DECODING EMOTIONS IN SOCIAL MEDIA: A LINGUISTIC ANALYSIS OF IMPLICIT EMOTIONS AND EVENTS

HELENA YAN PING LAU

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

DECODING EMOTIONS IN SOCIAL MEDIA: A LINGUISTIC ANALYSIS OF IMPLICIT EMOTIONS AND EVENTS

HELENA YAN PING LAU

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2019

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Signed)

HELENA YAN PING LAU (Name of student)

Abstract

This thesis aims to discover meanings expressed in text by focusing on implicit emotion and to investigate the interaction between implicit emotions and event types. Over the past few decades, the web has revolutionized how information is stored, published, and transmitted. Social media in particular has become one of the most influential communication tools, connecting billions of people around the world in a global communications network which is constantly evolving. This allows new information to spread faster and farther to a wider audience, revealing with greater permanence and clarity to the observer the kinds of emotions that are triggered in individuals through the implicit nature of their response to different events. Previous attempts at emotion analysis have focused mainly on the examination of explicit emotions, either in terms of linguistic syntactic and semantic characteristics or their identification and classification within the field of natural language processing. Explicit emotion refers to the emotion-related information denoted directly by the presence of emotion keywords, such as HAPPINESS, ANGER, and SURPRISE. In this thesis, I focus on an important, yet underdeveloped, branch of emotion analysis, implicit emotion. Implicit emotion refers to the presence of emotion-related information conveyed through inference or connotation instead of emotion keywords. I argue that an in-depth analysis of implicit emotion is a necessary component of emotion analysis. As corpus data has shown, the majority of emotions expressed are implicit in nature and there is a clear

gap in existing emotion research, which this current work aims to address.

By exploring implicit emotions in responses to different events posted on an online social media website, I attempt to address the following questions: How are such implicit emotions expressed in text? What kinds of events trigger different implicit emotions? In this study, an annotated Chinese event-comment corpus retrieved from Sina Weibo is constructed. With the empirical data gathered from the corpus, a comprehensive analysis on emotion expressions is carried out at the semantic, syntactic and discourse levels. Drawing on the insight of Pavlenko (2008), emotions expressed at a word level are studied in terms of the use of emotion words, emotion-related words, and emotion-laden words. An emotion expressed at the semantic level, whether implicit or explicit, can be identified with ease, by modifying an emotion taxonomy and proposing a list of emojis, emotion-related words and emotion-laden words.

I also study the syntactic structures that are frequently used to convey emotions. I claim that words of different parts-of-speech can serve as a good emotion indicator when there are no other linguistic clues found in text. The same word formed in different syntactic structures may express different emotions. Findings show that rhetorical questions are a relatively productive means applied in emotion expressions. At the discourse level, the atypical use of emojis is examined. When the emotion expressed in text and the emotion denoted by the emoji are at odds, the overall emotion is determined primarily by the one expressed in text.

Apart from the linguistic features of implicit emotion, the correlation between emotions and events are studied. In lieu of using existing event type classification models, I make use of language resources including TimeML (Sauri et al. 2009), WordNet (Miller 1995) and FrameNet (Baker et al. bamv1998) for the markup of events, event classification and the annotation of frame elements, respectively. Based on the annotated data, I summarize a list of event types which show a preference to a particular emotion and are statistically significant. Furthermore, I also investigate the interplay between emotion, event and semantic role. I confirm the hypothesis that HAPPINESS and ANGER is generally evoked by doers of events associated with that emotion. I also conclude that SADNESS is sometimes elicited by undergoers of events associated with SADNESS, and sometimes triggered by doers of events when the situation leaves the doers no option.

The linguistic account of implicit emotion directly helps to paint a fuller picture of the forms and representations of implicit emotions. First of all, a Chinese event-comment corpus is constructed, which provides valuable resources for emotion studies from the linguistic and computational perspectives. Second, the proposed linguistic cues and the syntactic structures may serve as the features for computational models and classifiers.

This thesis aims to shed light not only on the inference and identification of implicit information, but also on the automatic classification and detection of implicit emotions.

Acknowledgements

The past three years have been full of ups and downs. The journey was long that there were moments that I felt like it was easier to give up, and yet it was short that I did not have enough time to say a few words to the great people I have met along the journey. I am immensely grateful to many people for their love and support during my Ph.D. studies.

First and foremost, I feel deeply indebted to my chief supervisor, Dr. Sophia Lee Yat Mei for her generous guidance, support, and care at every step of my studies. She has been a great source of inspiration to me since 2013. Her unwavering enthusiasm for linguistics and emotion studies keeps me engaged with my research. Her optimism makes me feel alive and motivated. Whenever I feel down, she has always looked out for me and never fails to be my guiding light. One simply could not ask for a better mentor.

I offer my sincerest gratitude to my co-supervisor, Prof. Huang Chu-ren, who has provided invaluable guidance to me throughout my graduate education. His creativity, his vision, and his modesty have deeply impressed and inspired me. It was a great privilege to work under his supervision. I am extremely grateful for his unceasing support and help.

My gratitude also goes to Dr. Roxana Fung Suk Yee who has always been supportive of me. Whenever I need help, she is always there to help without any

hesitation. I am thankful that she saved my life during the 30th North American Conference on Chinese Linguistics (NACCL-30) in the U.S.

I profoundly thank my buddy Xuefeng Gao for accompanying me to the destination of this journey. He offered help to me when I was in need; he took good care of me when I was critically ill; he comforted me when I was going through a difficult time. He is a precious gift from God. He will be forever missed.

I would like to extend my full appreciation to Dr. Xu Hongzhi for helping me in many ways and being generous with his time. He offered some fresh insights into the topic and helped sharpen the thesis. It is a blessing to have a friend like him. Thanks also go to Dr. Wang Zhongqing and his student Wang Qifa for their tremendous help.

I am so grateful to all my friends who played an important part in seeing me through the Ph.D. research: Jasper Wu, Yike Yang, Jessica Leung, Winnie Yeung, Pia Cheng, Junfu Wong, Changwei Zhang and Peter Siew. I would like to offer my special thanks to Jason Yeung for his love, care, and support. I thank him for coming into my life and brightening up my days when life gets hard.

Lastly, I am truly thankful for my family, especially my grandpa and grandma. It is a pity that they would not be able to see me graduate again, but I know they have just become stars twinkling in the sky and bringing light to my heart.

I sincerely thank you all for being on this journey with me!

Table of Contents

List of Figures

List of Tables

CHAPTER 1
INTRODUCTION

1.1 Overview

Emotion has long been a well-studied topic for research across various disciplines, among the philosophy, psychology, sociology, computer science, and linguistics. As the foci of cross-disciplinary researches begin to converge on a post-technological understanding of emotions and how they are expressed online, we find ourselves increasingly looking to social media as a beacon for revealing users' thoughts and feelings on any given number of topics.

Despite the differences in the definition of emotion among the various fields, most emotion theories agree that emotion is a cognitive state that induces bodily reactions to external events (James 1884, Cannon 1927, Plutchik 1962, Ortony et al. 1988, Harkins and Wierzbicka 2001). As such, emotion is a pivot event that interacts with its associated events, namely pre-events (i.e. emotion causes) and post-events (i.e. emotion reactions). Moreover, emotion theories generally regard emotion causes as an integral part of emotion elicitation (James 1884, Plutchik 1980, Wierzbicka 1999). These studies highlight the significance emotion causes play in an emotion expression. Although work has been conducted on emotion-eliciting events, no study has conjured a clear picture of what kinds of events are more likely to trigger implicit emotions, and how such implicit emotions are expressed in text.

In this thesis, I will construct a Chinese event-comment corpus using Turner's (2000) emotion classification of basic and complex emotions. The data is taken from Sina Weibo, which is one of the most popular social media sites in Mainland China and provides a robust source of data for emotion studies. It serves as a platform for users to disseminate all kinds of information, and allow them to instantly respond to events that they are interested in. As emotion cause events are an integral part of emotion elicitation, incorporating both the posts and their corresponding comments may help improve the performance of implicit emotions identification as most comments are made on the events mentioned in the post. I will adopt the TimeML annotation guidelines (Sauri et al. 2009) for the markup of events in posts, map those events to WordNet (Miller 1995) for the event classification, and assign a semantic role for each core argument or adjunct mentioned in the events using the semantic frames and frame elements provided by FrameNet (Baker et al. 1998). In doing so, I will address the two major research questions with the annotated event-comment corpus:

a. How are implicit emotions expressed in text?
b. What kinds of events trigger different emotions?

1.2 Concept of Explicit and Implicit Emotions

Emotion identification in text has been a great challenge faced by researchers in the fields of linguistics and natural language processing (NLP). The performance of

existing approaches for emotion detection is still far from satisfactory. This can be attributed to the fact that emotions are frequently expressed in an implicit way without using any emotion keywords. Most attempts at emotion analysis in the field of linguistics and natural language processing have been dedicated to the examination of explicit emotions, and the majority of emotion models do not deal with emotions expressed without using any emotion keywords. From the linguistic perspective, explicit emotion is studied in terms of its semantic and syntactic characteristics. From the computational perspective, it is studied in terms of identification and classification. The term explicit emotion refers to emotion-related information denoted by the presence of emotion keywords. For instance, the emotion keyword "sad" in "I'm so sad that you can't come" is explicitly used to describe the emotion of the experiencer. This thesis differs from previous studies in its aim to explore the characteristics and representations of implicit emotion. Implicit emotion refers to emotion-related information which is inferred by readers, instead of being conveyed through emotion keywords. An example of such a sentence is "I don't want to hear that anymore!" which connotes the emotion of ANGER, despite none of the individual words in the sentence expressing such an emotion. Even though most emotions expressed are implicit in nature, little work has been done on investigating implicit emotions. This thesis will aim to bridge that gap. I will examine the semantic and syntactic features of implicit emotions found in the corpus and propose various linguistic cues and structures to identify potential implicit emotions. On top of linguistic findings, this thesis also aims to unveil the

interaction between emotions and events in greater detail.

1.3 Objectives and Goals of the Study

The majority of emotion expressions are implicit in nature. Despite the significant role implicit emotion plays in emotion studies, relatively fewer studies have been done on the linguistic features and characteristics of implicit emotion. Therefore, I argue that an in-depth analysis of implicit emotion is a necessary component of emotion analysis.

Although there remains little consensus among most emotion theories on how emotion should be defined, researchers studying emotions in different disciplinary fields generally agree that emotion is a cognitive state that induces bodily reactions to external events (James 1884, Cannon 1927, Plutchik 1962, Ortony et al. 1988, Harkins and Wierzbicka 2001). Some even highlight emotion causes as an integral part of emotion elicitation (James 1884, Plutchik 1980, Wierzbicka 1999). Moreover, previous studies show that implicit emotions can be inferred from emotion cause events. Therefore, such events are marked up and classified into different event types, which are then mapped to the emotions identified in the corresponding comments. The investigation into the interaction between different kinds of emotion cause events and implicit emotions not only paints a fuller picture of the characteristics of implicit emotions, it also offers useful information for the identification of implicit emotions in text.

To sum up, the present work aims to add to the linguistic account of implicit emotion through detailed analysis of emotion concepts and the characteristics of each emotion. It will also consolidate the framework for the development of computational models and classifiers for emotion identification and detection.

1.4 Organization of the Thesis

The thesis is organized as follows. Chapter 2 reviews relevant literature on emotion and event classification. Chapter 3 gives a detailed description of the Chinese event-comment corpus. Research methodology, including the markup of events, semantic roles labelling, and the classification of events, is presented. In Chapter 4, explicit and implicit emotions at the semantic level are discussed. Explicit emotions are mainly expressed by means of emotion keywords (i.e. emotion words) and emojis; implicit emotions are expressed by means of emotion-related words and emotion-laden words. Chapter 5 introduces the expressions of implicit emotions at both the syntactic and discourse level. At the syntactic level, an in-depth analysis of the use of certain syntactic structures and rhetorical questions will be given. At the discourse level, the atypical use of emojis will be discussed. Chapter 6 deals with the correlation between event types and emotions by mapping different event types to emotions identified in their corresponding comments. I propose a list of event types that are both strongly correlated with a certain emotion and are statistically significant. I also investigate the interplay of emotion, event and semantic role.

Chapter 7 concludes with a summary reiterating the main contribution of the current work and outlines future work.

CHAPTER 2
LITERATURE REVIEW

This chapter deals with previous work on some important issues pertaining to emotions and events. Section 2.1 introduces the concepts of explicit emotion and implicit emotion. Section 2.2 reviews literature on implicit emotions from different perspectives, namely linguistics, computer science, and psychology. Section 2.3 presents previous work on the classification of events. Events are categorized into linguistic events and real-world events. The former is discussed in terms of the classification of verbs and emotion verbs, and the latter is discussed in terms of the classification proposed in existing language resources.

2.1 Emotions: Explicit vs Implicit

In order to draw a clear distinction between *explicit* and *implicit* emotion to study the characteristics of *implicit* emotion, I first discuss the relation between language and emotion, i.e. how scholars view linguistic expressions of emotion.

Scholars generally agreed that there are two perspectives one can adopt, though named with different terms. Bamberg (1997) differentiated between *emotion talk/ talks about emotions* and *expression of emotion*. The former refers to the situations people engage in to talk about emotions using emotion terms, while the latter refers to the behavioural act of expressing emotions in communication. Grondelaers and Geeraerts (1998) adopted the terms *language about emotion* and

language as emotion. The former studies emotions through the denotation of linguistic expressions, while the latter focuses on the expressions that do not denote emotions literally but are of emotive values. Kövecses (2000) classified emotion words into *descriptive* and *expressive emotion words*. He considered *descriptive emotion words* are similar to assertive speech acts that describe emotions such as *anger, joy, sadness etc.*, whereas *expressive emotion words* as the analogue of expressive speech acts which express emotions directly such as *shit*, *wow* and *yuk*. However, the terms *descriptive* and *expressive* are not equivalent to *explicit* and *implicit* emotions. Kövecses (2000) further indicated that a descriptive emotion word can be used in an expressive way to both describe and express an emotion, as in "*I love you!*".

Fiehler (2002) termed them *thematization* and *expression* in a broader sense. Different from other researchers who only take emotion terms into consideration, Fiehler (2002) defined *thematization* as verbal labelling of experiences and emotions, description of experiences and emotions, designation or description of the events and circumstances relevant to the experience and description or narration of the situational circumstances of an experience. Not limited to verbal communication, *expression* includes all behaviours and involuntary physiological reactions, and it is "conceptualized from the outset in terms of its communicative function within interaction (Fiehler 2002: 87)".

Bednarek (2008) put forward the terms *emotion talk* and *emotional talk*. She suggested that *emotion talk* refers to expressions that directly denote a particular

emotional response such as *fear*, while *emotional talk* refers to expressions that can be related to some kinds of emotional experience. In Bednarek (2009: 11), she further defined *emotional talk* as "including all sorts of human behaviour that signal emotion without the resource to linguistic expressions that explicitly denote emotion (*emotion talk*)". That is, while *emotion talk* includes all emotion terms that denote emotions, *emotional talk* includes other expressions such as intonation, punctuation, interjections, inversion, exclamation, pronoun use, emphatic particles, intensifiers, swearwords, etc.

Foolen (2012) named the two ways of communicating emotions *symbolic* and *symptom*. The former uses words that are context-independent, such as *I find the food disgusting*; the latter shows how the speaker feels at the time of speaking, such as *Yuk*! He noticed that expressive linguistic forms are less studied than conceptual-descriptive emotion lexicons. He suggested that emotional interjections and many other forms or constructions such as *a bear of a man* in the 'an N of an N' pattern (Foolen 2004), *nandao*-interrogative in Chinese (Jing-Schmidt 2008) etc. should also be regarded as emotive/ expressive language. Drawing on the insight of the *descriptive* versus *performative* use of speech act verbs and *indirect* versus *direct* speech proposed in Verstraete (2001), Foolen (2012) found that emotions communicated in an expressive way are like *performative* utterances or *direct* speech, with the speaker being personally involved.

In fact, the selection of emotion words and the way how researchers define emotion varies. On top of emotion terms (used in descriptive ways) and

interjections (used in expressive ways), other linguistic expressions that are closely related to emotions are sometimes taken into account and sometimes being entirely neglected by different researchers. It leads to the difference in the size of emotion lexicons across languages. For example, some languages only have a few emotion words such as Chewong in Malaysia (Howell 1981), while some other languages contain more than a thousand emotion words, such as English (Wallace and Carson 1973). Pavlenko (2008) proposed three types of words that are connected with emotions, namely *emotion words*, *emotion-related words* and *emotion-laden words*. The three types are defined in terms of their functions (Pavlenko 2008: 148):

> *"EMOTION WORDS are seen as words that directly refer to particular affective states ("happy", "angry") or processes ("to worry", "to rage"), and function to either describe ("she is sad") or express them ("I feel sad"). In some contexts, these words may also elicit emotions and in others they may function just like abstract words. This definition does not include EMOTION-RELATED WORDS ("tears", "tantrum", "to scream") that describe behaviors related to particular emotions without naming the actual emotions... EMOTION-LADEN WORDS are seen here as words that do not refer to emotions directly but instead express ("jerk", "loser") or elicit emotions from the interlocutors ("cancer", "malignancy"). The following subcategories are commonly differentiated among emotion-laden words: (a) taboo and swearwords or expletives ("piss", "shit"), (b) insults ("idiot", "creep"), (c) (childhood) reprimands ("behave", "stop"), (d) endearments ("darling", "honey"), (e) aversive words ("spider", "death"), and (f) interjections ("yuk", "ouch")."*

Although the definitions seem to help select and distinguish lexicons of a particular emotion from those of the others, Pavlenko (2008) acknowledged that the selection of emotion words should also depend on the context. For example, some

swearwords which usually function as insults may be used in a friendly way to express affection, while some words that are not regarded as *emotion-laden words* such as "liberal" may appear as insults in some context. Ng et al. (2019) explored language specificity in the organization and distribution of emotion words in Mandarin Chinese. They extracted Chinese emotion words and classified them into the three categories proposed in Pavlenko (2008), namely emotion words, emotion-laden words and emotion-related words. Those words are tagged with frequency, valency, intensity and part-of-speech. It is found that verbs occupied the biggest percentage in both emotion words and emotion-related words categories. A template is also suggested for the identification of emotion words.

Following the definitions of descriptive emotion words and expressive emotion words proposed in Kövecses (2000), the study of descriptive emotion has been the focus of most research on emotion in linguistics as explicit emotion must be expressed with the presence of a descriptive emotion word, though it can be encoded in an expressive way simultaneously. The example given in Kövecses (2000) is "*I love you!*" where the descriptive emotion word *love* is used to describe and express the emotion of *love* at the same time. It proves that an explicit emotion may be expressed in a descriptive way or it may be expressed in both descriptive and expressive ways with a descriptive emotion word being present. As for implicit emotion, it can be expressed either in an expressive way or it can be expressed in a descriptive way at the sentence level without using any descriptive emotion words as in "*I thought you would come to my birthday party*" which expresses a SADNESS

(i.e. *disappointed*) emotion in a descriptive way without using any descriptive emotion words. Therefore, I argue that the term descriptive and expressive should be extended to the understanding of emotion at the sentence level instead of the lexical level. This claim is supported by Kövecses (2000: 6) who proposed a summary of types of emotion language as shown in Figure 2.1 (Kövecses 2000: 6).

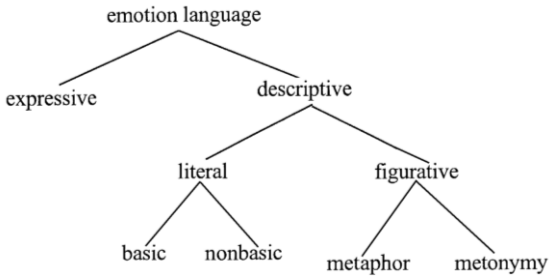


Figure 2.1: Summary of Types of Emotion Language (Kövecses 2000: 6)

As shown in Figure 2.1, three types of emotion language are presented, namely expressive terms, terms that literally denote emotions, and figurative expressions denoting certain aspects of emotions. In fact, figurative expressions which belongs to the descriptive emotion language are often used to express emotions implicitly. For example, the figurative expression "*I feel like a caged animal*" descriptively denotes a SADNESS emotion without using any terms that literally denote emotions. That means, the SADNESS emotion is implicitly expressed in a descriptive way. Therefore, I argue that the distinction between descriptive and expressive is not adequate to distinguish explicit and implicit emotions, and the only difference lies in the presence of descriptive emotion words. To avoid confusion, the terms *explicit*

emotion and *implicit* emotion are used in this study. This work attempts to investigate implicit emotions in both the descriptive and expressive ways (including but not restricted to *emotion-related* and *emotion-laden words* in Pavlenko's work (2008)). Lee (2015: 186) defined *explicit* emotion as "the presence of emotion-related information denoted by emotion keywords", and *implicit* emotion as "the emotion-related information that requires inference or connotation instead of being conveyed by emotion keywords". By *explicit* emotion, it refers to the emotion terms that directly denote one's emotional state as what Pavlenko (2008) termed as *emotion words*. For example, in the sentence "*I'm happy that everything is working out for you*", whereby the word *happy* explicitly refers to the emotion state of HAPPINESS. As for *implicit* emotion, it refers to the use of other linguistic cues without the presence of *emotion words*. An example of such a sentence is "*go away and leave me alone*" which denotes the ANGER emotion while none of the individual words in the sentence expresses such an emotion.

2.2 Previous Studies on Implicit Emotions

2.2.1 From the Linguistic Perspective

Although most emotions are expressed implicitly, little work has been done to examine implicit emotions. Greene and Resnik (2009) proposed an approach to classify implicit emotion based on grammatically relevant semantic features that characterize the interface between syntax and lexical semantics. They conducted

both semantic property ratings and sentiment ratings in order to validate the hypothesis that there is a connection between syntactic choice and implicit sentiment from the readers' perspective. As for semantic property ratings, they constructed sentences using 11 verbs of *killing*, which can be categorized into the groups of transitive verbs (i.e. externally caused change-of-state) and ergative verbs (i.e. internally caused change-of-state). These sentences are compiled in two forms, one of which contains a human agent as the subject, and the other contains a nominalization of verb as the subject. Participants were asked to answer questions with respect to the 13 semantic properties such as volition, agency, kinesis etc. as proposed in Dowty (1991) on the traditional thematic roles of agent and patient, and Hopper and Thompson (1980) on transitivity. They found that volition, among the 13 properties, has the strongest correlation with the *sympathy* emotion, followed by sentiment, and kinesis/movement. As for sentiment ratings, they examined whether syntactic forms of event description affect readers' perceptions of the author's emotion. They provided three forms of headlines including *transitive*, *nominalized*, and *passive form*, and asked participants to rate on a 1-to-7 scale based on how sympathy they perceive the headline to be toward the perpetrator. Examples taken from the work of Greene and Resnik (2009: 505) as in (1) – (3):

(1) Man suffocates 24-year-old woman (Transitive)

(2) Suffocation kills 24-year-old woman (Nominalized)

(3) 24-year-old woman is suffocated (Passive)

Results show that the overall effect of linguistic forms was significant. The *transitive* form was significantly lower in sympathy ratings than the *nominalized* and *passive* one.

Apart from examining implicit emotions at the syntactic level, Lee (2015) attempted to investigate implicit emotions at the semantic level by figuring out linguistic cues that connote emotions implicitly. She constructed a Chinese emotion annotated corpus for the linguistic analysis of implicit emotion. She indicated that approximately 47% of emotions are expressed implicitly among all the emotional posts, which proves that implicit emotion is an important component of emotion analysis. She extracted instances from social media and manually classified them into five primary emotions. A list of linguistic cues including adjectives and adverbs is proposed for each emotion type. For example, the adverb 終於 '*eventually*' is frequently found in posts conveying a *happiness* emotion. It is concluded that some cues have certain semantic orientation pointing to a particular emotion, and they are of great value for implicit emotion detection when no other notable lexical cues are found in text.

In addition to the identification of emotion-laden words, the detection of emotion cause may also help identify implicit emotions. Lee et al. (2010) first proposed the emotion cause detection. They constructed a Chinese emotion-cause annotated corpus for the purpose of extracting emotion causes. They identified seven groups of linguistic cues and two sets of linguistic rules that can be used for

emotion cause detection. Based on the linguistic rules proposed, Lee et al. (2013b) developed a rule-based system for the detection of emotion cause. Drawing from the insight of Lee et al. (2010, 2013b), a few studies (Gui et al. 2014; Li and Xu 2014; Gao et al. 2015) extended the rule-based method to the detection in informal text. These studies are introduced in Section 2.2.2 as they are basically done from the computational perspective. Based on most emotion theories which regard emotion cause event as an integral part of emotional elicitation (Descartes 1649; James 1884; Plutchik 1980; Wierzbicka 1999), Lee et al. (2013a, 2014) constructed another Chinese event-based emotion corpus with both pre-events and post-events annotated. Lee et al. (2013b) constructed a Chinese emotion cause corpus and identified seven groups of linguistic cues, namely causative verbs, reported verbs, say verbs, epistemic markers, prepositions, conjunctions, and others. They developed a rule-based system for emotion cause detection with the two sets of linguistic rules generalized. Results show that the system yielded a promising performance for cause occurrence and cause event detection. They suggested that there are significant interactions between emotions and pre-events as well as that of between emotions and post-events. It is believed that the study could be beneficial to the inferences of implicit information such as implicit emotion detection, as well as the discovery of new information based on cause-event relation.

2.2.2 From the Computational Perspective

Computer scientists have attempted to process and identify affective information,

including sentiment and emotion analysis. At first, attention has been paid to sentiment analysis only. Sentiment analysis aims to examine people's sentiment toward various objects, and it is composed of sentiment classification and opinion mining (Pang and Lee 2008). Sentiment analysis has been extensively studied, especially sentiment classification, which classifies texts according to their semantic polarity, i.e. positive and negative. However, this polarity-driven approach is often criticized as too general to satisfy some real-world applications, such as public consultation. In view of this, researchers began to explore more fine-grained affective information, i.e. emotions. Emotion analysis has gone beyond the binary classification used in sentiment analysis. The number of basic emotions varies from two to ten, such as *happiness*, *sadness*, *fear*, *anger*, and so on (Sabini and Silver 2005, Lee 2010, Keltner et al. 2014, Scheff 2015). It provides a more precise picture of people's mental state, and has even greater potential for various challenging tasks, such as improving marketing strategies, or making political decisions etc.

Emotion detection can be done using different approaches. Kozareva et al. (2007) adopted a statistical approach. They hypothesized that words that frequently co-occur with a particular emotion are highly likely to express that emotion, just as "birthday" appears more often with *joy*. Hence, they computed the Mutual Information scores of the content words of a news headline and six basic emotions. However, it is found that the approach does not work well in emotion classification task which may be attributed to the fine-grained classification.

Lei et al. (2014) used a lexicon-based approach. They built an emotion

detection system for news based on three modules, namely document selection, part-of-speech tagging, and emotion lexicon generation. They constructed a lexicon that contains both explicit and implicit emotion words for emotion prediction. It is suggested that all modules have a positive effect on emotion detection. However, Agrawal and An (2012) believed that lexicon-based approach is inadequate for implicit emotion detection. They also noticed that word order may also affect the overall emotion of a phrase. For example, in the phrase "joyless cheer", "joyless" denotes a *sadness* emotion whereas "cheer" a *happiness* emotion. As "joyless" influences the emotion vector of "cheer", and the whole phrase should be tagged as *sadness*, which will simply be labelled as neutral in tradition keyword-based approach.

Chaumartin (2007) used a knowledge-based approach to annotate news headlines with sentiment and emotion. They indicated the importance of head word and proposed a set of rules stating that when a noun is inherited from a given synset, certain emotions need to be boosted.

In addition to the traditional approaches abovementioned, emotion detection can also be done with the help of the collection of emotion-eliciting events. Tokuhisa et al. (2008) first defined emotion-provoking event. They collected massive examples of emotion-eliciting events in Japanese from the web by using an emotion lexicon and lexical patterns to train the classifiers for sentiment and emotion classification tasks. Vu et al. (2014) constructed a dictionary of emotion-eliciting events by using a self-reported approach. They have aggregated similar

events, and plenty of the emotion-eliciting events are mundane events such as meeting friends, being insulted by someone, thinking about the futures etc. The idea of emotion cause detection is first proposed in Lee et al. (2010). They constructed a Chinese emotion-cause annotated corpus, aiming to extracting emotion causes. They identified seven groups of linguistic cues and two sets of linguistic rules that can be used for emotion cause detection. Extending the rule-based approach to informal text, Gui et al. (2014) extracted from data from Chinese Weibo to construct an emotion cause corpus. Ghazi et al. (2015) utilized the emotions-directed frames in FrameNet to build an English emotion cause (or stimulus) corpus. They used the Conditional Random Fields, CRFs for the detection of emotion causes. Gui et al. (2016) built a dataset using SINA city news. They proposed a new event-driven emotion cause extraction method using multi-kernel SVMs. Ding and Riloff (2016) aimed to automatically acquire knowledge of stereotypically positive and negative events for sentiment analysis. Rashkin et al. (2018) proposed Event2Mind as a supporting evidence to common sense inference on events. They mainly concentrated on intents and reactions in informal texts. Liu et al. (2019) constructed a Cause-Emotion-Action Corpus, with emotions, pre-events, and post-events being manually annotated. They put forward the (cause, emotion, action) extraction task, and the emotion detection task based on the Corpus.

2.2.3 From the Psychological Perspective

Researchers have attempted to investigate emotions from different perspectives.

This section introduces one of the most influential emotion theories, i.e. Appraisal Theory.

Calvo and D'Mello (2010) reported serval psychological theories of emotions that view emotions as expressions, embodiments, outcomes of cognitive appraisal, social constructs, products of neural circuitry, and psychological interpretations of basic feelings. They explained the reason for certain episodes leading to a particular emotional state. Among these theories, the appraisal theory emphasizes that emotions are elicited and differentiated based on a person's subjective evaluation of a situation, object or event (Scherer 1999; Ellsworth and Scherer 2003).

The term "appraisal" is first introduced by Arnold (1960) to explain the elicitation of differentiated emotions. She believed that events are appraised according to three dimensions, namely *beneficial vs. harmful*, *presence vs. absence of some objects*, and *easy vs. difficult to approach or to avoid*. Lazarus (1966) claimed that stress and emotion are evoked by two stages of appraisal. The primary stage appraises the positive or negative significance of an event for one's well-being, and the secondary one appraises the ability to deal with the consequences of an event. Scherer (1999) made a comparison of the appraisal criteria proposed by various theorists (Frijda 1986; Roseman, 1984, 1991; Scherer 1984a, 1984b, 1986, 1988; Smith & Ellsworth, 1985) and claimed that "theorists in this tradition postulate that specific profiles of appraisal outcomes on these criteria determine the nature of the ensuing emotion" (Scherer 1999: 638). Examples are given in Table 2.1 (Scherer 1999: 639):

Table 2.1: Specific Profiles of Appraisal Outcomes (Scherer 1999:639)

Stimulus evaluation checks	Anger/rage	Fear/panic	Sadness
Novelty			
• Suddenness	High	High	Low
• Familiarity	Low	Open	Low
• Predictability	Low	Low	Open
Intrinsic pleasantness	Open	Open	Open
Goal significance			
• Concern relevance	Order	Body	Open
• Outcome probability	Very high	High	Very high
• Expectation	Dissonant	Dissonant	Open
• Conduciveness	Obstruct	Obstruct	Obstruct
• Urgency	High	Very high	Low
Coping potential			
• Cause: agent	Other	Other/nature	Open
• Cause: motive	Intent	Open	Chance/neg
• Control	High	Open	Very low
• Power	High	Very low	Very low
• Adjustment	High	Low	Medium
Compatibility with standards			
• External	Low	Open	Open
• Internal	Low	Open	Open

Note: Open-different appraisal results are compatible with the respective emotion.

Scherer (1999) summarized four strategies that theorists frequently used to unveil the relation between particular configurations of appraisal results and the ensuing emotion reaction. First, subjects were asked to recall emotion experiences, and the outcome of antecedent evaluation processes. Second, verbal reports on the appraisal process were obtained by inducing subjects' emotions using naturally occurring events. Third, emotion words judged according to the appraisal implications could be obtained. Fourth, subjects were asked to imagine and indicate the emotional reactions they might experience in scenarios that are manipulated based on the appraisal-related dimensions. However, Scherer (1999) reported that some researchers (Frijda 1993, Parkinson 1996, 1997) have extensively criticized the use of self-report of emotion-antecedent appraisal. These researchers argued that subjects are not likely to be able to report upon the antecedent processes which often

occur without awareness. Moreover, subjects may construct a rationale for their emotion responses. As for the analysis of narratives or interviews, researchers had already imposed an interpretative scheme according to a particular theory.

Moors (2018) also noticed that the use of self-reports of appraisal criteria and emotion labels in questionnaires has received severe criticism. First, there is no evidence yielded for the causal relations between appraisal criteria and emotion labels, therefore, the use of correlational studies may not be appropriate. Second, an emotion label is not a component of the emotion, thus, emotion labels should not be used to measure emotions. In addition, appraisal and emotion labels are conceptually related, thus the correlation may reflect conception instead of causal relations (Frijda & Zeelenberg 2001, Parkinson 1997). Third, appraisal process often occurs in an unconscious or automatic way whereas some appraisal criteria require consciousness, the integrated form they used may make it hard to be analysed (Scherer 2009), and that they are likely to be replaced by the stereotypic scripts about appraisals and emotions (Robinson and Clore 2002).

In fact, appraisal theory emphasizes that one of its major strengths is that it accounts for individual differences of emotion reactions to the same event (Smith and Lazarus, 1990; Smith and Pope, 1992). Therefore, the appraisal is highly subjective as the same events may trigger disparate emotions in different individuals due to appraisal bias. Moreover, the evaluation of emotion reactions including emotion type and emotion intensity may also vary across cultures (Mesquita et al., 1997). As Scherer (1999: 650) suggested, “even if emotion were

to be considered a relatively universal psychobiological mechanism, one can assume that the nature of the eliciting events and the type and intensity of emotional reactions to similar events would be highly different across different cultures". Apart from that, appraisal theory takes the idea that stimuli are constantly reappraised. Therefore, it might be difficult to distinguish one emotion-generative cycle from another.

Despite the criticism, some computational models were developed and derived from appraisal theories for emotion detection. These models attempt to predict emotions with the aid of the large number of emotion-eliciting events collected. In this section, some emotion detection models developed based on the appraisal theory are discussed.

Scherer (1993) constructed a GENESE expert system on knowledge-based that mapped appraisals to different emotions. The GENESE system attempted to address the following questions: (1) what are emotions? (2) what is an emotional event? (3) what is the role the computer system plays? It is shown that the system achieved a high accuracy predicting target emotions for emotional episodes described by subjects as a series of responses to situational questions.

Balahur et al. (2011) believed that emotions are mostly expressed not through specific words but by evoking situations, they therefore proposed a resource named EmotiNet for the detection of emotion using commonsense knowledge, which is built on the appraisal theory. They divided situations taken from the International Survey of Emotional Antecedents and Reactions (ISEAR) into a sequence of action

links (i.e. action chains), and described each link with a "4-tuple" (actor, action type, patient, emotional reaction). The action links as well as the corresponding emotion reactions are then stored to EmotiNet using an ontological representation. EmotiNet was then extended with VerbOcean. They conducted an experiment to evaluate the performance of the proposed approach and concluded that the approach is appropriate for emotion detection in text. Yet, the precision and recall measures remained quite low, and additional information and extra knowledge should be included in the model.

2.3 Previous Studies on Event Classification

Regarding events, Rosen (1999) suggested that events can be discussed in two ways, namely linguistic events and real-world events. The former refers to the linguistic representations of things that happen in the real world, and the latter refers to things that happen in the real world. I discuss linguistic events in terms of verbs in Section 2.3.1. Section 2.3.2 introduces real-world events by introducing some existing language resources.

2.3.1 Linguistic Events

2.3.1.1. Verbs

Instead of classifying events into a fine-grained list of event types, most linguistic work on event classification has been extensively attempted to classify verbs into a

small number of event types (Aristotle 1984, Vendler 1967, Dowty 1979). That is, the primitive elements of linguistic events.

In early work on event classification, researchers have developed the idea that verbs can be decomposed into a structured representation of an event. Aristotle (1984) made the first attempt to categorize verbs using an event-based approach. He proposed to distinguish states from events and made further distinction between culminating events and non-culminating events. He named states *actuality* and defined it as "the existence of the things". Culminating events and non-culminating events are termed *action* and *movement*, respectively.

Vendler's (1957, 1967) proposal on event classification has been one of the most influential and representative work in the linguistic field. He proposed a fourfold classification of verbs "to describe the most common time schemata implied by the use of English verbs (Vendler 1967: 98-99)". He suggested that verbs can be categorized into four types as follows:

(a) States: non-dynamic situations which hold for some period of time, such as *believe*, *desire* etc.

(b) Activities: events that go on for a time without a terminal endpoint, such as *walk*, *swim* etc.

(c) Accomplishments: events that proceed to an inherent endpoint, such as *draw a picture*, *build a house* etc.

(d) Achievements: instantaneous events that take place in a single moment, such as *find*, *die* etc.

Dowty (1979) put forward three atomic operators that can map onto the four categories Vendler (1957) proposed. These operators are DO, BECOME, CAUSE. A sentence without an atomic operator is a state; a sentence with a DO operator is an activity; a sentence with a BECOME operator is an achievement; and a sentence with the three atomic operators is an accomplishment. For instance, the verb break can be represented as DO CAUSE BECOME DAMAGED.

Vendler (1967) and Dowty (1979) has laid the foreground for the classification of verbs. Pustejovsky (1991) distinguished three basic event types: states, processes, and transitions. Unlike previous studies, he proposed a complex subeventual structure of event types. In addition to the four categories proposed by Vendler (1967), Smith (1991) added another class called semlfactives. Smith (1991) claimed that achievements and semlfactives are different in terms of the telicity. While semlfactives is defined as instantaneous events without a culminating point (i.e. atelic), achievements are instantaneous culminating events (i.e. telic). Examples of semlfactives include cough, sneeze etc.

The classification of Chinese verbs is similar as compared to English verbs. Ma (1981) and Teng (1986) adopted Vendler's four-way classification of situation types in Chinese. Tai (1982) indicated three categories of verbs pertaining to the notion of time, namely *state*, *activities*, and *result*. Tai (1984) further explained that accomplishment verbs in Chinese are expressed by means of resultative verb compounds. While Tai (1984) studied the issue with a focus on verb classification, Smith (1990) placed more emphasis on event types. Smith (1990) argued against

Tai's (1984) proposal that Mandarin Chinese does not have accomplishments but result aspect only. He claimed that resultative verb compounds in Chinese are the same as accomplishment verbs in English as they both represent a telic constellation.

Huang et al. (2000) proposed another method called the module-attribute representation of verbal semantics (MARVS) to represent event structures. MARVS is composed of four components of the models: event module, event-internal attributes, role module, and role-internal attributes, among which the first two are of concern to event classification. Event modules represent the properties pertaining to the aspectual composition of an event with five primitives of events, namely Boundary [·], Punctuality [/], Process [/////], State[___], and Stage [^^^^]. The five primitives can be combined to form complex event types. For example, the verb 凋谢 'wither' can be represented by mean of [· ^^^^ ·].

2.3.1.2 Emotion Verbs

Among all the word categories such as verb, adjective, noun, adverbs etc., emotion verbs are most extensively studied. Levin (1993) stated that psych-verbs (i.e. termed emotion verbs in this study) typically take two arguments which are most frequently characterized as experiencer and stimulus. He suggested that it is possible to classify emotion verbs into four subclasses in terms of argument expressions: (1) amuse verbs are transitive verbs whose subject is the cause[1] and

[1] The word "cause" is used instead of "stimulus" as some verbs allow the subject to receive an agentive interpretation, but some do not.

the object is the experiencer of the emotion. He described the bringing about of a change in emotional state as in "*the clown amused the children* (Levin 1993: 190)" Other members include *annoy*, *surprise*, *upset*, *worry* etc. (2) *admire* verbs are transitive verbs whose subject is the experiencer and the object is the stimulus of the emotion as in "*the tourists admire the paintings* (Levin 1993: 191)". Other members include *enjoy*, *like*, *envy*, *fear*, *hate* etc. (3) *marvel* verbs are intransitive verbs whose subject is the experiencer of the emotion, and the stimulus or object appears in a prepositional phrase as in "*Megan marveled at the beauty of the Grand Canyon* (Levin 1993: 193)". Other members include *sadden about*, *fear for*, *hurt from* etc. (4) *appeal* verbs are intransitive verbs whose subject is the stimulus, the experiencer appears in a prepositional phrase as in "*the painting appeals to Malinda* (Levin 1993: 193)". Other members include *niggle at*, *grate on*, *jar on* etc.

