是-*deshi4* 'say': 的說-*deshuo1* 'at': 於-*yu2* 'can': 能-*neng2*

5.2.1. Causative Verbs

Causative verbs are verbs that involve in expressions of causing or forcing a patient to perform certain actions or to be in certain state. In Chinese, the common ones are 使-*shi3*, 令-*ling4*, 讓-*rang4*, as correspond to the English equivalents "to make" or "to cause". Table 4 shows the position distribution of causative verbs relative to cause events. Results indicate that causative verbs show no preference in appearing to the left of the cause event (44%) or to the right (47%). Examples of both constructions are given in (9) and (10) respectively.

Table 4: Position of Causative Verbs Relative to Cause Events

Emotions	Causative Verbs (%)			Total (#)
	Left	Focus	Right	
Happiness	42.1	10.2	47.7	197
Sadness	39.5	5.0	55.5	119
Fear	48.1	7.4	44.4	81
Anger	42.6	14.7	42.6	129
Surprise	45.6	7.1	47.3	226
Total	44	9	47	752

(9) [*01n]兩者之間的因果關係[*02n]，**令人**<emotionword id=0>哀傷<-emotionword>。
Liang3zhe3-two parties *zhi1jian1*-between *de*-DE *yin1*-cause *guo3*-result *guan1xi1*-relation, *ling4*-cause *ren2*-people *ai1shang1*-sad
"The cause-result relationship between the two made people sad."

(10)**令**外界<emotionword id=0>詫異</emotionword>**的是**[*01e]在聯合記者會上，葉爾辛 體力與精神狀況都欠佳[*02e]。
Ling4-cause *wai4jie4*-outside world *cha4yi4*-surprise *de*-DE *shi4*-is *zai4*-at *lian2he2*-joint *ji4zhe3hui4*-press conference shang4-up, *ye4er3xin1*-Yeltsin *ti3li4*-physical *yu3*-and *jing1shen2*-mental *zhuang4kuang4*-condition *dou1*-both *qian4jia1*-bad
"What surprised the outside world was that Yeltsin was in very poor physical and mental condition at a joint press conference."

When looking into the development data, we find that most of the causative verbs, on the one hand, appear to the right of the cause events indicate the end of cause events as in (9). On the other hand, the ones appearing to the left of the cause events mark the beginning of cause events, in which case the emotion keywords are usually followed by words such as 的是-*deshi1*-is that, as seen in (10).

5.2.2. Perception Verbs

Perception verbs are verbs which indicate perception or sensation. The perception verbs marking the cause events are thinking and talking verbs. Statistics clearly show that the perception verbs tend to appear to the left of the cause events (59% vs. 20% (focus) and 21% (right)), as can be seen in Table 5. In other words, perception verbs mainly indicate the beginning of cause events, as shown in (11).

Table 5: Position of Perception Verbs Relative to Cause Events

Emotions	Perception Verbs (%)			Total (#)
	Left	Focus	Right	
Happiness	45.7	30.4	23.9	46
Sadness	68.8	12.4	18.8	32
Fear	87.5	6.25	6.25	16
Anger	50.0	22.2	27.8	18
Surprise	40.9	31.8	27.3	22
Total	59	20	21	134

(11)**談到**[*01n]「寶來國際金融機場」[*02n]，寶來證券集團董事長白文正難掩<emotionword id=0>興奮</emotionword>之情。
Tan2dao4-talk about *bao3lai2*-Polaris *guo2ji4*-international *jin1rong2*-finance *ji1chang3*-airport， *bao3lai2*-Polaris *zheng4quan4*-security *ji2tuan2*-group *dong3shi4zhang3*-chairman *bai2wen2zheng4*-Wayne Pai *nan2*-hard *yan3*-conceal *xing1fen4*-excited *zhi1*-POSS *qing2*-feeling
"Talking about 'Polaris International Financial Airport', the Chairman of Polaris Securities Group Wayne Pai could not conceal his excitement."

It is not surprising that the perception verbs, in particular, are highly collocated with cause events of *fear* (87.5%) since fear are often triggered by potential causes that require certain degree of perception of possible events that may happen. With the perception verbs, the cause events are usually recalled from the past experience. *Surprise*, in contrast, are usually triggered by the sudden change of the situation instead of recalling the past experience, thus it is less collocated with perception verbs (40.9%).

