Abstract. This paper presents a linguistic analysis of emotions by introducing some concrete linguistic rules in identifying the important elements of an emotion, which are the experiencer and the cause. It specifically discusses two primary emotions, happiness and surprise, in Chinese. These rules can be very useful for emotion detection and classification. We also examine the features of a cause event according to its degree of transitivity, namely agentivity, kinesis, and event participation. We find that the emotion classes (happiness vs. surprise) override the emotion verb types (change-of-state verb vs. homogeneous verb) in terms of cause event features. We believe that this study will have some implications for the linguistic account of causal events as well as the theory of emotions in general. We also hope that this work will shed lights in the automatic detection and classification of emotion in language technology.

Keywords: emotion, cause event, experiencer, happiness, surprise.

1 Introduction

Emotions have been widely studied in and across various disciplines. Various linguistic and psychological theories of emotions were proposed previously (e.g. Plutchik 1980, Frijda 1986, Weiss and Cropanozano 1996). They are generally considered the least tangible aspects of human experience, but exert powerful influence upon our thoughts and actions, and even physical appearance and physiological processes occurring within our bodies (Harkins and Wierzbicka 2001). With its complex nature, there have been considerable debates among the theories on the definition as well as the classification of emotions. Nonetheless, most theories agree that an emotion is composed of or stimulated by a cause event and that the cause events are the most tangible component of emotion (Wierzbicka 1996). As no objective criteria have been offered yet to define and classify cause events empirically, this study aims to examine the interaction between emotions and cause events.

As for real world applications, the information of who the experiencer of an emotion is and what triggers such an emotion is an essential element for NLP. Such information is particularly important for economic and political evaluations as well as product reviews. However, identifying the experiencer and the causes of an emotion is very challenging in NLP. We believe that a linguistic analysis of cause events may help generalize some useful patterns for automatic emotion detection and classification.

In this paper, we present a linguistic analysis of emotions in terms of their experiencers and causes. Causes refer to the explicitly expressed arguments or events that trigger the presence of the corresponding emotions, which are usually expressed by means of verbs, nominalizations,
and can sometimes be event nominals (Sauri et al. 2006). We specifically look at two primary emotions, happiness and surprise, in Chinese. Section 2 discusses the previous theories concerning emotion studies and Section 3 describes the data used in the study. Section 4 examines how the experiencers and the causes of happiness and surprise can be identified linguistically by introducing some concrete linguistic rules. We also discuss the correlations between the causes and the corresponding emotions based on the degree of transitivity in Section 5. Section 6 concludes the paper.

2 Emotion Theories

Researchers have attempted to study the emotion classification with different approaches in different fields, varying from linguistics, neuropsychology, to computer science, as well as varying from language to language. Yet, there is no consensus on the classification of emotions. In this section, we discuss the emotion classification that there was the most agreement on, and the emotion verb classes in Chinese.

2.1 Happiness and Surprise as Primary Emotions

It is assumed that a small number of emotions are considered primary emotions, and that other emotions are secondary emotions which are the mixtures of the primary emotions. Many different lists of basic emotions have been proposed, there is, nonetheless, no agreement neither on the number of basic emotions nor the classes of emotions. For example, Kemper (1987) suggests that there are at least four basic emotions, i.e. fear, anger, sadness, and satisfaction. The rationales for considering them as basic are that (1) they can be observed in most animals; (2) that they are universally found in all cultures; (3) they appear early in the course of human development; (4) they are outcomes of power and status interactions; (5) they are associated with distinct automatic patterns of physiological changes. Similar to Kemper, Ortony and Turner (1990) also give reasons for assuming the existence of basic emotions: (1) some emotions appear to exist in all cultures; (2) some can be identified in higher animals; (3) some have characteristic facial expressions; and (4) some seem to increase the chances of survival.

Various researchers have attempted to propose the list of basic emotions, which varies from two to ten basic emotions. 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.