Different from Levin (1993) who classified verbs in terms of argument expressions, Talmy (2000) classified English affective verbs into two valence types in terms of subject-selection, namely *experiencer-as-subject* (e.g. *fear/ like*) and *stimulus-as-subject* (e.g. *frighten/ please*). He pointed out that although verbs may select either the experiencer or stimulus as the default subject, it is possible for each type of verbs to switch to another type via grammatical-derivation. For instance, "*it frightens me*" selects the stimulus as the default subject, and the sentence can be derived as "*I'm frightened of it*" which selects the experiencer to be the subject. He noticed that the *stimulus-as-subject* type predominates in the case of English. The derivation is especially productive in deriving an experiencer subject with a

stimulus-as-subject verb than deriving a stimulus subject with an *experiencer-as-subject* verb.

In addition to emotion verbs, Jackendoff (2007) also included adjectival "psychological" predicates and proposed even finer morphological and semantic distinctions for the predicates in English. The variants include (1) Exp-Adj as in "*I'm bored*", (2) Exp-Adj-Stim as in "*I'm bored with this*", (3) Exp-Verb-Stim as in "*I detest this*", (4) Stim-Verb-Exp as in "*this bores me*" and (5) Stim-Adj-(Exp) as in "*this is boring (to me)*". Following Ekman and Davidson (1994), Jackendoff (2007) acknowledged that certain emotional experiences are pure feelings that are independent of surroundings which do not require the presence of a stimulus, such as *happy*, *sad*, *calm*, *scared*, and *upset*, while most are directed feelings that require the presence of a stimulus at which the affect is directed such as being *attracted*, *disgusted*, and *interested*. He named the former *inherent* feelings, and the latter *directed* feelings.

With regard to Chinese emotion verbs, Yan et al. (2006) realized that agents and patients of different emotion verbs may experience different emotions. They proposed a semantic analyser for emotion recognition. After selecting eight emotion predicates including 高興 'happy', 表揚 'praise', 討厭 'hate', 愛上 'fall in love', 苦惱 'distressed', 尊敬 'respect', 嘲笑 'mock', 喜愛 'like', some rules are proposed to assign an emotion to each verb, and to different semantic roles involved. For example, while the *joy* emotion is assigned to the agent of 嘲笑 'mock', the *sorrow* emotion is assigned to its patient. They extracted 10 sentences

for each predicate for examination. Decision Tree classifier is adopted to assign the semantic dependency relations between headwords and dependents. After combining both the semantic information and the rules proposed for the 8 emotion predicates, the analyser assigned emotions to a sentence as exemplified in Figure 2.2.

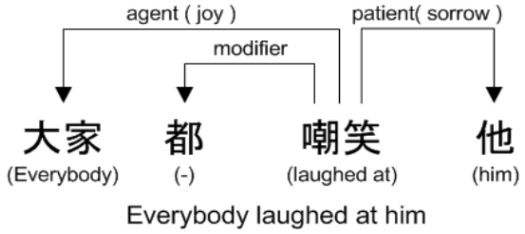


Figure 2.2: Emotion Assigned by the Analyzer (Yan et al. 2006:899)

They believed that the approach has the potential to determine the emotions of the agents, as well as those of the others.

Similar to English emotion verbs, emotion verbs in Chinese can also be classified into *experiencer-as-subject verbs* and *stimulus-as-subject verbs*. A number of studies on emotion lexicons in Mandarin Chinese focus on the syntax-semantics interface shown in the *experiencer-as-subject verbs*. Tsai et al. (1998) discussed the distinctive features between the near-synonym pair of "happy" 高興 and 快樂. They showed that the two words have different distributional behaviours in four aspects: 1) 高興 can take a sentential object, while 快樂 cannot, 2) 高興 can be followed by a perfective aspect marker 了, while 快樂 cannot, 3) 高興 constitutes evaluative sentences as in "*這件事很值得高興*", but never forms wish

sentences; 快樂 occurs in wish sentences as in "*祝你快樂*", but never in evaluative sentences, and 4) 高興 can form imperative sentences, while 快樂 cannot. In Tsai et al. (1999), they further investigated the distinctions between 高興 and 快樂. They found that 高興 shows a higher frequency in predicative use, eventive adverbials, and causal complements. It is proposed that the two near-synonyms differ in their states. 高興 represents an *inchoative state* (i.e. *change-of-state*) with higher degree of control, whereas 快樂 represents a *homogeneous state* with less volitional control.

Following Tsai et al. (1998, 1999)'s work, Chang et al. (2000) proposed 7 sets of the most frequent emotion verbs namely happy, depressed, sad, regretful, angry, afraid, and worried. They identified a number of verbs in the seven types of emotion verbs, each with a frequency of over 40 in the Sinica Corpus. The emotion verbs are then classified into two groups, namely change-of-state verbs, and homogenous state verbs in terms of five criteria. The dichotomy of emotion verbs and the frequency is shown as in Table 2.2.

Table 2.2: The Dichotomy of Emotion Verbs (Chang et al. 2000:64)

Subtype	Group A	Group B
HAPPY	**gaoxing 高興(669)**	**kuaile 快樂(942)**
	kaixin 開心(152)	yukuai 愉快(271)
		xiyue 喜悅(156)
		huanle 歡樂(141)
		huanxi 歡喜(107)
		kuaihuo 快活(48)
		tongkuai 痛快(40)
DEPRESSED	**nanguo 難過(232)**	**tongku 痛苦(443)**
	tongxin 痛心(48)	chenzhong 沉重(83)
		jusang 沮喪(62)
SAD	**shangxin 傷心(134)**	**beishang 悲傷(52)**
REGRET	**houhui 後悔(102)**	**yihan 遺憾(198)**
ANGRY	**shengqi 生氣(307)**	**fennu 憤怒(112)**
		qifen 氣憤(49)
AFRAID	haipa 害怕(261)	**kongju 恐懼(149)**
		weiju 畏懼(40)
WORRIED	**danxin 擔心(609)**	**fannao 煩惱(199)**
	danyou 擔憂(64)	kunao 苦惱(45)
	youxin 憂心(46)	

They noticed that there are distributional differences between the two groups of verbs as shown in Table 2.3. They further argued that the differences can be ascribed to their morphological differences, in which the *change-of-state verbs* are non-VV compounds and *homogenous state verbs* are VV compounds. Unlike non-VV compounds which usually elaborate on the denoted event, the two verbs in VV compounds combine two similar events or link two antonyms or synonyms to form the concept of "kind" or "property". Therefore, it is natural for VV compounds to be chosen to indicate a homogenous state.

Table 2.3: Syntactic Differences between Two Types of Verbs (Chang et al. 2000)

	Change-of-State Verbs	Homogeneous State Verbs
Distribution of grammatical functions	Mostly used as predicates	Mostly used as nominalization or nominal modifiers
Co-occurrence restriction on the head they modified	Can only modify a very restricted set of nouns or verbs	Can modify a number of nouns or verbs
Appropriateness in the imperative and evaluative constructions	All verbs appear in imperative or evaluative constructions	Only one verb (i.e. *fannao* 'to be worried') appears in imperative or evaluative constructions
Verbal aspect	More often associated with inchoative state	More often associated with homogeneous state
Transitivity	They take causes or goals as direct objects	They do not take causes or goals as direct objects

Apart from the binary classification, emotion verbs can also be classified in an even finer way. Liu and Hong (2008) used a frame-based approach proposed by Fillmore and Atkins (1992) and corpus-based approach to classify Mandarin emotion verbs. They adopted the frames proposed in the emotion domain in FrameNet and explored equivalents of those English emotion verbs. Given that the structure of English is different from Chinese, and the categorization in FrameNet is based on syntactic patterns only, some adjustments had been made to facilitate the Chinese data. After removing four frames that are irrelevant to Chinese and renaming the **Emotion_active** as **Cause_to_experience**, they came up with 9 frames to account for the syntax-to-semantic variations among Mandarin emotion verbs, as illustrated in Table 2.4.

Table 2.4: Frames of Mandarin Emotion Verbs (Liu and Hong 2008:116)

Frame Name	Lemma
Feeling	感覺、覺得、感到、感受
Emotion_directed	窘困、尷尬、羞慚、激動、煩躁、悲哀、痛苦、悲 傷、哀痛、高興、快樂、苦惱、不安、羞愧、窘迫、 困窘、惱火、悲痛、受屈、恐懼、吃驚、驚訝、困惑、煩、鎮定、平靜、振奮、目眩、眼花、消沈、失望、為難、洩氣、沮喪、生氣、陶醉、憂愁、著急
Emotion_active	擔心、顧慮、擔憂、掛心
Contrition	懊悔、後悔、悔恨、自責、惋惜
Experiencer_subj	愛、喜愛、喜歡、愛好、熱愛、酷愛、恨、討厭、 厭惡、痛恨、怕、害怕、畏懼、懼怕、恐懼、羨慕、妒忌
Experiencer_obj	安慰、吸引
Cause_to_experience	折磨、打擾、誘惑、激怒、惹惱
Judgement	欽佩、尊敬、欣賞、感謝、感激
Forgiveness	寬恕、赦免、原諒

They further generalized the 9 frames into two major types in terms of grammatical properties and semantic correlations, namely *complement-requiring verbs* (i.e. a complement is required), and *emotion-predicating verbs* (i.e. a complement is not necessary or required) with the former being further divided into *emotion-taking verbs* and *complement-taking verbs*. The structure of the classification is demonstrated in Figure 2.3.

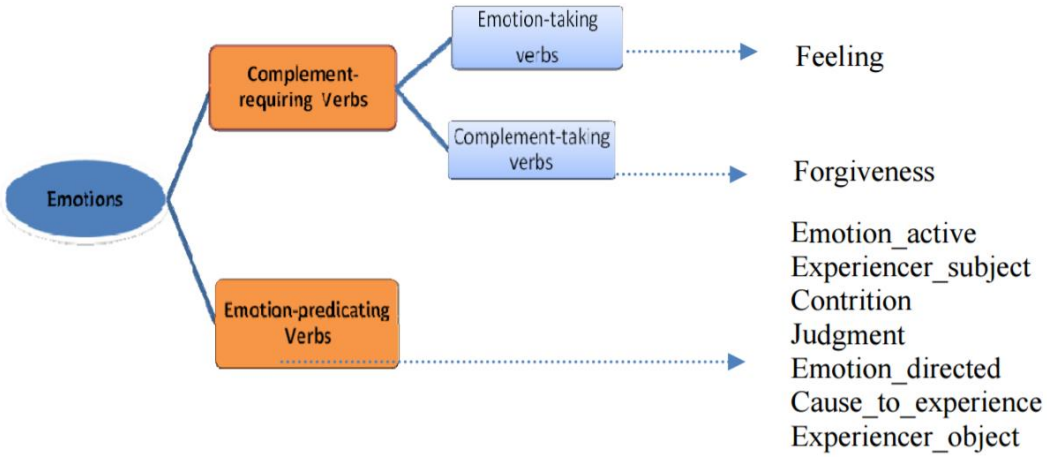


Figure 2.3: Classification of Emotion Verb Frames (Liu & Hong 2008:117)

The emotion-taking verbs share the features of [-很] and [+comp-emotion], and the distinctive pattern is "experiencer + feeling + emotion-predicating [ADJ]" as in "媽媽感到難過". As for complement-taking verbs, they are [-很] and [+comp] (either NP or CL), and the basic pattern is "judge + forgive + evaluee or offense [NP]" as in "他還是原諒了我". Emotion-predicating verb contains the largest numbers of frames, sharing the characteristics of [+很] and an optional complement. The pattern is "experiencer + (很) + emotion-predicating + (Comp [NP/CL])" as in "我很後悔(做了那件事)".

More recently, Liu (2016) attempted to look further into the typological variations in lexicalization patterns of emotions by exploring the interaction between lexical (i.e. semantics) and constructional (i.e. syntax) form-meaning mapping relations in Mandarin emotion predicates. She proposed that the lexical-constructional variations display in Mandarin emotion predicates uncover that affector, experiencer and stimulus are prominent in emotion predication. She argued

that affector is different from stimulus as the former profiles a higher degree of volitional impact in a more dynamic and eventive manner. Therefore, a three-way distinction can be made for emotion predicates, namely *experiencer-as-subject*, *stimulus-as-subject*, and *affector-as-subject*. The lexical-constructional features of the 3 lexical types are summarized in Table 2.5:

Table 2.5: Lexical-constructional Features with 3 Lexical Types (Liu 2016:39)

	Experiencer-Subj Verbs	Stimulus-Subj Verbs	Affector-Subj Verbs
Lexical meaning	Internal state of the Exp.	Property of the Stimulus	Impact by the Affector
Eventivity	Highly stative	Stative or eventive	Highly eventive
Syntactic pattern	Intransitive/ transitive	Intransitive/ transitive	Transitive only
Lexical status	Lexical	Lexemes/phrasal	Lexical
Morphological make-up	VV, VO, M(anner)-V	VV, VO, MV, Some syntactically derived	V-R(esultative)
Constructional association w/ representative verbs	Stative-Causative alternation 高興 *gaoxing* 'glad' (int) 羨慕 *xianmu* 'envy' (tr)	Stative intransitive Stative/eventive transitive 枯燥 *kuzao* 'dull' (int) 吸引 *xiyin* 'attract' (tr)	Inchoative Bei-passive Ba-construction 激怒 *jinu* 'infuriate'

The proposed three-role scheme laid the groundwork for lexical semantic categorization and cross-linguistic comparison, as languages may vary in subject role selection. She found that Chinese emotion lexicon differs from European

languages in two aspects. First, *stimulus-subject* verbs in Chinese are not as common as they are in English. For instance, there is no Chinese equivalent for some English emotion verbs, such as *please*, *excite*, and *frighten*. Instead of being lexically encoded, *stimulus-subject* predication in Chinese can only be done at the syntactic level by means of causative pattern. Second, the same emotion predicates in Mandarin may be associated with multiple subject roles and grammatical functions, leading to form-meaning mismatches.

2.3.2 Real-world Events

Previous work discussed above classified events using a small number of primitives or features, not many attempts have been made to categorize concrete events in a fine-grained way. In Doddington et al. (2004), the Automatic Content Extraction (ACE) program provides annotated data, evaluation tools, and evaluation exercises for various information extraction tasks which included event detection and characterization. In that task, annotators were asked to identify all event instances, event attributes (i.e. temporal, locative, and others like instrument or purpose), and event arguments (i.e. agent, object, source and target). ACE 2005 (LDC 2009) defined an event as something that happens, and it can frequently be described as a change of state. However, The ACE model does not annotate all kinds of events, instead, only certain kinds of events were taken into consideration. The predefined list of events is shown in Table 2.6.

Table 2.6: ACE Event Classification

Type	1) Life	2) Movement	3) Transaction	4) Business
Subtype(s)	1. Be-Born 2. Marry 3. Divorce 4. Injure 5. Die	1. Transport	1. Transfer-Ownership 2. Transfer-Money	1. Start-Org 2. Merge-Org 3. Declare-Bankruptcy 4. End-Org
Type	5) Conflict	6) Contact	7) Personnel	8) Justice
Subtype(s)	1. Attack 2. Demonstrate	1. Meeting 2. Phone-Write	1. Start-Position 2. End-Position 3. Nominate 4. Elect	1. Arrest-Jail 2. Release-Parole 3. Trial-Hearing 4. Charge-Indict 5. Sue 6. Convict 7. Sentence 8. Fine 9. Execute 10. Extradite 11. Acquit 12. Appeal 13. Pardon

Given that the number of concrete events may be rather large, WordNet (Miller 1995) seems to be a more appropriate resource due to its extensive coverage. WordNet is a lexical database which groups lexical words (i.e. verbs, nouns, adjectives, and adverbs) into sets of synonyms called synsets. Synsets are interlinked according to conceptual-sematic relation and lexical relation. Members of a synset are presented in a hierarchical structure. The dataset contains 155,327 words organized in 175,979 synsets for 207,016 word-sense pairs.

As for the annotation of events in text, Time Markup Language (TimeML) (Sauri et al. 2009) provides a standard guideline for the markup of events in English. TimeML defines events as “situations that happen, occur, hold, or take place”. Events can be punctual or last for a period of time, they can also be states or circumstances in which something holds true. TimeML was developed in 2002 during the Time and Event Recognition for Question Answering Systems Workshop (TERQAS). The TimeML project aims to create a standard markup language for events and temporal expressions in natural language. According to Pustejovsky et al. (2005: 2-3), the TimeML project addresses four issues in events and temporal expression markup:

(a) Time stamping of events (identifying an event and anchoring it in time);
(b) Ordering events with respect to one another (lexical versus discourse properties of ordering);
(c) Reasoning with contextually underspecified temporal expressions (temporal functions such as last week and two weeks before);
(d) Reasoning about the persistence of events (how long does an event or the outcome of an event last).

Event-event relations are of great importance to the problems TimeML aims to address. TimeML classified events into 7 types in TimeML annotation scheme, namely reporting, perception, aspectual, i_action (i.e. intentional action), i_state,

state, and occurrence. The classification is designated mainly for the sake of the annotation of event-event relation, including temporal relation, and subordinate relation, and aspectual relation which may be inadequate in dealing with all kinds of real-world events.

CHAPTER 3

CHINESE EVENT-COMMENT CORPUS

The present work presents a corpus-based study on emotion responses to different events. It deals with both constructions describing explicit emotions and implicit emotions with the major focus being placed on the implicit ones. Instead of illustrating with introspective examples, I construct a Chinese event-comment corpus retrieved from social media. Not only does it provide natural occurring data for a qualitative analysis, it also allows me to conduct a quantitative analysis with a large amount of data as supporting evidence. Section 3.1 presents the data collection. Section 3.2 briefly discusses my research questions and research methodology. Section 3.3 describes the three phases of event annotation task, namely the markup of events, the labelling of frame elements (i.e. semantic roles), and the classification of events. Section 3.4 proposes detailed annotation guidelines for emotion annotation. Section 3.5 introduces the annotation tool and analysis tool used in the current work.

3.1 Data Collection

The data used in this study is taken from Sina Weibo. Sina Weibo is one of the most popular social media sites in the Mainland China. Sina Weibo is launched by Sina Corporation in 2009, with over 465 million monthly active users as of the first quarter of 2019 as reported by Weibo Corporation in the Weibo Reports First

Quarter 2019 Unaudited Financial Results[2]. According to Alexa Internet, an American web traffic analysis company wholly owned subsidiary of Amazon, Sina Weibo ranks 19 in global internet traffic and engagement, and 8 in China on June 25, 2019. The statistics further stress the importance of the site around the globe and highlight the potential influence of Weibo in the community.

Not only does Sina Weibo attract individual users, it also draws the attention of celebrities and different organizations including the media, businesses, government departments, non-government organizations etc. as it provides a platform for users to disseminate and receive all kinds of information, and to instantly respond to events in which they are interested. The comments mostly convey writers' emotions to some kinds of events. Therefore, Sina Weibo provides a good source of data for both quantitative and qualitative studies on emotion.

The Chinese event-comment corpus was made up of 200 trending Weibo posts on news released by digital journalisms from April 2018 to June 2019. After extraction, we removed garbled comments, duplicated comments, comments that users made to respond to another comment, and short comments that contain less than 4 words in Chinese. Each post includes 150 comments, which adds up to 30,000 comments for the entire corpus. Of the 30,000 comments, 10,000 were manually annotated with emotion information. The 10,000 annotated comments consist of 245,651 words including punctuations.

[2] http://ir.weibo.com/news-releases/news-release-details/weibo-reports-first-quarter-2019-unaudited-financial-result

The data preparation task is composed of two phases. In Phase I, a web crawler is developed for the extraction of posts and comments. To ensure that all the extracted posts are trending events, only those with more than 500 comments are selected. 362 posts together with the Uniform Resource Locator (URLs) of each post are manually collected. Both the content of the post and the corresponding comments of each post are retrieved. In order to collect the corresponding comments of each post, the web crawler visited the list of URLs and archived 50 Hypertext Markup Language (HTML) pages for each post. As each HTML page includes 10 comments, 500 comments could be retrieved from 50 pages which were assumed to be enough even after removing garbled comments, duplicated comments, comments that users made to respond to another comment, and short comments that contain less than 10 characters, i.e. less than 4 words in Chinese. As for duplicated comments, one is removed, and one is kept in the dataset. As for comments made to respond to another comment, they all begin with 回覆 'Reply'. All of them are removed as most of them are responses to others' comments but not to events mentioned in a post. An example is given in (1).

(1) 回复: 你说的对，小编打错字了吧!
huifu: ni shuo de dui xiaobian da cuo zi le ba!
reply: 2.SG say right curator type wrong word ASP SFP!
'Reply: You're right. The content curator must have made a typo!'

After removing all invalid comments, only 157 out of 362 posts contain more than 200 comments; 209 posts contain more than 150 comments; 233 posts contain

more than 100 comments. The 209 posts containing more than 150 comments are then extracted for manual check. I read through all the 209 posts and removed 22 posts which contain garbled codes, abstract concepts or hypothetical situations as in (2), and a heading that fails to summarize events mentioned in a post as in (3). This is because a heading of a post will be used for event annotation.

(2) 【冰葬！遗体零下 200 度冷冻后被震成骨灰网友：这就叫“粉身碎骨”】瑞典一位生物学家研究出一种新的埋葬方式 “冰葬”：先将遗体放在特制的仪器中，利用液氮以-200 摄氏度左右低温冷冻人体，晃动人体直到遗体被“震成骨灰”。之后无法被分解的牙齿等组织，都会交由亲友处理。研究者表示这有助于减缓气候变化，并让人体真正回归到大自然。11 月 18 日 22:50

【*bing zang! yi ti ling xia 200 du leng dong hou bei zhen cheng gu hui wang you: zhe jiu jiao “fen shen sui gu”*】

rui dian yi wei sheng wu xue jia yan jiu chu yi zhong xin de mai zang fang shi “bing zang”: xian jiang yi ti fang zai te zhi de yi qi zhong, li yong ye dan yi -200 she shi du zuo you di wen leng dong ren ti, huang dong ren ti zhi dao yi ti bei “zhen cheng gu hui”. zhi hou wu fa bei fen jie de ya chi deng zu zhi, dou hui jiao you qin you chu li. yan jiu zhe biao shi zhe you zhu yu jian huan qi hou bian hua, bing rang ren ti zhen zheng hui gui dao da zi ran. 11 yue 18 ri 22:50

【ice bury! corpse sub-zero 200 degree freeze PAS shaken bone ash netizen: DET is call “powder body break bone”】

Sweden one CL biologist discover one CL new bury method “ice bury”: first put remain put in tailor-made equipment inside, use liquid nitrogen with -200 Celsius degree around low temperature freeze human body, shake human body until remain PAS “shatter become bone ash”. Then cannot PAS decompose tooth etc. tissue, all give relatives handle. researcher point out DET beneficial to relieve climate change, and let human body truly go back to mother nature. November 18th 22:50

‘[Ice burial! Corpses are frozen in minus 200 degrees to be vibrated to dust. Netizens, “This is called ‘shattered to ashes’”] A Swedish biologist has found out a new way of burial. To perform “Ice burial”, the remain has to be kept in a specially made device. Using liquid nitrogen, the remain would be frozen in the extreme low temperature of minus 200 degrees Celsius.

Through vibration, the remain would be turn into dust. The remaining parts (such as teeth) that cannot be broken down will be returned to the relatives of the deceased. The researcher pointed out that this is beneficial to relieving climate change and it also lets the deceased truly embrace mother nature.'

(3) 【丈夫被判赔 19 万元并刑拘】

11 月 14 日，广西一公园内，一名黑衣男子手拿棍棒反复击打躺倒在地怀抱幼童的妇女。据了解，该女子正是他的妻子，目前已被刑拘 10 日。南宁卢某曾因怀疑妻子出轨，将她打得双耳失聪，被判处赔付 19 万。???11 月 19 日 19:04

【*zhang fu*	*bei*	*pan*	*pei*	*19 wan yuan*	*bing xing ju*】
11 yue	*14 ri,*	*guang xi*	*yi gong yuan nei,*	*yi ming*	
hei yi nan zi	*shou*	*na*	*gun*	*bang*	*fan fu*
ji da	*tang dao*	*zai di*	*huai bao*	*you tong*	*de*
fu nü.	*ju liao jie,*	*gai*	*nü zi*	*zheng shi*	*ta de*
qi zi,	*mu qian*	*yi*	*bei*	*xing ju*	*shi ri.*
Nan Ning	*lu mou*	*ceng*	*yin*	*huai yi qi zi*	*chu gui,*
jiang	*ta*	*da de*	*shuang er*	*shi cong,*	*bei*
pan chu	*pei fu*	*19 wan.*	*11 yue*	*19 ri*	*13:39*
【husband	PAS	sentence	compensate	190k dollars	and detained】
November	14th,	Guangxi	one park	in,	one CL
man wearing	black	hand	grab	bat	repeatedly
hit	lie	on floor	cuddle	young child	POS
woman.	Reportedly,	DET	woman	exactly	3.SG-POSS
wife,	currently	already	PAS	detained	10 days.
Nan Ning	Lu	once	because	suspect wife	cheat,
put	3.SG	beat-DE	two ear	deaf,	PAS
sentence	compensate	190k.	November	19th	13:39

'[Husband sentenced to a fine of 190k yuan in detainment] On 14th November in Guangxi, a man wearing in black was repeatedly beating a woman cuddling a young child on the ground in a park. Reportedly, the woman was the wife of the man. Up to this point, the husband has been detained for 10 days. Nan Ninglu has once beaten up his wife since he suspected that his wife was cheating on him. His wife lost hearing in both ears as a result of the violence. The husband was sentenced to a fine of 190k yuan. 13:39, 19 November'

After removing all the invalid posts, only 187 posts contain more than 150 comments, and they are all collected for further process. Since the corpus should be composed of 200 posts, the second phase of extraction has been conducted. Following the same procedures in Phase I, another batch of data is collected in Phase II.

In Phase II, 20 posts together with their comments are collected following the same procedures in Phase I. The only difference lies in the number of pages archived for each post. Given that a large number of comments have been removed in Phase I, 150 HTML pages are archived for each post in Phase II to ensure that all the posts retrieved contain more than 150 comments even after data removal. 13 posts are randomly extracted for manual check to confirm all the posts are eligible for event annotation.

3.2 Research Questions and Research Methodology

With the event-comment corpus, I will be able to investigate the interplay between the event types and the implicit emotion expressions and address the two major research questions raised in the proposal.

a. How are implicit emotions expressed in text?
b. What kinds of events trigger different emotions?

To answer (a), I will first draw a clear distinction between explicit and implicit

emotions. Explicit emotions can only be expressed at word level. Therefore, I will briefly discuss how explicit emotion can be expressed by means of emotion keywords and emojis. Based on the distinction made, I will then focus on the implicit emotions and investigate various linguistic characteristics and linguistic cues of implicit emotions in terms of semantic and syntactic structures. Semantically, I will examine two types of words, namely emotion-related words and emotion-laden words. According to Pavlenko (2008), emotion-related words are defined as behaviours related to particular emotions, that is, the post-events of an emotion. Emotion-laden words do not describe the emotion state directly but express or elicit emotions from the interlocutor without using emotion keywords.

Syntactically, implicit emotions can be studied at phrase level and discourse level. At phrase level, I will explore the characteristics of implicit emotions in terms of syntactic structures. For example, the pattern "萬一/要是 '(what) if'……" is quite often found in comments, that imply a FEAR emotion. I will investigate whether some frequently occurring patterns can be linked to a particular emotion. Apart from that, my preliminary observation shows that emotions are quite often expressed using rhetorical questions. Rhetorical question refers to utterance that has the structure of a question but does not expect an answer. I will examine the relationship between emotions and different types of rhetorical questions including both open class questions and close class questions. Various syntactic structures of rhetorical questions that can be used to identify different emotions will also be proposed. At discourse level, although some emojis may serve as an emotion

indicator of an emotion, emojis are sometimes used ironically. That is, the emotion expressed via text is different from the one expressed via an emoji. When the two channels are at odds, it may be the emoji that determines the overall emotion or the text that determines the expressed emotion. Therefore, I will study the interaction between text and emojis.

To answer (b), I will study the emotion cause that triggers an emotion. My preliminary observation shows that the provoked emotions often relate to the event type; for example, events concerning "death" often provoke SADNESS. With the aid of the annotated corpus, I will be able to figure out what kinds of events are strongly related to a particular emotion and are statistically significant at the same time. In addition, I will also examine the interplay of emotion, event and semantic role.

3.3 Event Annotation

3.3.1 Event Markup – TimeML

Three language resources are used in the event annotation task. The TimeML annotation guidelines (Sauri et al. 2009) is adapted for the event annotation for two reasons. The first reason is because TimeML provides a standard guideline for the markup of events in English; the second reason is because the use of TimeML is highly motivated theoretically. TimeML is generated by the Generative Lexicon (GL) (Pustejovsky 1991) which is a theory of lexical semantics that focuses on the distributed nature of semantic compositionality. It emphasizes that word meanings

are not listed in the lexicon but generated in real linguistic usage by context. According to the GL theory (Pustejovsky 1995), a lexicon item is composed of four levels, namely lexical typing structure, argument structure, event structure and qualia structure. Although the GL theory is not designed for emotion analysis, Lee (2010) extended the event structure to the property of phrases and sentences to study the interactions between events and emotions. Lee (2010) regarded emotion state as a pivot event that links to two sequentially structure events, i.e. the cause event (pre-event) and the elicited event (post-event). Given that Lee's framework is adopted in the current study, using TimeML for event annotation is theoretically supported and highly motivated.

TimeML was conceptualized in 2002 during the Time and Event Recognition for Question Answering Systems Workshop (TERQAS). It aims to enhance the performance of natural language question answer systems and answer temporally-based questions about events and entities in news articles. The goal of TimeML is to create a markup language for temporal and event expressions, including time stamping of events, ordering events with respect to another, reasoning with contextually underspecified temporal expressions, and reasoning about the persistence of events. Event-event relations are of great importance to the problems TimeML aims to address. TimeML classified events into 7 types in TimeML annotation scheme, namely *reporting*, *perception*, *aspectual*, *i_action* (i.e. intentional action), *i_state, state*, and *occurrence*. The classification is designated for the sake of the annotation of event-event relation, including temporal relation,

and subordinate relation, and aspectual relation. However, this is not the major focus of the present work. The event classes seem to be inadequate for the study of event types and emotions as most events, even of different polarities, would fall into the category of *occurrence*. Therefore, I will only follow the TimeML annotation guidelines for the markup of events, but not the classification of events.

Following the TimeML annotation guidelines (Saurí et al. 2009: 3), events are "situations that happen, occur, hold, or take place". Events can be punctual or last for a period of time, they can also be states or circumstances in which something holds true. For the event annotation, only the headings of the 200 posts are annotated as a heading generally summarizes the focus of a post. Given that one single sentence often consists of several events, the event annotation task is done at the word level. As proposed in the TimeML annotation guidelines, events can be denoted by verbs, nouns, adjectives, prepositional phrases, or other elements such as locative adverbs. However, the guidelines are compiled based on English data. It is observed that events in Chinese may be encoded in a different way in terms of parts-of-speech. Similar to English, verbs, nouns, and adjectives are quite commonly used to denote events. However, rarely does a prepositional phrase or an adverb in Chinese denote an event. Therefore, I only consider events denoted by verbs, nouns, and adjectives.

3.3.1.1 Events denoted by Verbs

As for events denoted by verbs, all verbal predicates excluding the copula verb 是

'to be' denote an event, and they are marked up as such. According to Vendler (1967), verbs can be classified into four types, namely activity verbs, stative verbs, achievement verbs, and accomplishment verbs. An example of each type is exemplified as in (4) - (7). Note that not all the events tagged in the corpus are shown as only one type of verb is emphasized in each example. Therefore, only one of the verbal events is underlined in each example.

(4) 【小伙无故踹打八旬老太，警方：提请逮捕】

【*xiao huo*		*wu gu*	*chuai da*	*ba xun lao tai,*
jing fang:	*ti qing*	*dai bu*】		
【Young man		no reason	beat up	80-year-old old lady,
police:	filing	arrest】		

'[A lad beat up an 80-year-old lady without any reasons. Police: we're filing for an approval for arrest]'

(5) 【大巴上有杀人嫌犯司机拖时间等来警察】

【*da ba*	*shang*	*you*	*sha ren xian fan*	*si ji*
tuo	*shi jian*	*deng lai*	*jing cha*】	
【bus	on	have	murder suspect	driver
delay	time	wait for	police】	

'[A murder suspect got on a bus. The driver was stalling to buy time for the police to come]'

(6) 【编剧都不敢这么编!女子刮蹭小车后留纸条,车主竟打电话来感谢:车丢一个多月终于找着了】

【*bian ju* *dou bu gan* *zhe me bian!*
nu zi *gua ceng* *xiao che* *hou* *liu zhi tiao,*
che zhu *jing* *da dian hua* *lai gan xie:*
che *diu* *yi ge duo yue* *zhong yu* *zhao zhe le*】
【screenplay writer cannot even that make up!
woman scratch car after leave note,
car owner surprisingly dial telephone to thank:
car lost over a month finally find-ASP】
'[Not even a screenplay writer could have made up a story like that! A woman scratched a car and left a note. Surprisingly, the car owner called her to thank her, saying, "the car was lost for over a month. Now I finally got it back"]'

(7) 【双 11 后快递遭暴力分拣:有的被乱踢、扔飞,有的被踩碎】

【*shuang 11 hou* *kuai di* *zao* *bao li* *fen jian:*
you de *bei* *luan* *ti,* *reng* *fei,*
you de *bei* *cai sui*】
【Double 11 after express delivery PAS violently sort:
some PAS mess kick, throw fly,
some PAS step crumble】
'[After Double 11, packages are sorted in an immensely chaotic order: some packages were kicked and hurled; some stormed by foot and smashed]'

In the present study, the four types of verbs are considered as one of the ways to denote verbal events. The verbal event in (4) is denoted by an activity verb 踹打 'beat up' which stands for an event that goes on for a period of time without a clear terminal point. 踹打 'beat up' is a parallel verb compound which is constituted by two verbs that are synonymous or signal the same type of predicative notions (Li and Thompson, 1981). The event in (5) is denoted by a stative verb 有 'have' which describes a non-dynamic situation that holds for some time. As for the event in (6), it is denoted by an achievement verb 丢 'lost' which depicts the event takes

place in a single moment. (7) shows an example of an event being denoted by an accomplishment verb 踩碎 'smashed by foot'. An accomplishment verb indicates events that have an inherent endpoint which is realized in the second predicate in a resultative verb compound.

In addition to verbal compounds including parallel verb compounds and resultative verb compounds, other compound verbs are also marked up as a single event. These compound verbs include verb-object compounds such as 打电话 'make a phone call', 上厕所 'go to the toilet' etc., adverb-verb compounds such as 怒摔 'throw angrily', 苦劝 'convince earnestly' etc., and noun-verb compounds such as 刑拘 'detention', 拳打 'punch' etc. For verb-object compounds, they are usually marked as a single event due to the polysemous nature the verbs have. As verbs such as 打 can combine with different nouns to form different types of events as in 打手电筒 'shine a torch into', 打牌 'play cards', 打表 'run a meter' and so on, marking up the entire unit is to facilitate the classification of event they denoted as the meaning of the construction cannot be derived from the meaning of the constituents. Additionally, verb-object compounds allow aspect markers, measure phrases, or other modifiers of the object constituent to intervene between the verb constituent and the object constituent, as in 上不了火车 'unable to get on the train', 连上 13 小时班 'non-stop working for 13 hours'. Verb-object compounds also allow the object constituent to precede the verb constituent, as in 晨会开一半 'during the morning meeting'.

3.3.1.2 Events denoted by Nouns

Following TimeML annotation guidelines (Sauri et al. 2009) which consider copulative predicate 'to be' as an element to be marked up, the copula verb 是 'to be' is also regarded as a verb to be marked up under certain circumstances in the present study. The markup of the copula verb 是 'to be' depends on the property of noun that follows it. Nominal events can be denoted by event-denoting nouns or sortal states (Sauri et al. 2009). The copula verb 是 'to be' is often used to introduce a predicative complement; however, not all kinds of predicative complements but only those expressing a sortal state should be marked up as an event. As proposed by Sauri et al. (2009), sortal states are generally expressed by agentive nominals who participate in certain activities or actions. It can also be expressed by nouns that denote professions, roles or positions, or terms that refer to the same entity across the world. Therefore, the copula verb 是 'to be' is marked up only if the noun following it meets the requirements of a sortal state. Consider (8) and (9).

(8) 【男子持刀行凶他第一个站出来…他，曾是军人！】

【*nan zi*	*chi*	*dao*	*xing xiong*	*ta*	*di yi ge*
zhan	*chu lai…*	*ta*	*ceng*	*shi jun ren*！】	
【man	hold	knife	commit crime	3SG	first-CL
stand	out…	3.SG	once	is	soldier！】

'[A man held a knife committing an assault. He was to first to stand up. He was a soldier!]'

(9) 【这名交警执法硬气，网友怒赞：他不就是“李云龙”嘛！】

【*zhe* *ming* *jiao jing* *zhi fa* *ying qi*
wang you *nu* *zan:* *ta* *bu jiu* *shi* *Li Yunlong*
ma! 】

【DET CL traffic police enforce law bold
Netizen angry praise: 3.SG not is Li Yunglong
SFP! 】

‘[This traffic police officer boldly enforcing the law. Netizens offered huge praise, saying, “Isn’t he Li Yunlong?”]’

The copula verb 是 ‘to be’ in (8) is marked up as it is followed by a sortal state expressed by the role 军人 ‘soldier’. However, the copula verb is not marked up in (9) as the agentive nominal 李云龙 ‘Li Yunlong’ does not indicate any activities or actions that the traffic officer participated in.

As for event-denoting nouns, Sauri et al. (2009) proposed four rules and claimed that a noun should be compatible with at least two of them to be regarded as an event-denoting one. Since the rules are compiled based on English data, I revise the rules as follows:

(a) NOUN 持續了數秒/分鐘/日/年/……

‘NOUN lasted for several seconds/ minutes/ days/ years/ …’

(b) NOUN (將)在 TEMPORAL EXPRESSION 發生/舉行

‘NOUN took/ takes/ will take place in TEMPORAL EXPRESSION’

(c) NOUN 在 TEMPORAL EXPRESSION 開始/持續/結束

‘NOUN began/ continued/ ended in TEMPORAL EXPRESSION’

If a noun satisfies at least one of the three conditions, it is regarded as an event-

denoting noun in the present study. These rules prove that some nouns do coerce an activity or action even without literally implying that. For example, the noun 會議 'meeting' is compatible with all the above-mentioned rules, whereas the noun 電腦 'computer' is incompatible with any of them. The reason for that is because 會議 'meeting' coerces the meaning of having a meeting which is an activity, whereas the noun 電腦 'computer' does not have such a coercion. Some other examples found in the corpus include 暴雨 'rainstorm', 家暴 'domestic abuse', 亂象 'chaos', 事故 'accident' and so on.

Event-denoting nouns acting as prenominal modifiers are not marked up as an event. Consider (10).

(10) 日本向中国提新大熊猫租借请求

ri ben	*xiang*	*zhong guo*	*ti*	*xin*	*da*	*xiong*	*mao*
zu jie	*qing qiu*						
Japan	towards	China	raise	new	big	panda	
rental	request						

'Japan made a proposal to China regarding renting a new panda.'

In (10), 租借 'rent' is a noun that can be regarded as an event-denoting noun. However, it acts as a prenominal modifier which modifies another noun 請求 'request'. In that case, the prenominal modifier should never be annotated as an event, and only the noun should be.

3.3.1.3 Events denoted by Adjectives

As for events denoted by adjectives, they usually denote a stative event. An

adjective can be classified as an attributive one or a predicative one. The former acts as a pre-modifier of a noun, whereas the latter functions as a predicative complement of a verb. Consider the phrases in (11) and (12).

(11) 最安静的守护

zui	*an jing de*	*shou hu*
most	quiet-POSS	protection

'The quietest protection.'

(12) 外卖小哥满脸委屈

wai mai	*xiao ge*	*man*	*nian*	*wei qu*
delivery	lad	full	face	wronged

'The delivery guy looked as if he was wronged.'