5.2.3. Epistemic Markers

Epistemic markers are verbs which mark the cognitive awareness of emotion cause events (Lee and Huang 2009). Four types of epistemic markers are identified, including seeing, hearing, knowing, and existing. It is noted in Table 6 that epistemic markers tend to appear before the cause event, marking the beginning of cause events in different constructions as presented in (12)-(14).

Table 6: Position of Epistemic Markers Relative to Cause Events

Emotions	Epistemic Markers (%)			Total (#)
	Left	Focus	Right	
Happiness	42.1	41.8	16.1	935
Sadness	54.8	25.0	20.2	347
Fear	47.8	30.5	21.7	387
Anger	50.9	30.0	19.1	446
Surprise	42.7	35.9	21.4	571
Total	48	32	20	2,686

(12) 蘭妮公主**聽到**[*01e]自己永遠不能恢復人形[*02e]，
<emotionword id=0>傷心</emotionword>極了。
Lan2ni1-Lanny *gong1zhu3*-princess *ting1dao4*-hear *zi4ji3*-oneself *yong3yuan3*-forever *bu4*-not *neng2*-can *hui1fu4*-return *ren2xing2*-human figure，*shang1xin1*-sad *ji2*-extreme *le*-ASP
"Princess Lanny was very sad to hear that she would never be returned to a human being."

(13) **發現**[*01e]國王已經餓死在那了[*02e]，大家都十分
<emotionword id=0>傷心</emotionword>。
Fa1xian4-discover *guo2wang2*-king *yi3jing1*-already *e4*-hungry *si3*-die *zai4*-at *na4*-there *le*-ASP, *da4jia1*-everybody *dou1*-all *shi2fen1*-very *shang1xin1*-sad
"Everyone was very sad to find that the King was starved to death."

(14) 我實在太<emotionword id=0>高興</emotionword>**看到**
[*01e]民間主動成立這個委員會[*02e]。
Wo3-1.SG *shi2zai4*-indeed *tai4*-very *gao1xing4*-happy *kan4dao4*-see *min2jian1*-non-governmental circles *zhu3dong4*-initiate *cheng2li4*-form *zhe4*-DET *ge4*-CL *wei3yuan2hui4*-committee
"I am very happy to see that non-governmental circles took the initiative to set up this committee."

The insignificant difference of epistemic markers appearing at the left position and the focus position, i.e. within the cause event, is attributed to the fact that some verbs of seeing, hearing, knowing, and existing are actually part of the cause event, instead of serving as the indicator marking the cause events. Examples are given in (15) and (16).

(15) 我[*01e]看到這件作品[*02e]很<emotionword id=0>震驚</emotionword>。
Wo3-1.SG *kan4dao4*-see *zhe4*-DET *jian4*-CL *zuo4pin3*-piece *hen3*-very *zhen4jing1*-surprise
"When I saw this piece, [I] was very surprised."

(16) 農夫[*01e]聽到這句話[*02e]，心裡非常<emotionword id=0>高興</emotionword>。
Nong2fu1-farmer *ting1dao4*-hear *zhe4*-DET *ju4*-CL *hua4*-word, *xin1li3*-heart *fei1chang2*-very *gao1xing4*-happy
"When the farmer heard this, [s/he] was very happy."

The cause events in (15) and (16) refer to the actual seeing and hearing actions involved rather than functioning as cause event markers. Therefore, they are marked as cause events, but not epistemic markers.

In addition, the comparatively low percentage of the epistemic markers occurring before the cause event is partially due to the considerable noises resulted from the inclusion of the verb of existing in extracting causes such as (12) and (13). Therefore, when extracting these cause events, the verb 有-*you3* should be excluded.

5.2.4. Prepositions

As can be seen in Table 7, most prepositions apparently appear to the left of cause events (58% on average vs. 19% and 23%), and mark the beginning of cause events across all primary emotions. Examples are given in (17) and (18).