In our study, we adopt Turner’s emotion classification (Turner 2000), which identifies five primary emotions, namely happiness, sadness, fear, anger, and surprise, and various complex emotions. In this classification, surprise is considered optional. The inclusion of surprise in our study are mainly attributed to the fact that, in Chinese, jing “surprise” appears to be a productive emotion word which has a strong ability to compose complex emotion words such as jingxi “surprise+happiness” and jingxia “surprise+fear” (Chen et al. 2009).

2.2 Emotion Verb Classes in Chinese

In the literature, there has been some work done on the classification of Chinese emotions. Based on a corpus study using Academia Sinica Balanced Corpus (Sinica Corpus), which contains five million Chinese words, Chang et al. (1999) proposed seven sets of most frequent emotion verbs namely happy, depressed, sad, regret, angry, afraid, and worried. They identified a total of 33 verbs in the seven types of emotion verbs, each with a frequency of over 40 in the Sinica Corpus. They then proposed a dichotomy of emotion verbs in the seven subtypes of Chinese emotion verbs, as shown in Table 1:
Table 1: The Dichotomy of the Emotion Verbs in Chinese

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Change-of-State</th>
<th>Homogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>Gaoxing 高興(699)</td>
<td>Kuaile 快樂(942)</td>
</tr>
<tr>
<td></td>
<td>Kaixin 開心(152)</td>
<td>Yukuai 愉快(271)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Xiyou 喜悅(156)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huanle 欢樂(141)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huanxi 欢喜(107)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kuaihuo 快活(48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tongkuai 痛快(40)</td>
</tr>
<tr>
<td>Depressed</td>
<td>Nanguo 難過(232)</td>
<td>Tongku 痛苦(443)</td>
</tr>
<tr>
<td></td>
<td>Tongxin 痛心(48)</td>
<td>Chenzhong 沈重(83)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jusang 沮喪(62)</td>
</tr>
<tr>
<td>Sad</td>
<td>Shangxin 傷心(134)</td>
<td>Beishang 悲傷(52)</td>
</tr>
<tr>
<td>Regret</td>
<td>Houhui 後悔(102)</td>
<td>Yihan 遺憾(198)</td>
</tr>
<tr>
<td>Angry</td>
<td>Shengqi 生氣(307)</td>
<td>Fennu 憤怒(112)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qifen 氣憤(49)</td>
</tr>
<tr>
<td>Afraid</td>
<td>Haipa 害怕(261)</td>
<td>Kongji 恐懼(149)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weiju 懦弱(40)</td>
</tr>
<tr>
<td>Worried</td>
<td>Danxin 擔心(609)</td>
<td>Fannao 擔懼(199)</td>
</tr>
<tr>
<td></td>
<td>Danyou 擔憂(64)</td>
<td>Kunao 苦惱(45)</td>
</tr>
</tbody>
</table>

For each subtype, there are two types of emotion verbs: change-of-state emotion verbs and homogeneous state emotion verbs. Chang et al. found that change-of-state and homogeneous state verbs behave differently in terms of five criteria: 1) their grammatical functions; 2) their cooccurrence restrictions; 3) the appropriateness in the imperative and evaluative construction; 4) the verbal aspect; and 5) the transitivity.

Since surprise was not examined in Chang et al.’s analysis, we test the emotion verbs of surprise in Chinese according to the five criteria based on the Sinica Corpus (which will be discussed in Section 3). We find that the two most frequent verbs of surprise are jingya and zhenjing, and that jingya behaves more like a change-of-state emotion verb whereas zhenjing is more of a homogeneous emotion verb.

With this account of emotion verbs, we examine the most frequent verbs of happiness and surprise, one from each group, i.e. jingya and zhenjing for surprise, and gaoxing and kuaile for happiness. Based on this dichotomy, we explore whether the causes of the two groups of emotion verbs are expressed differently, i.e. jingya and gaoxing vs. zhenjing and kuaile.

3 Data

We give a corpus-based study on emotion analysis. It provides a systematic analysis of authentic examples of Chinese from a variety of genres. Instead of illustrating with introspective examples, corpus data serve as evidence from naturally occurring data. By observing the text in the corpus with the help of data collection tools, we can make relevant generalizations.