In (11), the adjective 安静 'quiet' is in an attributive position which will never be annotated as an event, while the adjective 委屈 'grievance' is in a predicative position and will be marked up as an event. According to TimeML, only predicative adjectives which denote a non-persistent property of the noun they modified should be annotated as an event, i.e., the change of state. For example, the deliveryman in (12) changes its emotion state from not feeling aggrieved to feeling aggrieved. Sauri et al. (2009) suggested that for an adjective to be annotated as an event, it should satisfy at least one of the following conditions: (1) the adjective should denote a non-persistent property of the noun it modifies, (2) the adjective should be a state that is temporally bound to a particular point or a period of time, (3) the adjective should indicate an opinion, knowledge, someone's belief, or a matter under discussion. Therefore, I follow the TimeML annotation guidelines and only mark up those adjectives that satisfy at least one of the three conditions.

3.3.1.4 Exceptional Cases

As mentioned in the previous sections, events can be denoted by verbs, nouns and adjectives in Chinese. If they satisfy the above-mentioned conditions, all of them are marked up as an event. However, there are cases that an eligible word may not be marked up, such as subjective evaluations written by the writer of the post, or informative questions. Consider (13) and (14).

(13) 【女生太多！高校征用男厕改女厕 男生：能体谅】

【*nü sheng*	*tai*	*duo!*	*gao xiao*	*zheng yong*	*nan ce*
gai	*nü ce*		*nan sheng:*	*neng*	*ti liang*】
【girl	too	many!	high school	requisition	male toilet
change	female toilet		male:	can	understand】

'Too many girls! A high school is requisitioning male bathrooms to make room for more female bathrooms. Male students commented, "we can understand that".'

(14) 【暖心！民工路边晕倒，路过女孩上前施救还抹泪：担心自己没做好】

【*nuan xin!*		*min gong*	*lu bian*	*yun dao,*	*lu guo*		
nü hai	*shang*	*qian*	*shi qiu*	*hai mo*	*lei:*	*dan*	*xin*
zi ji	*mei*	*zuo*	*hao*】				
【Heart-warming!		labour	street side	faint,		bypass	
girl	up	front	rescue	also wipe	tear:	worry	
self	not	doing	well】				

'Heart-warming news! A worker passed out on the street. The by-passing girl came to rescue. She was wiping her teardrops, stating that she feared she didn't do well enough.'

It is observed that some posts contain an exclamatory sentence in the heading, which is either a subjective or objective evaluation. For instances, the first sentence 女生太多！'too many girls!' in (13) is an objective evaluation which describes the state of the school having too many girls. It is a stative event which causes the

following events 征用 'requisition' and 改 'change'. However, 暖心 'heart-warming' in (14) is the writer's own evaluation. In this thesis, I only mark the objective ones as events but not the subjective ones. This is because the former provides further information to readers, but the latter does not. Therefore, the adjective 多 'many' in (13) is annotated but the one 暖心 'heart-warming' in (14) is not.

Moreover, a question in a heading should be dealt with more carefully, be it in a main clause or in an embedded one. Some questions are raised in the heading as a thought-provoking question as in (15), some function as an information-seeking question which has yet to be revealed as in (16), and some are rhetorical questions which do not aim to elicit an answer but to make a statement as in (17).

(15) 【女子写淫秽小说卖钱被判入狱 10 年，对还是错？】

【*nü zi* *xie* *yin hui* *xiao shuo* *mai qian* *bei* *pan* *ru yu* *10 nian,* *dui hai shi cuo?*】

【woman write obscene novel sell PASS sentence prison 10 year, right or wrong? 】

'[A woman was sentenced a 10-year imprisonment for writing and selling obscene novels: is this right or wrong?]'

(16) 【重庆高空项目安全绳突然脱落官方：正调查是营销还是疏漏】

【*Chongqin* *gao kong* *xiang mu* *an quan sheng* *tu ran* *tuo luo* *guan fang: zheng* *diao cha* *shi ying xiao hai shi* *shu lou*】

【Chongqin high attitude project safety rope suddenly fall off official: now investigate is marketing or oversight】

'[Safety rope of a high attitude facility in Chongqin broke off without warning. Official is investigating the incident. Is this a marketing stunt of an oversight?]'

(17) 【中学生午休上厕所违反校规被处分？教育局：由纪检牵头调查核实】
【*zhong xue sheng wu xiu shang ce suo wei fan xiao gui bei chu fen? jiao yu ju: you ji jian qian tou diao cha he shi*】
【Secondary school student afternoon break go toilet break school rule PAS punish? Education Bureau: from disciplinary forces lead investigate confirm】
'[Were Secondary school students punished for going to the bathroom during afternoon break? The Education Bureau stated that the disciplinary forces were leading the investigation to validate the incident]'

In (15) - (17), the questions are in bold face. In (15) and (16), the events are not marked up even if they satisfy the conditions to be an event-denoting word. In (17), the event-denoting words are all marked up as they do indicate the subevents that happened, namely 午休 'afternoon break', 上厕所 'go to the toilet', 违反 'violate', and 处分 'punish'.

3.3.2 Event Classification – WordNet

After marking up all the events, the events are then translated from Chinese to English in order to map to categories proposed by WordNet (Miller 1995). WordNet is an English lexical database which groups words into a set of synonyms named synsets. Therefore, the mapping can be done by searching for an appropriate category using the word translated from Chinese to English or using its synonyms. It is observed that some of the mappings contain more than one potential category due to the polysemous nature a Chinese word may have. For example, the Chinese word 丢 can be understood as the act of "throwing" or "losing someone or

something". Moreover, the classification in WordNet is rather fine-grained. There may be several categories representing similar concepts with just a slightly difference in meaning. Thus, the mappings should be done manually. For example, by searching the word "throw", there are already five items containing the word "throw" in the name of the categories as shown in Table 3.1.

Table 3.1: WordNet Categories Pertaining to "Throw"

ID	Name	Definition
105359	(throw)	the act of throwing (propelling something with a rapid movement of the arm and wrist); "the catcher made a good throw to second base"
1248165	(throw)	casting an object in order to determine an outcome randomly; "he risked his fortune on a throw of the dice"
92054	(discard, throwing_away)	getting rid something that is regarded as useless or undesirable
7365795	(throw, stroke, cam_stroke)	the maximum movement available to a pivoted or reciprocating piece by a cam
14509574	(throw)	a single chance or instance; "he couldn't afford $50 a throw"

In view of this, the mappings are manually done case by case. Take the lexical item 守 as another example, it can be interpreted as "guarding", "obeying" or "waiting" which depends on the context. Consider (18) and (19).

(18) 【泪目！主人被撞身亡小狗原地守 80 天盼主人归】
【*lei mu!* *zhu ren* *bei* *zhuang* *shen wang*
xiao gou *yuan di* *shou* *80 tian* *pan* *zhu ren* *gui*】
【tear in eye! master PAS hit die
puppy in place wait 80 day hope master return】
'[It's heart-breaking! A pet owner was hit and killed. The puppy waited at one spot for 80 days, waiting for the return of its master]'

(19) 【山东寿光一超市老板不收消防官兵钱，子弟兵守纪律返回付款】
【*Shandong Shouguang yi chao shi lao ban bu shou xiao fang guang bin qian, zi di bing shou ji lü fan hui fu kuan*】
【Shandong Shouguang one supermarket owner not collecting fireman money, trooper obey regulation back pay】
'[The owner of a supermarket in Shouguang, Shandong decided not to collect any money from the firemen (for their purchase). But the civil servants obeyed the regulations and went back (to the supermarket) to make the payment]'

The verb 守 in (18) refers to the meaning of "waiting", and the one in (19) refers to "obeying". Thus, they should be classified into different types of events.

The total number of events marked up in the 200 posts is 732. Of the 732 events, some belong to the same category in WordNet, and thus the total number of WordNet categories (i.e. event types) identified is 504. The conditional probability, the count and the entropy of each event type will be computed for further analysis on the correlation between emotions and events. Details will be further discussed in Section 6.1.

3.3.3 Annotation of Frame Elements – FrameNet

After event classification, FrameNet (Baker et al. 1998) is employed to label the semantic roles of the arguments or adjuncts mentioned in the event. FrameNet is an electronic lexical database developed based on the theory of frame semantics proposed by Fillmore (1982) who was the project leader of FrameNet in 1997.

FrameNet is considered a useful resource. It contains over 13,000 word senses annotated with examples showing the meaning and usage of a lexical item. It also contains more than 1,200 semantic frames which provide computational linguists a set of training data for sematic role labelling. The FrameNet project began in 1997 and has been influential in the field of both linguistics and natural language processing. It is often adopted in studies on automatic semantic role labelling. In view of this, FrameNet is adopted in the present study to label the semantic roles of the arguments or adjuncts mentioned in the event, which is named as frame elements in FrameNet. That is, the basic unit of a frame is composed of frame elements which are frame-specific defined semantic roles of an event. In the following sections, the terms "semantic role" and "frame element" are used interchangeably.

To label semantic roles of an event with the aid of FrameNet, I first translated all the annotated events into English and did a search for the most suitable frame for each event through FrameNet Search[3]. For example, the event 出生 'born' is translated as 'born'. I used the word 'born' to search for a frame that represents the event 出生 'born'. The result is shown in Figure 3.1.

[3] https://framenet.icsi.berkeley.edu/fndrupal/framenet_search

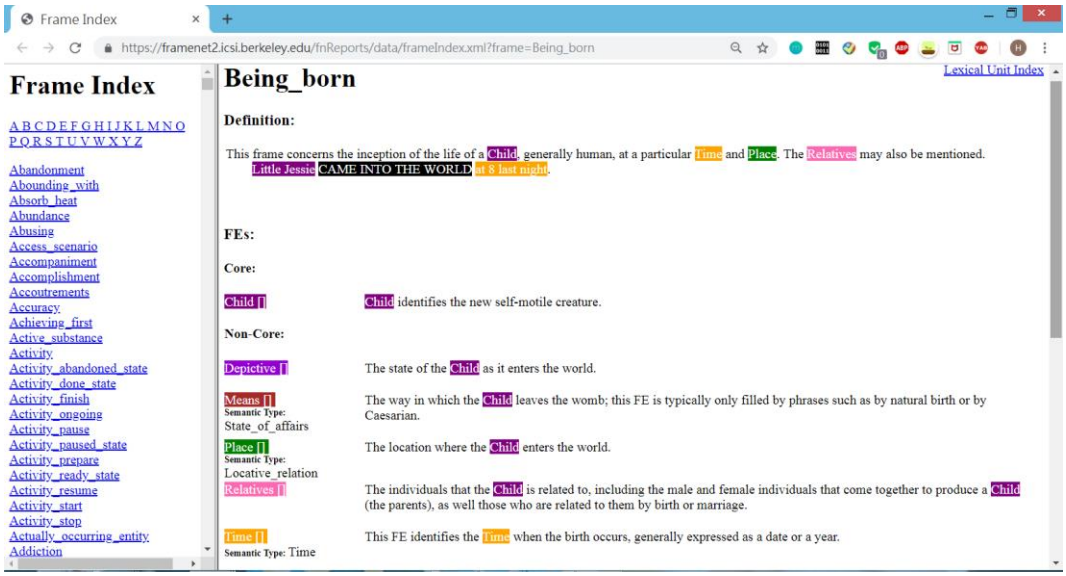


Figure 3.1: The Frame “Being_born” in FrameNet

Figure 3.1 shows that FrameNet provides the definition of a frame and different frame elements that involved in that specific frame. These elements may include participants, props, conceptual roles etc, and they are mainly classified into core and non-core frame elements. Core frame element refers to elements that are important to the meaning of a frame, such as the *child* in the event of ‘being_born’. Non-core frame element refers to some peripheral elements, such as *time*, *place*, *means* etc. In the present study, all the core frame elements are annotated if they can be found in either the heading or the content of a post. As for non-core elements, only those that appear in the heading are annotated. It is believed that a peripheral element may be of great importance to the event if it does appear in the heading. An example is exemplified in Figure 3.2.

<Text>【佛门净地 [e1-出生] 108 个“罗汉娃”，住持说：做好事不要怕别人说】十年前，汶川地震发生时，四川什邡市妇幼保健院成危房，经当地政府协调，大批临产孕妇被转移至隔壁古刹罗汉寺，在这里，108 个娃诞生了。十年后，108 个“罗汉娃”重聚在罗汉寺，共同庆祝同一个“生日”。???05 月 06 日 09:37

(1)- Being_born_e1: 108 个“罗汉娃”: Child
(2)- e1: 四川什邡市古刹罗汉寺: Place
</Text>

Figure 3.2: An example of Semantic Role Labelling

Figure 3.1 shows that the frame “being_born” is composed of one core element *child*, and five non-core elements, namely *depictive*, *means*, *place*, *relatives*, and *time*. Figure 3.2 shows that the core element ‘child’ is tagged, and ‘place’ is the only non-core element being tagged. This is because the place 佛门净地 ‘Buddhist temple’ (i.e. 四川什邡市古刹罗汉寺 ‘Sichuan Shifang Luohan Temple’) is the only non-core elements mentioned in the heading.

Regarding to the annotation of events for semantic role labelling, some points must be clarified. First, an event may be mentioned more than once in a heading, only one of them should be annotated with semantic roles. Consider (20) - (22).

(20) 【国庆首日稻城亚丁堵车从白天堵到黑夜】
【*guo qing* *shou ri* *Daochengyading* *du che*
cong *bai tian* *du* *dao* *hei ye*】
【Nation Day celebration first day Yading congestion
from day congest until night】
‘[On the first day of National Day celebration, the roads in Yanding were congested the entire day]’

(21) 【中学生午休上厕所违反校规被处分？教育局：由纪检牵头调查核实】
【*zhong xue sheng wu xiu shang ce suo wei fan xiao gui bei chu fen? jiao yu ju: you ji jian qian tou diao cha he shi*】
【Secondary school student afternoon break go toilet break school rule PAS punish? Education Bureau: from disciplinary forces lead investigate validate】
'[Were Secondary school students punished for going to the bathroom during afternoon break? The Education Bureau stated that the disciplinary forces were leading the investigation to validate the incident]'

(22) 【环卫阿姨救下残疾小狗，5 年后小狗又救了阿姨】
【*huan wei a yi jiu xia can ji xiao gou, 5 nian hou xiao hou you jiu le a yi*】
【Huanwei auntie save-ASP disable puppy, 5 year later puppy too save-ASP auntie】
'[A female sanitation worker saved a disabled puppy. Five years later, the puppy saved her in return]'

In (20), the two events 堵車 'congestion' and 堵 'congest' refer the same event. The first half of the heading focuses on the event of congestion, while the second half focuses on the duration of the congestion. The two events should be regarded as a single event in which the same set of semantic roles is involved. Example (17) is repeated here in (21). Although the meanings of the two verbs 调查 'investigate' and 核实 'validate' are not exactly the same, they are highly related to each other in semantics. The purpose of an investigation is to validate the event. Moreover, they share the same set of semantic roles (i.e. 纪检 'disciplinary forces' is the agent of both 调查 'investigate' and 核实 'validate', and the event of 中学生午休上厕所违反校规被处分 'secondary school students being punished for going to the bathroom during afternoon break' is what to be investigated and validated.

Therefore, they are regarded as a single event in the annotation of frame elements.

As for (22), although 救下 'save' and 救 'save' are seemingly the same event, they should be regarded as two events. This is because the two events do not refer to the same event happening at the same time as indicated by 5 年后 '5 years later'. Besides, the two events do not share the same set of semantic roles. The sanitation worker is the agent in the first rescuing event and the patient in the second rescuing event. Therefore, while the events of 'congestion' in (20) should be regarded as a single event, the events of 'rescue' in (21) should not.

Second, some verb-object constructions should also be marked as a single event as they refer to the same event and share the same set of semantic roles as in (23) and (24). Example (21) is repeated here in (24).

(23) 【前妻办完离婚就随他人离开前夫当街开车冲撞打砸】
【*qian qi ban wan li hun jiu sui ta ren li kai qian fu dang jie kai che chong zhuang da za*】
【ex-wife handle complete divorce then follow others leave ex-husband in the public drive hit smash 】
'[Ex-wife left the husband and was meeting someone else soon after the divorce. The husband made a scene by crashing things as he drove]'

(24) 【中学生午休上厕所违反校规被处分？教育局：由纪检牵头调查核实】
【*zhong xue sheng wu xiu shang ce suo wei fan xiao gui bei chu fen? jiao yu ju: you ji jian qian tou diao cha he shi*】
【Secondary school student afternoon break go toilet break school rule PAS punish? Education Bureau: from disciplinary forces lead investigate validate】
'[Were Secondary school students punished for going to the bathroom during afternoon break? The Education Bureau stated that the disciplinary forces were leading the investigation to validate the incident]'

Although 办 ‘handle’ and 离婚 ‘divorce’ in (23) are marked up as two events following the TimeML annotation guidelines, they should be regarded as one single event for semantic role labelling. This is because the two events are referring to the same event (i.e. getting a divorce), and there are no points labelling the semantic roles separately. Unlike the event-denoting object 离婚 ‘divorce’ in (23), the object 厕所 ‘toilet’ in (24) is not an event-denoting noun. However, it is also marked up since the verb 上 ‘go’ is polysemous. It can be linked with various objects, such as 上班 ‘go to work’, 上學 ‘go to school’, 上車 ‘get on the car’ etc. Therefore, the entire construction is marked up as an event in order to understand what the event is.

All the 200 posts in the corpus are labelled with semantic roles following the abovementioned guidelines. A number is assigned to each frame element and the numbers are to be used in the annotation of opinion target. Details will be discussed in Section 3.4.4.

3.4 Emotion Annotation

After marking up events in the 200 posts, a maximum of 10,000 comments (i.e. 50 comments of each post) will be annotated with emotion information. Emotion information includes emotion types, the use of rhetorical questions, emotion expressions, emotion causes (i.e. pre-events), emotion reactions (i.e. post-events), and opinion targets (i.e. semantic roles that trigger a particular emotion). Emotion

types, the use of rhetorical questions and opinion targets are annotated by two annotators. When annotator 1 and annotator 2 disagree with each other on the annotation, annotator 3 would make the final decision.

3.4.1 Emotion Type

For the annotation of expressed emotions, five basic emotions are classified, namely HAPPINESS, SADNESS, ANGER, FEAR and SURPRISE. Regardless of the emotion types, emotions can be expressed either in an explicit or implicit way. Explicit emotions are expressed by means of emotion keywords such as 開心 'happy', and implicit emotions refers to the emotion-related information conveyed through inference or connotation without any emotion keywords. The annotation of explicit emotions will be done automatically based on the Chinese emotion taxonomy proposed in Lee (2010). The taxonomy is a robust and versatile emotion annotation scheme based on cognitive emotion theories. It lists a group of emotion keywords of each emotion type as shown in Table 3.2.

Table 3.2: Chinese Emotion Taxonomy (Lee 2010)

Primary Emotions	Variations in Intensity			First-Order Emotions	Second-Order Emotions
	High	Moderate	Low		
HAPPINESS/ 喜-*xi3*	痛快，振奮，亢奮，興奮	欣慰，高興，愉悅，欣喜，歡欣，樂，歡暢，開心，康樂，歡快，快慰，歡，舒暢，快樂，快活，歡樂，暢快，舒心，舒坦，歡娛，如意，喜悅，順心，歡悅，爽心，曉暢，鬆快，歡愉，歡喜	閒適，怡和，放鬆，自在	+FEAR: Pride/傲：自傲，驕橫，驕慢，驕矜，驕傲，自負，自信，自豪，自滿，自大，自狂，狂，炫耀，得意 Relief/安心：安心，寬心，放心 +ANGER: Appeased/解狠：解狠 +SADNESS: Moved/感動：感動	
SADNESS/ 哀-*ai1*	悲慟，悲痛，哀傷，悲哀，哀痛，沉痛，痛心，悲涼，悲淒，悲切，悲傷，悲愴，哀戚，絕望	感傷，傷心，傷感，心酸, 沉悶，憋氣，鬱悒，苦悶，無聊，鬱悶，乏味，沉鬱，憋悶，憂鬱，陰鬱，悵惘, 沮喪，消沉，頹喪，頹唐，煩悶	灰心，喪氣	+FEAR: Misery/悲愁：悲愁，哀愁，愁悶，惆悵 Remorseful/後悔：後悔，慚愧，抱歉，抱愧，對不起，羞愧，背悔，懊惱，懊悔，悔恨，懊喪 +ANGER: Aggrieved/委屈：委屈，冤冤枉，抱委屈，哀怨 Dissatisfied/不滿：不滿，不快，不悅 +SURPRISE: Disappointment/失望：失望，心寒 Embarrassed/窘：窘 Panic/驚恐：驚恐，驚駭，驚惶，驚□，嚇人，慌張，驚慌，惶惑，慌亂	+fear, anger: Guilt/疚：疚，內疚，負疚

Table 3.2: Chinese Emotion Taxonomy (Lee 2010), *continued*

Primary Emotions	Variations in Intensity			First-Order Emotions	Second-Order Emotions
	High	**Moderate**	**Low**		
FEAR/ 恐-*kong3*	惶恐，恐慌，恐懼	畏怯，心虛，心慌， 害怕，怕，畏懼，發慌， 發怵	羞涩，羞怯，羞慚，害羞，害臊， 遲疑，為難	+ANGER: Envy/嫉：嫉妒，妒嫉，妒忌，忌妒，嫉狠，眼紅，忌狠 +SADNESS: Anxious/急：焦慮，焦渴，焦急，焦躁，焦炙，心浮，心焦，揪心，心急，心切，□急 Worry/愁：愁，苦惱，愁苦，憂愁，發愁，擔憂，擔心，犯愁，憂慮，緊張，困惑	
ANGER/ 怒-*nu4*	憤怒，忿恨，激憤，憤懣，憤慨，忿怒，悲憤，暴怒，蔑視，瞧不起， 輕蔑，鄙夷，鄙薄，鄙視，歧視，自卑，痛恨，怨恨，憎惡，憤恨，厭煩，膩煩，惱恨	生氣，窩火，火, 厭倦，討厭，厭惡，反感，敵視，衝動	煩，煩躁，煩亂，煩心，煩人，煩惱，煩雜，浮躁	+HAPPINESS: Rudeness/瘋狂：瘋狂 +FEAR: Suspicion/疑：疑，懷疑，疑心，疑惑 +SADNESS: Bitterness/辛酸：辛酸，酸辛	
SURPRISE/ 驚-*jing1*	駭怪，駭異，震驚	詫異，吃驚，愕然，驚訝， 驚奇，驚詫，驚愕	奇怪	+HAPPINESS: Delighted/驚喜：驚喜	

The automatic annotation is manually checked by the annotators to ensure the accuracy of the annotated explicit emotions. As for implicit emotion, it is loosely defined in the present study. Before conducting the implicit emotion annotation, annotators should go through the taxonomy in Table 3.2 to familiarize themselves with the classification of emotions. As shown in Table 3.2, emotion keywords referring to the same emotion may vary in their intensity and complexity. Complex emotions are formed by the combinations of primary emotions. They are further classified into first-order emotions (i.e. the combination or two primary emotions) and second-order emotions (i.e. the combination or three primary emotions). Annotators should only label the primary emotion tag regardless of the intensity and the complexity of the emotion. For example, 感動 'MOVED' is a complex emotion that is composed of a greater amount of HAPPINESS and a lesser amount of SADNESS as illustrated in Table 3.2. Given that only the major one should be tagged in this study, 感動 'MOVED' should be tagged as HAPPINESS only as it is the major component of that complex emotion. If annotators are unsure about the emotion expressed in a comment, they should skip the comment and go to the next one. Some important points are as follows:

a. An emotion can be expressed and interpreted at clause level, sentence level, or even document level. In the present study, emotions should be interpreted at clause level if two clauses convey different emotions. Consider (25).

(25) 虽然有点可怜，但是好想笑啊😂

sui ran	*you dian*	*ke lian,*	*dan shi*	*hao*	*xiang*	*xiao*	*a*
although	a bit	pitiful,	but	very	want	laugh	SFP

'It's a bit pitiful, but (I) want to laugh so badly😂'

In (25), the first clause expresses a SADNESS emotion as hinted by the adjective 可怜 'pitiful', and the second one expresses a HAPPINESS emotion as indicated by the post-event 笑 'laugh' as well as the emoji. Therefore, annotators should tag both the SADNESS and HAPPINESS emotions in that single comment.

b. Emotions expressed in a comment can either be the writer's emotion(s) or the others' emotion(s). Annotators should only annotate the writer's emotion(s). Consider (26).

(26) 哎 换个角度想想也觉得好可怜，活的时间少，别的狗狗天天开开心心的，它们随时随地都要工作

ai	*huan*	*ge*	*jiao du*	*xiang xiang*	*ye*	*jue de*
hao	*ke lian,*	*huo de shi jian*	*shao,*	*bie de*	*gou gou*	
tian tian	*kai kai xin xin de,*	*ta men*	*sui shi*	*sui di*	*dou yao*	
gong zuo						
Alas	change	CL	perspective	think	too	feel
very	poor,	live	time	short,	other	dogs
everyday	happy,		3.PL	anytime anywhere	have	
work						

'Alas, if you switch to another perspective and think about it, you'd think these dogs are actually quite pitiable. Dogs don't have a long life expectancy – while other dogs are living a happy life, they have to stay alert and be ready for work anytime and anywhere'

(26) conveys both the HAPPINESS and SADNESS emotion, but only the SADNESS emotion should be tagged because it was not the writer but 别的狗狗 'other dogs' who feel happy. Annotators should only annotate the writers' emotions.

c. It is possible that a single comment may contain multiple emotions as shown in (27).

(27) 看了那个脸上被砍到的孩子，太可怜了，那么小的孩子，丧尽天良，管理制度问题

kan le	*na*	*ge*	*lian shang*	*bei*	*kan dao de*
hai zi,	*tai*	*ke lian le,*		*na me*	*xiao de*
hai zi,	*sang jin tian liang,*		*guan li*	*zhi fu*	*wen ti*
see-ASP	DET	CL	face	PAS	chop
kid,	too	piteous	SFP,	such	small
kid,	outrageous,		manage	system	problem

'Having seen the scarred face kid, (I thought) the kid was too piteous. The kid is so small! This is outrageous! The governance has been problematic'

On the one hand, the writer expresses a SADNESS emotion as the little kid got hurt by the knife. The SADNESS emotion is indicated by the adjective 可怜 'pitiful'. On the other hand, the writer vents his/ her *anger* on the agent who hurt the little kid by using another adjective 丧尽天良 'outrageous' to describe the action.

d. Apart from emotion keywords, emotions can also be expressed through emojis. Emojis may serve as an emotion indicator that drops a hint for emotion annotation. Emojis in Weibo are originally pictorial representations of facial expressions, but all these small images are not compatible with the txt files encoded by Extensible Markup Language (XML). Each emoji in Weibo has a corresponding name, such as  [哈哈]. Weibo users can type in an emoji by clicking on an image shown in the website or they can also type in the name in between square brackets which will automatically turn into the corresponding

image after posting the comment. As images are not supported in the txt files, all the emojis are displayed by words instead of images. A list of emojis and their corresponding names are given to the annotators to ensure that the replacement would not affect the annotation. The list is attached in Appendix I. It is observed that some emojis are highly associated with a particular emotion. For example,  typically conveys a HAPPINESS emotion in most cases etc. However, it is also possible for an emoji to convey an unexpected emotion. An example is exemplified in (28).

(28) 对于这种人，最好的办法就是以暴制暴，我又不是圣人，大家也都是第一次做人，干嘛我要受委屈 

dui yu zhe zhong ren, zui hao de ban fa jiu shi yi bao zhi bao, wo you bu shi sheng ren, da jia ye dou shi di yi ci zuo ren, gan ma wo yao shou wei qu

to DET CL people, best method is an eye for an eye, 1.SG also not saint, everyone too also first time be human, why 1.SG need suffer

'The best way to deal with these people is to treat them with an eye for an eye approach. I am not a saint. We all have only been through being a mortal the first time. Why should I take the suffering?  '

In (28),  represents HAPPINESS as indicated by its happy face. However, the emotion conveyed in text is not aligned with the one expressed by the emoji. It demonstrates that emojis may be used in an ironic or sarcastic way, and annotators should not solely rely on emojis when annotating emotions. Instead, they must judge if it is the text or the emoji that expresses a writer's emotion.

To clarify how each emotion should be annotated, some examples are given for discussion. (29) – (33) explain how HAPPINESS should be annotated.

(29) 我们大鄂尔多斯的民警棒棒哒👍👍👍
wo men *Daeerduosi de* *min jing* *bang bang da*
1.PL Daeerduosi-POS police awesome
'The Daeerduosi police force is awesome👍👍👍'

(30) 现在真是年纪大了，看这样的新闻都会不禁有泪目的冲动。❤️
xian zai *zhen shi* *nian ji* *da le,* *kan* *zhe* *yang de* *xin wen*
dou *hui* *bu jin* *you* *lei mu* *de* *chong dong.*
now really age big-ASP, watch DET kind of news
always have cannot help have tear in eye POSS impulse.
'(I'm) now getting old. Each time as I come across these news, I cannot help but burst into tears❤️'

(31) 这才是法治应有之精神！
zhe *cai* *shi* *fa zhi* *ying you* *zhi* *jing shen*
DET only is rule of law should have POS spirit
'This is the spirit of the rule of law!'

(32) 老板脑子好
lao ban *nao zi* *hao*
boss brain good
'The boss has a sharp mind.'

(33) 这样的女局长很少往往也很令人惊艳
zhe yang de *nü* *ju zhang* *hen* *shao* *wang wang* *ye* *hen* *ling ren jing yan*
DET kind of female director very rare oftentimes too very impressive
'Female directors like such are rare. But they're often impressive.'

(29) is a typical example that conveys a HAPPINESS emotion. It implies that the writer was happy with what the police did. The emotion (30) conveys is MOVED, which is a complex emotion that is composed of HAPPINESS and SADNESS. As HAPPINESS is the major emotion of MOVED, it should only be tagged as HAPPINESS. Similar to (29), (31) is marked with as HAPPINESS as the comment reveals the

writer's stance that he/she was happy and satisfied with the judgement. However, (32) and (33) are positive statements but they should not be annotated as HAPPINESS. The writer of (32) states that the boss has a sharp mind, but it has no concern with the writer's emotion. As for (33), the word 令人惊艳 'impressive' may possibly trigger SURPRISE. However, 令人惊艳 'impressive' in (33) is a general statement that female directors are often impressive, but not that he/she is now impressed by the director. Therefore, it does not represent the writer's emotion. (34)- (37) are examples exemplifying the expressions of SADNESS.

(34) 祈祷吧，心里特别不舒服
qi dao ba *xin* *li* *te bie* *bu* *shu fu*
pray SFP heart in especially not well
'Pray, mentally speaking, I don't feel fine (about something).'

(35) 活着真累，但也没办法
huo zhe *zhen* *lei* *dan ye* *mei* *ban fa*
living really tiring, but too no solution
'Living is exhausting, you can do nothing but to deal with it'

(36) 如果不是遇到烦心事和真困难，谁能这样？
ru guo bu shi *yu dao* *fan xin shi* *he* *zhen* *kun nan,* *shei* *neng* *zhe yang?*
if not encounter trouble matter and real difficulty, who can DET?
'If it wasn't because of the hardship or difficulties, who would want to be in that situation?'

(37) 警察就不去查一查这事情的原因吗？
jing cha *jiu bu qu* *cha* *yi* *cha* *zhe* *shi jian* *de* *yuan yin* *ma?*
police not investigate once investigate DET incident POS reason SFP?
'Aren't the police going to investigate further on what was going on with this incident?'

When a writer shows compassion to people, the writer is undoubtedly expressing a SADNESS emotion. For example, the writer of (36) shows compassion to the person who has difficulties. As for (34), the SADNESS emotion is indicated by the phrase 心里特别不舒服 'not feeling fine (about something)'. Unlike (34), the emotion in (35) is not directly described, but it can be inferred by the whole sentence expressing that the writer felt helpless as he/ she found living exhausting but still have to deal with it regardless. (37) is a rhetorical question judging that the police should have investigated further on this incident. In other words, the writer is dissatisfied with the police for not investigating the matter. Therefore, it should not be annotated as SADNESS but ANGER. The annotation of ANGER is illustrated as in (38) – (41).

(38) 有钱喝酒 没钱吃饭

you *qian* *he jiu* *mei* *qian* *chi fan*
have money drink mo money eat

'So you have the money to get yourself boozed, but not the money to eat'

(39) 呵，南京还不是能允许日本人进去参观吗？

he, *Nanjing* *hai bu shi* *neng* *yun xu* *riben ren* *jin qu* *can guan* *ma?*
INTJ, Nanjing still not can allow Japanese enter visit SFP?
'Huh (showing despise), we're still letting Japanese in Nanjing for a visit, are we not?'

(40) 我怎么听他们的口气还很爽！扫射的那么多人睡得安稳吗！

wo zen me ting ta men de kou qi hai hen
shuang! sao she de na me duo ren shui de an wen ma!
1.SG how come listen 3.PL POSS tone still very
happy! shooting that many people sleep-DE steady SFP!
'From what I am hearing from them, it sounded like they were having a fun time! How can they even get to sleep when they are shooting this many people?'

(41) 不租不借，滚

bu zu bu jie, gun
not rent not lend, get out
'Not renting, not lending. Bugger off'

All the above examples express an ANGER emotion. When the writer shows despise in an expression, it should be annotated as ANGER as in (38) and (39). The writer of (40) condemns those who shot dead many people. As for (41), the ANGER emotion is expressed in an expressive way. The imperative 滚 'bugger off' shows that the writer was ANGRY at the interlocutor. (42)- (45) show how FEAR is expressed.

(42) 啊，好恐怖啊

a, hao kong bu a
INTJ, very scary SFP
'Ah, it's so scary.'

(43) 知道的假奶粉是 8 吨 吃下去了几吨 在售还剩几何

zhi dao de jia nai fen shi 8 dun chi xia qu le ji dun zai shou hai sheng ji he
known fakemilk powder is 8 ton eat down-ASP how many ton sale still left how many
'The amount of fake formula milk powder we know of is 8 tons. But what about the amount of fake milk powder that has been consumed or left in the market for sale?'

(44) 我觉得还是一枪打死好一点，不然要是毒贩没有失去行动能力 心生怨恨 伤害人质怎么办

wo jue de hai shi yi qiang da si hao yi dian, bu ran yao shi du fan mei you shi qu xing dong neng lixin sheng yuan hen shang hai ren zhi zen me ban

1.SG reckon still one shot kill better, or else if drug dealer not lost mobility resentment hurt hostage then what

'I still reckon (someone) should have shot the drug dealer to death. What if the drug dealer hadn't lost his mobility and turned to the hostages for revenge? Then what?'

(45) 幕后老板不会是华人吧？

mu hou lao ban bu hui shi hua ren ba?

The boss behind not really is ethnical Chinese SFP?

'The power behind the throne isn't really a ethnic Chinese, right?'

Example (42) is obviously expressing a FEAR emotion as indicated by the adjective 恐怖 'scary' which directly leads to 'FEAR'. The writer of (43) is WORRIED (i.e. FEAR) as he/ she does not know the amount of fake milk powder that has already been consumed and left on the market for sale. In (44), the writer was WORRIED about the hostage as indicated by the syntactic pattern "要是……怎麼辦" 'what if …'. The writer of (45) suspects the manipulator is Chinese. SUSPICION is a complex emotion consists of ANGER and FEAR, as ANGER is the major emotion, and FEAR is the peripheral one, only ANGER should be tagged. Examples of SURPRISE are given in (46)- (48).

(46) 居然敢卖房

ju ran gan mai fang

surprisingly dare sell estate

'(He is) going so far as to sell the estate?'

(47) 这些家长去骂奶茶店？怪奶茶店？不是脑子有泡？自己孩子管理不好，怪别人？人家开店，又不犯法，你 tm 国家总理？管的到人家？沙币家长

zhe xie jia zhang qu ma nai chi dian? guai nai cha dian? bu shi nao zi you pao? zi ji hai zi guan li bu hao, guai bie ren? ren jia kai dian, you bu fan fa, ni tm guo jia zong li? guan de dao ren jia? sha bi jia zhang

DET parents go scold milk tea shop? blame milk tea shop? not brain has bubble? own kid manage bad, blame others? others operate store, again not illegal, you bloody Prime Minister? manage others? cuckoo parents

'These parents are blaming the bubble tea shop? What? Have they gone utterly insane? They're blaming others for the misbehavior of their kids? Who is to be blamed for the improper behavior of their own children? What makes them qualified to judge the shop? They're not a bloody Prime Minister!'

(48) 罂粟都可以误种？下次我来误制一点冰毒可以不？

ying su dou ke yi wu zhong? xia ci wo lai wu zhi yi dian bing du ke yi bu?

opium poppy also can mistakenly plant? next time 1.SG come mistakenly produce some crystal meth can not?

'You're planting opium poppy by accident? How about me making some meth by accident too?'

The writer of (46) is SURPRISED by the action done by the person who sold the estate as indicated by the adverb 居然 'surprisingly'. SURPRISE is quite often co-occurred with other emotions, such as ANGER as in (47) and (48). The writer of (47) is SURPRISED that the parents blamed the milk tea store. He/ she feels ANGRY at the same time as the action of blaming the store sounded unreasonable to the writer. The writer of (48) is SURPRISED and ANGRY about the ridiculous claim that the poppies were mistakenly planted.

3.4.2 Rhetorical Question

Questions can roughly be classified into two types, namely information-seeking questions and rhetorical questions. In this work, I only focus on rhetorical questions. Unlike information-seeking questions which generally aim to elicit an answer, rhetorical questions expect no answer but to achieve a pragmatic goal, such as to emphasize, to persuade, or to show emotions.

For the identification of rhetorical questions, a question mark is not necessary. Once a question is identified, be it a question in a main clause or in an embedded clause, annotators need to judge whether the question is seeking information or not. If it is an information-seeking question, the question should not be marked. If it is a rhetorical question, the question should then be annotated with its question type following 14 types proposed in Lau and Lee (2018) as in Table 3.3:

Table 3.3: The Definition and Example(s) of Question Types

Question Type	Definition/ Example(s)
A. Series of Questions	When more than one question appears in a single comment and that the questions are rhetorical questions, these questions should be tagged as Series of Questions.
B. A-not-A	A-not A refers to questions that form with an affirmative and its negative counterpart juxtaposed with the A-not-A pattern, such as 是不是，有沒有
C. Alternative	Alternative questions explicitly provide two or more possible options which are mostly connected by the morpheme 還是/ 或者 "or"
D. Echo	Echo question refers to questions that have the form of a declarative sentence but end with a question mark in the written form. e.g. 人家承认侵略我们我们还要点赞？
E. Particle	Particle questions refer to questions that end with a sentence-final particle, such as 嗎， 呢， 吧. Some of the question words are often replaced with a netizen transformation, such as 嘛， 麼. They should also be tagged as a particle question even though the question word is not in its standard form.
F. Others	*Others* includes questions formed with rhetorical interrogation markers, such as 難道， 豈， 何必， 何苦 etc.
G. How	如何， 怎樣， 怎麼，and 是有多
H. How many/ much	多少
I. What	什麼
J. Which	哪些，哪个
K. Who	谁
L. Why	为什么， 为何， 怎麼 （肿么， 咋， 为神马， 为嘛，为毛，为啥，干嘛）
M.Where	哪，哪里
N. When	什么时候

There are some important points to be noted:

a. Some question words may belong to more than one question type, annotators should tag the question type according to the context. For instance, the word 怎麼 in 我高考的时候怎么没这样 'why it wasn't like that when I was having

the college entrance examination' expresses the meaning of why while it expresses the meaning of how in 都是要做妈妈的人，怎么忍心对孩子下手 'as a mother-to-be, how could she be that cruel to the kid'. Therefore, the question should be tagged based on the meaning the question word conveys.

b. As for the question word 怎么办, it can either be interpreted as what and how. In this study, those questions formed with 怎么办 should be tagged as what questions for consistency.

c. If a rhetorical question contains more than one question word that belongs to different question types, annotators should consider which question(s) is raised by the writer as only the writer's question(s) should be annotated.

d. If a rhetorical question contains more than one question word, annotators should choose the question type that plays a more important role in determining the question type. For example, in 求死勇气那么大，为什么不好好活着呢, 為什麼 '(you) have great courage to commit suicide, why don't you live well, why' belongs to a why question, and 呢 belongs to a particle question. However, the question should be annotated as a why question instead of a particle question. This is because even without the particle 呢, the question is still well-formed with the meaning of why.