Table 7: Position of Prepositions Relative to Cause Events

Emotions	Prepositions (%)			Total (#)
	Left	Focus	Right	
Happiness	58.4	20.3	21.3	567
Sadness	57.4	17.3	25.3	284
Fear	59.3	17.2	23.5	378
Anger	62.3	15.9	21.8	446
Surprise	51.9	23.5	24.6	378
Total	58	19	23	2,053

(17) 他**對**[*01e]謠言散布之快[*02e]感到<emotionword id=0>驚訝</emotionword>。
Ta1-3.SG.M *dui4*-for *yao2yan2*-rumor *san4bu4*-spread *zhi1*-POSS *kuai4*-fast *gan3dao4*-feel *jing1ya4*-surprise
"He was surprised by the fast spreading of rumors."

(18) **對於**[*01e]能有機會經營如此出色的俱樂部[*02e]，我感到非常<emotionword id=0>快樂</emotionword>。
Dui4yu2-for *neng2*-can *you3*-have *ji1hui4*-opportunity *jing1ying2*-run *ru2ci3*-such *chu1se4*-great *de*-DE *ju4le4bu4*-club, *wo3*-1.SG *gan3dao4*-feel *fei1chang2*-very *kuai4le4*-happy
"I am very happy for having the oppportunity to run such a great club."

It is, however, noted that the preposition 以-*yi3-as* frequently occurs in the corpus, and is not a reliable cue to mark cause events, therefore, it is removed from the list of linguistic cues.

5.2.5. Conjunctions

Conjunctions are words that link two phrases or clauses. For cause event detection, three types of conjunctions are identified: *because*, *so*, and *but*. The overall statistics show that most conjunctions (68% on average) appear before the cause events.

Table 8: Position of Conjunctions Relative to Cause Events

Emotions	Conjunctions (%)			Total (#)
	Left	Focus	Right	
Happiness	71.9	3.9	24.2	128
Sadness	69.1	7.4	23.5	68
Fear	66.0	3.4	30.6	147
Anger	69.5	2.1	28.4	95
Surprise	59.2	2.8	38.0	71
Total	68	4	28	509

(19) 但 很 快 地 他 們 又 <emotionword id=0> 高 興 </emotionword> 起來 了，**因為** [*01e]他們是唯一了解日本過去歷史的人 [*02e] 。
Dan4-however *hen3*-very *kuai4*-fast *de*-DE *ta1men2*-3.PL *you4*-again *gao1xing4*-happy *qi3lai2*-begin *le*-ASP, *yin1wei4*-because *ta1men2*-3.PL *shi4*-is *wei2yi1*-only one *liao3jie3*-understand *ri4ben3*-japan *guo4qu4*-past *li4shi3*-history *de*-POSS *ren2*-people
"However, they became happy soon, because they were the only people who understand the history of Japan."

(20) **由於**[*01e]愛滋病出現[*02e]引起大家的<emotionword id=0>恐慌</emotionword>。
You2yu2-because *ai4zi1bing4*-AIDS *chu1xian4*-appear *yin3qi3*-cause *da4jia1*-everybody *de*-POSS *kong3huang1*-fear
"Because of the existence of AIDS, people are frightened."

Examples (19) and (20) shows that the *because*-conjunctions effectively mark the beginning of cause events. However, we find that the *so*- and *but*-conjunctions do not in the corpus data. They are thus removed from the list.

5.2.6.   Others
There are some useful cues that do not fall into any of the five groups of linguistic cues as given above. They are grouped under "Others", including 的是-*deshi4*-is, 的說 -*deshuo1*-say, 於-*yu2*-at, 能-*neng2*-can. These cues are shown to be effectively marking the beginning of cause events as they tend to occur at the left position of cause events (58% on average). Examples are given in (21) and (22).

Table 9: Position of Other Cues Relative to Cause Events

Emotions	Others (%)			Total (#)
	Left	Focus	Right	
Happiness	50.2	27.1	22.7	321
Sadness	54.8	27.9	17.3	104
Fear	60.1	11.3	28.6	168
Anger	55.2	15.7	29.1	134
Surprise	67.6	14.0	18.4	222
Total	58	19	23	949

(21) 我<emotionword id=0>驚訝</emotionword>**的是**, [*01e]居然有候選人以這種方式來合理化對女性參政的歧視[*02e]。
Wo3-1.SG *jing1ya4*-surprise *deshi4*-is that, *ju1ran4*-unexpectedly *you3*-exist *hou4xuan3ren2*-candidate *yi3*-as *zhe4*-DET *zhong3*-CL *fang1shi4*-way *lai2*-to *he2li3hua4*-rationalize *dui4*-for *n ü 3xing4*-female *can1zheng4*-participate in politics *de*-POSS *qi2shi4*-discrimination
"I was surprised that there were candidates who rationalized the discrimination against women in politics in this way."