The data used in this paper are taken from the Sinica Corpus, which is a tagged Mandarin corpus containing a total of ten million words. The texts in the corpus are collected from different sources and in various topics such as Philosophy, Science, Arts, etc. We aim at analyzing 100 emotional sentences with cause events for each group of emotion verbs of happiness and surprise.

The Chinese Word Sketch generates linguistically meaningful collocations automatically from the Sinica corpus, and provides rich lexicon-based grammatical information with stochastic information. Using Chinese Word Sketch, we randomly extract sentences of each of emotion verbs from the Sinica Corpus, together with their contexts in which the causes of the
emotions appear. Please note that sentences containing emotion verbs do not necessarily emotional. Hence, we filter out non-emotional sentences as well as emotional sentences with no explicit cause. Examples of emotional sentences containing corresponding cause events are given in Table 2:

Table 2: Example Sentences with the Emotions and the Corresponding Causes

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Change-of-State</th>
<th>Homogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>Gaoxing高興</td>
<td>Kuaile快樂</td>
</tr>
<tr>
<td></td>
<td>今天很高興說了這麼多中國話。</td>
<td>幾天後，就可看到我的傑作了，真是快樂呀！</td>
</tr>
<tr>
<td></td>
<td>“I am very happy that I was able to speak much Chinese today.”</td>
<td>“I am happy that I can see my masterpiece few days later!”</td>
</tr>
<tr>
<td>Surprise</td>
<td>Jingya驚訝</td>
<td>Zhenjing震驚</td>
</tr>
<tr>
<td></td>
<td>他很驚訝台灣竟然有這麼大型又美麗的鳥。</td>
<td>沒想到她所說的都是真話，讓他震驚不已。</td>
</tr>
<tr>
<td></td>
<td>“He was surprised that there is such a large and beautiful bird in Taiwan.”</td>
<td>“What she said was, unexpectedly, the truth, he was very surprised.”</td>
</tr>
</tbody>
</table>

4 Identification of Emotion Experiencers and Causes

By examining the 100 emotional sentences of each emotion verb extracted from Sinica Corpus, we generalize some rules for identifying the experiencer and the cause of the corresponding emotion verb. We consider the cause as a proposition. It is generally assumed that a proposition has a verb which optionally takes a noun occurring before it as the subject and a noun after it as the object. However, as mentioned earlier, a cause can also be expressed as a nominalization or nominal. We also identify the position of the proposition as well as the experiencer, which can occur in the clause containing the emotion verb (focus clause), the clause before the focus clause, or the clause after the focus clause. The abbreviations used in the rules are given as follows:

P = Proposition/cause event
E = Experiencer
K = Keyword/emotion verb
B = Clause before the focus clause
F = Focus clause/the clause containing the emotion verb
A = Clause after the focus clause

In the following sub-sections, we summarize the generalized rules for each of the emotion verbs in question, i.e. zhenjing, jingya, kuaile, and gaoxing. The rules are presented in order of specificity, and each is illustrated with an example. For simplicity, an example of the rule description is given in Rule 1 of zhenjing.

4.1 Zhenjing

Rule 1:

i) P (B/F) + 讓/令/該‘to cause’(F) + E(F) + K(F)

ii) E = the nearest Na/Nb/Nc/Nh after 讓/令/該 in F

iii) P = the nearest (N)+V+(N) before 讓/令/該 in F/B

Rule 1 indicates that the experiencer appears to be the nearest Na (common noun)/Nb (proper noun)/Nc (place noun)/Nh (pronoun) after the preposition 讓,令 or 該 in the focus clause, while
at the same time comes before *zhènjīng*. Besides, the cause (P) comes before 讓, 令 or 使. We simplify P as a structure of (N)+V+(N), and P is very likely to contain the cause event. In identifying P, we first look for the nearest verb occurring before 讓, 令 or 使 in the focus sentence or the clause before the focus clause, and consider this verb as an anchor. From this verb, we search to the left for the nearest noun, and consider it as the subject; we then search to the right for the nearest noun until the presence of 讓, 令 or 使, and consider it as the object. Please note that the bracketed nouns in the rules are optional. The detected subject, verb, and object form the cause event. An example is given in (1):

(1) [P]伊拉克細菌武器秘密的曝光， 使 [E]聯合國大為[K]震驚。

“The reveal of Iraq's secret bacteriological weapons shocked the United Nations.”