3.4.3 Emotion Expression, Pre-event and Post-event

This section introduces the annotation of three parts, (a) sentence(s) that expresses emotion, (b) pre-event(s) that triggers the emotion, and (c) post-event(s) that is

evoked by the emotion (i.e. an action or reaction of the experiencer). It should be labelled in the form of "<*emo*>*Sentence(s)$Pre-event$Post-event*<*emo*>". The instructions are as follows:

a. Each annotated emotion should have its own line of code, i.e. "<*emo*> *Sentence(s)$Pre-event$Post-event*<*emo*>". The emotion tag <emo> refers to the annotated emotion, the label should be as in Table 3.4:

Table 3.4: Emotion Labels

Emotion	Emotion Label (<emo>)
Happiness	<H>
Sadness	<S>
Anger	<A>
Fear	<F>
Surprise	<U>

For example, if a comment is tagged as HAPPINESS, annotators should put <H>*Sentence(s)$Pre-event$Post-event*<H>.

b. Some emotions are expressed explicitly (i.e. with emotion keywords), and some are expressed implicitly (i.e. without emotion keywords). Annotators should identify the sentence(s) and emoji(s) that express emotions implicitly and put it right after the first emotion tag <emo>.

c. Pre-event is loosely defined in the present study. According to Talmy (2000), an emotion cause should be an event itself. A pre-event does not necessarily need to be an actual cause of an emotion or what directly leads to an emotion. It refers to the cause of the emotion which can be the actual trigger event, or the perception of the trigger event as in (49).

(49) 多一些这样普及历史的纪录片会更好！虽然很残忍😢

duo yi xie zhe yang pu ji li shi de ji lu pian
hui geng hao! sui ran hen can ren
more one CL DET general history POSS documentary
be better! although very cruel
'It's good to have more of these documentaries for general history. Though (history) can be cruel…😢'

The phrase 很残忍 'very cruel' in (49) is marked as the pre-event of SADNESS even though it is not an actual event that literally denotes an action or activity. However, the phrase 很残忍 'very cruel' directly adds information and explains the reason for the emotion triggered (i.e. SADNESS). It is observed in the corpus that implicit emotions are often expressed by a short judgemental phrase without mentioning the actual cause of emotion, such as 太没品了 '(someone is) so poorly educated' which implies an ANGER emotion. In that case, the perception of the trigger event 太没品了 '(someone is) so poorly educated' is the only possible pre-event to be marked. Therefore, the pre-event of (49) should be 很残忍 'very cruel' for the sake of consistency.

d. For the markup of pre-events, annotators should use a dollar sign "$" to mark the beginning and the end of a pre-event. If there is more than one pre-event mentioned in a single comment, use two strokes "//" to separate them. If a pre-event is not mentioned in a comment, annotators should simply put 'none' in between the two dollar signs.

e. A pre-event may be expressed by means of perception verbs, nouns, verbs, rhetorical questions, and anaphoric expressions as illustrated in (50) – (54).

(50) Perception verb:

我不管他们道不道歉，骂就完事了，我只是看到"日本"俩字就上火

wo bu guan ta men dao bu dao qian, ma jiu wan shi le, wo zhi shi kan dao "ri ben" liang zi jiu shang huo

1.SG don't care 3.PL apologize-not-apologize, scold then done-ASP, 1.SG just see-ASP "Japan" two words then angry

'I don't care if they have apologized or not. I'm done scolding them, but I just got so furious every time when I see the word "Japan" '

(51) Noun:

我们必须送日本人去见上帝

wo men bi xu song ri ben ren qu jian shang di

1.SG must send Japanese go see God

'We must send the Japanese to God.'

(52) Verb:

我怎么听他们的口气还很爽！扫射的那么多人睡得安稳吗！

wo zen me ting ta men de kou qi hai hen shuang! sao she de na me duo ren shui de an wen ma!

1.SG how come listen 3.PL POSS tone still very happy! shooting that many people sleep steady SFP!

'From what I am hearing from them, it sounded like they were having a fun time! How can they even get to sleep when they are shooting this many people?'

(53) Rhetorical Question:

我怎么听他们的口气还很爽！扫射的那么多人睡得安稳吗！

wo zen me ting ta men de kou qi hai hen shuang! sao she de na me duo ren shui de an wen ma!

1.SG how come listen 3.PL POSS tone still very happy! shooting that many people sleep steady SFP!

'From what I am hearing from them, it sounded like they were having a fun time! How can they even get to sleep when they are shooting this many people?'

(54) Anaphoric Expression:
不敢看这片子，只是感到中国人对待日本战俘太仁慈了！
bu gan kan zhe pan zi, zhi shi gan dao zhong guo ren dui dai ri ben zhan fu tai ren ci le!
not dare watch DET movie, just feel Chinese treat Japanese person of war too kind SFP!
'I dare not to watch this movie. I just felt that the Chinese were being too merciful to the persons of war from Japan.'

f. Pre-events are sometimes expressed by means of a request, a negation and so, in these cases, annotators should not annotate those as a pre-event as only causes that are explicitly or directly stated should be considered. For example, in "还我土地！妈的" '*return the land to me! Shit*' , the ANGER emotion is triggered by someone not returning the land to me, however, it should not be tagged as a pre-event as it is not expressed in a direct way (i.e. you owned me the land).

g. Post-events are defined as the action(s) of the experiencer which is induced by an emotion. It can either be a physical reaction or an action that is triggered by the emotion. A post-event does not necessarily need to be a past or present event; it could also be a future action that the experiencer will or will not take because of the emotion, such as 不買日貨 'not buying Japanese products', 不去日本旅遊 'not traveling to Japan'.

h. Only concrete actions should be annotated as post-events, such as 呼籲 'urge', 要求 'request', 决定 'decide', 勸 'convince' etc. Abstract actions that no actual action can be carried out such as 希望 'hope' etc., they should not be marked.

i. If the writer directly uses an expression to scold or condemn the person for doing something in an event as in 真不要脸 'shameless', 可耻 'shameful',

丢人啊 ‘shameless’, the sentence should not be regarded as a post-event. However, if the action is clearly stated, such as 罵了他一頓 ‘scolded him’, it should be tagged as a post-event.

j. Apart from textual information, a post-event can also be expressed by means of an emoji, such as 😭 [泪], 😂 [哈哈]. However, only those actions/ reactions that experiencers will do in daily life should be considered as a post-event. For example, the emoji 🤧 [挖鼻] is usually used to show that someone does not care, and sometimes used to show that someone looks down on another person. As this is not an emotion reaction people usually do in their real life, it should not be labelled as a post-event. If a post-event is not mentioned in a comment, just put ‘none’ as in *<emo>Sentence(s)$Pre-event$none<emo>*. If there is more than one post-event, use two strokes ‘//’ to separate them.

3.4.4 Opinion Target

Opinion target refers to a (frame) element by which an emotion is triggered. The goal of the annotation is to see whether an emotion is highly related to a particular person/ element who has conducted an event/ some events. In the dataset, each post consists of at least one subevent, and they are marked with *e1*, *e2* etc. Each subevent contains a set of frame elements that provide information to the sematic structures of an event. To figure out the opinion target(s), annotators need to read the subevents and see which subevent(s) does trigger the annotated emotion(s). Consider the event and comments in Figure 3.3:

Event:

【日本士兵[e1-承认]南京大[e2-屠杀]CG[e3-还原]枪杀现场】日本电视台于5 月 14 日播出了南京大屠杀的调查纪录片《南京事件 2——检验历史修正主义》。在纪录片中，日本士兵描述了 1937 年 12 月 16、17 日如何杀害中国俘虏，承认当时杀死数万中国人。并用 CG 动画还原了令人心痛的枪杀现场。???05 月 15 日 19:52

(1)- Reveal_secret_ e1: 日本士兵: Speaker
(2)- e1: 南京大屠杀: Topic
(3)- e1: 纪录片: Medium
(4)- e1: 杀死数万中国人: Information
(5)- Killing_ e2: 日本士兵: Killer
(6)- e2: 中国俘虏: Victim
(7)- Duplication_ e3: 枪杀现场: Original
(8)- e3: CG 动画: Copy

Comment:	Opinion Target:
1. <A>日本有我们学习的地方，但是性质是真坏，会为他们的行为付出代价的$他们的行为$none<A>	5
2. <H>只能說很欣慰$日本终于肯承认自己犯下的罪行$none<H> <A>我们没资格谈原谅，也没什么好感谢$这本来就是日本应该做的$勿忘国耻<A>	1
3. <A>他们真的敢还原吗？这算还原吗？真实的比这些更残忍！这只是冰山一角！$这算还原吗？//真实的比这些更残忍！这只是冰山一角！$none<A>	5,8
4. <A>呵，南京还不是能允许日本人进去参观吗？$南京还不是能允许日本人进去参观$none<A>	0

Figure 3.3: The Annotation of Opinion Target(s)

In Comment 1, what triggers an ANGER emotion is the behaviour of the Japanese which refers to the *killing* event (e2). Although the frame elements (1) and (5) both refer to Japanese soldiers, frame element (1) focuses on the *reveal_secret* event (*e1*), while (5) focuses on the *killing* event (*e2*). Therefore, only frame element (5) should be tagged. As for comment 2, both the HAPPINESS and ANGER emotions are elicited by the *reveal_secret* event (*e1*) done by Japanese soldier, thus, only frame element

(1) is tagged. Comment 3 is about the *killing* event (*e2*) as well as the *duplication* event (*e3*). The ANGER emotion is triggered by both Japanese soldier (i.e. frame element (5)) and the computer-generated animation (i.e. frame element (8)). As for Comment 4, the cause of the ANGER emotion is that Nanjing (Massacre Museum) still allows Japanese to pay a visit to. As none of the frame element is related to Nanjing, annotators should put a "0" to indicate that it is an external cause that evokes the ANGER emotion.

3.4.5 Inter-annotator Agreements

The Chinese Event-comment Corpus is composed of 200 posts. 50 comments of each post are annotated with emotion-related information, which add up to a total number of 10,000 annotated comments in the corpus. The emotion annotation task is done by two annotators. To verify the quality of the annotation, the two annotators are asked to annotate the same set of data (i.e. 20 posts) for inter-annotator agreements. Cohen's Kappa κ is used to evaluate the inter-annotator agreement for emotion annotation:

$\kappa = \frac{p_o - p_e}{1 - p_e}$, where

p_o is the observed agreement between two annotators, and p_e is the expected agreement between them. Let n_{ij} denotes the number of examples where annotator 1 labels as category i and annotator 2 labels as category j; $n_{i.}$ denotes the number of examples annotator 1 labels as category i, $n_{.j}$ denotes the number of examples annotator 2 labels as category j; n denotes the number of categories

and N the number of annotated examples. Then, p_o and p_e can be calculated as follows.

$$p_o = \frac{\sum_{i=1}^{n} n_{ii}}{N}$$

$$p_e = \sum_{i=1}^{n} \frac{n_{i\cdot}}{N} \times \frac{n_{\cdot i}}{N}$$

The inter-annotator agreement for emotion type annotation is 0.839 which verified the reliability of the data.

Unlike the set of emotion types which is composed of five categories, the set of opinion targets (types of event) is open. Therefore, the agreement for the annotation of opinion target is evaluated with Accuracy (i.e. percentage). The calculation of Accuracy is the same as p_o in the abovementioned equation. The agreement is 0.863.

As for the text scope of emotions, pre-event and post-event of emotions, I use a similarity threshold (0.95) between the texts annotated by the two annotators. If the similarity is above the threshold, then it is counted as 'agree/same', otherwise, 'disagree/different'. I then differentiate a partial agreement when the similarity of the annotated texts by two annotators is above 0.3. I output the number of different types of agreements with codes defined as follows, and the agreements are shown as in Table 3.5.

Agreement code:

0: agree on non-empty expression
1: agree partially on non-empty expression
2: agree on empty expression
3: disagree as (non-empty, non-empty)
4: disagree as (non-empty, empty)
5: disagree as (empty, non-empty)
6: disagree on emotion

Table 3.5: Inter-annotator Agreements for Emotion Expression, Pre-event, and Post-event

	Emotion Expression	Pre-event	Post-event
Agreement:	0: 0.882	0: 0.810	0: 0.180
	1: 0.028	1: 0	1: 0.003
	2: 0.054	2: 0.168	2: 0.798
	3: 0.005	3: 0.001	3: 0
	4: 0.012	4: 0	4: 0.002
	5: 0.003	5: 0.004	5: 0.001
	6: 0.017	6: 0.017	6: 0.017

Table 3.5 shows that the annotation of emotion expressions and pre-event agree on the non-empty expression. Given that most comments do not contain post-events, the two annotators also largely agree on empty expression for the annotation of post-events. The high agreements for the three annotation tasks indicate that the annotation scheme is adequate, and the quality of the annotation is reliable.

3.5 Annotation Tool and Analysis Tool

The Chinese Event-Comment Corpus is composed of 200 posts. Each post is saved in a txt file encoded by Extensible Markup Language (XML). For the event annotation, the markup of events and the annotation of frame elements are inserted

in the post of the txt files. Figure 3.4 illustrates the coding of a post.

```
<?xml version="1.0" encoding="UTF-8"?>
<Root>
  <Post>
  <Text>:【日本士兵[e1-承认]南京大[e2-屠杀]CG[e3-还原]枪杀现场】日本电视台于 5 月 14 日播出了南京大屠杀的调查纪录片《南京事件 2——检验历史修正主义》。在纪录片中，日本士兵描述了 1937 年 12 月 16、17 日如何杀害中国俘虏，承认当时杀死数万中国人。并用 CG 动画还原了令人心痛的枪杀现场。???05 月 15 日 19:52
(1)- e1:  日本士兵: Speaker
(2)- e1:  南京大屠杀: Topic
(3)- e1:  纪录片: Medium
(4)- e1:  杀死数万中国人: Information
(5)- e2:  日本士兵: Killer
(6)- e2:  中国俘虏: Victim
(7)- e3:  枪杀现场: Original
(8)- e3: CG 动画: Copy
</Text>
  </Post>
```

```
  <Comment ID="33">
      <Emotion>Anger</Emotion>
      <Emotion_Keyword>None</Emotion_Keyword>
      <Tool><A>卧槽  那个士兵的话语中还是透露着变态的思想$那个士兵的话语中还是透露着变态的思想$none<A></Tool>
      <OpinionTarget>1</OpinionTarget>
      <RhetoricalQuestion Type="None">None</RhetoricalQuestion>
      <Text>卧槽  那个士兵的话语中还是透露着变态的思想</Text>
      <QuestionEmotion> </QuestionEmotion>
  </Comment>
```

Figure 3.4: Coding of a Post

Figure 3.4 is an example taken from the corpus which contains two parts. The upper part illustrates the post (i.e. event(s)), and the bottom part shows the comments. For each post, 50 comments of each post are annotated with emotion-related information, but only one comment is shown in Figure 3.4 as an example. <Text> marks the beginning of the entire post while </Text> marks the end of it. In each

post, both the heading and the content are included. The keywords that denote an event is marked with [e1-*keyword*], in which "e" refers to an event, and "1" marks the index of the event. As shown in the heading of the post in Figure 3.4, the post contains three sub-events, namely [e1-承认], [e2-屠杀], and [e3-还原]. To facilitate the annotation of opinion target, the frame elements of each event are listed right below the content of the post as listed in (1) to (8). Each element is assigned with a number, followed by its corresponding event number, e.g. *e1*. The argument/ adjunct is placed right after the event number, followed by the semantic role (i.e. frame element) it plays in that event. For example, (1) is assigned to 日本士兵 'Japanese soldier' who plays a role in [e1-承认] as a speaker. When an emotion expressed in a comment is triggered by the Japanese soldier concerning [e1-承认], annotators should tag the corresponding number assigned to the Japanese soldier (i.e. 1) as the opinion target.

As for the bottom part, <Comment ID="*x*"> marks the index of each comment. 50 comments of each post are annotated with emotion-related information, namely emotion type, emotion keyword, emotion expression, pre-event(s) and post-event(s), opinion target(s), rhetorical question, and its corresponding question type. The emotion type(s) is marked with the set of <Emotion>. If a comment contains an emotion word(s) or emoji(s) that express an emotion explicitly, the keyword(s) would be marked with the set of <Emotion_Keyword>. <Tool> marks the emotion expression, pre-event(s) and post-event(s), and the </Tool> marks the end of it. The three items are annotated in the form of <*emo*> *emotion expression* $ *pre-event* $

post-event <*emo*>. The <emo> should be tagged according to the emotion type. For example, the expressed emotion of the comment in Figure 3.4 is ANGER, and therefore the <emo> tags should be <A>. The frame element(s) that triggers an emotion is marked in between the set of <OpinionTarget>. If a rhetorical question does exist, the question type would be marked with <RhetoricalQuestion Type="*x*">, and that the question would be placed in between the <RhetoricalQuestion Type="*x*"> and </RhetoricalQuestion>. The emotion(s) expressed by via rhetorical question is marked with the set of <QuestionEmotion>. The above-mentioned information is derived from the content of the comment which is marked with the set of <Text>. Figure 3.5 demonstrates the annotation tool has been created for the emotion annotation task.

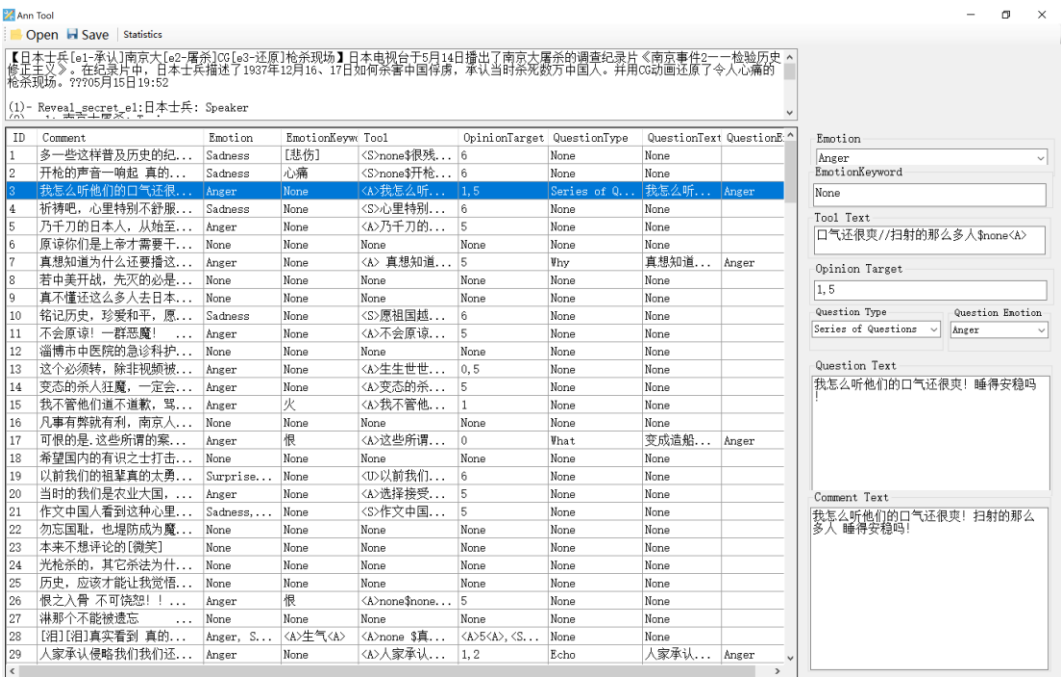


Figure 3.5: Annotation Tool

Figure 3.5 shows the annotation tool for emotion annotation. The post is on the top of the tool and the comments are placed below. When users click on one of the comments, the entire comment will be displayed in the text box of "Comment Text" on the bottom right of the tool. On the right side of the tool, there are several text boxes. The emotion type, emotion expression pre-event and post-event, opinion target, rhetorical question and its corresponding question type, and the emotion it expresses can be typed in through this tool. After entering all the 6 categories, they will be automatically coded as shown in the comment part in Figure 3.4.

In addition to the annotation tool, and analysis tool is also created for data analysis. The analysis tool is illustrated in Figure 3.6.

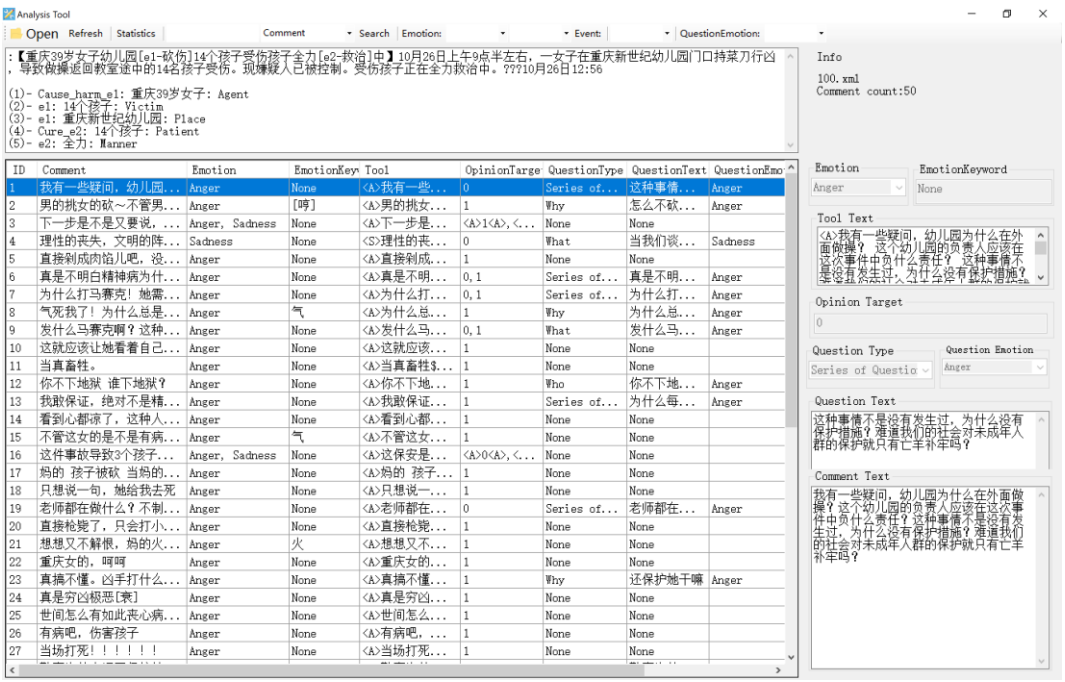


Figure 3.6: Analysis Tool

Figure 3.6 shows the analysis tool for data analysis. It can be used to open one single file or multiple files. The post is shown on the top of the tool if only one file is selected; if multiple files are selected, only the post of the first file will be shown. The purpose of opening multiple files is to study the linguistic features of implicit emotions, regardless of the event type. When users click on one of the comments, the corresponding details will be shown in the right side of the tool.

The analysis tool supports several functions as in Figure 3.7.

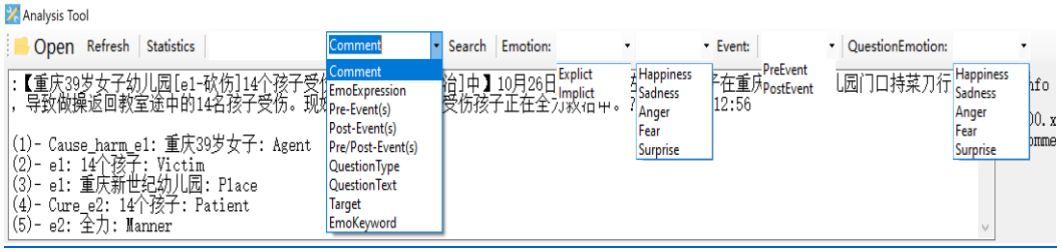


Figure 3.7: Functions of the Analysis Tool

As illustrated in Figure 3.7, the analysis tool supports keyword search. Apart from search the entire comment, users may choose to do the search within the emotion expression scope, pre-event scope etc. Furthermore, a second search can also be done based on the result of the previous search. For instance, if a user wants to search those comments that contain both 傷心 'sad' and 哭 'cry' regardless of the sequence, the user can type in 傷心 and press the 'search' button, comments that contain 傷心 will then be shown. After that, the user can make the second search of 哭, and the user will get all the comments containing both keywords. As for patterns such as 为什么.....不 'why.....not' in which the sequence of the two keywords matters, and there are words in between them, the user can simply key in

“为什么 不” with a space between the two phrases. Moreover, users can do a search of tokens that express a certain emotion by means of rhetorical question using the button on the right side of the tool. Users can also select question type by typing in a question type, such as “why”, all the rhetorical questions being tagged as a why question will then be shown. It is also applicable to the search of opinion targets.

The tool also supports the display of comments in terms of emotion types. Users can select an emotion type to see all the comments that are used to convey a particular emotion. They can also select the ‘explicit’ or ‘implicit’ button to see all the comments that are expressed with or without an emotion keyword/ emoji.

CHAPTER 4

EMOTION EXPRESSIONS: A SEMANTIC PERSPECTIVE

Emotions can be conceptualized at different linguistic levels: word, sentence, and discourse levels. This chapter deals with emotions conceptualized at the lexical level from a semantic perspective. According to Pavlenko (2008), emotion lexicons can be classified into three types, namely emotion words, emotion-related words, and emotion-laden words. Emotion words are words that directly refer to particular emotional states. They are used to form expressions of explicit emotions. Emotion-related words are words that do not refer directly to emotional states but to behaviours that are related to particular emotions, such as tears, scream etc. Emotion-laden words are words that express emotions without using emotion-bearing words or words that elicit emotions from interlocutors. Both are used to form expressions of implicit emotions.

Section 4.1 presents the expressions of explicit emotions, including both emotion words and emojis which directly refer to a particular emotion. Section 4.2 discusses two types of expressions that convey emotion implicitly, namely emotion-related words and emotion-laden words. Section 4.3 gives a summary of the main points addressed in the chapter.

4.1 Expressions of Explicit Emotions

This section presents the explicit emotion expressions. Explicit emotions can only be expressed at the semantic level in terms of emotion words and emojis. Section 4.1.1 discusses the use of emotion words, and Section 4.1.2 discusses the use of emojis.

In Chapter 1, I place much emphasis on implicit emotions and hypothesized that most emotions expressed are implicit in nature. According to the annotated data, the distribution of explicit emotions and implicit emotions in all the 10,000 comments are as in Table 4.1.

Table 4.1: Distribution of Explicit and Implicit Emotions

	Explicit	Implicit
Happiness	486	1,410
Sadness	324	1,456
Anger	424	5,306
Fear	107	466
Surprise	20	678
Total	1,361 (12.7%)	9,316 (87.3%)

Table 4.1 shows that the distribution of explicit emotions and implicit emotions are 12.7% and 87.3%. The distribution does not only confirm the hypothesis that a considerable amount of emotions is expressed in an implicit way, it also highlights the importance of implicit emotions, on which this present work mainly focuses.

Given that implicit emotions do play a vital role in emotion expressions, an in-depth analysis of implicit emotions is considered a necessary component of

emotion studies that should not be neglected or overlooked. Section 4.2 turns to the two means that are used to express emotions implicitly, namely emotion-related words and emotion-laden words. The former is introduced in Section 4.2.1, and the latter in Section 4.2.2.

4.1.1 The Use of Emotion Words

The term *explicit emotion* refers to the presence of emotion-related information denoted by emotion keywords. For example, in the sentence "he looked at me in surprise", whereby the word "surprise" directly refers to the SURPRISE emotion. In this work, the term "emotion keyword" is interchangeable with the term "emotion word" as defined by Pavlenko (2008).

Explicit emotion has long been the focus of most previous studies on emotions. As explicit emotions can be easily detected with the help of emotion lexicons due to the presence of emotion words, a great deal of research has attempted to construct emotion lexicons in different languages. Most existing models of Chinese emotion words relied on expert judgement, among which some created emotion lexicon from Chinese dictionary, and some translated emotion words from other languages into Chinese.

In the current work, I adopted Lee's (2010) emotion taxonomy which is a robust and versatile emotion annotation scheme based on cognitive emotion theories. The taxonomy is generated by mapping Chinese emotion words to the model of English emotion words proposed in Turner (2000), with some

modifications made based on Plutchik (1980). The emotion taxonomy proposed by Lee (2010) contains five basic emotion types, namely HAPPINESS, SADNESS, ANGER, FEAR, and SURPRISE. Each emotion type has three intensity level, i.e. high, moderate, and low. Moreover, complex emotions which consist of more than one primary emotion are listed as first-order emotions. For example, the complex emotion MOVED is composed of HAPPINESS and SADNESS, with the former being the major component. Therefore, when a comment is expressing a MOVED emotion, it should be tagged as expressing the primary emotion, i.e. HAPPINESS. Additionally, second-order emotions which consist of three emotions are also listed in the taxonomy. To deal with the annotation of a complex emotion, be it a first-order or a second-order emotions, only the major component of a complex emotion is annotated in this work.

All the explicit emotions expressed by means of emotion keywords in the Chinese Event-comment Corpus have been automatically annotated based on Lee's (2010) emotion taxonomy. In order to perfect the taxonomy, I read through all the 10,000 annotated comments to add some additional emotion words and remove some ambiguous ones.

As for the additional emotion words, they are selected according to their word meaning, usage, as well as their emotion orientation. An emotion word should be defined as a word that directly refers to an emotional state. In other words, words describe a situation of a person or a thing should not be classified as an emotion word even if they do evoke an emotion. Consider (1).

(1) 这孩子好**可怜**

zhe	*hai zi*	*hao*	***ke lian***
DET	child	very	**poor**

'Such a really poor kid'

In (1), the adjective 可怜 'poor' has a semantic orientation pointing to the SADNESS emotion. However, it should not be regarded as an emotion word as it is used to describe situations that trigger the SADNESS emotion. Therefore, it can only be considered an emotion-laden word of SADNESS. Details will be discussed in Section 4.2.2.

Instead of modifying situations that trigger a particular emotion, an emotion word has to refer to an emotion state directly. For example, 悲傷 'sad' is an emotion word which explicitly indicates an emotion state. In addition, an emotion word must show a tendency towards a particular emotion in most cases. I utilize the Chinese Gigaword Corpus for the verification purpose. Chinese Gigaword Corpus is a comprehensive archive of newswire text data that has been acquired from Chinese news sources, namely Agence France Presse, Central News Agency, Taiwan, Central News Service, Guangming Daily, Peoples Daily, Peoples Liberation Army Daily, Xinhua News Agency and Zaobao Newspaper. Given that the data in this work is retrieved from Sina Weibo, I only use the sub-corpus of Chinese Gigaword Corpus – Gigaword_XIN to do the search for emotion words. Gigaword_XIN includes news texts from the Xinhua News Agency of Beijing, and it contains 382,881,000 tokens. When a potential emotion word is found in the corpus but not in the list of Lee's (2010) emotion taxonomy, I first judged whether or not the word is directly referring to an emotion state. If the prerequisite is fulfilled,

I then searched the word in Gigaword_XIN and randomly extracted 10 instances containing that word. If 8 or more than 8 instances do express such an emotion, the word should be added to the taxonomy as an emotion word. Take the word 气愤 'angry' as an example. There are nine comments containing 气愤 'angry' found in the corpus, among which eight of them are expressing an ANGER emotion. The remaining one is not annotated as ANGER as 气愤 'angry' in that case is not describing the writer's emotion. Therefore, 气愤 'angry' is regarded as a potential emotion word to be added to the taxonomy. As 气愤 'angry' is defined as "furious/ indignant" which directly refers to an emotion state, I did a search in the Gigaword_XIN and randomly extracted 10 instances. Of the 10 instances, all of the them conveys an ANGER emotion. An example taken from Gigaword_XIN is given in (2).

(2) 张从顺闻讯后十分气愤，直奔山寨
Zhang congshun wen xun hou shi fen qi fen, zhi ben shan zhai
Zhang Congshun hear news later very angry, straight run cottage
'Zhang Congshun was very angry after hearing the news and went straight to the cottage'

Example (2) shows the typical use of 气愤 'angry'. The word directly and explicitly expresses an emotion state of the subject. The emotion cause as well as the emotion reaction are clearly stated. As all the 10 instances show that 气愤 'angry' has an obvious semantic orientation pointing to ANGER, it should be added to the taxonomy.

As for words to be removed from the taxonomy, I first considered their

semantic meanings. If the definition of an emotion word does not refer to an emotion state, they will be directly removed. For example, the word 炫耀 'SHOW OFF' is a complex emotion composed of HAPPINESS and FEAR. However, the word itself refers not to an emotion state but a behaviour that is intended to attract attention or admiration. The action of showing off often elicits an ANGER emotion from the others. This claim is supported by the corpus data as in (3).

(3) 这是值得炫耀的事吗
zhe shi zhi de xuan yao de shi ma
DET is worth show off POSS event SFP
'Is that something worth showing off?'

The writer of (3) expresses an ANGER emotion by questioning the person who showed off. Instead of expressing a HAPPINESS emotion, all the 3 comments containing the word 炫耀 'show off' in the corpus indeed elicits an ANGER emotion from the interlocutors. Therefore, it should be removed from the taxonomy and be regarded as an emotion-laden word.

In addition to words that do not refer to an emotional state, some ambiguous emotion words should also be removed from the taxonomy even if they do express an emotional state explicitly. Consider (4) and (5).

(4) 有关人员因无此先例而颇感为难。
you guan ren yuan yin wu ci xian li er po gan wei nan.
concerned personnel because no such precedent so a bit feel embarrassed.
'The personnel concerned are a bit embarrassed as there is no such precedent.'

(5) 别为难自己了

bie *wei nan* *zi ji* *le*

don't make things difficult oneself SFP

'Stop making things harder for yourself'

Examples (4) and (5) illustrate how 为难 'embarrass/ make things difficult' can be used in different context. In (4), it explicitly describes the emotional state of FEAR. However, 为难 'embarrass/ make things difficult' in (5) refers not to an emotional state but to the action of creating difficulties. Of the 10 instances extracted from Gigaword_XIN, only 6 of them convey the meaning of embarrassed. It suggests that the occurrence of 为难 'embarrass/ make things difficult' does not often associated with a FEAR emotion, and it should therefore be removed from the taxonomy due to the ambiguity.

With the help of Gigaword_XIN Corpus, I summarized the emotion words to be added and removed as in Table 4.2.

Table 4.2: Emotion Words to be Added and Removed

Emotion	Emotion Word(s) to be added	Emotion Word(s) to be removed
Happiness	N/A	放松，炫耀，狂
Sadness	心痛/心疼，难过，无奈，悲	无聊
Anger	恨，气愤	疯狂
Fear	N/A	为难
Surprise	N/A	N/A

I revised the Chinese emotion taxonomy based on Table 4.1, and I also moved all the first-order emotions under PANIC from SADNESS to FEAR as PANIC is a primary emotion belong to the FEAR emotion, according to Turner's (2000) taxonomy. The revised Chinese emotion taxonomy is shown as in Table 4.3.

Table 4.3: Revised Chinese Emotion Taxonomy

Primary Emotions	Variations in Intensity			First-Order Emotions	Second-Order Emotions
	High	Moderate	Low		
HAPPINESS/ 喜-*xi3*	痛快，振奮，亢奮，興奮	欣慰，高興，愉悅，欣喜，歡欣，樂，歡暢，開心，康樂，歡快，快慰，歡，舒暢，快樂，快活，歡樂，暢快，舒心，舒坦，歡娛，如意，喜悅，順心，歡悅，爽心，曉暢，鬆快，歡愉，歡喜	閒適，怡和，自在	+FEAR: Pride/傲：自傲，驕橫，驕慢，驕矜，驕傲，自負，自信，自豪，自滿，自大，自狂，得意 Relief/安心：安心，寬心，放心 +ANGER: Appeased/解恨：解恨 +SADNESS: Moved/感動：感動	
SADNESS/ 哀-*ai1*	悲慟，悲痛，哀傷，悲哀，哀痛，沉痛，痛心，悲涼，悲淒，悲切，悲傷，悲愴，哀戚，絕望，心痛/心疼	感傷，傷心，傷感，心酸, 沉悶，憋氣，鬱悒，苦悶，鬱悶，乏味，沉鬱，憋悶，憂鬱，陰鬱，悵惘, 沮喪，消沉，頹喪，頹唐，煩悶，難過	灰心，喪氣，無奈	+FEAR: Misery/悲愁：悲愁，哀愁，愁悶，惆悵 Remorseful/後悔：後悔，慚愧，抱歉，抱愧，對不起，羞愧，背悔，懊惱，懊悔，悔恨，懊喪 +ANGER: Aggrieved/委屈：委屈，冤枉，抱委屈，哀怨 Dissatisfied/不滿：不滿，不快，不悅 +SURPRISE: Disappointment/失望：失望，心寒 Embarrassed/窘：窘	+fear, anger: Guilt/疚：疚，內疚，負疚

Table 4.3: Revised Chinese Emotion Taxonomy, *continued*

Primary Emotions	Variations in Intensity			First-Order Emotions	Second-Order Emotions
	High	**Moderate**	**Low**		
FEAR/ 恐-*kong3*	惶恐，恐慌，恐懼	畏怯，心虛，心慌，害怕，怕，畏懼，發慌，發怵，驚恐，驚駭，驚惶，驚懼，嚇人，慌張，驚慌，惶惑，慌亂	羞涩，羞怯，羞慚，害羞，害臊，遲疑	+ANGER: Envy/嫉：嫉妒，妒嫉，妒忌，忌妒，嫉狠，眼紅，忌狠 +SADNESS: Anxious/急：焦慮，焦渴，焦急，焦躁，焦炙，心浮，心焦，揪心，心急，心切，着急 Worry/愁：愁，苦惱，愁苦，憂愁，發愁，擔憂，擔心，犯愁，憂慮，緊張，困惑	
ANGER/ 怒-*nu4*	憤怒，忿恨，激憤，憤懣，憤慨，忿怒，悲憤，暴怒，蔑視，瞧不起， 輕蔑，鄙夷，鄙薄，鄙視，歧視，自卑，痛恨，怨恨，憎惡，憤恨，厭煩，膩煩，惱恨，恨	生氣，氣憤，窩火，火，厭倦，討厭，厭惡，反感，敵視，衝動	煩，煩躁，煩亂，煩心，煩人，煩惱，煩雜，浮躁	+FEAR: Suspicion/疑：疑，懷疑，疑心，疑惑 +SADNESS: Bitterness/辛酸：辛酸，酸辛	
SURPRISE/ 驚-*jing1*	駭怪，駭異，震驚	詫異，吃驚，愕然，驚訝，驚奇，驚詫，驚愕	奇怪	+HAPPINESS: Delighted/驚喜：驚喜	

4.1.2 The Use of Emojis at the Semantic Level

In addition to emotion words, explicit emotions can also be expressed by means of emoticons and emojis. The term emoticon is a contraction of "emotion" and "icon" which initially refers to a series of text characters such as punctuations and symbols to represent a facial expression or gesture, such as :) as a happy face and :(as a sad face. Emoticons have been steadily replaced by emojis nowadays. For example, when someone types in the happy face :), the emoticon is automatically transformed into its corresponding emoji 😊. The term emoticon is thereafter used in a broader sense to include both text characters representing facial expressions and gestures (i.e. emoticons) and colorful pictograph representing facial expressions and gestures (i.e. emojis). Given that emoticons are rarely found in the corpus, I only focus on the use of emojis provided by Sina Weibo in the present work. Emojis play an important role in text-based cyber communication. While some emojis do have an obvious orientation pointing to a certain emotion in most cases, some are rather ambiguous as they can be used to express different emotions depending on the context. To indicate which Weibo emojis have a strong relationship with a particular emotion, I selected a number of emojis according to their frequency and unambiguousness in the emotion corpus. The selected emojis should have at least 5 occurrences in the corpus, among which 80% of the total occurrences should be expressing a particular emotion. If an emoji occurs less than 5 times in the corpus, it is not selected even if all the tokens containing that emoji express a certain emotion. The purpose of setting such a condition is to avoid selecting emojis of

high ambiguity. Moreover, there is little likelihood that a selected emoji is linked to a particular emotion just by coincidence. Following the criteria, a list of Weibo emojis are proposed as in Table 4.4.