(22) 我實在太<emotionword id=0>高興</emotionword>**能**[*01e]贏得這個獎[*02e], 讓我備感光榮！
Wo3-1.SG *shi2zai4*-indeed *tai4*-too *gao1xing4*-happy *neng2*-can *ying2de2*-win *zhe4*-DET *ge4*-CL *jiang3*-award, *rang4*-cause *wo3*-1.SG *bei4gan3*-feel *guang1rong2*-honour
"I am very happy that I could win this award which I feel honoured!"

We notice that the word 的說-*deshuo1*-say may occur before or after the emotion keywords as in (23) and (24) respectively, while other members can only occur after the emotion keywords.

(23) 我<emotionword id=0>心酸</emotionword>**道**：「可是[*01e]他們好可憐[*02e]喔！」
Wo3-1.SG *Xin1suan1*-sad *dao4*-say: *ke3shi4*-but *ta1men2*-3.PL *hao3*-very *ke3lian2*-pity *wo1*-PART
"I said sadly, 'But they are pitiful!'"

(24) 那個男人**說**：「[*01e]像我這樣好條件，更好的女孩多的是[*01e]」。這次喬依絲哭的很<emotionword id=0>傷心</emotionword>。
Na4-DET *ge4*-CL *nan2ren2*-man *shuo1*-say: *xiang4*-like *wo3*-1.SG *zhe4yang4*-such *hao3*-good *tiao2jian4*-quality, *geng4hao3*-better *de*-POSS *n ü 3hai2*-girl *duo1*-many *de*-DE *shi4*-is 。 *Zhe4*-DET *ci4*-CL *qiao2yi1si1*-Joyce *ku1*-cry *de*-DE *hen3*-very *shang1xin1*-sad
"The man said, 'There are many desirable girls for a man with good qualities like me." Joyce cried sadly.

Due to the different marking behaviour, 的 說 -*deshuo1*-say should be separated from the list of others and form another group of linguistic cues called say verbs with its variants 說-*shuo1* and 道-*dao4*.

5.3 Overview of Cause Event Markers
After the investigation of corpus data, seven groups of linguistic cues that are highly collocated with cause events are finalized, as shown in Table 10.

Table 10: Seven Groups of Cause Event Markers

Group	Cue Words
Causative Verbs	'to cause':讓-*rang4*,令-*ling4*, 使- *shi3*
Perception Verbs	'to think about': 想到-*xiang3dao4*, 想起 -*xiang3qi3*, 一想-*yi1xiang3*, 想來-*xiang3 lai2* 'to talk about': 說到-*shuo1dao4*, 說起 -*shuo1qi3*, 一說-*yi1shuo1*, 講到-*jiang3dao4*, 講起 -*jiang3qi3*, 一 講 -*yi1jiang3*, 談 到 -*tan2dao4*, 談起-*tan2qi3*, 一談-*yi1tan2*, 提到 -*ti2dao4*, 提起-*ti2qi3*, 一提-*yi1ti2*
Say words	'to say': 說-*shuo1*, 的說-*deshuo1*,道-*dao4*
Epistemic Markers	'to see': 看 -*kan4*, 看 到 -*kan4dao4*, 看 見 -*kan4jian4*, 見到-*jian4dao4*, 見-*jian4*, 眼看 -*yan3kan4*, 瞧見-*qiao2jian4* 'to hear': 聽 -*ting1*, 聽 到 -*ting1dao4*, 聽說 -*ting1shuo1* 'to know': 知道-*zhi1dao4*, 得知-*de2zhi1*, 得悉-*de2xi1*, 獲知-*huo4zhi1*, 獲悉-*huo4xi1*, 發現-*fa1xian4*, 發覺-*fa1jue2* 'to exist': 有-*you3*
Prepositions	'for' as in 'I will do this for you': 為-*wei4*, 為了-*wei4le* 'for' as in 'He is too old for the job': 對-*dui4*, 對於-*dui4yu2*
Conjunctions	'because': 因 -*yin1*, 因 為 -*yin1wei4*, 由 於 -*you2yu2*
Others	'is': 的是-*deshi4* 'at': 於-*yu2* 'can': 能-*neng2*