Rule 2:

i) E(F) + 對 ‘preposition’(F) + P(F) + K(F)
ii) E = the nearest Na/Nb/Nc/Nh before 對 in F
iii) P = the nearest N+N+V/V+N/V+的+N/N+之+N before K in F

(2) [E]國際間對[P]北韓糧荒之嚴重深感[K]震驚。

“The international community was deeply shocked by serious food shortages in North Korea.”

Rule 3:

i) E(F) + K(F) + 於 ‘preposition’(F) + P(F)
ii) E = the nearest Na/Nb/Nc/Nh after K in F
iii) P = the nearest N+V+N after 於 in F


“While I was still shocked by my own firm answer, his laughter made me speechless.”

Rule 4:

i) K(F) + E(F) + 的 ‘possession’(F) + P(F)
ii) E = the nearest Na/Nb/Nc/Nh after K in F
iii) P = the nearest N+V+N+的+N after 的 in F

(4) 一九七九年，台灣爆發了[K]震驚[E]海內外的[P]「美麗島事件」。

“In 1979, ‘Formosa Incident’ broke out in Taiwan, which shocked within and abroad the country.”

Rule 5:

i) P(B) + K(F) + E(F)
ii) E = the nearest Na/Nb/Nc/Nh after K in F
iii) P = the nearest (N)+V+(N) before K in B

(5) [P]這個怪異的情景立刻[K]震驚了[E]全市民...

“The bizarre scene shocked the whole community at once…”

Rule 6:

i) P(B) + E(F) + K(F)
ii) E = the nearest Na/Nb/Nc/Nh before K in F
iii) P = the nearest N+V+N before K in B

(6) [P]金帝完顏亮南侵，[E]宋朝廷[K]震驚，一時群情激憤。

“Emperor Jin, Wanyan Liang, invaded the south, Song court was shocked and enraged.”
4.2 Jingya

Rule 1:
i) 令/使/讓/教 ‘to cause’(F) + E(F) + K(F) + P(F)
ii) E = the nearest Na/Nb/Nc/Nh after 令 in F
iii) P = the nearest N+V+N after K in F/A
   “But what is surprising is that the old man is actually illiterate.”

Rule 2:
i)  E(F) + K(F) + 的是 ‘is that’ + P(F)
ii) E = the nearest Na/Nb/Nc/Nh before K in F
iii) P = the nearest N+V+N after 是 in F/A
   “I was surprised that there were candidates who rationalized the discrimination against women in politics in this way.”

Rule 3:
i)  E(F) + K(F) + 發現 ‘to find’(F) + P(F)
ii) E = a: the nearest Na/Nb/Nc/Nh before K in F; b: the first Na/Nb/Nc/Nh in B
iii) P = the nearest N+V+N after 發現 in F
   “He was surprised to find that all the people were talking about Asia.”

Rule 4:
i)  E(F) + 對 ‘preposition’(F) + P(F) + K(F)
ii) E = the nearest Na/Nb/Nc/Nh before 對 in F
iii) P = the nearest N+V+N after 對 in F
   “He was surprised by the fast spreading of rumors.”

Rule 5:
i)  P(B) + E(F) + K(F)
ii) E = the nearest Na/Nb/Nc/Nh before K in F
iii) P = the nearest N+V+N before K in B
   “Perhaps I sounded rude so that s/he was surprised and puzzled.”

4.3 Kuaile

Rule 1:
i)  E(B) + P(B) + K(F)
ii) E = the first Na/Nb/Nc/Nh in B
iii) P = the nearest (N)+V+N before K in B
   “The moment that I heard the music, I felt very happy.”