Table 4.4: Emojis as a Representation of Emotions

Emotion	Emojis and their Built-in Names			
Happiness	 [中国赞]	 [赞]	 [赞啊]	 [good]
	 [偷笑]	 [嘻嘻]	 [笑 cry]	 [鼓掌]
	 [可爱]	 [哈哈]	 [太开心]	 [爱你]
	 [坏笑]			
Sadness	 [悲伤]	 [可怜]	 [失望]	 [伤心]
	 [蜡烛]			
Anger	 [怒]	 [鄙视]	 [哼]	 [怒骂]
	 [费解]	 [白眼]	 [微笑]	 [挖鼻]
	 [吃瓜]	 [吐]		
Fear	N/A			
Surprise	N/A			

Table 4.4 demonstrates some emojis and their built-in names and their corresponding emotions. More than 80% of the use of each emoji express a particular emotion. Emojis indicating a HAPPINESS emotion are mostly a smiling face or a thumbs-up gesture; emojis of SADNESS show a wince, a broken heart or a burning candle to mourn the death of the loved ones; emojis of ANGER mostly show an angry face, a face turning up its nose or a face showing disgust. It should be noted that the emoji 🙂 [微笑] having a literal sense of HAPPINESS is more typically used to express ANGER. Of all the 90 occurrences, 83.3% are used to convey ANGER and the remaining ones convey HAPPINESS. It is suggested that the emoji 🙂 is more often used to express ANGER than HAPPINESS. An example of the typical use of 🙂 in given in (6).

(6) 哦 听懂了向中国人的钱包道歉🙂

o	*ting dong le*	*xiang*	*Zhong guo ren*	*de*
qian bao	*dao qian*			
oh	listen-understand-ASP	to	Chinese-	POSS
wallet	apologize			

'Oh, I got it. Apologized (to Chinese) for their money 🙂'

Although this emoji shows a smiling face, it is typically used to express an ANGER emotion. (6) is a comment taken from a post concerning the founders of Dolce and Gabbana apologized to the Chinese for using racist in an advertising campaign. The writer is irritated by the company for being racist. He/ she believes that the apology made by the founders of Dolce and Gabbana is not a genuine one. The company

apologized only because Chinese consumers are responsible for a large number of luxury market sales. 🙂 in (6) is used to show disdain. When 🙂 is used to convey ANGER, it can function as a sneer, a chuckle, a snigger, or a sly smile instead of a genuine smile. The atypical use of this emoji will be discussed in detail in Section 5.3.2.

From Table 4.4, it is surprising that no emoji of SURPRISE and FEAR meets the criteria for unambiguousness and frequency. For example, although 😯 [吃惊] occurs 10 times in the corpus, half of them are used to express SURPRISE and the other half are used to express FEAR as in (7) and (8).

(7) 搬地方啦😯😯

ban *di fang* *la*
move place SFP

'(It's) moved to another place 😯😯'

(8) 😯敲响警钟，以后走路得多抬头少低头了

qiao xiang *jing zhong*，*yi hou* *zou lu* *de* *duo* *tai tou* *shao* *di tou* *le*
ring alarm bell, future walk need more raise head less bow head SFP

'😯 (It) rings alarm bells (in my mind), got to walk more carefully in the future'

The emoji 😯 in (7) shows that the writer is SURPRISED as something has moved to another place, and the one in (8) expresses a FEAR emotion. The background information of (8) is that a person stumbled because he/ she was playing with his/ her phone, and the event gives the writer a real scare. Given that 😯 does not show

an obvious tendency towards a particular emotion, it is not included in the proposed list. Another example is the emoji [疑问]. Although all the 4 tokens containing express SURPRISE, it is excluded due to its low occurrence.

It is observed that emojis are typically placed at the end of a sentence and are very often placed at the end of the last sentence. Emojis are mainly used to strengthen the emotion expressed in text, or to portray an emotion when there are no other cues that convey such an emotion as in (9) and (10), respectively.

(9) 太过分了，

tai	*guo fen*	*le*
so	go too far	SFP

'That's gone too far, '

(10) 人身伤害这已经构成犯罪了

ren sheng shang hai zhe	*yi jing*	*gou cheng*	*fan zui*	*le*
bodily injury	DET	already	constitute	crime	SPF

'(Causing) bodily injury is already an offence '

In (9), although there is no emotion word in the first sentence, it expresses an ANGER emotion implicitly by means of the word 过分 'go too far'. Therefore, the emojis are simply used to strengthen the intensity of ANGER. As for (10), the text per se is a general statement which does not convey any emotions at all. It is the emoji that portrays the writer's ANGER emotion.

Although those emojis listed in Table 4.4 are highly associated with a particular emotion, they may sometimes be used in an atypical way. That is, the emotion expressed via the emoji and the one expressed via its accompanying

linguistic text are at odds. For example, the emoji  showing a slightly smiling face is typically regarded as an emoji expressing HAPPINESS. However, it is observed that  is frequently used ironically to express an ANGER emotion. Given that the overall emotion depends not only on the emoji but on the interaction between the emoji and its accompanying text, the atypical use of emojis is beyond the semantic level. Therefore, it will be discussed in detail in Section 5.3 which deals with implicit emotion expression at the discourse level.

4.2 Expressions of Implicit Emotions

Different from explicit emotion which is expressed by means of emotion words or emojis representing a certain emotion, implicit emotion refers to emotion-related information conveyed through inference or connotation. At the semantic level, words that implicitly express emotions can be classified into two types, one is emotion-related words, and the other one is emotion-laden words.

According to Pavlenko (2008: 148), emotion-related words refer to "behaviors related to particular emotions without naming the actual emotions". Although Pavlenko (2008) did not further clarify the term, she gave a couple of examples, namely tears, tantrum and to scream. All the examples given are the actions or reactions taken by experiencers after experiencing a certain emotion, i.e. post-events of emotions. Ng et al. (2019) further classified these behaviors into facial expression, bodily symptoms, and action tendencies. Emotion-laden words

are defined in Pavlenko (2008: 148) as words that "do not refer to emotions directly but instead express ("jerk", "loser") or elicit emotions from the interlocutors ("cancer", "malignancy")".

4.2.1 The Use of Emotion-related Words

Unlike emotion words, emotion-related words are not as unique as emotion words to particular emotions. For example, the action of crying is generally deemed to be a post-event of SADNESS. However, it is not uncommon that crying can also be the post-event of HAPPINESS or FEAR. As for the HAPPINESS emotion, people may cry tears of joy, or cry because they are MOVED. As for FEAR, people may cry because they are worried about something or scared by something. Thus, one should note that most emotion-related words proposed in this section do not solely have a close link to a particular emotion, but only show a tendency towards a particular emotion. In order to know which emotions are more likely to have their post-events stated in an emotion expression, Table 4.5 shows the occurrence of post-events of each emotion.

Table 4.5: Occurrence of Post-events of each Emotion

	Occurrence of Post-events	Total no. of comments	Percentage
Happiness	778	1,896	41.0%
Sadness	316	1,780	17.8%
Anger	458	5,730	8.0%
Fear	94	573	16.4%
Surprise	54	698	7.7%
Total	1,700	10,677	15.9%

Of the 10,677 comments, only 15.9% of comments contain a post-event as shown in Table 4.5. Among the five emotions, comments expressing a HAPPINESS emotion are more likely to contain a post-event accounting for 41.0% of all the comments expressing HAPPINESS, followed by SADNESS (17.8%), FEAR (16.4%), and ANGER (8.0%). To my great surprise, comments expressing SURPRISE is least likely to have a post-event stated in the expressions with only 7.7% of all the 698 comments. To further investigate what kinds of post-events (i.e. emotion-related words) are more likely to link with a certain emotion, I summarize a list of emotion-related words for each emotion in Table 4.6- Table 4.9. However, as the SURPRISE emotion is least likely to have its post-events being stated in emotion expressions, there are insufficient statistics to support any claims regarding post-events of SURPRISE. Table 4.6 summarizes emotion-related words of HAPPINESS:

Table 4.6: Emotion-related Words of Happiness

Category	Emotion-related Words	Happiness	Total	Percentage
A. Showing Admiration	點讚/[讚啊][讚]/	358	369	97.0%
	[中國讚]/[good]			
	[鼓掌]	19	21	90.5
	打 call/ 打電話	6	7	85.7%
Total		383	397	96.5%
B. Laughing	想笑	11	11	100%
	[壞笑]	5	5	100%
	[笑 cry]	56	68	82.4%
	[偷笑]	9	11	81.8%
	笑死	17	21	81.0%
	[哈哈]/ [笑哈哈]	35	45	77.8%
Total		133	161	82.6%
C. Showing Gratitude	謝謝/ 感謝	22	28	78.6%
Total		22	28	78.6%
D. Crying	想哭	10	14	71.4%
Total		10	14	71.4%
E. Blessing	祝/ 祝福	32	48	66.7%
Total		32	48	66.7%
F. Supporting	支持	19	32	59.4%
Total		19	32	59.4%

As shown in Table 4.6, a post-event can be either stated using emojis, or simply using words. There are mainly six types of post-events that show a tendency towards HAPPINESS, as compared to the other four emotions.

The most frequent post-events of HAPPINESS are to show admiration. It is often found that writers tend to show admiration to people for their selfless devotion to duty by using the emojis of thumbs-up or clapping hands. Consider (11).

(11) 咱们巴中人民真是棒🥹🥹

zan men	*ba zhong ren min*	*zhen shi*	*bang*
1.PL	Bazhong people	really	awesome

'We Bazhong people are really awesome🥹🥹'

The background information of (11) is that 50 Bazhong people collaborated to save the driver who was injured in an accident. As a Bazhong person, the writer is so 'proud' of them because of what they did. The emojis 🥹🥹 are a post-event of HAPPINESS to show admiration for the 50 Bazhong people who saved the driver's life. Regarding the supporting event 打 call / 打電話 'cheers', it should be noted that the event refers not to its literal meaning of being on the phone but the action of audience showing admiration and supports to performers performing on stage with actions such as screaming, clapping hands, moving one's body following the rhythm. It has now been extended to a broader sense of showing admiration to someone. An example is given in (12).

(12) 为大海安打 call 😁

wei	*da* *hai an*	*da call*
for	Hai'an	make a call

'Cheers for the (the traffic police officer) in Hai'an😁'

Example (12) is a comment taken from a post concerning a traffic police officer in Hai'an who enforced the law boldly. As the writer is HAPPY with what the police officer did, he/she cheers for the police officer to show his/her admiration or approval.

As for the action of laughing, one may typically think that the action of laughing must be very closely associated with HAPPINESS. However, it is observed that approximately 16.5% of the laughing action are not related to HAPPINESS but to other emotions. Part of the reason is because some of them are used to express a complex emotion of SURPRISE 驚喜 'DELIGHTED' and another reason is because the action of laughing can be used ironically or in an unkind way to express an ANGER emotion as in (13).

(13) 调查中 调查中 真的笑死人了 真真真是笑死人了 看来现在这个社会还是做恶人的好……
diao cha zhong diao cha zhong zhen de xiao si ren le
zhen zhen zhen shi xiao si ren le kan lai xian zai
zhe ge she hui hai shi zuo e ren de hao
investigating investigating really laugh die person SFP
really really really laugh die person SFP seem now
DET CL society still be bad person good
'Still investigating, still investigating. That's so ridiculous. That's so so so ridiculous. It seems that being a bad person in this society nowadays is even better…'

The action of laughing in (13) is not a typical post-event of HAPPINESS. The writer repeats the phrases 调查中 'investigating' and 笑死人 'laugh to death' twice. The former is the emotion cause that triggers his/ her ANGER emotion; the latter indicates that he/ she was ridiculed for the event still being under investigation. Therefore, the action of laughing in that case is not a post-event of HAPPINESS but an action that the writer took to imply his/ her ANGER. It suggests that the action of laughing is typically not necessarily a post-event of HAPPINESS.

The actions of showing gratitude, crying, blessing, and supporting are other typical post-events of HAPPINESS. As for the crying action, it only includes the action

of wanting to cry but not an actual cry. It is observed that when people only state that they want to cry but are not actually crying, they are experiencing a MOVED emotion, which belongs to HAPPINESS. Consider (14).

(14) 听到这句话，感动得想哭🫵

ting dao zhe ju hua, gan dong de xiang ku
hear DET CL speech, moved want cry

'*Hearing these words, I am so moved that I want to cry*🫵'

The writer of (14) directly indicates that he/ she is so MOVED. The emotion cause (i.e. pre-event) is the event of hearing these words, and the post-event is the reaction of wanting to cry.

A summarization of post-events triggered by sadness is shown as in Table 4.7:

Table 4.7: Emotion-related Words of Sadness

Category	Emotion-related Words	Sadness	Total	Percentage
A. Begging	求/ 求求	4	6	66.7%
Total		4	6	66.7%
B. Praying	祈祷/[祈祷]/祈福	45	69	65.2%
Total		45	69	65.2%
C. Crying	😭[泪]/ 哭/ (眼)泪	119	214	55.6%
Total		119	214	55.6%
D. Showing Compassion	🥹 [允悲]	79	165	47.9%
Total		79	165	47.9%

As shown in Table 4.7, there are four kinds of post-events of SADNESS, namely the actions of begging, praying, crying, and showing compassion. For the begging

event, it only includes the informal event of begging but not the official requests indicated by 請求 'request' or 要求 'request'. The action of praying is often found when the posts are concerned with an injured person. Writers are SAD to see people suffering from a great deal of misfortune. Therefore, they can do nothing but to pray for them. As for the action of crying, more than a half of them (55.6%) are related to SADNESS. Most of the remaining ones are found in the emotion expressions of HAPPINESS which are the reaction of people experiencing the complex emotion MOVED. It is found that the crying events evoked by SADNESS are more likely to be expressed by means of emojis, while the crying events elicited by HAPPINESS are more likely to be expressed in a descriptive way, such as 熱淚盈眶 'tears in one's eyes', 淚流不止 'shed tears', 淚目 'shed tears'. The reason for the use of descriptive expression may be due to the complexity of the MOVED emotion which can hardly be expressed simply by emojis.

As for the category of showing compassion, the emoji 🥲 [允悲] is officially defined as 'allowing me to show a sad face though I really want to laugh'. However, users generally use an emoji without digging deeper into the definition of an emoji. As the emoji 🥲 shows a facial expression of both a smile and tears, and an act of covering its face with its hand, it conveys emotions of being HAPPY, EMBARRASSED, ANNOYED, or DISAPPOINTED about something. Thus, 🥲 is a post-event that can only be summarized as having a tendency but not solely linked to SADNESS.

Comparing to the other four emotions, emotion expressions of ANGER have a wider variety of post-events as in Table 4.8.

Table 4.8: Emotion-related Words of Anger

Category	Emotion-related Words	Anger	Total	Percentage
A. Condemning	骂/ [怒骂]	7	7	100%
	想骂/不想骂/可以骂人嗎	5	5	100%
Total		12	12	100%
B. Humph	[哼]/ [右哼哼]	9	9	100%
Total		9	9	100%
C. Beating	(想) 打 / 揍 / 抽 / 扇 / 給…巴掌/回敬…巴掌	31	32	96.9%
Total		31	32	96.9%
D. Requesting	要求/請求	7	8	87.5%
Total		7	8	87.5%
E. Questioning	問	22	26	84.6%
Total		22	26	84.6%
F. Smiling	[微笑]	75	90	83.3%
Total		75	90	83.3%
G. Rejecting/ not doing something	拒絕	5	6	83.3%
	不會……	8	11	72.7%
	不可…….	7	7	100%
	不想……	8	11	72.7%
Total		28	35	80.0%
H. Suggesting/Urging	呼吁	4	5	80%
	建议	15	19	78.9%
Total		19	24	79.2%

Table 4.8 illustrates 8 types of post-events which most frequently found in expressions of ANGER. When a person conducts the action of condemning or has an intent to scold someone, that person is very likely experiencing an ANGER emotion. To vent one's ANGER, one may also make a sound of humph. Another frequent post-event of ANGER is the action of beating. It is one of the most common actions that

one may take when an outburst of ANGER is provoked. Moreover, making a request is another category that is often found to be connected to the ANGER emotion. The action of making a request mentioned in Table 4.8 is different from the action of begging mentioned in Table 4.7. Making a request is a formal or official request indicated by words such as 要求 'request' and 請求 'request', but begging is just an informal action asking for something. As for the action of questioning, writers often use the pattern "(我)(只/就)想問……" '(I) (just) want to ask' or "請問……" 'may I know…' to raise a question(s) triggering their ANGER emotion. Consider (15).

(15) 我就想问这豆腐渣工程，有关部门的官员到底贪了多少？
wo jiu xiang wen zhe dou fu zha gong cheng, you guan bu men de guan yuan dao di tan le duo shao?
I just want to ask DET jerry-built project, relevant department official in fact embezzle how much?
'I just want to ask: how much have the relevant officials embezzled from this jerry-built project?'

It is observed that implicit emotions are often expressed via its pre-event or post-event. For example, the ANGER emotion in (15) is implicitly expressed by the action of questioning as well as information given in the question. The action of questioning is the post-event of ANGER. It is formed with the structure "我就想問……" 'I just want to ask…' to bring up the pre-event(s) provoking an ANGER emotion, i.e. 豆腐渣工程 'jerry-built project', and 官员贪了 'corruption'.

As I mentioned in Section 4.1.2, 🙂 [微笑] is more typically used to express ANGER instead of HAPPINESS. Although 🙂 [微笑] shows a smiling face, the emoji 🙂 [微笑] tends to be a post-event of ANGER. Consider (16).

(16) 欺负学生 这女的我能把她头发扯烂🙂🙂

qi fu *xue sheng* *zhe nu de* *wo* *neng* *ba* *ta tou fa* *che lan*

bully student this female I can make her hair tear broken

'This woman bullied (her students), I can pull her hair 🙂🙂'

The writer of (16) is ANGRY with the woman for bullying her students. He/ she explicitly states that he/ she wants to pull the woman's hair. The emojis 🙂🙂 in (16) is a contempt smile. A contempt smile is composed of a mixture of disgust and resentment. The emojis 🙂🙂 in (16) may function as a sneer to show disrespect to the woman who bullied her students. Although the smile is disconcertingly similar to a genuine smile, the tightened lips indeed indicate that the smile has a secretive attribute and it is often a post-event of ANGER. Other categories such as rejecting or being reluctant to do something and suggesting/urging are also good indicators of ANGER.

Due to the small number of comments expressing FEAR, there are not many post-events of FEAR to be summarized in Table 4.9.

Table 4.9: Emotion-related Words of Fear

Category	Emotion-related Words	Fear	Total	Percentage
A. Trembling	瑟瑟发抖	3	3	100%
	背脊发凉	1	1	100%
	头皮发麻	2	2	100%
Total		6	6	100%
B. Dare not do something	不敢……	27	30	90.0%
Total		27	30	90.0%

Table 4.9 shows two types of post-events of FEAR, namely the action of trembling, and the state of dare not do something. Both types of post-events are highly likely to be associated with the FEAR emotion. This is because it is very common for one to tremble with FEAR as well as avoid doing something that they have a FEAR of.

As for emotion expressions of SURPRISE, only 55 post-events are found in the corpus. Therefore, it is rather difficult to summarize a list of emotion-related words. Even if an action/ reaction occurs several times as the post-event of SURPRISE, the same action still occurs more frequently as the post-event of other emotions. For example, the action of wanting to know formed with "想知道…" 'want to know…' occurs 3 times as the post-event of ANGER. However, the total occurrence of the action is 10, meaning that the action of wanting to know does not show a significant tendency towards the SURPRISE emotion. The only conclusion to be made is that emotion-related words are inadequate for the identification of SURPRISE.

Apart from emotion-related words, emotion-laden words are another type of lexical items that convey emotion information without directly naming the emotion. In the following section, the use of emotion-laden words for the expressions of each emotion type will be presented.

4.2.2 The Use of Emotion-laden Words

According to Pavlenko (2008: 148), emotion-laden words are defined as words that "do not refer to emotions directly but instead express ("jerk", "loser") or elicit emotions from the interlocutors ("cancer", "malignancy")". She further classified emotion-laden words into six types, namely (1) taboo and swearwords or expletives, (2) insults, (3) (childhood) reprimands, (4) endearments, (5) aversive words, and (6) interjections. Emotion-laden words are especially important for implicit emotion identification as some categories are a component of an emotion expression and some are a component of a pre-event. First, taboo and swearwords or expletives, (childhood) reprimands, insults, and interjections are often used to express emotion implicit. For example, "*Behave yourself!*" connotes an ANGER emotion with none of the individual words in the sentence expressing such an emotion. The example is an emotion expression, and the word "behave" is an example of (childhood) reprimand given in the work of Pavlenko (2008). Second, insults and aversive words are a component of a pre-event triggering the emotion. In this work, a pre-event is loosely defined as the cause of an emotion. Most emotion theories regard the recognition of emotion cause as an integral part of emotion elicitation (James 1884; Plutchik 1980; Wierzbicka 1999). Previous studies generally regarded emotion cause (i.e. pre-event) as the cause that evoke an emotion (Lee et al. 2010, 2013a, 2014; Gui et al. 2014, 2016, 2017; Liu et al. 2019). While Lee et al. (2010) defined it as an immediate cause of the emotion, which can be an actual trigger event or a perception of the trigger event, Gui et al. (2017) placed more emphasis

on the relation between the emotion word and emotion cause. Following Lee et al. (2010), pre-events are loosely defined in this work for two reasons. First, the dataset collected in this work include both a post (i.e. event) and comments, and the comments were made about the post. The actual cause that triggers an emotion is often referred to in the post and is not directly mentioned in most of the comments. Consider (17).

(17) 从未见过如此厚颜无耻之徒!

cong wei	*jian guo*	*ru ci*	*hou yan wu chi*	*zhi*	*tu!*
never	see	such	shameless	POSS	person!

'Have never seen such a shameless person!'

Example (17) expresses an ANGER emotion and the pre-event (i.e. emotion cause) of the emotion is 厚颜无耻之徒 'a shameless person'. Although the actual cause is not mentioned in the comment, the ANGER emotion is hinted by the adjective 厚颜无耻 'shameless'. The adjective 厚颜无耻 'shameless' is an emotion-laden word of ANGER. In (17), it is used to describe the person who has done something unethical but still being nominated for a "Good Person Award" as mentioned in the post. Although the direct cause of the ANGER emotion is not exactly the 厚颜无耻之徒 'a shameless person' but the behaviour done by that person, the nominal is the perception of the trigger event. Given that the actual cause of an emotion is not mentioned in most comments, pre-events are loosely tagged in the present work. Therefore, emotion-laden words are often marked as part of a pre-event as they indicate how writers evaluate an event. Therefore, they serve as a linguistic cue for the identification of implicit emotions.

Second, the present study investigates both explicit and implicit emotions with the major focus being placed on implicit emotions. Due to the absence of emotion keywords, implicit emotions are often expressed by means of a pre-event as in (17). In order to investigate how implicit emotions can be expressed by means of emotion-laden words, Table 4.10 shows the occurrence of pre-events of each emotion.

Table 4.10: Occurrence of Pre-events of each Emotion

	Occurrence of Pre-events	Total no. of comments	Percentage
Happiness	1,609	1,896	84.9%
Sadness	1,549	1,780	87.0%
Anger	4,726	5,730	82.5%
Fear	449	573	78.4%
Surprise	616	698	88.3%
Total	8,949	10,677	83.8%

As shown in Table 4.10, 83.8% of the total number of comments contain a pre-event. This suggests that an emotion mostly occurs with its pre-event explicitly expressed in text, which confirms the prominent role of pre-events in expressing an emotion. Among the five emotions, pre-events of SURPRISE are most frequently being stated in the comments (88.3%), followed by SADNESS (87.0%), HAPPINESS (84.9%), ANGER (82.5%), and FEAR (78.4%). This section briefly explains the reason why pre-events is loosely defined and annotated in the present work. Some kinds of emotion-laden words such as insults and aversive words are often (but not necessarily) the perception of the trigger event of an emotion. These words unveil how writers evaluate an event and give a hint about the implicit emotion expressed

in a comment. In addition, some kinds of emotion-laden words such as interjections, taboo and swearwords or expletives etc. are also the component of an emotion expression. Therefore, studying the relation between emotion-laden words and emotions are of vital importance to the identification of implicit emotion. Whenever a certain emotion-laden word appears in a comment, the corresponding emotion can be easily identified even without the presence of any emotion keywords.

According to Pavlenko (2008), emotion-laden words can be further categorized into six types, namely (1) taboo and swearwords or expletives, (2) insults, (3) (childhood) reprimands, (4) endearments, (5) aversive words, and (6) interjections. It is somewhat confusing as these terms are not clearly defined in Pavlenko (2008). For example, “piss” and “shit” are the examples of taboo and swearwords or expletives and “behave” and “stop” are the examples of (childhood) reprimands given by Pavlenko (2008). However, it is hard to tell the differences between the two subtypes even with the examples given. Pavlenko (2008) confessed that the boundaries of these subtypes are fuzzy for two reasons: (1) some words may belong to more than one category, (2) some words can be regarded as emotion-laden words only if they appear in certain contexts. To avoid confusion, I redefined all the terms so that an emotion-laden word does not have to be repeatedly placed under different categories. All the categories proposed in Pavlenko (2008) remain in this work. Furthermore, it is found that such a classification is insufficient for the five basic emotions. Among the six types of emotion-laden words, four of them are specific to negative emotions, in particular the ANGER emotion, namely

taboo and swearwords or expletives, insults, (childhood) reprimands, and aversive words. In other words, endearments and interjections are the only types that are possibly related to the HAPPINESS emotion. Furthermore, emotion-laden words connoting HAPPINESS are not necessarily endearments or interjections. It is found that some evaluation words connoting a HAPPINESS emotion are unable to be categorized into any of the six types, such as 可愛 'adorable', 暖心 'thoughtful' etc. Similar to the case of HAPPINESS, emotion-laden words of SADNESS, SURPRISE or FEAR do not necessarily need to be aversive words that shows a strong emotion of dislike. In view of this, I propose two additional types of emotion-laden words to cater to the need of other emotions, namely 'favourable words' and 'unfavourable words'. Each category is defined as in Table 4.11.

Table 4.11: Definition of the Types of Emotion-laden Words

Category	Definition
1. Interjections	A word used to show a short and sudden expression of emotion
2. Endearments	A word that is used to show your love to someone
3. Taboo and Swearwords or Expletives	A foul language/ speech that is considered offensive and rude
4. Insults	An offensive remark used to name or refer to a person
5. (Childhood) Reprimands	A word that is used in an expressive way to express to someone your strong disapproval of them
6. Aversive Words	An unpleasant event or noun that objectively elicits an emotion from interlocutors
7. Favourable Words	A positive adjective used to subjectively evaluate an event or a person that elicits an emotion from interlocutors
8. Unfavourable Words	A negative adjective used to subjectively evaluate an event or a person that elicits an emotion from interlocutors

For a word to be considered an emotion-laden word of a particular emotion, 80 percent of the tokens containing that word have to be an emotion expression of that particular emotion. Table 4.12 illustrates the emotion-laden words of each emotion type.

Table 4.12: Emotion-laden Words of each Emotion Type

Emotion	Types of Emotion-laden Words (Percentage)
Happiness	**A. Favourable words** 大快人心 (100), 好样的 (100), 表扬 (100), 模范 (100), 暖心 (100), 舒适 (100), 可爱 (93.5), 懂事 (90.9), 棒 (88.4), 感人/感触 (83.3), 善举 (83.3), 天使 (80.0)
Sadness	**A. Unfavourable words** 悲剧 (100), 惨痛 (100), 苦逼 (100), 飞来横祸 (100), 自愧不如 (100), 惋惜 (100), 无助 (100), 可怜 (91.8), 倒霉 (90), 可惜 (89.2) **B. Interjections** 唉 (81.8)
Anger	**A. Unfavourable words** 蛮横 (100), 恶心 (100), 恶毒/歹毒 (100), 厚颜无耻 (100), 丧心病狂 (100), 无耻 (100), 缺德/没道德 (100), 没素质 (100), 无法无天 (100), 可笑 (100), 可恨 (100), 可恶 (100), 可耻 (100), 令人发指 (100), 恶毒 (100), 毫无底线 (100), 无良 (100), 过分 (97.6), 无语 (95.2), 丢人/丢人现眼 (94.3), 无知 (92.3), 笑话 (90.9), 丢脸 (90.0), 腐败 (90.0), 自私 (88.9), 嚣张 (84.4), 顽皮/调皮 (83.3) **B. Taboo and swearwords or expletives** 尼玛 (100), 特么/TM/tm (93.6), 特妈的/他妈的/妈的/ tm 的/tmd/TMD/md/MD (85.4), 我擦 (83.3) **C. Insults** 恶魔 (100), 禽兽 (100), 垃圾(excluding the literal meaning of 'garbage') (100), 老毛子(100), 神经病 (100), 死孩子 (100), 母狗 (100), 巨婴 (100), 傻逼 (100), 人渣/渣渣 (100), 小人 (100), 走狗 (100), 畜生 (97.7),泼妇 (92.9), 恶人 (92.3), 熊孩子 (92.1), 奇葩 (91.7), 流氓 (90.0), 弱智/智障(85.7), 败类 (84.6), 无赖 (83.3), 家伙 (81.8), 变态 (81.0) **D. (Childhood) reprimands** 放屁(100), 不要脸 (98.0), 有病 (96.3), 滚/滚蛋 (95.7), 该死 (94.4), 活该 (80.9) **E. Aversive words** 人贩子 (100), 贪污/贪腐 (100), 屠杀 (100), 屁事 (100), 强奸 (100), 抢劫 (100), 狡辩 (100), 歹徒 (100), 酒驾 (100), 钓鱼执法 (100), 豆腐渣工程 (100), 造谣 (100), 炫耀(100), 害人 (91.4), 报复 (88.8), 戾气 (87.5), 坑 (85.7), 耍赖 (83.3), 杀人犯 (81.8), 质疑 (81.8), 背锅 (81.3), 猥亵 (80.0) **F. Interjections** 呸 (100), 呵/呵呵 (93.7) **G. Favourable words** 理直气壮 (100)
Fear	**A. Unfavourable words** 细思恐极 (100), 提心吊胆 (100), 触目惊心 (100), 胆战心惊 (100), 兢兢战战 (100)
Surprise	N/A

As shown in Table 4.12, HAPPINESS is expressed by favourable words instead of endearments. Endearments are not found as the comments in the corpus are made on news articles by writers but not a direct message sent to a person they know. The favourable words of HAPPINESS are used to evaluate an event or a person who did something. Events or people being evaluated are the emotion cause (i.e. pre-event). Consider (18).

(18) 大快人心啊，无视法律必被法律无视！
da kuai ren xin *a* *wu shi* *fa lu* *bi* *bei* *fa lu* *wu shi*
fabulous INTJ ignore law must PAS law ignore
'Fabulous! Who broke the law must be governed by the law!'

The favourable word 大快人心 'fabulous' in (18) implies a HAPPINESS emotion. It is an evaluation made on the event of a person who broke the law first and being governed and punished by the law later. Thus, whenever 大快人心 'fabulous' is used, the HAPPINESS emotion as well as the actual cause of the emotion can easily be identified. However, a favourable word does not necessarily link to a positive emotion. For example, the favourable word 理直氣壯 'speak justly with a compelling argument' is an emotion-laden word connoting an ANGER emotion as in (19).

(19) 现在的人自己做错事情都这么理直气壮了？
xian zai de ren *zi ji* *zuo cuo* *shi qing dou zhe me* *li zhi qi zhuang* *le?*
now person oneself do wrong thing all that speak assuredly SFP?
'People nowadays do not even feel ashamed about what they did, and they can still argue with the others like that?'

The favourable word 理直气壮 'speak assuredly' in (19) is positive adjective but often being used to describe someone who is completely unreasonable but still argues with the other with seemingly compelling arguments. Therefore, there is higher probability that the word implies an ANGER emotion than a HAPPINESS one.

As for the SADNESS emotion, some unfavourable words and the interjection 唉 'sigh' is found. Again, the unfavourable words are used to evaluate an event of a person. For example, the word 惨痛 'agonizing' is used to describe an event causing extreme physical or mental pain. When a writer uses it to describe an event, he/ she often shows a compassion to the affected party.

The ANGER emotion has the widest variety of emotion-laden words as compared to the other four emotions. An ANGER emotion can be identified by certain types of emotion-laden words, in particular unfavourable words, insults and aversive words. An example of an insult is given in (20).

(20) 什么垃圾交警!
shi me *la ji* *jiao jing!*
what trash traffic police officer!
'What trash the traffic police officer is!'

The writer of (20) uses the insult 垃圾 'trash' to describe the traffic officer as he/ she was ANGRY with the judgement pronounced by the traffic officer. According to the percentage shown in Table 4.12, whenever the word 垃圾 'trash' is used to refer to a person, it is certain that the writer is probably expressing an ANGER emotion. As for aversive words, they are events or nouns that elicits an emotion from interlocutors, such as 强奸 'rape', 歹徒 'gangster' etc.

All the unfavourable words of FEAR connote a meaning of “scaring”. The words do not directly express an emotion but describe how such an event frightens the writers.

Although some words such as such as 神奇 ‘magical’, 厲害 ‘fabulous’ are often found in the expression of SURPRISE, they do not show an obvious tendency (greater than 80%) towards the SURPRISE. It is observed that pre-events of SURPRISE are often expressed by rhetorical questions instead of emotion-laden words. Therefore, it can be concluded that SURPRISE is not productively expressed at the semantic level.

4.3 Summary

This chapter investigates how different explicit and implicit emotions are expressed at the semantic level. Explicit emotions are generally encoded with emotion words or emojis. I revise the Chinese emotion taxonomy by adding additional emotion keywords to the taxonomy and removing those unreliable or ambiguous ones from it. I also propose a list of emojis used in Weibo which have an obvious orientation pointing to a particular emotion.

Statistics confirm the claim that implicit emotion does play an important role in emotion expressions. It is a necessary yet underdeveloped component in emotion studies. Therefore, I focus on implicit emotion in Section 4.2. Implicit emotions are typically expressed by means of emotion-related words and emotion-laden words.

Emotion-related words refer to behaviours related to a particular emotion, i.e. post-events of emotions. I summarize the most frequent post-events of each emotion and group them into different categories. The post-events are some emotion-related words that show a tendency towards a particular emotion. As for emotion-laden words, they are words that express emotions or elicit emotions from the interlocutors without naming an emotion directly. An emotion-laden word is often a component of a pre-event which gives a hint about the implicit emotion expressed in text. In addition to the six types of emotion-laden words proposed by Pavlenko (2008) which particularly favour the ANGER emotion, I add two types for the sake of other emotions. The two types are 'favourable words' and 'unfavourable words'. Results show that the ANGER emotion is most frequently being expressed by means of emotion-related words and emotion-laden words, whereas the SURPRISE emotion is rarely expressed by emotion-related words or any kinds of emotion-laden words at the semantic level.

CHAPTER 5

EMOTION EXPRESSIONS: A SYNTACTIC AND DISCOURSE PERSPECTIVE

The previous chapter discusses emotions expressed at the semantic level. This chapter examines emotions at the syntactic and discourse level. At the syntactic level, I first focus on syntactic patterns that are designated for a particular emotion. I examine words of different parts-of-speech that are seemingly not semantically related to emotions but are used to express a particular emotion from a syntactic perspective. I then conduct an in-depth analysis on the use of rhetorical questions in emotion expressions. Based on the corpus data, I argue that rhetorical questions are a rather productive means used to convey emotion implicitly. I further classify rhetorical questions into 14 different types of questions and show how frequent an emotion is expressed by means of different types of rhetorical questions.

At the discourse level, I examine the interplay between emoji and its accompanying linguistic text when the two channels are at odds. In general, certain emojis are typically regarded as an explicit way to convey a certain emotion. However, it is observed that even emojis of low ambiguity may not explicitly convey a corresponding emotion in a message. I argue that the interplay between an emoji and its accompanying text may also influence the expressed emotions. It is an important component of implicit emotion studies that should by no means be overlooked.

Section 5.1 proposes various syntactic structures that are frequently used to express a particular emotion. Some words per se do not have a clear semantic orientation pointing to a particular emotion, but when they interact with certain components in a sentence, these words become a good emotion indicator. Section 5.2 discusses the use of rhetorical questions in emotion expressions. Rhetorical questions are classified into 14 subtypes, including both open class questions and close class questions. The interaction between different types of rhetorical questions and emotions are explored. Section 5.3 presents the the interplay between emoji and its accompanying linguistic text when the two channels are at odds. A summary of the main findings is given in Section 5.4.

5.1 Syntactic Structures Used in Emotion Expressions

In addition to the three types of emotion-bearing words mentioned in the previous chapter, emotions can also be expressed by means of syntactic structures. Although some words per se are emotionless in nature, they are associated with a particular emotion when they interact with certain components in a sentence. In the following subsections, I discuss some of the syntactic patterns that contain certain words of different parts-of-speech.

Adverbs are words that give more information about a verb, an adjective, or another adverb. Although some of them do not have any association with an emotion, it is observed in the corpus that adverbs do play an essential role in implicit

emotion expressions when they interact with other components in the sentence. This claim is also supported by Lee (2015) who indicates that adverbs are of great importance in implicit emotion studies. Table 5.1 shows the adverbs that I found to be extensively used in the expressions of a particular emotion.

Table 5.1: The Occurrence of Adverbs as an Emotion Indicator

	Occurrence (Normalized Frequency per 1,000,000)				
Adverbs	Happiness	Sadness	Anger	Fear	Surprise
明明	0 (0)	2 (37.3)	31 (200.2)	0 (0)	0 (0)
簡直	1 (27.4)	4 (74.6)	35 (226.0)	2 (140.7)	0 (0)
根本	1 (27.4)	8 (149.3)	39 (251.9)	2 (140.7)	0 (0)
居然	2 (54.8)	3 (56.0)	23 (148.5)	0 (0)	35 (2492.2)
竟然	4 (109.6)	5 (93.3)	14 (90.4)	1 (70.3)	18 (1281.7)
原來	1 (27.4)	6 (112.0)	11 (71.0)	1 (70.3)	20 (1424.1)

Table 5.1 shows that there is a stronger correlation between certain adverbs and a particular emotion. For example, 明明 'obviously' has the strongest connection with ANGER emotion as compared with other emotions. As the total numbers of comments containing different emotions vary, the occurrence of an adverb should undergo the normalization process in order to make a comparison among them. To normalize, I treat all the comments of each given emotion as an individual dataset. The data size of each emotion is shown in Table 5.2.

Table 5.2: Data Size of Each Emotion

	Happiness	Sadness	Anger	Fear	Surprise
No. of Tokens	1,896	1,780	5,730	573	698
No. of Words	36,490	53,593	154,850	14,219	14,044

The normalized frequency is generalized by:

(1) $F_N=F_O(10^6)/C$

where, F_N is the normalized frequency, F_O is the occurrence, and C is the total number of words of each given emotion.

According to the normalized frequencies shown in Table 5.1, three adverbs including 明明 'obviously', 簡直 'absolutely', and 根本 'at all' are more frequently used to express an ANGER emotion as compared to the other four emotions. The adverb 明明 'obviously' can be used to express ANGER, SADNESS, and SURPRISE based on the corpus data, but the normalized frequency is the highest in the ANGER dataset. Consider (2).

(2) 明明受苦受难的是巴勒斯坦人，却有那么多支持加害者的，什么情况？
ming ming shou ku shou nan de shi ba le si tan ren, que you na me duo zhi chijia hai zhe de, shen me qing kuang?
clearly suffered is Palestinian, but has that many support perpetrator, what situation?
'It was the Palestinian who's been suffering. Why on earth are there so many people supporting the perpetrator?'

It is observed that the adverb 明明 'obviously' is often found in the structures of "明明 'obviously' ……卻/ 偏 'but'……" and "明明 'obviously' ……怎麼/ 為

什麼 'why' ……". The proposition placed right after the adverb 明明 'obviously' is an obvious fact or belief to the writer, but the following clause indicates that what happened afterwards falls short of the writer's expectations. Therefore, the interaction between 明明 'obviously' and the following proposition usually leads to an ANGER emotion. For example, it is obvious to the writer of (2) that it is the Palestinians who suffered. However, the writer finds it unreasonable that there are so many people supporting the perpetrator. Therefore, the writer is expressing an ANGER emotion as what has happened does not live up to his/ her expectation. In that case, the adverb 明明 'obviously' serves as an indicator introducing an emotion cause (i.e. pre-event) of *anger*. 明明 'obviously' does not necessarily need to co-occur with 卻/ 偏 'but' or 怎麼/ 為什麼 'why'. When 明明 'obviously' occurs alone, it still implies the meaning that something falls short of the writer's expectations and it sounds unreasonable to the writer. One may argue 卻 or 偏 'but' may also imply the meaning of "out of expectations". However, 卻 or 偏 'but' only implies the meaning of "beyond expectations" instead of "fall short of one's expectations". Consider (3).