The seven groups include causative verbs, perception verbs, say verbs, epistemic markers, conjunctions, prepositions, and others. Each group of linguistic cues serves as an indicator marking the cause events in different structures of emotional constructions, in which causative verbs specifically mark the end of cause events while the other six groups mark the beginning of cause events.

6. Conclusion and Future Work

In this paper, we build a corpus of emotion cause event which can be used for designing automatic systems of emotion cause detection and emotion classification. Through analyzing the corpus data, we confirm that cause event is an important component of emotion.

We also examine the position of cause events as well as various linguistic cues for emotion detection and classification. Specifically, seven groups of linguistic cues are presented and analyzed in detail, showing that these linguistic cues, to a great extent, provide a transparent environment for emotion causes.

In our future work, we intend to develop a rule-based system for cause event detection based on our findings, as well as automatic annotation of cause events for emotion processing in text.

References

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

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

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray.

Descartes, R. (1649). The Passions of the Soul. In J. Cottingham et al. (Eds), *The Philosophical Writings of Descartes*. Vol. 1, 325-404.

Ekman, P. (1984). Expression and the Nature of Emotion. In Scherer, K. and P. Ekman (Eds.), *Approaches to Emotion*. Hillsdale, N.J.: Lawrence Erlbaum. 319-343.

James, W. (1884). What is an Emotion? *Mind*, 9(34):188–205.

Kozareva, Z., B. Navarro, S. Vazquez, and A. Nibtoyo. (2007). UA-ZBSA: A Headline Emotion Classification through Web Information. In *Proceedings of the 4th International Workshop on Semantic Evaluations*.

Kövecses, Z. (2000). *Metaphor and Emotion: Language, Culture and Body in Human Feeling*. Cambridge: Cambridge University Press.

Lee, S. Y. M. and Chu-Ren Huang. (2009). Explicit Epistemic Markup of Causes in Emotion Constructions. The Fifth International Conference on Contemporary Chinese Grammar (ICCCG5). Hong Kong. November 27- December 1.

Masum, S. M., H. Prendinger, and M. Ishizuka. (2007). Emotion Sensitive News Agent: An Approach Towards User Centric Emotion Sensing from the News. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*.

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

Ortony A., G. L. Clone, and A. Collins. (1988). *The Cognitive Structure of Emotions*. New York: Cambridge University Press.

Picard, R.W. (1995/2000). Affective Computing. Cambridge. MA: The MIT Press.

Plutchik, R. (1962). *The Emotions: Fact, Theories and a New Model*. New York: Random House.

Plutchik, R. (1980). *Emotions: A Psychoevolutionary Synthesis*. New York: Harper & Row.

Saurí, R., J. Littman, R. Knippen, R. Gaizauskas, A. Setzer, and J. Pustejovsky. (2004). *TimeML Annotation Guidelines*. http://www.timeml.org.

Spinoza, B. (1985). *Ethics*. In E. Curley, *The Collected Works of Spinoza*. Princeton, N.J.: Princeton University Press. Vol 1.

Talmy, L. (2000). *Toward a Cognitive Semantics*. Vol. 1and 2. Cambridge: MIT Press.

Tokuhisa, R., K. Inui, and Y. Matsumoto. (2008). Emotion Classification Using Massive Examples Extracted from the Web. In *Proceedings of COLING*.

Turner, J. H. (1996). The Evolution of Emotions in Humans: A Darwinian-Durkheimian Analysis. *Journal for the Theory of Social Behaviour*, 26:1-34.

Turner, J. H. (2000). *On the Origins of Human Emotions: A Sociological Inquiry into the Evolution of Human Affect*. California: Stanford University Press.

Wierzbicka, A. (1999). *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge: Cambridge University Press.