Rule 2:
i)  P(B) + E(F) + K(F)
ii) E = the nearest Na/Nb/Nc/Nh before K in F
iii) P = the nearest (N)+V+N before K in B
When I saw my father's smile, I was very happy.

The eagle was very happy, because his children could fly.

I am very happy that I could win this award which I feel honoured.

We are very happy that the first issue was finally released.

When the farmer heard this, he was very happy.

They gave me a piece of dried meat as the prize. My father was very happy.

5 Cause Event Features

On top of identifying the position and the syntactic structure of the cause of an emotion, it is significant to examine how the cause affects the experiencer of the emotion. In order to do so, we give an in-depth analysis of causes based on its degree of transitivity (Lee and Huang 2009).

5.1 Transitivity of Causes

Transitivity is traditionally understood to be a property of verbs that determines whether or not a verb can take direct objects. This is often classified as transitive verbs such as kick and beat that take a direct object, and intransitive verbs such as cry and sleep that cannot take a direct
object. However, transitivity is sometimes considered to be a continuum rather than a binary category. This approach takes into account the degree to which an action affects its object. For example, the verb *hear* is described as having "lower transitivity" than the verb *hit*.

Instead of simply focusing on the presence of an object of the verb, Hopper and Thompson (1980) argue that transitivity involves a number of components. They identify the ten parameters that determine the transitivity of a sentence, as shown in Table 3:

<table>
<thead>
<tr>
<th>Table 3: Transitivity Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
</tr>
<tr>
<td>A. Participants</td>
</tr>
<tr>
<td>B. Kinesis</td>
</tr>
<tr>
<td>C. Aspect</td>
</tr>
<tr>
<td>D. Punctuality</td>
</tr>
<tr>
<td>E. Volitionality</td>
</tr>
<tr>
<td>F. Affirmation</td>
</tr>
<tr>
<td>G. Mode</td>
</tr>
<tr>
<td>H. Agency</td>
</tr>
<tr>
<td>I. Affectedness of Object</td>
</tr>
<tr>
<td>J. Individuation of Object</td>
</tr>
</tbody>
</table>

The ten parameters determine whether an action is transferred from an agent to a patient. To give an in-depth analysis of the causes based on its degree of transitivity, we specifically focus on the three components of causes, namely *agentivity, kinesis,* and *participation,* with some modifications.

The agentivity of the subject of the cause event shows the intention of the subject to affect the experiencer. We identify the subject of the cause event as one of the two basic prototypical role types, namely Agent-Proto role or Patient-Proto role, according to a list of Proto-role properties proposed by Dowty (1991). Dowty (1991) argues that role types are not discrete categories; instead, arguments may have different degree of membership in a role type. Hence, he believes that it is efficient to have only two role types to describe argument selection, which are Proto-Agent and Proto-patient. The Proto-Agent and Proto-patient argument roles have different list of contributing properties, as given in (19) and (20).

(19) Contributing Properties for the Agent Proto-Role:
   a. volitional involvement in the event or state
   b. sent[ience (and/or perception)
   c. causing an event or change of state in another participant
   d. movement (relative to the position of another participant)
   e. (exists independently of the event named by the verb)

(20) Contributing Properties for the Patient Proto-Role:
   a. undergoes change of state
   b. incremental theme
   c. causally affected by another participant
   d. stationary relative to movement of another participant
   e. (does not exist independently of the event, or not at all)

The subject of the cause is considered more agent-like or patient-like according to the number of Agent or Patient Proto-role properties they fulfill. The more agent-like the subject is, the higher intention the subject is to affect the experiencer.
Kinesis refers to whether the cause event is an action or non-action. Actions can be transferred from one participant to another, therefore, an action cause event is more likely to directly affect the experiencer. Non-action events, on the other hand, tend to affect the experiencer more indirectly.