(3) 简陋得不行的生日，却感动死了
jian lou de bu hang de sheng ri, *que gan dong* *si le*
simplest birthday, but moved die-ASP
'(It's) the simplest birthday celebration, but (I was) deeply moved.'

Example (3) indicates that 卻 'but' can be used to introduce a positive statement. It confirms that what introduced by 卻 or 偏 'but' can either be better or worse than expected without the presence of 明明 'obviously'. However, it can only be

worse than expected when 明明 'obviously' co-occurs in the same sentence. It is the interaction between 明明 'obviously' and the following proposition that highlights the reason for the cause of ANGER.

As for 簡直 'absolutely', it can be used to express four emotions, but its linkage with ANGER is the strongest according to the normalized frequencies shown. Although 簡直 'absolutely' per se does not connote any emotions, it adds strong force to a proportion to express a strong emotion as in (4).

(4) 简直是个畜牲

jian zhi	*shi*	*ge*	*chu sheng*
absolutely	is	CL	animal

'Such a brute!'

It is observed that 簡直 'absolutely' often interacts with a negative word in a sentence, such as 畜牲 'animal', 笑話 'joke', 放屁 'fart' etc. All these words have a literal meaning and an underlying meaning. When 簡直 'absolutely' precedes one of these negative words, these words are used to convey its underlying meaning. For example, the literal sense of (4) is to describe someone as a brute/ an animal. If that is what the writer truly wants to convey, he/ she does not have to use the word 簡直 'absolutely' to raise the tone. Thus, the underlying meaning of (4) is to condemn someone as an ill-bred, rough and violent person in an expressive way. Although 簡直 'absolutely' does not directly express an ANGER emotion, the occurrence of 簡直 'absolutely' triggers readers to interpret the following word in a non-literal way. Words following 簡直 'absolutely' often connote an underlying meaning, and it reflects the ANGER of writers. For instance, when 簡直 'absolutely'

interacts with 笑話 'joke', 笑話 'joke' refers to the meaning of "ridiculous". When it interacts with 放屁 'fart', 放屁 'fart' refers to the meaning of "nonsense".

根本 'at all' is often formed in the structure of "根本 + negation marker". It is used to express an ANGER emotion towards a situation that the writer holds firm to, as in (5).

(5) 报警根本没用，警察说外地的没法查…
bao jing gen ben mei you, jing cha shuo wai di de mei fa cha…
calling the police essentially useless, police say outside the country no way investigate…
'Calling the police won't help. They told us that they can't investigate oversea cases…'

It is observed in the corpus that 根本 'at all' co-occurs with a negation marker such as 沒(有) 'not', 不 'not' in most cases. The function of 根本 'at all' is to lay particular stress on a negated statement. For example, the negation marker 没 'not' is used to negate the statement of 报警有用 "calling the police is practical". Since the writer of (5) wants to convey that he/ she is strongly against the statement for some reasons, he/ she uses 根本 'at all' before 没 'not' to emphasize that calling the police is completely useless. When a writer negates a statement assuredly with 根本 'at all', 根本 'at all' is expressive of ANGER. It is not 根本 'at all' per se but the interaction between 根本 'at all' and the negation marker that connotes an ANGER emotion.

For 居然 'surprisingly' and 竟然 'surprisingly', it is not surprising that they are closely related to the SURPRISE emotion as both adverbs directly refer to

the emotion of SURPRISE as in (6) and (7).

(6) 怪不得国足烂！五百米都跑不了的货居然是运动员兼教练。

guai bu de	*guo zu*		*lan*	*wu bai mi*	*dou*
pao bu le	*de*	*huo*		*ju ran*	*shi*
yun dong yuan	*jian*	*jiao lian*			
no wonder	national football team		suck	500 meter	all
run-not-ASP	POSS	commodity		actually	is
athlete	and	coach			

'No wonder the National Football team sucks! It's unbelievable that the guy who can't even run a good 500 meters is the football player and the coach (of the team).'

(7) 刚看了一个新闻，郝医生竟然大度的原谅了那个大一女学生？！

gang	*kan le*	*yi ge*	*xin wen,*	*hao yi sheng*	*jing ran*
da du de	*yuan liang le*	*na ge*	*da yi*	*nü xue sheng?!*	
just	read-ASP	one CL	news,	Dr. Hao	unexpectedly
generous	forgive-ASP	DET	freshman	female student?!	

'Just read some surprising news - Dr. Hao was this forgiving and forgave that female freshman.'

Although the two adverbs 居然 'surprisingly' and 竟然 'surprisingly' in (6) and (7) intrinsically convey the meaning of SURPRISE, the position of the adverbs help spot the pre-event(s) of the emotion. The noun preceding the adverb is the entity that evokes the SURPRISE emotion, and the clause following the adverb details the actual trigger event of the emotion. For example, 郝医生 'Dr. Hao' in (7) is the entity who evokes the writer's SURPRISE emotion, and the actual trigger event is because 大度的原谅了那个大一女学生 'Dr. Hao was so forgiving that he/ she forgave the female freshman.' Therefore, the two adverbs do not only help identify the SURPRISE emotion, they also help locate the entity and the actual trigger event of the emotion.

As for the adverb 原來 'suddenly realize', it introduces a statement that a

writer suddenly realizes. An example is given in (8).

(8) 原来外国也做假货。
yuan lai *wai guo* *ye* *zuo* *jia huo.*
in any case foreign country also make counterfeits.
'Didn't expect that foreign countries also make counterfeits.'

原來 'suddenly realize' in (8) implies that the writer did not expect that foreign countries also make counterfeits. The adverb 原來 'suddenly realize' per se and the SURPRISE emotion are not directly correlated, but a SURPRISE emotion is connoted only when it is used to introduce a clause. In addition to the meaning of 'suddenly realize', 原來 'suddenly realize' may also refer to the meaning of 原本 'originally'. An example is derived from (8) as in (9).

(9) 外国原来也做假货。
wai guo *yuan lai* *ye* *zuo* *jia huo.*
foreign country originally also make counterfeits.
'Didn't expect that foreign countries also make counterfeits. /Foreign countries originally make counterfeits too.'

The adverb 原來 'suddenly realize/ originally' in (9) can either refer to the meaning of 'suddenly realize' or 'originally'. When it refers to the former, a SURPRISE emotion is expressed; when it refers to the latter, the sentence is a general statement with no emotion is expressed. Therefore, it is important to work out the meaning 原來 'suddenly realize/ originally' conveys in a sentence. It is found that when 原來 refers to 'originally', it cannot be moved to the beginning of the sentence as in (8). It does not express a SURPRISE emotion unless there are other linguistic cues found in text. However, if it refers to 'suddenly realize', it can either precede the subject of the sentence or follow the subject of the clause. The sentence

is likely expressing a SURPRISE emotion.

In addition to adverbs, conjunctions also help identify implicit emotions. For example, the conjunction 既然 'now (that)' is found to be closely related to the ANGER emotion. Table 5.3 shows the occurrence and the normalized frequencies of the conjunction.

Table 5.3: The Occurrence of Conjunctions as an Emotion Indicator

	Occurrence (Normalized frequency per 1,000,000)				
Conjunction	Happiness	Sadness	Anger	Fear	Surprise
就算	0 (0)	10 (186.6)	34 (219.6)	1 (70.3)	1 (71.2)
既然	0 (0)	4 (74.6)	22 (142.1)	0 (0)	0 (0)

Among the five emotions, the conjunction 就算 'even if' is more frequently used in expressions of ANGER and SADNESS. (10) and (11) show how the conjunction 就算 'even if' is used to convey an ANGER emotion and a SADNESS emotion.

(10) 就算是临时请的幼师，也不能如此不负责任

jiu suan	*shi*	*lin shi*	*qing de*	*you shi,*
ye	*bu neng*	*ru ci*	*bu fu ze ren*	
even if	is	temporary	hire	kindergarten teacher,
still	cannot	this	irresponsible	

'Even if he's just a substitute kindergarten teacher, he shouldn't be this irresponsible.'

(11) 就算真的是送气球，也不能哄抢啊，哎……

jiu suan	*zhen de*	*shi*	*song*	*qi qiu,*	
ye	*bu neng*	*hong qiang*	*a,*		*ai……*
even if	really	is	giving out	balloon,	
still	cannot	rushing to grab	SFP,		INTJ……

'Even if the balloons are indeed given out for free, (you) still shouldn't be rushing to grab them.'

As illustrated in the examples, the structures of (10) and (11) are very similar, with the structure of "就算 'even if' ……，也不能 'shouldn't' ……". It is proved that it is not the structure that affects the emotion expressed but the context that matters. The emotion depends on the clause following 也不能 'still should not'. For example, an ANGER emotion is triggered in (10) as the writer sees a kindergarten teacher being so irresponsible. A SADNESS emotion is expressed in (11) as the writer sees someone rushes to grab the balloons given out for free. The writer further uses the sigh 哎 'sigh' to express SADNESS. The occurrence in Table 5.3 includes but does not restrict to the structure of "就算 'even if' ……，也不能 'shouldn't' ……". However, if 就算 'even if' appears in the sentence but is not functioning as a conjunction as in (12), the occurrence should not be counted.

(12) 让运营方的头头去替换马试试，能坚持一天就算他赢。
rang yun ying fang de tou tou qu ti huan ma,
shi shi neng jian chi yi tian jiu suan ta ying.
let operator-POSS representative go replace horse,
try try can insist one day then count 3.SG win.
'Let the representative give this a try. If he swoops his position with the horse and could insist that for a day, we'd just say he won'

As for the conjunction 既然 'now (that)', it appears more frequently in expressions of ANGER, followed by SADNESS. Generally speaking, 既然 'now (that)' is used to give an explanation of a new situation. It is found that 既然 'now (that) is frequently formed in the pattern of "既然 'now (that)' ……就 'go'……" and "既然 'now (that)'……*why*……". An example is given in (13).

(13) 既然查到就应该重罚，使它从此不敢再做出这样没良心的事。

ji ran cha dao jiu ying gai zhong fa, shi ta cong ci bu gan zai zuo chu zhe yang mei liang xin de shi.

now investigate-ASP then should heavy punishment, make it from now on not dare again do such unconscionable thing.

'Since (it) was found, it should be heavily punished, so that it would never dare to do such unconscionable things from now on.'

In "既然 'now (that)' ……就 'go'……" , the conjunction 既然 'now (that)' puts forward a premise or a ground, and 就 'then' introduces the writer's suggestion about what to be or not to be done (by the others). Take (13) as an example. The background information of (13) is a restaurant being reported for poor hygiene practices. The writer thinks that the restaurant should be heavily punished given that it has been reported. Therefore, what the conjunction 既然 'now (that)' introduces is often the pre-event of ANGER or SADNESS. When the sentence containing 既然 'now (that)' expresses ANGER, the suggestions given by the writer are in a severe tone. However, when it expresses SADNESS, the writer may soften the tone when making suggestions by using words such as 希望 'hope', 但求 'hope'. Therefore, the interaction between the conjunction 既然 'now (that)' and the accompanying component(s) may drop a hint about which emotion is expressed.

Apart from adverbs and conjunctions, verb is another word class to be discussed for the identification of implicit emotions. It is found that syntactic structures do play an important role in emotion expressions.

I found that 謝謝 'thank' is either used to express HAPPINESS or ANGER, and the position of 謝謝 'thank' in a sentence may affect the emotion expressed as

illustrated in Table 5.4.

Table 5.4: The Occurrence of "謝謝" in Different Syntactic Structures

	Occurrence (Normalized Frequency per 1,000,000)	
Syntactic Structure	Happiness	Anger
"……謝謝"	0 (0)	11 (71.0)
"謝謝 + noun"	11 (301.5)	0 (0)

As shown in Table 5.4, when 謝謝 'thank' occurs in different positions of a sentence, it affects the expressed emotion directly. 謝謝 'thank' is typically used to express one's gratitude conveying a HAPPINESS emotion. It is observed that when 謝謝 'thank' is placed before a noun, it is a genuine thank you message expressing HAPPINESS as in (14).

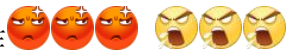
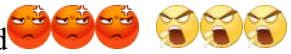
(14) 谢谢暖心的师傅!
xie xie *nuan xin de* *shi fu!*
thank you warm-hearted driver!
'Thank you, the warm-hearted driver!'

(14) is taken from a post on a taxi driver who stopped a girl from committing suicide. The writer of (14) is thankful for what the driver has done. As indicated by the adjective 暖心的 'warm-hearted', the writer is MOVED by the driver's action and is therefore expressing a HAPPINESS emotion. The action of showing gratitude is one of the hints that connotes HAPPINESS. When 謝謝 'thank' precedes a noun who is the recipient of the action, the thank you message is considered a genuine one and the writer is truly happy or grateful because of something. However, when 謝謝 'thank' is placed at the end of the response, it is likely an ironic use. 謝謝 'thank'

is no longer a way to express gratitude but a way to refuse a request, to clarify a statement or even to make a request as in (15) – (17).

(15) 不借，谢谢
bu *jie* *xie xie*
not lending thanks
'(I'm) not lending it. Thanks.'

(16) 我们衡中从来没有站着吃饭 拒绝黑 谢谢
wo men *heng zhong* *cong lai* *mei you* *zhan zhe* *chi fan* *ju jue* *hei* *xie xie*
1.PL Hengshui High School never no standing eat reject scandalize thanks
'None of us from Hengshui High School had ever eaten meals while standing up. (We) don't accept being scandalized. Thanks.'

(17) 气炸  请判刑 谢谢!!
qi zha *qing* *pan xing* *xie xie*
mad please sentence thanks
'I'm so mad  Please sentence (that person). Thank you!!'

In (15) – (17), 謝謝 'thanks' is placed at the end of the sentence, be it with or without punctuations. Unlike those tokens that are truly used to express gratitude to a particular person, the writers do not have a specific person to address to in all these cases. The ironic use of 謝謝 'thanks' often implies that the writer is offended. For example, (15) is a comment taken from a post concerning Japan wants to borrow pandas from China in order to improve their relationship. Therefore, it may sound ridiculous and the Chinese may feel offended by the request which lacks logic. The writer of (15) refuses the request and use 谢谢 'thank you' to express ANGER implicitly. As for (16), the writer clarifies the rumour about his/ her school

which irritates him/ her. 谢谢 'thank you' in that case is used ironically to connote the message of "mind you own business!". As for (17), the writer expresses ANGER explicitly as a young man punched an elderly for no reasons. The writer urged the judge to sentence the person. 谢谢 'thank you' is to strengthen his/ her advice that the young man must be punished.

5.2 The Use of Rhetorical Questions

Section 5.1 discusses how words of different parts-of speech play a role in emotion expressions from the syntactic perspective. In addition to words of different parts-of-speech, rhetorical questions are observed to be another significant means used to express emotions. Generally speaking, questions can roughly be classified into two types, namely information-seeking questions and rhetorical questions. The former refers to questions that aim to elicit an answer, and the latter are questions that aim to achieve a pragmatic goal, such as to emphasize, to persuade, to show emotions etc. (Frank, 1990; Roberts & Kreuz, 1994) without having expectations for answers. As a form of figurative language, rhetorical questions often imply complicated meanings that goes beyond the literal. Although a rhetorical question may occasionally lead to a discussion or elicit a response from the hearer(s), the questioner who poses a rhetorical question does not mean to get an answer to the question. For example, the rhetorical question "*can't you just stay away from me?*" conveys the intended meaning of "*you ought to stay away from me"* instead of

asking whether the hearer can stay away from him/ her.

Previous work suggests that rhetorical questions are a rather productive means of expressing or evoking emotions, in particular the negative ones (Roberts and Kreuz 1994; Gibbs et al. 2002; Lee 2018). Although a great deal of work has investigated the syntactic structures of rhetorical questions, little work has been done on the linguistic features of rhetorical questions used in emotion expressions. Recently, Lau and Lee (2018) have attempted to explore the interaction between rhetorical questions and emotions from the linguistic perspective. Drawing from the insight of their work, I expand the scope and size of the dataset and further examine the linguistic features of rhetorical questions in emotion expressions. With a larger number of tokens, I show a fuller picture of the use of rhetorical questions in emotion expressions.

In the present work, rhetorical questions are annotated at clause level. That is, both main interrogative and embedded interrogatives are identified and annotated. Rhetorical questions are classified into 14 subtypes, some are open questions, and some are closed questions (Lau and Lee 2018). Open questions include some *wh-word* questions, such as what, why, how etc.; closed questions contain *echo questions*, *A-not-A questions*, *alternative questions*, *particle questions*, and *others*. *Echo questions* are questions that have a declarative form but end with a question mark in the written form. *A-not-A questions* refer to questions formed with an affirmative preceding its negative counterpart, such as 會不會. *Alternative questions* provide two or more than two options connected by 還是/ 或 'or'.

Particle questions are questions formed with a Chinese sentence-final particle, such as 嗎, 呢, and 吧. As for rhetorical interrogation markers such as 難道, 何必, 豈, etc. are all categorized as *others*.

5.2.1 Corpus Data

Of the 10,000 annotated comments, 10,677 emotions are identified. That means, some posts contain more than one emotion. Among the five emotions, *anger* has the highest frequency (53.7%), followed by *happiness* (17.8%), *sadness* (16.7%), *surprise* (6.5%) and *fear* (5.4%). The total number of rhetorical questions identified is 2,727. As a single rhetorical question may express more than one emotion, the total number of emotions expressed by those rhetorical questions is 2,796 which account for 26.2% of all the annotated emotions.

Figure 5.1 shows the distribution of emotions expressed using rhetorical questions, which is calculated relative to the total number of rhetorical questions identified.

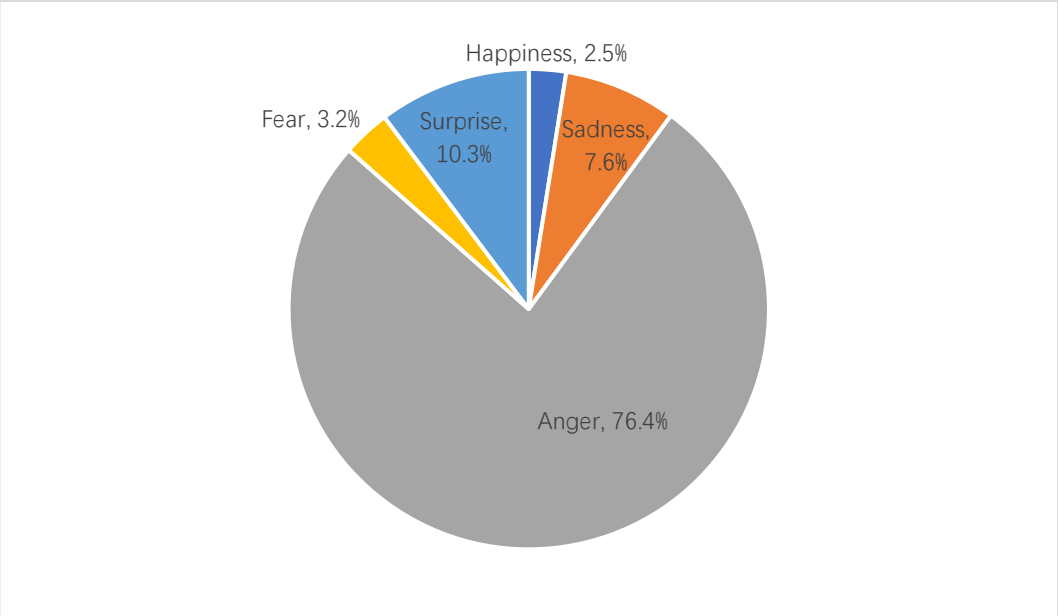


Figure 5.1: Emotions Expressed using Rhetorical Questions

It is observed that 76.4% of rhetorical questions express ANGER. The remaining 23.6% are used to express SURPRISE (10.3%), SADNESS (7.6%), FEAR (3.2%), and HAPPINESS (2.5%). A preliminary remark to be made is that the positive emotion HAPPINESS is least likely to be expressed by means of rhetorical questions. This is in line with findings of previous studies. To further examine whether the strongest connection between rhetorical questions and ANGER is due to the large number of comments containing ANGER, I calculate the distribution of rhetorical questions relative to the total number of comments of a given emotion type. Figure 5.2 illustrates the distribution of rhetorical questions per emotion in all post.

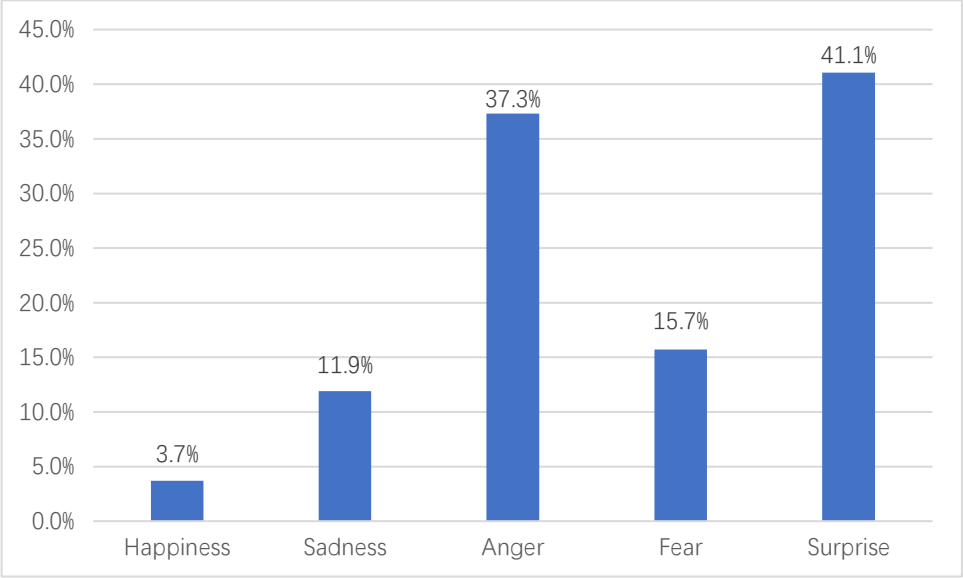


Figure 5.2: Distribution of Rhetorical Questions per Emotion in All comments

Figure 5.2 is calculated relative to the total number of comments of a given emotion type. It illustrates that rhetorical questions are a rather productive means in emotion expressions. Among all the five emotions, SURPRISE has the greatest tendency (41.1%) to be expressed through rhetorical questions, followed by ANGER (37.3%), FEAR (15.7%), SADNESS (11.9%), and HAPPINESS (3.7%). Different from the claim proposed in previous studies that rhetorical questions are most frequently used to express negative emotions, statistics illustrate that rhetorical questions are even more tightly associated with the neutral emotion, SURPRISE (41.1%). Therefore, rhetorical questions are not only particularly productive in evoking negative emotions such as ANGER and FEAR, but also in evoking the neutral emotion SURPRISE. In the following subsection, I further examine the interaction between question types and emotions. Various distinctive syntactic features of rhetorical questions will be proposed for the identification of emotions.

5.2.2 Data Analysis

In the previous subsection, it is proved that rhetorical questions are a rather productive means in expressing emotions. In order to show the correlation between question type and emotion, I further classified rhetorical questions into 14 different subtypes. As mentioned in Section 5.2, there are 2,727 rhetorical questions identified in the corpus. The distribution of rhetorical questions is shown in Figure 5.3.

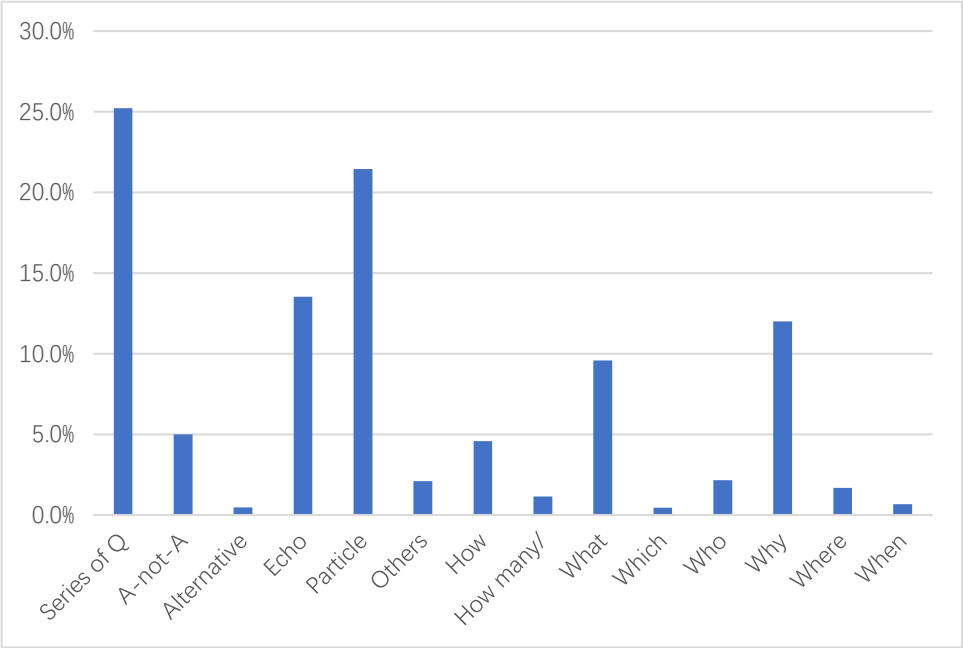


Figure 5.3: The Distribution of Rhetorical Questions

Figure 5.3 shows that the most frequently used type of rhetorical questions in the corpus is a *series of questions*, followed by *particle questions*, *echo questions*, *why questions*, *what questions*, *A-not-A questions*, and *how questions*. Other categories such as *others*, *who questions*, *where questions*, etc. are not often found in the

corpus. Each of them only accounts for less than 3% of the total number of rhetorical questions. To further investigate whether particular types of questions have a preference towards a particular emotion, Table 5.5 shows the distribution of each type of rhetorical questions used to express the five emotions.

Table 5.5: The Distribution of Each Type of Questions In Emotion Expressions

	Close Class Question								Open Class Question						
	Series of Q	A-not-A	Alternative	Echo	Particle	Others	How	How many/ much	What	Which	Who	Why	Where	When	**Total**
Happiness	4%	11%	0%	4%	31%	1%	9%	0%	9%	0%	3%	27%	0%	0%	100%
Sadness	17%	5%	1%	3%	22%	2%	8%	3%	12%	1%	5%	17%	3%	1%	100%
Anger	28%	5%	0%	13%	21%	2%	4%	1%	9%	0%	2%	11%	2%	1%	100%
Fear	7%	10%	0%	11%	28%	2%	8%	0%	29%	0%	1%	3%	0%	1%	100%
Surprise	17%	1%	0%	36%	20%	3%	2%	0%	6%	0%	0%	12%	1%	1%	100%
All Emotions	25%	5%	0%	15%	21%	2%	5%	1%	9%	0%	2%	12%	2%	1%	100%

Table 5.5 is calculated relative to the total number of comments of a given emotion, but not the total number of a particular question type. For example, *a series of questions* makes up 4% of the comments expressing HAPPINESS by means of rhetorical questions. Table 5.5 shows that different types of rhetorical questions have different preferences for a particular emotion. Among all 14 question types, *a series of questions* is the question type that writers most often used to express the ANGER emotion. Frank (1990: 734) explains the use of rhetorical questions as "…in each case the question is re-stated for emphasis, in slightly different form. This makes for a stronger impact on the hearer; a strategy that most likely would be unnecessary if these were simply informational questions, but is a highly effective device for persuasion…". Regarding emotions, rhetorical questions are regarded as

a device writers use to draw readers' attention to their strong emotions (Lee 2018), and a tactic writers use to vent their ANGER to the one who triggers that emotion in them (Lau and Lee 2018). A typical example is exemplified in (18).

(18) 这些家长去骂奶茶店？怪奶茶店？不是脑子有泡？自己孩子管理不好，怪别人？

zhe xie jia zhang qu ma nai chi dian? guai nai cha dian? bu shi nao zi you pao? zi ji hai zi guan li bu hao, guai bie ren?

DET parents go scold milk tea shop? blame milk tea shop? not brain has bubble? own kid manage bad, blame others?

'These parents are blaming the bubble tea shop? What? Have they gone utterly insane? They're blaming others for the misbehaviour of their kids? Who is to be blamed for the improper behaviour of their own children?'

(18) shows how *a series of questions* is typically raised. Although the four questions are literally different, they are indeed semantically related to each other as proposed by Frank (1990). The first two questions only used different verbs 骂 'scold' and 怪 'blame' to question about those parents' behaviour. The writer restates the question to emphasize that he/ she finds the behaviour completely unreasonable. Thus, the writer directly criticizes the parents with the third question, asking whether they have gone insane. The fourth questions are to further support the claim stated in the previous questions. Although the writer does not directly describe his/ her ANGER through any emotion-bearing words, he/ she intentionally uses *a series of questions* to imply that he/ she is annoyed with the parents' behaviours as the parents are fully responsible for the improper behaviours of their own kids. By repeating questions regarding the same issue, the emotion intensity can be increased to a certain extent that can never be reached by stating the question only once. Given

that they are mostly used by writers to notify readers of their ANGER emotion by increasing the intensity of that emotion, the use of a series of rhetorical questions is a good hint that an ANGER emotion is likely expressed.

Among all types of closed questions, *particle questions* in general are most frequently used to express emotions. It accounts for 21% of the total number of rhetorical questions. As observed in Table 5.5, when writers want to express HAPPINESS and FEAR, they tend to use particle questions instead of other types of closed questions. However, HAPPINESS and FEAR are least likely to be expressed by means of rhetorical questions as previously mentioned. It is rather difficult to make any assumptions about the syntactic features of particle questions designated for the expressions of HAPPINESS or FEAR. However, it is observed that some patterns forming a particle question are rarely found in the expressions of HAPPINESS and FEAR but often found in the expressions of other emotions. These patterns are believed to be able to distinguish the expressions of HAPPINESS and FEAR from the other three emotions when using *particle questions*. One of the patterns is "……了吧/了嗎/了麼", as shown in (19).

(19) 处罚太轻了吧？

chu fa	*tai*	*qing*	*le*	*ba*
penalty	too	light	SFP	SFP

'The penalty was too light?'

Comment (19) expresses an ANGER emotion by posing a rhetorical question to complain about the light penalty. This syntactic pattern is often used to convey an ANGER emotion, by stating what the writer believes before 了吧. In (19), it is the

writer's purpose to call into question the decision made by the judge about the penalty with a rhetorical question. Besides, using a rhetorical question can help stress his/ her emotion and let the others know he/ she is not happy with that. The effect cannot be reached by simply using a declarative sentence like "it is such a light penalty". Among all the 65 rhetorical questions containing the form "……了吧/了嗎/了麼"[4], 50 tokens are associated with the ANGER emotion, followed by SURPRISE (7 tokens), SADNESS (6 tokens), FEAR (3 tokens), and HAPPINESS (1 token).

Another syntactic structure that is often found in ANGER and SURPRISE, but seldom in HAPPINESS and FEAR is a particle question formed with a negation marker as in "不……嗎/麼". Some tokens containing that structure are not counted as the negation marker is not a component used to form a rhetorical question as in (20).

(20) 如果这事发生在宁夏，你看你们敢说一个不字吗

ru guo	*zhe*	*shi*	*fa sheng*	*zai*	*nin xia,*
ni	*kan*	*ni men*	*gan*	*shuo*	*yi ge*
bu	*zi*	*ma*			
if	DET	incident	occur	in	Ninxia,
2.SG	see	2.PL	dare	say	one-CL
no	word	SPF			

'If it happened in Ninxian, see if you dare to say the word "no"?'

Of all the 124 tokens formed with this pattern, 85 tokens are used to express ANGER, 20 tokens are used to express SURPRISE, 10 tokens are used to express SADNESS, 6 tokens are used to express FEAR, and none of them is used to express HAPPINESS. Another example is given in (21).

[4] Those rhetorical questions containing the form "……了吧/了嗎/了麼" but belong to *a series of questions* are not included.

(21) 不会管好孩子吗？

bu	*hui*	*guan*	*hao*	*hai zi*	*ma?*
not	know	govern	well	child	SFP?

'Don't know how to discipline (your) child?'

(21) is, again, used to express ANGER. The writer is not asking whether someone knows how to teach a child to behave well but indicating that a person did not discipline his/ her child to behave well. By using a negation marker "不" in the sentence, the writer can increase the intensity of his/ her ANGER emotion. The proposition being negated is often a positive one, such as 管好孩子 'discipline your child' in (21). By negating a positive proposition, one may be able to express a negative emotion. Therefore, the pattern is rarely found in the expression of a positive emotion, i.e. HAPPINESS.

In addition to *particle questions*, *echo questions* are another type of rhetorical questions that is often found to express emotion. *Echo questions* are not formed with question words; instead, they have a declarative structure in the written form and should be raised in a rising tone. Therefore, it is not surprising that *echo questions* are the most frequently used question type in expressing SURPRISE (36%), with 104 tokens. It is observed that out of the 104 tokens that express a SURPRISE emotion using *echo questions*, 37 tokens simultaneously express an ANGER emotion using the same question. Certain adverbs are very often found in the comments expressing both SURPRISE and ANGER, including 才, 也, 还 as the examples shown in (22) – (24).

(22) 让人家眼部受伤，才拘留 10 日？

rang	*ren jia*	*yan bu*	*shou shang,*	*cai*	*ju liu*	*10 ri*
let	others	eye	injured,	only	custody	10 days

'(He's) only receiving a 10-day custody for injuring the eye of another person?'

(23) 这也能评选好人？？

zhe	*ye*	*neng*	*ping xuan*	*hao ren??*
Det	also	can	nominate	Good Person Award??

'People like that can still be nominated for the Good Person Award??'

(24) 上厕所还违规？

shang	*ce suo*	*hai*	*wei gui?*
Go	toilet	still	against school rules?

'Even going to the toilet is against school rules?'

In (22) and (23), the adverbs 才 'only' and 也 'too' are used to express a SURPRISE emotion about the proposition following the adverbs. For example, the writer of (22) finds the duration of the detaining event too short as connoted by 才 'only'. Part of the reason for the trigger of emotions is clearly stated in the first sentence. As someone has caused harm to the eye of another person, the writer is SURPRISED as well as ANGRY that he/ she was just detained for 10 days which is just a light penalty in the writer's opinion. As for (23), the person who is nominated for a "Good Person Award" has done something unethical as mentioned in the post. Therefore, the writer is experiencing both SURPRISE and ANGER as he/ she finds it unjustifiable that the person can still be nominated as a "Good Person" even after doing something unethical.

The writer of (24) is SURPRISED and IRRITATED that even going to toilet is against school rules. The adverb 还 'still' is used to indicate that he/she finds it unreasonable or even ridiculous as everyone has the right or freedom to go to toilet.

As for those *echo questions* which only convey SURPRISE, they do not often have a specific pattern. However, it is observed that the *echo questions* usually repeat event details mentioned in the post, as in (25). (25a) is an excerpt of a post, and (25b) is a comment made on the post.

(25) a. 24 日，宁夏银川一住户在家中抓住一小偷，嫌疑人陈某撬开后门进入，在现场逗留 3 个小时，因太饿翻出 15 个鸡蛋做了顿饭，还把户主儿子的暑假作业给撕了。
24 ri, nin xia yin chuan yi zhu hu zai jia zhong zhuo zhu yi xiao tou, xian yi ren Chen mou qiao kai hou men jin ru, zai xian chang dou liu 3 ge xiao shi, yin tai e fan chu 15 ge ji dan zuo le dun fan, hai ba hu zhu er zi de shu qi zuo ye gei si le.
24th, Ninxia Yinchuan one household in home catch one thief, suspect Chen-someone crack open back door enter, in scene stay 3-CL hours, because too hungry rake out 15-CL egg make-ASP CL-meal, and make master's son-POSS summer assignments let tear SFP.
'24th. In Ninxia Yinchuan, a thief was caught in a household. The suspect, Chen, cracked open the back door to enter the house. He stayed in there for 3 hours. (During this period) He was so hungry that he raked out 15 eggs to prepare himself a meal. He even tore the summer vacation assignment of the son of the owner of the house.'

b. 吃十五个鸡蛋啊？!!
chi shi wu ge ji dan a?!!
eat fifteen CL ji dan SFP?!!
'(He) ate 15 eggs?'

In (25), the comment is annotated as conveying a SURPRISE emotion. The content of the comment 十五个鸡蛋 '15 eggs' is already mentioned in the post concerning the thief having 15 eggs while thieving. The writer is replicating a particular point mentioned in the post at which he/ she is surprised. Therefore, when a rhetorical question repeats information that has already been stated in the post, it is very likely

that the comment is expressing a SURPRISE emotion.

Among the six types of open class questions, three of them have a stronger connection with comments containing emotions. *Why questions* make up 12 % of all types of rhetorical questions. *Why questions* are typically formed with different forms of question words referring to the meaning of why, such as 為什麼, 為啥, 怎麼, 干嘛, 咋[5] etc. Data illustrates that "*why*……這麼/那麼……" such as "為什麼……這麼/那麼……" is one of the most frequently used patterns. The pattern is particularly effective in expressing HAPPINESS. Although the pattern shows a tendency to HAPPINESS, it does not necessarily link to that emotion. However, it drops a hint to the emotion expressed. Consider (26) and (27).

(26) 看川普咋觉得那么搞笑呢😂😂😂

kan	*chuan pu*	*za*	*jue de*	*na me*	*gao xiao*	*ne*
see	Trump	why	feel	that	funny	SFP

'Why does it seem so entertaining when watching Trump? 😂😂😂'

(27) 滴滴司机咋就那么恶心呢！

Didi si ji	*za jiu*	*na me*	*e xin*	*ne!*
Drivers of Didi	why	that	gross	SFP!

'Why are Didi drivers this gross?'

As shown in the examples, the pattern "why……這麼/那麼……" can be used to express different emotions, such as HAPPINESS in (26) and ANGER in (27). It is the adjective following 這麼/那麼 that determines the emotion. As in (26), when 搞笑 "entertaining" is not used ironically, it is a positive word that may lead to a

[5] 咋 is widely used in Northeastern dialect which can refer to the meanings of "why" and "how".

HAPPINESS emotion. As for (27), the adjective 恶心 "gross" is an emotion-laden word of ANGER as I proposed in Section 4.2.2. Therefore, an ANGER emotion is expressed towards the taxi driver.

Certain structures are mostly used to form comments expressing an ANGER emotion. One of the structures is "*why* (adv/ noun) 不……", and the other one is "*why* (adv/ noun) 要……". Of all the 70 tokens formed with the "*why* (adv/ noun) 不……", 62 of them are related to ANGER, 4 to SADNESS, 3 to SURPRISE, and 1 to HAPPINESS. As for the latter, 47 tokens are found in the corpus, in which 36 of them express ANGER, 6 SADNESS, and 5 SURPRISE. The two patterns show an obvious tendency to the ANGER emotion. An example of each structure is given in (28) and (29).

(28) 这样的老师不合格，为什么不开除，把学生都教成野蛮人
zhe yang de lao shi bu he ge, wei shen me bu kai chu,
ba xue sheng dou jiao cheng ye man ren
like such teacher fail, why expel,
make student all teach-become barbarian
'Teachers like (him/her) are unqualified. Why doesn't the school just fire the teacher? School children have become barbarians (because of him/her).'

(29) 心真狠啊，不想养，为什么要生呢
xin zhen hen a, bu xiang yang,
wei shen me yao sheng ne
heart really brutal SFP, not want raise,
why have give birth SFP
'(He/she) is so cold-blooded. Why would (he/she) give birth to (the child) when (he/she) doesn't want to raise up (the child)?'

In (28), the writer believes that such a bad teacher should be made redundant for influencing students to act like barbarians. However, the teacher is not fired as indicated by the rhetorical question. Therefore, the question is raised by the writer

to vent his/ her ANGER towards the reality of not giving the teacher a sack. In (29), the writer blames the parent for giving birth to their child but not wanting to raise up the child. In these cases, people tend to use the two patterns to blame someone for not doing something or to blame someone for having done something which does not live up to the writer's expectation. Therefore, the emotion expresses through these two patterns are likely ANGER.

Another pattern found to be vastly associated with the ANGER emotion is the use of 干嘛 in "(noun) + 干嘛……" or "……干嘛". It should be noted that the question word 干嘛 either refers to the meaning of *why* or *what*. Regardless of the question type it refers to, the use of 干嘛 shows a tendency towards ANGER. Consider (30) and (31).