Participation indicates whether the experiencer is involved in the cause event. We examine whether the experiencer of the emotion is one of the participants of the cause event.

5.2 Results and Analysis

A summary of the cause feature tendency of each emotion is presented in Tables 4 and 5:

| Table 4: Distributional Tendency of Cause Features of Change-of-State Emotion Verbs |
|-----------------------------------------------|---------------|---------------|---------------|---------------|
| Emotions | Agentivity | Kinesis | Participation |
|          | Proto-agent | Proto-patient | Motion | Non-motion | Yes | No |
| Gaoxing  | * (high) | * | * |
| Jingya   | * (low) | * | * |

| Table 5: Distributional Tendency of Cause Features of Homogeneous Emotion Verbs |
|-----------------------------------------------|---------------|---------------|---------------|---------------|
| Emotions | Agentivity | Kinesis | Participation |
|          | Proto-agent | Proto-patient | Motion | Non-motion | Yes | No |
| Kuaile   | * (high) | * | * |
| Zhenjing | * (low) | * | * |

As shown in Tables 4 and 5, we find that there are some correlations between the types of emotions and the cause features. Both gaoxing and kuaile tend to have proto-agent subjects with high agentivity in the cause event. They also tend to involve motion events. In addition, the emotion experiencers usually participate in the cause events. Same as verbs of happiness, jingya and zhenjing mostly have proto-agent subjects in the cause events, however, they are usually associated with low agentivity subjects. In addition, the causes of both jingya and zhenjing tend to involve non-motion events, and the experiencers do not usually participate in the cause events.

As for the emotion verb types, i.e. change-of-state vs. homogeneous verbs, the results show no difference for both happiness and surprise. This indicates that the emotion classes (happiness vs. surprise) override the emotion verb types (change-of-state verb vs. homogeneous verb) in terms of cause event features.

The results allow us to rewrite the rules in Section 4 by incorporating the cause event features in (21), and an example of the revised rule for zhenjing is given in (22):

(21) Cause Event Features
Happiness:
P (gaoxing) = [+agent, +motion, +participation]
P (kuaile) = [+agent, +motion, +participation]
Surprise:
P (zhenjing) = [+agent, -motion, -participation]
P (kuaile) = [+agent, -motion, -participation]

(22) Zhenjing
Rule 1:
In this paper, we provide some concrete linguistic rules in identifying the important elements of an emotion (happiness and surprise), which are the experiencer and the cause. These rules can be very useful for emotion detection and classification. For example, a rule-based system can be developed using these rules; or the value of each rule serves as a feature in a statistical system. It is very common that an emotion is expressed without an emotion keyword in web data, such as web blog and online chat. The possible way to detect emotions in this kind of data is to rely on the detection of the cause events in the contexts.

We also examine the features of a cause event according to its degree of transitivity, namely agentivity, kinesis, and event participation. We find that the emotion classes (happiness vs. surprise) override the emotion verb types (change-of-state verb vs. homogeneous verb) in terms of cause event features. In order to automatically detect cause events based on these three features, more work need to be done, for instances, providing a verb list that takes a motion event, identifying the semantic roles of the nouns using a semantic parser, etc.

6 Conclusion and Future Work

In this paper, we provide some concrete linguistic rules in identifying the important elements of an emotion (happiness and surprise), which are the experiencer and the cause. These rules can be very useful for emotion detection and classification. For example, a rule-based system can be developed using these rules; or the value of each rule serves as a feature in a statistical system. It is very common that an emotion is expressed without an emotion keyword in web data, such as web blog and online chat. The possible way to detect emotions in this kind of data is to rely on the detection of the cause events in the contexts.

We also examine the features of a cause event according to its degree of transitivity, namely agentivity, kinesis, and event participation. We find that the emotion classes (happiness vs. surprise) override the emotion verb types (change-of-state verb vs. homogeneous verb) in terms of cause event features. In order to automatically detect cause events based on these three features, more work need to be done, for instances, providing a verb list that takes a motion event, identifying the semantic roles of the nouns using a semantic parser, etc.

References