(30) 不好好干就别干了，伤害小朋友干嘛
bu hao hao gan jiu bie gan le, shang hai xiao peng you gan me
not well-do then not do SFP, hurt child do-what
'If you don't want to do it, just back off. Why would you bring pain to the child?'

(31) 法律是用来干嘛的，总是有些令人看不懂的判法
fa lü shi yong lai gan me de, zong shi you xie ling ren kan bu dong de pan fa
law is use for what, always some make-people see-not-understand judgement
'What is law for? Judgements (made by the court) are often so difficult to comprehend.'

(30) is literally raised to question the reason why someone brought pain to the child. As a rhetorical question is not used to elicit an answer, the question in (30) is used to vent the writer's ANGER by blaming that person for bringing pain to the child.

Different from (30), the question word 干嘛 in (31) refers not to the meaning of *why* but to the meaning of *what*. In the word 干嘛, the second component 嘛 is the Chinese equivalent of *what*. The question in (31) is to indicate that the writer does not know what law is for. The reason is given in the following sentence which states that judgements made by the court are often so difficult to comprehend. Although 干嘛 can either refer to *why* or *what*, rhetorical questions formed with 干嘛 mostly express an ANGER emotion.

What questions are the second most frequent type of rhetorical questions that are used to express FEAR. The pattern "(要是)……怎么办/这怎么算/杂办/咋整/咋办" is often found to express FEAR. To distinguish whether the pattern is used to express FEAR or other emotions, one may consider whether the proposition preceding 怎么办 is a hypothetical situation as in (32), a situation that may happen in the future as in (33), or a situation that have already happened as in (34).

(32) 要是导致抢救病人延误了时间，该怎么办。
yao shi *dao zhi* *qiang jiu* *bing ren* *yan wu le* *shi jian,* *gai* *zen me ban.*
if cause rescue patient delay-ASP time, should what-do.
'If (that) caused the delay of patient rescue, what should be done?'

(33) 有这样的父母，孩子以后怎么办？
you *zhe yang de* *fu mu,* *hai zi* *yi hou* *zen me ban?*
have such parent, child later what-do?
'How is the child going to do when he has a parent like such?'

(34) 自己对自己都不负责任 还能怎么办 如果这是我孩子 我会打到他不认识

zi ji *dui* *zi ji* *dou* *bu fu ze ren* *hai neng* *zen me yang* *ru guo* *zhe* *shi* *wo hai zi wo* *hui da dao* *ta* *bu ren shi*

self to self even irresponsible still can how if DET is my kid I will beat-ASP 2.SG not recognize

'If you're not responsible to your own life, who's going to be responsible for that? If he/she were my kids, I would beat him/her black and blue.'

(32) and (33) express a FEAR emotion. (32) states a situation that might have happened. Although it did not really happen, the rhetorical question indicates that the writer was worried about it. As for (33), the writer is WORRIED about the child for having such parents. The writer of (34) expresses an ANGER emotion towards the kid who is irresponsible to his/ her own life. The writer put it bluntly that if that irresponsible person is his/her kid, he/ she would beat him black and blue.

In this section, I examine the use of rhetorical questions in emotion expressions. It is suggested that different emotions may have a preference to different question types. Apart from question types, I also propose certain syntactic structures of rhetorical questions that are frequently used to express a specific emotion. A summary is illustrated in Table 5.6.

Table 5.6: Syntactic Structures of Rhetorical Questions Highly Associated with a Specific Emotion

Question Type	Syntactic Structure	Emotion(s)
Particle	"……了吧/了吗/了么"	Anger
Particle	"不……吗/么"	Anger/ Surprise
Echo	"(……)才/还/也……"	Anger/ Surprise
Why	"*why* ……这么/那么……"	Happiness
Why	"*why (adv/ noun)* 不……"	Anger
Why	"*why (adv/ noun)* 要…"	Anger
Why/ What	"干嘛……/……干嘛"	Anger
What	"(要是)…… 怎么办/这怎么算/杂办/咋整/咋办"	Fear

5.3 The Use of Emojis at the Discourse Level

In Section 4.1.2, I have proposed a list of Weibo emojis that show a high tendency towards a particular emotion in Table 4.4. At the semantic level, an emoji explicitly expresses a particular emotion. For example, the grinning face with smiling eyes emoji 😁 expresses a HAPPINESS emotion. When it appears in a comment, it is likely that the comment is expressing a HAPPINESS emotion. However, it is observed in the corpus data that an emoji can either play the role as a complementary means in emotion expressions, or it can be used express an emotion conflicting with the one expressed in text. Some studies indicated that emojis can serve as an amplifier or modifier when an emoji and its linguistic text are in disagreement. Derks et al. (2008) explore whether the sentiment (i.e. positive and negative) of text can be altered by the presence of an emoticons. They indicate that when an emoticon and the accompanying text disagree with each other, the sentiment become closer to neutral, but not enough to switch from one polarity to another. The claim is

supported by Skovholt et al. (2014) who suggest that emoticons can either soften negativity or strengthen positivity. Novak et al. (2015) propose that text determines whether an emoji functions as an amplifier or a modifier. Riordan (2017) states that it is the verbal content that determines the emotion of a message, and an emoji only alters the intensity of that emotion. Tian et al. (2017) argue that when emojis and the accompanying linguistic texts express different emotions, the overall communicated meaning is not a sum of the two channels. They observe a significant amount of positive emojis in text expressing ANGER, meaning that the emojis are used ironically. This poses a real challenge to the implicit emotion identification task. In view of this, this section discusses the interplay between emojis and linguistic texts, with a special attention being placed on the situations that the two channels are at odds. In this section, I aim to address the following questions: (1) which channel determines the overall emotion when the emoji and its accompanying text disagree with each other? (2) If the overall emotion is determined by text, what is the function of the emoji?

5.3.1 The Definition of Typical and Atypical Use

Of the 10,000 annotated comments, 831 comments were found to contain at least one emoji that is highly associated with a particular emotion. After extracting all the 831 comments, annotators are asked to read through the comments and to decide whether the emojis are used in a typical way. The decision should be made on the basis of the general perception of an emoji. For example, emojis showing a smiling

face express a HAPPINESS emotion; a frowning face expresses a SADNESS emotion; a pouting face expresses an ANGER emotion etc. If the emotion expressed by emoji(s) is in line with the emotion expressed by the accompanying text, the emoji(s) is used in a typical way as in (35).

(35) 生活真的很难，想起我的爸爸😟😟

sheng huo	*zhen de*	*hen*	*nan ,*	*xiang qi wo*	*de*	*ba ba*
life	really	very	hard,	think of 1.SG	POSS	father

'Life is really hard, (I'm) thinking of my father😟😟'

The emojis 😟😟 in (35) show a teary-looking face representing SADNESS. It is a typical use as both the text and the emojis express a SADNESS emotion. In that case, 😟 expresses an emotion at the semantic level. This issue has been addressed in Section 4.1.2. However, emojis can also be dealt with at the discourse level. If the emotions expressed via the text and the emoji(s) are in disagreement, the use of emoji(s) should then be considered atypical. The atypical use of emojis is the major focus of this section. When the two channels are at odds, annotators must judge whether the overall emotion is determined by the emoji(s) or by its accompanying text. Consider (36).

(36) 快开学了 要写完了 被撕了也是很绝望🤔

kuai	*kai xue le*	*yao*	*xie wan le*	*bei*
si le	*ye*	*shi*	*hen*	*jue wang*
soon	start school-ASP	almost	write-finish-ASP	PAS
tear-ASP	also	is	very	hopeless

'School is about to start. It is really hopeless to see the homework being torn when it is almost done🤔'

The emoji 🤔 in (36) expresses HAPPINESS whereas the text expresses SADNESS as

indicated by the emotion word 绝望 'hopeless'. By looking at the context given in the post, it is not the writer but a child's homework being torn. The emotion word 绝望 'hopeless' is not the emotion experienced by the writer but his/ her prediction about the child's emotion state. Without the emoji 🥺, it seems that the writer feels a lot of sympathy for the child. However, when 🤭 is used, it indicates that the writer does not feel compassion for the child but finds the situation hilarious. Therefore, the overall emotion is determined by the emoji expressing HAPPINESS.

Of the 831 comments, 78.2% of emojis are used typically, and 21.8% atypically as shown in Table 5.7. In addition to the overall distribution, Table 5.7 also illustrates the distribution of the typical and atypical use of emojis of each emotion type.

Table 5.7: Typical and Atypical Use of Emojis

	Typical (%)	Atypical (%)	Total
Happiness	319 (66.6%)	160 (33.4%)	479
Sadness	88 (92.6%)	7 (7.4%)	95
Anger	243 (94.6%)	14 (5.4%)	257
Fear	0 (0%)	0 (0%)	0
Surprise	0 (0%)	0 (0%)	0
Total	650 (78.2%)	181 (21.8%)	831

To avoid situations that an emoji is correlated with a specific emotion just by coincidence, I only study emojis of more than 5 occurrences. Besides, each selected emoji should be unambiguously expressing a particular emotion so that one can distinguish its typical and atypical use. As there are no emojis associated with FEAR and SURPRISE have more than 5 occurrences, no statistics can be shown for the two

emotions. Among the other three emotions, emojis of HAPPINESS (33.4%) are most likely used in an atypical way, followed by SADNESS (7.4%) and ANGER (5.4%). It shows that emojis expressing a positive emotion (i.e. HAPPINESS) disagree with their linguistic text most frequently. In the following subsections, I examine the interplay between emojis of each emotion type and the accompanying text. As emojis of FEAR and SURPRISE are either ambiguous or of low frequency (i.e. less than 5), the following subsections discuss HAPPINESS, SADNESS, and ANGER only.

5.3.2 Emojis Expressing Happiness

There are 14 emojis that are found to express a HAPPINESS emotion. I manually read through all the extracted comments and summarized the distribution of their typical use and atypical use in Table 5.8. Typical use means that the emotion expressed by an emoji is consistent with the emotion expressed by text; atypical use means that the two channels conflict with each other.

Table 5.8: Typical and Atypical Use of Happiness Emoji

Emojis	Typical Use	Atypical Use		Total
		Determined by Emoji	Determined by Text	
 [微笑]	15	0	75	90
 [good]	81	1	8	90
 [赞]	67	0	5	72
 [笑 cry]	37	18	13	68
 [哈哈]	23	10	10	43
 [嘻嘻]	20	4	4	28
 [鼓掌]	19	0	2	21
 [爱你]	13	0	1	14
 [中国赞]	11	0	0	11
 [太开心]	9	1	1	11
 [偷笑]	6	3	2	11
 [可爱]	8	0	2	10
 [坏笑]	5	0	0	5
 [赞啊]	5	0	0	5
Total	319 (66.6%)	37 (7.7%)	123 (25.7%)	479

Table 5.8 suggests that 66.6% of emojis expressing HAPPINESS and they occur in text that also expresses HAPPINESS. The remaining 43.4% are comments that contain an emoji indicating HAPPINESS but express another emotion in text. It is observed in Table 5.8 that the overall emotion is more often determined by text, accounting

for 25.7%. Only 7.7% of comments are determined by emoji when the emoji(s) and text are at odds. In order to figure out whether it is a comment conveying a particular emotion that can override the HAPPINESS emotion expressed by emoji (s), Table 5.9 shows the distribution of the atypical use of HAPPINESS emojis.

Table 5.9: The Distribution of the Atypical Use of Happiness Emojis

Possibilities of the atypical use of emojis: (Emoji of Happiness + Emotion in Text → Overall)	Total
(1) Happiness + Sadness → Sadness	12
(2) Happiness + Sadness → Happiness	16
(3) Happiness + Anger → Anger	93
(4) Happiness + Anger → Happiness	12
(5) Happiness + Fear → Fear	3
(6) Happiness + Fear → Happiness	4
(7) Happiness + Surprise → Surprise	16
(8) Happiness + Surprise → Happiness	4

Table 5.9 shows that when a HAPPINESS emoji co-occurs with text conveying a SADNESS or FEAR emotion, there is only a fifty-fifty chance that the overall emotion is determined by one of the two channels. However, when a HAPPINESS emoji co-occurs with text conveying an ANGER or SURPRISE emotion, there is every likelihood that the overall emotion is determined by the text. An example of ANGER is given in (37).

(37) 是不是傻😂

shi bu shi *sa*
is-not-is stupid
'Are you stupid 😂'

Without the emoji 😂, (37) presents a rhetorical question conveying an intended meaning of "you are so stupid". It is often used when the writer is irritated by someone's stupidity. Therefore, the text itself expresses ANGER. However, with the presence of 😂, the writer is mocking the interlocutor instead of being annoyed with his/ her stupidity. The emoji 😂 indicates that the writer's purpose is to make fun of someone. Therefore, even though the rhetorical question itself expresses ANGER, the HAPPINESS emotion expressed by the emoji still overrides the ANGER emotion expressed by the text. However, a HAPPINESS emoji is more frequently being overrode by the ANGER emotion expressed in text. Consider (38).

(38) 不该是刑事拘留嘛🙂

bu gai *shi* *xing shi ju liu* *ma*
shouldn't is criminal detention SFP
'Shouldn't it be criminal detention🙂'

In (38), the text also expresses an ANGER emotion and the emoji expresses HAPPINESS. Similar to (37), a rhetorical question is raised to convey the underlying meaning that the punishment should be criminal detention. An ANGER emotion is connoted via the text as the writer thinks the penalty is too light. Thus, the emoji 🙂 in (38) is used ironically. In fact, there are heated discussions, not just in Mainland China, on what the slightly smiling emoji truly conveys. The reason for that is because the slightly smiling emoji looks a bit different from simple smiling

emojis. First, simple smiling emojis usually have a beaming or grinning face with an open mouth (e.g. 😊, 😄), but the slightly smiling emoji has a closed smile which gives the impression that someone is trying to fake a smile. Second, simple smiling emojis often show a pair of smiling eyes (e.g. 😁), while the slightly smiling emoji shows a pair of open eyes. Apart from offering a slight expression of friendliness and politeness, its tone can also be patronizing, passive-aggressive, or ironic. The emoji is therefore open to interpretation. Some people, particularly the older generation, interpret the slightly smiling emoji as a polite and friendly smile, while the younger generation interpret it as an ironic means to convey passive aggression. The ironic use of this emoji is popular with most Weibo users in Mainland China as shown in the Table 5.8. 75 out of 90 of the emoji are used in an ironic way to express ANGER. Therefore, the emoji in (38) is a contempt smile which is composed of both disgust and resentment. It functions as a faint sneer which implies that the writer is not satisfied with the judgement. Therefore, the writer fakes a smile in a passive-aggressive way to avoid expressing his/ her ANGER openly. In that case, the overall emotion is determined by the text and the emoji 🙂 functions as an ironic means to strengthen the intensity of ANGER by showing disrespect for the someone with a sneer. Consider (39) as another example of the slightly smiling emoji being used as an ironic means.

(39) 不租不借，滚🙂

bu zu	*bu jie,*	*gun*
not rent	not lent,	get out

'Not renting, not lending. Bugger off🙂'

Example (39) is taken from a post concerning Japan hoping to rent giant pandas from China. No doubt the text in (39) is obviously expressing an ANGER emotion as indicated by 滚 'bugger off'. However, the writer discordantly inserts a slightly smiling emoji after using a rude way of telling someone to go away. In that case, the writer is trying to look friendly with a supressed smile that the social norms compel him/ her to do, and the overall emotion is still ANGER. It can be concluded that when 🙂 appears in text that expresses ANGER, the overall emotion is highly likely an ANGER emotion.

As for the co-occurrence of a HAPPINESS emoji and a text expressing SURPRISE, it is more likely that the overall emotion is determined by text. Consider (40).

(40) 🤭居然是我们邯郸人，佩服

ju ran	*shi*	*wo men*	*han dan*	*ren,*	*pei fu*
surprisingly	is	1.PL	Handan	people,	marvellous

'🤭Surprisingly, it's a person from Handan (who achieved something). Remarkable'

In (40), the emoji 🤭 expresses a HAPPINESS emotion and the text expresses a SURPRISE emotion as indicated by 居然 'surprisingly'. (40) expresses a complex emotion 驚喜 'DELIGHTED' which is composed of SURPRISE as the primary emotion and HAPPINESS as the secondary. Although a complex emotion is comprised of two to three basic emotions, only the emotion of the greater amount is annotated. For example, the overall emotion expressed in (40) is the complex emotion 驚喜 'DELIGHTED', but the comment is annotated as SURPRISE instead of SURPRISE +

HAPPINESS. Therefore, the overall emotion follows the text. In that case, 🤔 functions as a modifier to enrich the emotion complexity.

It is also possible yet less frequent that the overall emotion is determined by the HAPPINESS emoji when it co-occurs with text expressing SURPRISE. Consider (41).

(41) 这小偷估计也是饿的不行了，十五个鸡蛋😂
zhe xiao tou gu ji ye shi e de bu hang le, shi wu ge ji dan
this thief probably also is hungry not good-ASP, fifteen-CL egg
'The thief was probably starving, (he/she had) 15 eggs 😂'

In (41), the overall emotion is not the complex emotion 驚喜 'DELIGHTED'. The text expresses SURPRISE as it is surprising that the thief broken into the house and had 15 eggs. In that case, the overall emotion is determined by the emoji (i.e. HAPPINESS) as the writer finds it hilarious. In sum, when the emoji and the text interact to express 驚喜 'DELIGHTED', the overall emotion is determined by the text. However, if their interaction does not lead to 驚喜 'DELIGHTED', the overall emotion should follow the emoji.

5.3.3 Emojis Expressing Sadness

There are 5 types of emojis that are found to express SADNESS. The distribution of the typical and atypical use is demonstrated in Table 5.10.

Table 5.10: The Distribution of Typical and Atypical Use of Sadness Emojis

Emojis	Typical Use	Atypical Use		Total
		Determined by Emoji	Determined by Text	
[悲伤]	25	0	4	29
[蜡烛]	27	0	2	29
[失望]	14	0	0	14
[伤心]	12	0	0	12
[可怜]	10	1	0	11
Total	88 (92.6%)	1 (1.1%)	6 (6.3%)	95

Table 5.10 shows that 92.6% of SADNESS emojis occur in text that also expresses SADNESS. There are only a few exceptional cases. All the possibilities are shown in Table 5.11.

Table 5.11: The Distribution of the Atypical Use of Sadness Emojis

Possibilities of the atypical use of emojis: (Emoji of Sadness + Emotion in Text → Overall)	Total
(1) Sadness + Happiness → Happiness	4
(2) Sadness + Happiness → Sadness	1
(3) Sadness + Anger → Anger	2
(4) Sadness + Anger → Sadness	0
(5) Sadness + Fear → Fear	0
(6) Sadness + Fear → Sadness	0
(7) Sadness + Surprise → Surprise	0
(8) Sadness + Surprise → Sadness	0

Table 5.11 illustrates that the atypical use of SADNESS emojis is mostly related to text expressing HAPPINESS. In most cases, it is the text that controls the overall emotion as in (42).

(42) 形式不重要，有朋友在就很温暖😢

xing shi	*bu*	*zhong yao,*	*you*	*peng you*	*zai*
jiu	*hen*	*wen nuan*			
form	not	important,	have	friend	around
then	very	warm			

'It's not the form that matters. Having friends around you is heart-warming 😢'

In (42), the emoji 😢 expresses SADNESS but the text implicitly conveys HAPPINESS. The SADNESS emoji 😢 interacts with the text to represent tears of joy. The sentence expresses the complex emotion, 感動 MOVED. The MOVED emotion is composed of HAPPINESS (primary) + SADNESS (secondary). Therefore, the comment should be tagged as HAPPINESS instead of SADNESS. It can be concluded that when a SADNESS emoji and text expressing HAPPINESS interact to express MOVED (i.e. HAPPINESS), the SADNESS emoji is not capable of overriding the HAPPINESS emotion expressed by text.

5.3.4 Emojis Expressing Anger

There are 9 emojis that correspond to the ANGER emotion. Table 5.12 illustrates the typical use and atypical use of these emojis.

Table 5.12: Typical and Atypical Use of Anger Emojis

Emojis	Typical Use	Atypical Use		Total
		Determined by Emoji	Determined by Text	
[怒]	103	0	0	103
[挖鼻]	30	0	5	35
[费解]	28	0	2	30
[吃瓜]	24	0	2	26
[哼]	18	1	3	22
[鄙视]	14	0	0	14
[怒骂]	12	0	0	12
[吐]	8	0	1	9
[白眼]	6	0	0	6
Total	243 (94.6%)	1 (0.4%)	13 (5.1%)	257

Unlike HAPPINESS emojis, not many ANGER emojis can be used in an atypical way. Although they are occasionally used in an atypical way, certain emojis are more often used, such as , , and . Table 5.13 demonstrates the distribution of the atypical use of ANGER emojis.

Table 5.13: The Distribution of the Atypical Use of Anger Emojis

Possibilities of the atypical use of emojis: (Emoji of Anger + Emotion in Text → Overall)	Total
(1) Anger + Happiness → Happiness	4
(2) Anger + Happiness → Anger	0
(3) Anger + Sadness → Sadness	4
(4) Anger + Sadness → Anger	1
(5) Anger + Fear → Fear	1
(6) Anger + Fear → Anger	0
(7) Anger + Surprise → Surprise	3
(8) Anger + Surprise → Anger	0

Table 5.13 shows that when an ANGER emoji co-occurs with text expressing other emotions, the overall emotion depends heavily on the text as in the cases of HAPPINESS, SADNESS, and SURPRISE. Consider (43).

(43) 小样，还收拾不了你了 🤔耍无赖也不看看地方

xiao yang, hai shou shi bu le ni le
shua wu lai ye bu kan kan di fang
look at you, still settle not-ASP 2.SG SFP
play rogue too not see-see place

'See how I get to you! 🤔 How dare you play rouge before us!'

In (43), the text expresses a HAPPINESS emotion as the writer eventually got to the person who play rouge before him/ her. As it is gratifying to see that person being punished, the writer uses 🤔 to show an undisguised contempt which reflects that the writer is gloating over this. The writer is experiencing APPEASED which is a

complex emotion consists of HAPPINESS and ANGER. Therefore,  in (43) interacts with the text expressing HAPPINESS to expressed APPEASED. It functions as a modifier to enrich the complexity of the HAPPINESS emotion.

In addition to text expressing HAPPINESS, it is also observed that ANGER emojis may interact with texts expressing SADNESS to express a complex emotion of DISSATISFIED as in (44).

(44) 马云双十一赚了那么多.... 也不帮助一下人家.....哎.... 

ma yun	*shuang shi yi*	*zuan le*	*na me*	*duo…*
ye	*bu*	*bang zhu yi xia*	*ren jia …*	*ai…*
Ma Yun	Nov 11	earn	that	much…
also	not	help	that person…	sigh…

'Ma Yun earned so much in Nov 11, but he still did not offer help to him, sigh…  '

In (44), 人家 'that person' refers to a disabled young man who still needs to take care of his mother. The text in (44) expresses a SADNESS emotion as indicated by the interjection 哎 'sigh'. However, the side eye emoji  represents ANGER.  reacts to the text and expresses DISSATISFIED which is composed of SADNESS as the primary emotion and ANGER as the secondary. The writer is DISSATISFIED with Ma Yun as he has made a lot of money but does not offer any help to the people in need. In that case, the emoji  serves as a modifier to complement the expression of a complex emotion.

As for the co-occurrence of an ANGER emoji and text expressing SURPRISE, it is always the text which determines the overall emotion as exemplified in (45).

(45) 车是怎么能被偷的😅😅😅

che	*shi*	*zen me neng*	*bei tou de*
car	is	how come	be stolen

'How come the car is stolen😅😅😅'

In (45), the text expressing SURPRISE co-occurs with the ANGER emojis 😅😅😅. Although the emoji 😅 is typically used to express ANGER, the three question marks indeed indicate that the emoji somewhat shows curiosity. When the writer of (45) is curious about how the car could be stolen, the emojis function as an amplifier of SURPRISE.

5.4 Summary

This chapter deals with the linguistic features of implicit emotions at a sentence level and a discourse level. At the sentence level, various syntactic structures containing words of different parts-of-speech are proposed, including adverbs, conjunctions and a verb. I present how implicit emotions are expressed using these structures; and show how the position of these words affects the expressed emotion. Given that rhetorical questions are found to play an important role in implicit emotion expressions, I examine how different emotions can be expressed by means of rhetorical questions. Statistics confirm that rhetorical questions are a rather productive means to express emotions, in particular SURPRISE, ANGER and FEAR. I also claim that different question types may have different preference towards a specific emotion.

At the discourse level, I examine the interplay between emoji and its accompanying linguistic text when the two channels are at odds. It is observed that HAPPINESS emojis are most often used in an atypical way, as compared the other emotions. Furthermore, HAPPINESS emojis are often used ironically to express an ANGER emotion. The overall emotion is mostly determined by the text (i.e. ANGER), meaning that the emojis representing HAPPINESS are frequently used ironically. When the two channels conflict with each other, the overall emotion is mostly determined by text regardless of the emotion type. It is suggested that (1) when the emoji is used in an ironic way or (2) the emoji expressing a secondary emotion of a complex emotion interacts with the text expressing a primary emotion of a complex emotion, the overall emotion should be determined by the text. When the overall emotion is determined by text, an emoji can function as an iconic means to strengthen the emotion expressed by text. It can also function as a modifier to enrich the complexity of the emotion.

CHAPTER 6
EMOTIONS AND EVENTS

In Chapter 1, I have raised two research questions: (1) How are implicit emotions expressed in text, and (2) what kinds of events trigger different emotions? In Chapter 4 and 5, I have addressed the first question by investigating various linguistic characteristics of implicit emotions and the relevant linguistic cues in terms of semantic, syntactic and discourse features. In this chapter, I address the second question by examining the interaction between emotion and event. Section 6.1 presents the correlation between emotion and event based on the annotated event-comment corpus. A list of event types that are strongly connected with a particular emotion is proposed. Section 6.2 discusses the interplay of emotion, event and semantic role. Section 6.3 sums up the contributions made in this chapter.

6.1 The Correlation between Emotions and Events

As mentioned in Section 3.3.1 TimeML (Sauri et al. 2009) is adopted for the markup of events. According to the TimeML annotation guidelines, events in English can be denoted by verbs, nouns, adjectives, prepositional phrases, or other elements such as locative adverbs. Given that Chinese typically encodes events by means of certain word classes, I only consider events denoted by verbs, nouns, and adjectives. After marking up all the events in the heading of the 200 posts, I translate the events from Chinese into English in order to map those events to the corresponding

categories in WordNet. The reason for adopting WordNet is because the coverage of WordNet categories is extensive. Moreover, WordNet categories are presented in a hierarchical structure which can be used to group similar events that trigger the same emotion. By doing so, I will be able to tell which event types (i.e. synsets) and their hyponyms are highly associated with a specific emotion. It is believed that WordNet is a good language resource for event classification.

In addition, all the events marked up in the 200 posts are annotated with semantic roles of the arguments or adjuncts (i.e. termed 'frame elements' in FrameNet) using FrameNet. In other words, each annotated event is assigned to a frame in FrameNet, and each frame corresponds to a category in WordNet. As FrameNet contains more than 1,200 semantic frames which provide a set of training data for sematic role labelling and it has been influential in the field of both linguistics and natural language processing in the past decade, I use the semantic frames proposed in FrameNet to label semantic roles of arguments and adjuncts mentioned in the events. The annotation of semantic roles is for the annotation of opinion target (i.e. emotion trigger). In this work, an opinion target is defined as a frame element by which an emotion is evoked. By annotating opinion target(s) of each comment, I will be able to reveal the relation between emotion and event type as well as the relation between emotion and a particular person/ element that has involved in an event/ some events. The relation between emotion and event types can be unveiled as each assigned element belongs to one of the events mentioned in a post. An emotion of a single comment can be elicited by different elements of

the same event, or it can be elicited by elements of different events that refer to the same person/ thing. To be simpler, an emotion in a comment may be triggered by one or more than one event. In order to show the relationship among event, opinion target and emotion, an example of a post and a comment is given in Figure 6.1.

Post:
<Text>【日本士兵[e1-承认]南京大[e2-屠杀]CG[e3-还原]枪杀现场】日本电视台于 5 月 14 日播出了南京大屠杀的调查纪录片《南京事件 2——检验历史修正主义》。在纪录片中，日本士兵描述了 1937 年 12 月 16、17 日如何杀害中国俘虏，承认当时杀死数万中国人。并用 CG 动画还原了令人心痛的枪杀现场。???05 月 15 日 19:52

①- e1: 日本士兵: Speaker
②- e1: 南京大屠杀: Topic
③- e1: 纪录片: Medium
④- e1: 杀死数万中国人: Information
⑤- e2: 日本士兵: Killer
⑥- e2: 中国俘虏: Victim
⑦- e3: 枪杀现场: Original
⑧- e3: CG 动画: Copy
</Text>

Comment:

```
<Comment ID="33">
  <Emotion>Anger</Emotion>
  <Emotion_Keyword>None</Emotion_Keyword>
  <Tool><A>卧槽 那个士兵的话语中还是透露着变态的思想$那个士兵的话语中还是透露着变态的思想$none<A></Tool>
  <OpinionTarget>1</OpinionTarget>
  <RhetoricalQuestion Type="None">None</RhetoricalQuestion>
  <Text>卧槽 那个士兵的话语中还是透露着变态的思想</Text>
  <QuestionEmotion> </QuestionEmotion>
</Comment>
```

Figure 6.1: An Example of a Post and a Comment

Figure 6.1 illustrates an example of a post and a comment taken from the corpus. As shown in the heading of the post in Figure 6.1, the post contains three sub-events,

namely [e1-承认], [e2-屠杀], and [e3-还原]. The frame elements of each sub-event are listed right below the content of the post as in (1) – (8) for the annotation of opinion target(s). For example, (1) is assigned to 日本士兵 'Japanese soldier' who plays a role in [e1-承认] as a speaker. The annotation of opinion target helps unveil the correlation between events and emotions. Take the comment in Figure 6.1 as an example. The comment expresses an ANGER emotion triggered by 日本士兵 'Japanese soldier'. However, 日本士兵 'Japanese soldier' is the speaker of [e1-承认], and the killer of [e2-屠杀]. As indicated by 那个士兵的话语 'the soldier's speech' in the comment, ANGER is triggered by the 日本士兵 'Japanese soldier' who conducted the event of [e1-承认] but not the event of [e2-屠杀]. Therefore, the opinion target (i.e. emotion trigger) should be tagged as ① which refers to the soldier in [e1-承认]. Therefore, the ANGER emotion expressed in this comment is triggered by [e1-承认] but not the other sub-events. After collecting all the comments annotated with opinion targets, the correlation between emotions and event types can be revealed.

As each sub-event in all the 200 posts has been mapped to the WordNet categories, the terms "event" and "synset" are used interchangeably. To explore the correlation between events and emotions, I calculate the synset entropy, the count (i.e. the occurrence of a synset being linked to an emotion), the total count (i.e. the occurrence of a synset being linked to the five emotions), and the conditional probability of each synset being linked to a particular emotion.

Instead of classifying events using a pre-determined event list, I look for event

types that have an obvious tendency towards a particular emotion based on these figures. There are drawbacks to the use of a pre-determined event list, be it fine-grained or coarse-grained. First, if I rashly generate a fine-grained list of event types, some event types may contain only several tokens which may not be statistically significant enough even if they all trigger the same emotion. Moreover, a fine-grained list of event types may not be of much use for emotion studies if similar events associated with a particular emotion are not grouped into the same event type. Second, if a coarse-grained list of event types is generated, an event type may not have an obvious tendency towards a particular emotion. For example, it would be rather difficult to show the correlation between an event type and emotion if events are simply classified into a binary classification such as *event* and *state*. This is also the reason why I do not make use of any existing pre-determined event lists for the event classification. With the help of the hierarchy of event types given in WordNet, the correlation between events and emotions can be revealed.

The synset entropy H of the emotion category variable E for given a synset $S=s$ is defined as follows.

$H(E|S=s) = -\sum_{e\in\Omega_E} P(E=e|S=s)\log_2 P(E=e|S=s)$, where

$\Omega_E = \{Anger, Fear, Happiness, Sadness, Surprise\}$

The synset entropy $H(E|S=s)$ measures how many bits are needed in order to describe its distribution among the five emotions. The smaller the entropy, the stronger the correlation between a synset and an emotion.

As for the count and the total count, the former is calculated by counting the

occurrence of a synset being linked to a specific emotion, while the latter is calculated by counting the occurrence of a synset being linked to the five emotion. The conditional probability $P(E = e|S = s)$ is calculated by counting the co-occurrence of a certain emotion $E = e$ and a synset $S = s$, and the total number of $S = s$, where the synset *s* has been annotated as associated with all possible emotions.

$$P(E = e|S = s) = \frac{count(E{=}e,\ S{=}s)}{\sum_{e \in S_E} count(E{=}e,\ S{=}s)}$$

Since one emotion can be targeted on or triggered by more than one event (corresponding to proper synsets), the counting is done by assigning an emotion instance a weight 1.0. If an emotion in a comment is triggered by multiple events, then the total weight 1.0 will be split equally to all the events. Thus, the count of $(E = e,\ S = s)$ could be non-integers. The larger the count is, the more significant an event type is statistically linked to a certain emotion.

To figure out the correlation between event types and emotions, I compute the synset entropy, the count, the total count and the percentage of all the event types that are associated with an emotion(s) in the 200 posts. The total number of event types in the 200 posts is 732. Among all the 200 posts, the maximum number of events identified is 7, and the minimum is 1. The distribution of events is shown as in Figure 6.2.

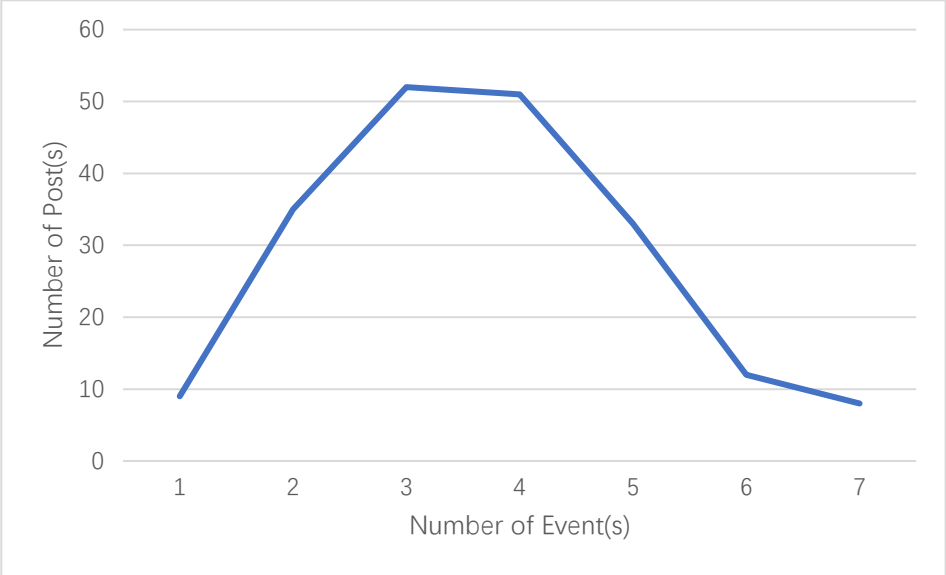
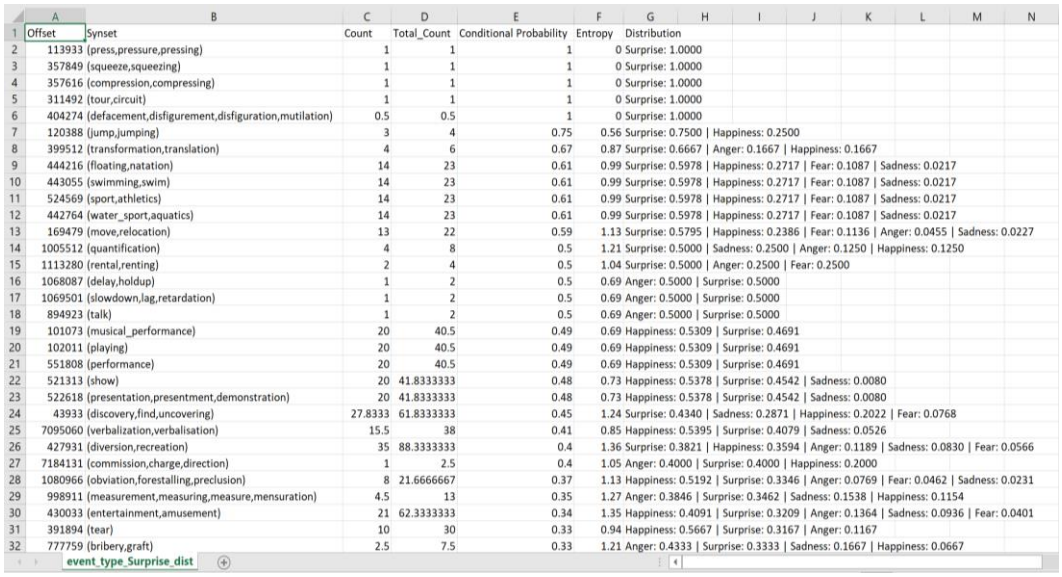


Figure 6.2: The Distribution of Events in all the 200 Posts

As shown in Figure 6.2, the distribution of events is in an inverted U-shaped curve, meaning that most of the 200 posts are composed of 3 to 4 events on average. As some events do not trigger any emotions at all and some events in different posts belong to the same WordNet category, the total number of event types that are associated with an emotion(s) is 504, including the hypernyms of events. The count, total count, entropy, conditional probability and emotion distribution of event types for each emotion are generated and arranged in an excel file, part of the list for the FEAR emotion is illustrated as in Figure 6.3.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Offset	Synset	Count	Total_Count	Conditional Probability	Entropy	Distribution							
2	113933	(press,pressure,pressing)	1	1	1	0	Surprise: 1.0000							
3	357849	(squeeze,squeezing)	1	1	1	0	Surprise: 1.0000							
4	357616	(compression,compressing)	1	1	1	0	Surprise: 1.0000							
5	311492	(tour,circuit)	1	1	1	0	Surprise: 1.0000							
6	404274	(defacement,disfigurement,disfiguration,mutilation)	0.5	0.5	1	0	Surprise: 1.0000							
7	120388	(jump,jumping)	3	4	0.75	0.56	Surprise: 0.7500 | Happiness: 0.2500							
8	399512	(transformation,translation)	4	6	0.67	0.87	Surprise: 0.6667 | Anger: 0.1667 | Happiness: 0.1667							
9	444216	(floating,natation)	14	23	0.61	0.99	Surprise: 0.5978 | Happiness: 0.2717 | Fear: 0.1087 | Sadness: 0.0217							
10	443055	(swimming,swim)	14	23	0.61	0.99	Surprise: 0.5978 | Happiness: 0.2717 | Fear: 0.1087 | Sadness: 0.0217							
11	524569	(sport,athletics)	14	23	0.61	0.99	Surprise: 0.5978 | Happiness: 0.2717 | Fear: 0.1087 | Sadness: 0.0217							
12	442764	(water_sport,aquatics)	14	23	0.61	0.99	Surprise: 0.5978 | Happiness: 0.2717 | Fear: 0.1087 | Sadness: 0.0217							
13	169479	(move,relocation)	13	22	0.59	1.13	Surprise: 0.5795 | Happiness: 0.2386 | Fear: 0.1136 | Anger: 0.0455 | Sadness: 0.0227							
14	1005512	(quantification)	4	8	0.5	1.21	Surprise: 0.5000 | Sadness: 0.2500 | Anger: 0.1250 | Happiness: 0.1250							
15	1113280	(rental,renting)	2	4	0.5	1.04	Surprise: 0.5000 | Anger: 0.2500 | Fear: 0.2500							
16	1068087	(delay,holdup)	1	2	0.5	0.69	Anger: 0.5000 | Surprise: 0.5000							
17	1069501	(slowdown,lag,retardation)	1	2	0.5	0.69	Anger: 0.5000 | Surprise: 0.5000							
18	894923	(talk)	1	2	0.5	0.69	Anger: 0.5000 | Surprise: 0.5000							
19	101073	(musical_performance)	20	40.5	0.49	0.69	Happiness: 0.5309 | Surprise: 0.4691							
20	102011	(playing)	20	40.5	0.49	0.69	Happiness: 0.5309 | Surprise: 0.4691							
21	551808	(performance)	20	40.5	0.49	0.69	Happiness: 0.5309 | Surprise: 0.4691							
22	521313	(show)	20	41.8333333	0.48	0.73	Happiness: 0.5378 | Surprise: 0.4542 | Sadness: 0.0080							
23	522618	(presentation,presentment,demonstration)	20	41.8333333	0.48	0.73	Happiness: 0.5378 | Surprise: 0.4542 | Sadness: 0.0080							
24	43933	(discovery,find,uncovering)	27.8333	61.8333333	0.45	1.24	Surprise: 0.4340 | Sadness: 0.2871 | Happiness: 0.2022 | Fear: 0.0768							
25	7095060	(verbalization,verbalisation)	15.5	38	0.41	0.85	Happiness: 0.5395 | Surprise: 0.4079 | Sadness: 0.0526							
26	427931	(diversion,recreation)	35	88.3333333	0.4	1.36	Surprise: 0.3821 | Happiness: 0.3594 | Anger: 0.1189 | Sadness: 0.0830 | Fear: 0.0566							
27	7184131	(commission,charge,direction)	1	2.5	0.4	1.05	Anger: 0.4000 | Surprise: 0.4000 | Happiness: 0.2000							
28	1080966	(obviation,forestalling,preclusion)	8	21.6666667	0.37	1.13	Happiness: 0.5192 | Surprise: 0.3346 | Anger: 0.0769 | Fear: 0.0462 | Sadness: 0.0231							
29	998911	(measurement,measuring,measure,mensuration)	4.5	13	0.35	1.27	Anger: 0.3846 | Surprise: 0.3462 | Sadness: 0.1538 | Happiness: 0.1154							
30	430033	(entertainment,amusement)	21	62.3333333	0.34	1.35	Happiness: 0.4091 | Surprise: 0.3209 | Anger: 0.1364 | Sadness: 0.0936 | Fear: 0.0401							
31	391894	(tear)	10	30	0.33	0.94	Happiness: 0.5667 | Surprise: 0.3167 | Anger: 0.1167							
32	777759	(bribery,graft)	2.5	7.5	0.33	1.21	Anger: 0.4333 | Surprise: 0.3333 | Sadness: 0.1667 | Happiness: 0.0667							

event_type_Surprise_dist

Figure 6.3: Count, Entropy and Emotion Distribution of Event Types

In order to generate a list of event types that are statistically significant and strongly correlated with a particular emotion, I marked the boundary of the acceptance region with critical values for the selection of event types that show a tendency to a certain emotion as in Table 6.1.

Table 6.1: Critical Values for the Selection of Event Types

	Critical Value
Conditional Probability	$\geq 80\%$
Synset Entropy	< 0.75
Count	≥ 25

The conditional probability must be greater than or equal to 80%, meaning that whenever a synset is the trigger of an emotion, at least 80% of the synset trigger a specific emotion.

As for the entropy value, it ranges from 0 to 1.46 according to the annotated data. It is observed that the emotion distribution of event types that have an entropy

higher than 0.75 is relatively scattered than those of an entropy lower than 0.75, as none of an event type of an entropy higher than 0.75 shows a tendency to a particular emotion of more than 80%.

As for the count value, it ranges from 0.33 to 4761.3. While the low count value means that a particular event type is not often an emotion trigger, the high value of count represents that an event type is statistically significant. However, selecting an event type of extremely high value of count does not necessarily mean the event type is of great value to emotion studies. The large number is basically due to the large number of hyponyms it contains. For example, the category that has a count of 4761.3 is "event". "Event" is the root of the hierarchy containing all the remaining categories as the root of the hierarchy. Even though it is most statistically significant, the conditional probability of it being linked to *anger* is only 0.56, and the synset entropy is 1.24. That means the distribution of emotions of "event" is rather scattered. Given that 50 comments of each post are annotated, I consider 25 or above the most reasonable count value to prove the statistical significance. This is because even if a synset occurs only once in a post in the entire corpus, the synset is still the major trigger of a specific emotion in that post. Having more than a half of annotated comments expressing a specific emotion means that it is not just a coincidence that the synset correlates with that emotion.

It is found that the synsets falling into the acceptance region only show a strong correlation with HAPPINESS, SADNESS, or ANGER, none of them strongly correlates with FEAR and SURPRISE. It may be attributed to the limited number of

comments expressing FEAR (573 instances) and SURPRISE (698 instances). Besides, comments expressing FEAR and SURPRISE are scattered all over the posts. Therefore, FEAR and SURPRISE are usually triggered by different sub-events in different posts. Even if a synset shows a high conditional probability linking to FEAR or SURPRISE, the count value of the synset is still lower than 25. Therefore, I only discuss synsets showing a strong correlation with HAPPINESS, SADNESS, or ANGER as shown in Table 6.2 – Table 6.4, respectively.

Table 6.2: Event Types Highly Associated with Happiness

	Synset	Count	Conditional Probability	Entropy	Event-denoting Word(s)
1	(move)	27	1	0	举动
2	(purchase)	26	1	0	买, 购票
3	(rescue,deliverance, delivery,saving)	192	0.94	0.35	救下, 救, 救救, 施救
4	(derring-do)	33	0.94	0.28	站出来
5	(first_aid)	30.5	0.94	0.27	插管急救
6	(cuddle,nestle, snuggle)	43	0.93	0.40	依偎
7	(law_enforcement)	29	0.91	0.40	执法
8	(birth,nativity, nascency,nascence)	41.2	0.90	0.41	生日, 出生, 生子
9	(disapproval)	26	0.90	0.40	(上演)批评
10	(defense,defence)	37	0.88	0.51	守护
11	(lecture,lecturing)	33.5	0.86	0.47	讲题
12	(assembly, assemblage,gathering)	27	0.82	0.69	带

Table 6.3: Event Types Highly Associated with Sadness

	Synset	Count	Conditional Probability	Entropy	Event-denoting Word(s)
1	(calamity,catastrophe, disaster,tragedy, cataclysm)	34	0.89	0.41	暴雨, 受灾
2	(vending,peddling, hawking,vendition)	28.5	0.89	0.42	卖
3	(sustenance,sustentation, sustainment,maintenance, upkeep)	29.5	0.87	0.69	吃饭, 吃
4	(martyrdom)	26.8	0.87	0.46	牺牲
5	(death,decease,expiry)	103.3	0.81	0.69	身亡, 心跳停止, 溺亡, 遇难, 殉职, 死亡, 去世

Table 6.4: Event Types Highly Associated with Anger

	Synset	Count	Conditional Probability	Entropy	Event-denoting Word(s)
1	(sprinkle,sprinkling, sparge)	42.5	1	0.09	泼向
2	(beat)	60	0.98	0.23	打
3	(throw)	46	0.96	0.2	扔
4	(annoyance,annoying, irritation,vexation)	71	0.95	0.33	耍酒疯, 无理取闹, 气炸
5	(ballup,balls-up, cockup,mess-up)	62.8	0.94	0.32	乱象
6	(affirmation,assertion, statement)	58.2	0.94	0.29	称, 承认, 发布声明
7	(threat)	54	0.94	0.31	威胁
8	(feeding,alimentation)	48	0.94	0.26	投食
9	(conversion)	46	0.94	0.48	策反, 为间谍
10	(restitution,return, restoration,regaining)	33	0.94	0.28	归还
11	(demand)	81.5	0.93	0.34	讨, 索要, 要求
12	(smack,smacking,slap)	138	0.92	0.37	掌掴, 扇耳光
13	(violation,infringement)	103	0.92	0.46	认账, 闯, 违反, 违规
14	(harassment, molestation)	70.5[6]	0.91	0.51	猥亵
15	(breakage,break, breaking)	60.8	0.91	0.39	踩碎, 踹碎, 怒摔, 破窗
16	(prick,pricking)	39.5	0.91	0.5	针扎, 扎
17	(climb,mount)	27	0.9	0.39	翻越, 爬, 爬坡
18	(sackcloth_and_ashes)	30.8	0.89	0.54	说对不起, 道歉, 致歉
19	(investigation, investigating)	26	0.88	0.45	调查核实, 介入调查, 安排暗访, 查, 不明, 调查
20	(larceny,theft,thievery, thieving,stealing)	142.8	0.87	0.49	盗窃, 偷
21	(kick,boot,kicking)	96.7	0.87	0.5	踹, 踹打, 踢

[6] As the synset (harassment,molestation) is the hypernym of (tease,teasing,ribbing,tantalization), the count of (harassment,molestation) also includes the count of (tease,teasing,ribbing,tantalization).

Table 6.4: Event Types Highly Associated with Anger, *continued*

	Synset	Count	Conditional Probability	Entropy	Event-denoting Word(s)
22	(maltreatment, ill-treatment, ill-usage, abuse)	42.8	0.86	0.58	骂
23	(insult,affront)	41.2	0.86	0.54	辱骂, 侮辱
24	(flow,stream)	39.5	0.86	0.55	流向
25	(tease,teasing,ribbing, tantalization)	34	0.86	0.66	骚扰
26	(denial)	168.2	0.84	0.61	拒, 拒绝, 拒载
27	(robbery)	103.3	0.84	0.53	抢
28	(arson,incendiarism, fire-raising)	43	0.84	0.5	纵火案
29	(swindle,cheat,rig)	82.5	0.83	0.73	骗, 碰瓷
30	(address,speech)	65.2	0.83	0.65	言论
31	(push,pushing)	44	0.83	0.6	推入
32	(blunder,blooper, bloomer,bungle,pratfall, foul-up,fuckup,flub, botch,boner,boo-boo)	34	0.83	0.64	失手
33	(miss,misfire)	41	0.82	0.6	错过
34	(damage,harm,hurt, scathe)	182.2	0.81	0.69	乱踢扔飞, 殴打, 脚踹拳殴, 围殴
35	(capture, gaining_control,seizure)	94	0.81	0.68	查获, 占, 霸
36	(contamination, pollution)	41.5	0.81	0.65	发黑发臭, 腐臭难闻, 发霉
37	(loss)	38.3	0.81	0.55	丢, 败给
38	(tirade,philippic, broadside)	29	0.81	0.65	大放厥词
39	(destruction, devastation)	66	0.8	0.67	破坏, 严重, 冲撞打砸
40	(domestic_violence)	44.7	0.8	0.67	家暴, 虐打
41	(indication, denotation)	35	0.8	0.58	点名

In Table 6.2- 6.4, I exclude some categories which are the hypernyms of certain event types as their hyponyms are already included in the list. These hypernyms do not have any corresponding event-denoting words. For example, no events in the dataset are annotated as the synset (stroke), but (stroke) is the hypernym of (beat), (smack,smacking,slap), and (kick,boot,kicking). All the three hyponyms are used to annotate events and have event-denoting words but (stroke) does not. Therefore, the synset (stroke) is removed from the list. Apart from that, I also remove those hypernyms with a count value lower than 25 after reducing the sum of the total count of their hyponyms. For example, the synset (display) is the hypernym of (sackcloth_and_ashes). The count value of both synsets are the same, meaning that the ANGER emotion is only triggered by the hyponym but not the hypernym. Therefore, the hypernym should be removed in that case.

As shown in Table 6.2-6.4, there are 13 event types of HAPPINESS, 5 of SADNESS, and 41 of ANGER found in the acceptance region. The corresponding event-denoting word(s) belonging to these event types are also listed. It is not surprising that most event-denoting words in the list of HAPPINESS are either positive or neutral events, except for some events such as 批評 'criticize'; the events in the list of SADNESS are either neutral or negative; and the events in the list of ANGER are mostly negative. Although one may be able to predict the emotion a synset is associated with in some cases, without the empirical data, it is impossible to generate a list of event types that are statistically significant for each emotion.

In order to summarize a shorter list of event types, I group event types that

express the same emotion using the WordNet hierarchy. I use all the event types listed in Table 6.2-6.4 and generate a simpler version of WordNet hierarchy as in Appendix II. To be more specific, when both a hypernym and its hyponym(s) are listed in Table 6.2-6.4, and they all show a tendency (≥ 70%) to the same emotion in the emotion distribution, the hypernym is taken as the category to group its hyponym(s). For example, the synset (death, decrease, expiry) is a close hypernym (level 6) of the synset (martyrdom) in the WordNet hierarchy. Both synsets are event types listed in Table 6.3 which have a strong correlation with SADNESS. To shorten the list and group similar events together, the hypernym (death, decrease, expiry) is taken as an event category in Table 6.5, and its hyponym (martyrdom) is under this category. However, if a hypernym is not in Table 6.2-6.4 but it contains at least two hyponyms, and all of them show a tendency (≥ 70%) to an emotion in the emotion distribution, I go one level higher and use the hypernym to group the hyponyms in Table 6.2-6.4. For example, the synsets of (beat), (smack,smacking,slap), and (kick,boot,kicking) in Table 6.5 are the hyponyms of (stroke). The synset (stroke) is not listed in Table 6.5 as it has no corresponding event-denoting word in the corpus. All the three synsets correlate with ANGER. As the hypernym (stroke) of the three synsets is also found to be highly associated with ANGER (0.91), the hypernym is therefore used to group the three synsets as demonstrated in Table 6.5. The list of event types showing a strong correlation to a specific emotion is illustrated as in Table 6.5

Table 6.5: The Summary of Event Types

	Emotion	Synset, Probability being Linked to that Emotion & Count (include synsets in Table 6.2-6.4 only)	Hyponym(s)	Event-denoting Words
1	Happiness	(move) Probability: 1; Count: 27	N/A	举动
2	Happiness	(purchase) Probability: 1; Count: 26	N/A	买, 购票
3	Happiness	(first_aid) Probability: 0.94; Count: 30.5	N/A	插管急救
4	Happiness	(derring-do) Probability: 0.94; Count: 33	N/A	站出来
5	Happiness	(rescue,deliverance,delivery,saving) Probability: 0.94; Count: 192	N/A	救下, 救, 救救, 施救
6	Happiness	(cuddle,nestle,snuggle) Probability: 0.93; Count: 43	N/A	依偎
7	Happiness	(law_enforcement) Probability: 0.91; Count: 29	N/A	执法
8	Happiness	(birth,nativity,nascency,nascence) Probability: 0.90; Count: 41.2	N/A	生日, 出生, 生子
9	Happiness	(disapproval) Probability: 0.90; Count: 26	N/A	(上演)批评
10	Happiness	(defense,defence) Probability: 0.88; Count: 37	N/A	守护
11	Happiness	(lecture,lecturing) Probability: 0.86; Count: 33.5	N/A	讲题

Table 6.5: The Summary of Event Types, *continued*

12	Happiness	(assembly,assemblage,gathering) Probability: 0.82; Count: 27	N/A	带
13	Sadness	(calamity,catastrophe,disaster,tragedy,cataclysm) Probability: 0.89; Count: 34	N/A	暴雨，受灾
14	Sadness	(vending,peddling,hawking,vendition) Probability: 0.89; Count: 28.5	N/A	卖
15	Sadness	(sustenance,sustentation,sustainment,maintenance,upkeep) Probability: 0.87; Count: 29.5	N/A	吃饭，吃
16	Sadness	(death,decease,expiry) Probability: 0.81; Count: 103.3	(martyrdom)	身亡，心跳停止，溺亡，遇难，殉职，死亡，去世，牺牲
17	Anger	(sprinkle,sprinkling,sparge) Probability: 1; Count: 42.5	N/A	泼向
18	Anger	(feeding,alimentation) Probability: 0.94; Count: 48	N/A	投食
19	Anger	(restitution,return,restoration,regaining) Probability: 0.94; Count: 33	N/A	归还
20	Anger	(conversion) Probability: 0.94; Count: 46	N/A	策反，为间谍
21	Anger	(demand) Probability: 0.93; Count: 81.5	N/A	讨，索要，要求
22	Anger	(stroke) Probability: 0.91; Count: 294.7	(beat) (smack,smacking,slap) (kick,boot,kicking)	打，掌掴，扇耳光，踹，踹打，踢

Table 6.5: The Summary of Event Types, *continued*

23	Anger	(climb,mount) Probability: 0.90; Count: 27	N/A	翻越, 爬, 爬坡
24	Anger	(violation,infringement) Probability: 0.92; Count: 103	N/A	认账, 闯, 违反, 违规
25	Anger	(mistreatment) Probability: 0.91; Count: 200.3	(annoyance,annoying,irritation,vexation) (harassment,molestation) (tease,teasing,ribbing,tantalization) (maltreatment,ill-treatment,ill-usage,abuse)	耍酒疯, 无理取闹, 气炸, 猥亵, 骚扰, 骂
26	Anger	(prick,pricking) Probability: 0.91; Count: 39.5	N/A	针扎, 扎
27	Anger	(investigation,investigating) Probability: 0.88; Count: 26	N/A	调查核实, 介入调查, 安排暗访, 查, 不明, 调查
28	Anger	(nonaccomplishment,nonachievement) Probability: 0.87; Count: 135.1	(ballup,balls-up,cockup,mess-up) (blunder,blooper,bloomer,bungle,pratfall,foul-up,fuckup,flub,botch,boner,boo-boo) (loss)	乱象, 失手, 丢, 败给
29	Anger	(flow,stream) Probability: 0.86; Count: 39.5	N/A	流向
30	Anger	(insult,affront) Probability: 0.86; Count: 41.2	N/A	辱骂, 侮辱
31	Anger	(miss,misfire) Probability: 0.82; Count: 41	N/A	错过

Table 6.5: The Summary of Event Types, *continued*

32	Anger	(contamination,pollution) Probability: 0.81; Count: 41.5	N/A	发黑发臭，腐臭难闻，发霉
33	Anger	(capture,gaining_control,seizure) Probability: 0.81; Count: 94	N/A	查获，占，霸
34	Anger	(destruction,devastation) Probability: 0.80; Count: 66	N/A	破坏，严重，冲撞打砸
35	Anger	(domestic_violence) Probability: 0.80; Count: 44.7	N/A	家暴，虐打
36	Anger	(propulsion,actuation) Probability: 0.77; Count: 90	(throw) (push, pushing)	扔，推入
37	Anger	(crime,criminal_offense,criminal_offence,law-breaking) Probability: 0.76; Count: 328.6	(larceny,theft,thievery,thieving,stealing) (robbery) (swindle,cheat,rig)	盗窃，偷，抢，骗，碰瓷
38	Anger	(change_of_integrity) Probability: 0.73; Count: 325. 7	(breakage,break,breaking) (arson,incendiarism,fire-raising) (damage,harm,hurt,scathe)	踩碎，踹碎，怒摔，破窗，纵火案，乱踢扔飞，殴打，脚踹拳殴，围殴
39	Anger	(speech_act) Probability: 0.73; Count: 440.4	(affirmation,assertion,statement) (sackcloth_and_ashes) (denial) (address,speech) (tirade,philippic,broadside) (indication,denotation) (threat)	称，承认，发布声明，威胁，说对不起，道歉，致歉， 拒，拒绝，拒载，言论，大放厥词，点名

As demonstrated in Table 6.5, the 13 event types of HAPPINESS can only be grouped into 12 types and the 5 of SADNESS can only be grouped into 4. It can be attributed to the small number of event types of the two emotions, which leads to the scattered distribution in the WordNet hierarchy. As for the case of ANGER, the number of event types has reduced from 41 to 23, which means the use of WordNet for event annotation and classification can effectively group similar events which are highly associated with a certain emotion. By grouping these hyponyms using this approach, one can predict the emotion associated with that hypernym and all its hyponyms more accurately. To sum up, I propose in Table 6.5 the synsets that have an obvious tendency towards an emotion and all of them are statistically significant at the same time.

Based on the event types listed in Table 6.5, I make use of some synsets to investigate the interplay of emotion, event type, and semantic role in the following section.

6.2 The Interplay of Emotion, Event and Semantic Role

In Section 6.1, I briefly mention the annotation of opinion targets. By the term 'opinion target', it refers to a (frame) element by which an emotion is triggered. The aim of the annotation of opinion target is to see whether an emotion is highly related to a particular person/ element who has conducted an event/ some events. I also generate a list of event types that show an obvious tendency to one of the five emotions. Statistics show that the synsets only show a strong correlation with HAPPINESS, SADNESS, or ANGER, none of them is highly assocaited with FEAR and SURPRISE. Therefore, I only discuss the interplay of emotion, event, and semantic role concerning HAPPINESS, SADNESS, or ANGER. It is observed that most events in the list of HAPPINESS are positive or neutral; the events in the list of SADNESS are either neutral or negative; and the events in the list of ANGER are mostly negative.

In this section, I further investigate how semantic role of an opinion target affects the emotion expressed. I hypothesize that (1) HAPPINESS is mainly triggered by the doer of an event, (2) SADNESS is mainly triggered by the undergoer of an event, and (3) ANGER is mainly triggered by the doer of an event. In the following sub-sections, I choose a synset of the highest count value from each of the three emotions to verify the three hypotheses.

6.2.1 Semantic Role that Triggers Happiness

The synset (rescue,deliverance,delivery,saving) is chosen from the list of event

types that are highly associated with HAPPINESS. Among all the synsets concerning HAPPINESS, it has the highest count. It contains several event-denoting verbs, including 救下, 救, 救救, 施救 'rescue'. I extract a post containing one of these verbs in the heading. The post and the semantic role labels are shown in Figure 6.4.

Post: 【环卫阿姨 [e1-救下]残疾小狗，5 年后小狗又[e2-救了]阿姨】 【*huan wei a yi [e1-jiu xia] can ji xiao gou, 5 nian hou xiao gou you [e2-jiu le] a yi*】 【female sanitation worker [e1-save-ASP] disabled puppy, 5 year later puppy also [e2-save-ASP] worker】 '[A female sanitation worker saved a disabled puppy. Five years later, the puppy saved her in return]' 11 月 15 日，辽宁沈阳。一只白色的小狗坐在一辆环卫车上，跟着主人上下班。环卫白阿姨称，这只狗 5 年前因为腿瘸被遗弃，她从狗贩子手里救下来。5 年后，自己脑出血发病，狗狗叫唤才惊动了邻居，最终自己被及时救治。???11 月 20 日 08:20 'On November 15, Shenyang, Liaoning. A white puppy sits in a sanitation truck and follows his master to work. A female sanitation worker said that the dog was abandoned 5 years ago because of a lame leg and she rescued the dog from the dog dealer. Five years later, the cerebral haemorrhage suddenly occurred in the worker, and the white puppy called to disturb her neighbour. She was finally treated in time. ??? November 20 08:20'
Semantic Role Labels: ①- Rescuing_e1: 环卫阿姨: Agent ②- e1: 腿瘸被遗弃, 狗贩子手里: Harmful_situation ③- e1: 残疾小狗: Patient ④- Rescuing_e2: 残疾小狗: Agent ⑤- e2: 脑出血发病: Harmful_situation ⑥- e2: 环卫阿姨: Patient

Figure 6.4: An Example of (rescue,deliverance,delivery,saving) Event

The post in Figure 6.4 contains two rescuing events concerning a sanitation work and a disabled puppy. According to FrameNet, a rescuing event involves an agent, a harmful situation and a patient. The agent of a rescuing event is the doer who volitionally helps someone out of a harmful situation; the patient is the one who undergoes the event of being rescued. In the event in Figure 6.4, the sanitation worker is the doer of the first rescuing event (*e1*) and the undergoer of the second rescuing event (*e2*), and the puppy is the undergoer of *e1* and the doer of *e2*.

Among the 50 annotated comments, 29 of them express a HAPPINESS emotion. Each comment is annotated with an opinion target(s) using the corresponding numbers of the semantic role labels. An example is given in (1).

(1) 为　善良的　人　点赞
wei　*shan liang de*　*ren*　*dian zan*
for　kind-hearted　person　thumbs-up
'Thumbs-up for this kind-hearted person'

(1) demonstrates that the writer is MOVED (i.e. HAPPINESS) by the sanitation worker who saved the disabled puppy. The emotion is triggered by the agent of the first rescuing event as indicated by the noun 善良的人 'kind-hearted person'. In that case, the emotion is tagged as HAPPINESS, and the opinion target is tagged as ① referring to 环卫阿姨 'a female sanitation worker' as the agent of the first rescuing event.

I summarize the opinion target(s) tagged in the 29 comments expressing HAPPINESS as in Table 6.6.

Table 6.6 Opinion Target Tagged in Comments of Rescuing Events

Opinion Target	⓪- Not in the event	①- Sanitation Worker as an agent of *e1*	④- Puppy as an agent of *e2*	⑥- Sanitation Worker as a patient of *e2*
Occurrence	1	15	23	1

Table 6.6 shows the HAPPINESS emotion expressed in the rescuing event is mainly triggered by both ④ the puppy when playing the agent role in *e2* and ① the sanitation worker when playing the agent role in *e1*. As for ⑥ the sanitation worker who is the patient of *e2*, it is only tagged in one comment expressing HAPPINESS as the writer is happy that the worker was reciprocated. The finding confirms the first hypothesis that HAPPINESS is mainly triggered by the doer of an event.

6.2.2 Semantic Role that Triggers Sadness

As for the synsets that are highly associated with SADNESS, the synset (death,decease,expiry) is chosen from due to its highest count value. It contains eight different event-denoting verbs all referring to the meaning of death, namely 身亡, 心跳停止, 溺亡, 遇难, 殉职, 死亡, 去世, 牺牲. An example is given in Figure 6.5.

Post:
【[e1-失联]的浙大博士生 [e2-遇难] 遗体[e3-漂浮]江中被[e4-打捞]】
【*[e1-shi lian de] zhe da bo shi sheng [e2-yu nan]*
yi ti [e3-piao fu] jiang zhong bei [e4-da lao]】
【[e1-missing] Zhejiang University doctoral student [e2-death]
remains [e3-float] river PAS [e4-salvage]】
'[The missing doctoral student from Zhejiang University was found dead. His remains were found floating in the river and has been salvaged]'

10 月 14 日，浙江杭州。公羊救援队称，失联的 2016 级浙大博士生侯某遗体被打捞，经家属确认身份，警方正处理。10 日网友发文，侯某写遗书欲投江轻生，校方确认其失联。???10 月 14 日 20:29
'October 14, Hangzhou, Zhejiang. The Rescue Team of Ram Union said that the remains of Hou, a doctoral student from Zhejiang University who was lost, was salvaged. His family has confirmed the identity. On the 10th, netizens made a post, Hou wrote a suicide note and wanted to commit suicide by jumping into the river, the school confirmed that he was lost. ??? October 14, 20:29'

Semantic Role Labels (only those concerning the event of death):

②- Death_e2: 浙大博士生: Protagonist

Figure 6.5: An Example of (death,decease,expiry) Event

As shown in Figure 6.5, the post contains 4 sub-events, and each of them has a set of frame element(s). ① is the frame element of [e1-失联], and ③-⑦ are the elements of [e3-漂浮] ot [e4-打捞]. Only ② is listed in Figure 6.5 as it is the only semantic role concerning the death event. 33 out of 50 comments indicate that it is the protagonist (i.e. the undergoer of the death event) 浙大博士生 'doctoral student' who triggers the SADNESS emotion. Apart from SADNESS, the protagonist is also the cause of ANGER (4 tokens) and SURPRISE (2 tokens). The ANGER emotion is triggered as some believe that the protagonist died of his own volition as stated in the content of the post. Therefore, he is not only the undergoer of the death event but also the underlying cause to the happening of the event. Therefore, some were

annoyed with the protagonist for giving up his life so easily. As for the SURPRISE emotion, the writers are wondering the reason for many people committing suicide recently.

The abovementioned event confirms that the SADNESS emotion is elicited by the undergoer who passively undergoes an event. However, when a synset of SADNESS does not only involve an undergoer in the event such as 卖 'sell' in the synset of (vending,peddling,hawking,vendition), the SADNESS emotion can also be triggered by the doer of the synset. In these cases, the agent is often forced by the environment that he/ she has to do something. Take an example from the corpus for further explanation. A post reported that an old man insists on selling his product in a chilly night as he has walked from a village to the city for 6 hours. There are some comments expressing a SADNESS emotion towards the old man who is the doer of the selling event. The reason is because the poor old man is left with no choice but to sell the products. Therefore, some writers show a compassion to him and express a SADNESS emotion. In sum, it is confirmed that SADNESS is often triggered by the undergoer of an event linked to SADNESS, but when an event does not involve an undergoer, it should be the doer who triggers SADNESS.

6.2.3 Semantic Role that Triggers Anger

As summarized in Table 6.5, several synsets are grouped under the synset (speech_act). The synset (speech_act) therefore has the highest count among the synsets of ANGER, and it is chosen from the list of event types that are highly

associated with ANGER. Table 6.7 illustrates the hyponyms of (stroke) and their corresponding event-denoting words.

Table 6.7: Hyponyms and Event-denoting Words of the Synset (speech_act)

Synset	Hyponym	Event-denoting Word(s)
(speech_act)	(affirmation,assertion,statement)	称, 承认, 发布声明
	(sackcloth_and_ashes)	说对不起, 道歉, 致歉
	(denial)	拒, 拒绝, 拒载
	(address,speech)	言论
	(tirade,philippic,broadside)	大放厥词
	(indication,denotation)	点名
	(threat)	威胁

I then extracted a post which contains one of these verbs in the heading. The heading and the semantic role labels are illustrated as in Figure 6.6.

Post:
【连[e1-上 13 小时班][e2-请假]遭[e3-拒] 员工[e4-破坏]360 万元机器给自己[e5-“放假”]】
【*lian [e1-shang 13xiao shi ban] [e2-qing jia] zao [e3-ju] yuan gong [e4-po huai] 360wan yuan ji qi gei zi ji [e5-fang jia]*】
【continuiouly [e1-take 13-hour shift] [e2-ask for leave] PASS [e3-reject] employee [e4-damage] 3.6 million dollar machine let oneself [e5-on leave]】
‘[Taking a 13-hour shift continuously, a leave request was still rejected. An employee damaged a machine worth 3.6 million dollar just to give himself a leave]’

最近，陕西铜川新区警方破获了一起破坏企业生产设备的案件。犯罪嫌疑人是厂里一名有着 10 年工龄的老员工，案发前连上 13 小时班，因身体不适想请假休息，不仅没有获得批准还被要求通宵加班。由于压力过大，便想人为制造设备意外故障放假休息，据悉被损坏的设备价格 360 万。目前，冯某已被检察机关批准逮捕。11 月 10 日 20:05
‘Recently, the police in Tongchuan New District of Shaanxi Province cracked a case of destroying the production equipment of the enterprise. The suspect was an employee with 10 years of service in the factory. Before the incident, he was continuously at work for 13 hours. He wanted to take a break due to physical discomfort. Not only did he not get the approval, he was also required to work overtime overnight. Due to the excessive pressure, he intentionally caused an “accidental failure” of the manufacturing equipment. It is reported that the price of damaged equipment is 3.6 million. Currently, Feng has been arrested by the procuratorial authority. November 10th 20:05’

Semantic Role Labels (only those concerning the event of denial):

⑤- Respond_to_proposal_e3: 厂: Speaker

⑥- e3: 冯某: Interlocutor

⑦- e3: 请假: Proposal

Figure 6.6: An Example of (denial) Event

The post in Figure 6.6 is composed of 5 sub-events. Each of them has a set of frame element(s). Since [e3-拒] is the major focus in this part, I only list those frame elements concerning the event of [e3-拒]. According to the annotated data, 39 out

of 50 annotated comments express an ANGER emotion, among which 38 of them is triggered by 厂 'the enterprise' who rejected the leave request made by the employee 冯某 'Feng'. The writers are annoyed as the employee had unsympathetic treatment from the enterprise. The enterprise is the agent of the denial event, and what the agent did is considered inhuman. Therefore, it confirms the third hypothesis which suggests that *anger* is mostly triggered by the doers of events associated with ANGER.

6.3 Summary

This chapter addresses the second research question: what kinds of events trigger different emotions? In Section 6.1, I make use of several language resources including TimeML, WordNet and FrameNet to look for event types showing a strong correlation with one of the fives emotions. I take into account the count of each synset linking to a particular emotion so as to rule out the probability of a synset being associated with an emotion just by coincidence. I first employ TimeML for the markup of events denoted by verbs, nouns and adjectives. I then map those events to the WordNet categories and assigned frame elements to the arguments and adjuncts of each event. The opinion target(s) (i.e. trigger of emotion) of each comment is annotated. By computing the conditional probability, count and entropy of each event type, I extract lists of event types that show a strong correlation with HAPPINESS, SADNESS and ANGER. I then group similar event types into one with the help of WordNet hierarchy. No synset is found to have an obvious tendency towards

FEAR and SURPRISE. The reason may be attributed to the limited number and the scattered distribution of FEAR and SURPRISE.

In Section 6.2, I further examine the interplay of emotion, event and semantic role. I hypothesize that (1) HAPPINESS is mainly triggered by doers of events associated with HAPPINESS, (2) SADNESS is mostly elicited by undergoers of events associated with SADNESS, and (3) ANGER is generally evoked by doers of events associated with ANGER. The first and the third hypotheses have been confirmed, while the second one is partly correct. It is observed that SADNESS can also be elicited by doers of events associated with SADNESS. When the doers of an event of SADNESS have no choice but to conduct that event, it is usually the doers who evoke SADNESS.

CHAPTER 7

CONCLUSION AND FUTURE RESEARCH

In the field of emotion studies, previous attempts at emotion analysis have mainly been dedicated to the examination of explicit emotions. However, as emotions are often implicit in nature, an in-depth analysis of the explicit emotion is necessary but not sufficient – there should be an examination of the implicit aspects as well. Due to the disregard of implicit emotions in previous work, the performance of existing emotion identification and classification models is still far from satisfactory. In view of this, this thesis investigates the essential yet underdeveloped branch of emotion analysis, i.e. implicit emotion. This work explores the nature of explicit and implicit emotions expressed in response to different events shared on social media. In doing so, a more detailed analysis of implicit emotion has been conducted, which contributes to a broader picture of the forms and representations of implicit emotions and unveils the correlation between different kinds of events and emotions in greater detail.

The definitions of both explicit emotion and implicit emotions have been discussed as a starting point for the introduction to the emotion identification. Following Lee (2015), explicit emotion is defined as emotion-related information denoted by the presence of emotion keywords and certain emojis; implicit emotion refers to emotion-relation information which is not encoded by means of emotion keywords but has connotations of an emotion inferred by the readers. In order to

conduct an in-depth and comprehensive analysis of implicit emotion, I construct a Chinese event-comment corpus retrieved from one of the most popular social media websites - Sina Weibo. Not only does Sina Weibo provide a platform for users to disseminate and receive information, it also allows users to respond instantly to particular events that they are interested in. Therefore, it serves as an invaluable resource for studies on the linguistic features of implicit emotions and the interaction between implicit emotions and events.

Emotions can be expressed either explicitly or implicitly. Explicit emotions can only be encoded by means of emotion keywords or emojis, while implicit emotions can be encoded at word level, sentence level, and discourse level. Corpus data shows that the majority of emotions are expressed implicitly, which highlights the significance of the current work. Drawing on the insight of Pavlenko (2008), emotion-related information refers to three types of words, namely emotion words, emotion-related words and emotion-laden words. Emotion words are words that describe an emotion directly, such as DELIGHTED, SORROW, etc. Chapter 4 first presents the use of emotion words in expressing emotions explicitly. I make adjustments to an existing emotion taxonomy proposed by (Lee 2010) which can be used for automatic explicit emotion detection. In addition, emojis that have an obvious orientation pointing to a particular emotion are identified. I then deal with the expressions of implicit emotions in terms of the use of emotion-related words and emotion-laden words. The former refers to the behaviours triggered by an emotion (i.e. post-events), and the latter refers to words that express emotions or

elicit emotions from the interlocutors without naming an emotion directly. It is observed that emotion-laden words are usually part of a pre-event (i.e. emotion cause) of an emotion. Based on the empirical data, a list of emotion-related words and a list of emotion-laden words are generated. The two lists can greatly help identify implicit emotions expressed at the semantic level.

Having discussed emotion expressions from the semantic perspective, I examine implicit emotions at both the sentence and the discourse levels. Various syntactic structures that are frequently used to express a certain emotion are proposed. A subsection is also dedicated to rhetorical questions which are very often found in emotion expressions on social media (Lee 2018, Lau and Lee 2018). Departing from previous studies, statistics illustrate that rhetorical questions are not only productive in evoking negative emotions, such as ANGER and FEAR, but also in evoking the neutral emotion SURPRISE. It is also suggested that different emotion types may be encoded with different types of rhetorical questions. At the discourse level, I investigate the interplay between emojis and linguistic text, with special attention being placed on the situations where the two indicators are at odds with each other. Emojis expressing HAPPINESS and ANGER are more likely to be used to express another emotion. When the emotion expressed in text is different from the emotion denoted by an emoji, the former is more reliable for the identification of the overall emotion expressed.

With the help of the Chinese event-comment corpus, the correlation between emotions and event types are investigated. I first make use of TimeML annotation

guidelines (Sauri et al. 2009) for the markup of events. As TimeML annotation guidelines are compiled based on English data, I revise the guidelines according to the way events are denoted in Chinese. The event-denoting words are then mapped to the WordNet categories (i.e. event type) (Miller 1995). Each event type is annotated with frame elements following the semantic frame provided by FrameNet (Baker et al. 1998) for the annotation of opinion target(s) of the comments. Opinion target refers to the semantic role(s) in an event that triggers the writers' emotion. It helps figure out event types that are strongly correlated with a particular emotion and statistically significant. I propose a list of event types that are highly correlated with a certain emotion. This list will be of great value to the identification of emotions expressed in responses to events on social media. Moreover, I also investigate the interrelationship between emotions and semantic roles of an event. It is suggested that (1) HAPPINESS is mainly triggered by doers of events associated with HAPPINESS, (2) ANGER is mainly triggered by doers of events associated with ANGER, and (3) SADNESS is often elicited by undergoers of events associated with *sadness*. However, when the doers of an event of SADNESS have no choice but to conduct that event, it is usually the doers who evoke SADNESS.

The linguistic account of implicit emotions lay the groundwork for the development of natural language processing. For instance, the proposed linguistic cues and the syntactic structures may be used as the features for computational models and classifiers for the automatic emotion detection task.

I believe that the current work will have some implications not only for the

linguistic account of inference and identification of implicit information, but also for the automatic classification and detection of implicit emotions.

Appendix I - Emojis and the Corresponding Names

									
[锦鲤]	[中国赞]	[二哈]	[抱抱]	[摊手]	[跪了]	[酸]	[给你小心心]	[大侦探皮卡丘微笑]	[小黄人高兴]
									
[钢铁侠]	[美国队长]	[喵喵]	[doge]	[爱你]	[允悲]	[悲伤]	[吃惊]	[偷笑]	[疑问]
									
[右哼哼]	[互粉]	[顶]	[污]	[害羞]	[可怜]	[失望]	[生病]	[憧憬]	[黑线]
									
[感冒]	[亲亲]	[并不简单]	[晕]	[吃瓜]	[打脸]	[可爱]	[汗]	[笑而不语]	[馋嘴]
									
[抓狂]	[太开心]	[坏笑]	[吐]	[色]	[微笑]	[笑cry]	[酷]	[衰]	[哼]
									
[思考]	[怒]	[鼓掌]	[钱]	[困]	[舔屏]	[拜拜]	[嘘]	[左哼哼]	[阴险]
									
[怒骂]	[费解]	[挤眼]	[闭嘴]	[嘻嘻]	[哈欠]	[委屈]	[泪]	[鄙视]	[哈哈]
									
[傻眼]	[挖鼻]	[睡]	[白眼]	[NO]	[赞]	[作揖]	[握手]	[ok]	[good]
									
[弱]	[加油]	[来]	[耶]	[伤心]	[心]	[haha]	[蜡烛]	[最右]	[威武]

Appendix II – Event Hierarchy

lv0.00001740.(entity)
lv1.00002137.(abstraction,abstract_entity)
lv2.00023280.(psychological_feature)
lv3.00029677.(event)
lv4.00030657.(act,deed,human_action,human_activity)
lv5.**[00067036]**.(nonaccomplishment,nonachievement)
lv6.**[00071785]**.(mistake,error,fault)
lv7.**[00076438]**.(ballup,balls-up,cockup,mess-up)
lv7.**[00075610]**.(blunder,blooper,bloomer,bungle,pratfall,foul-up,fuckup,flub,botch,boner,boo-boo)
lv6.**[00067456]**.(failure)
lv7.**[00068346]**.(loss)
lv5.**[00046648]**.(recovery,retrieval)
lv6.**[00094303]**.(rescue,deliverance,delivery,saving)
lv5.00045991.(propulsion,actuation)
lv6.**[00105359]**.(throw)
lv6.**[00113132]**.(push,pushing)
lv5.07175534.(speech_act)
lv6.**[07217786]**.(affirmation,assertion,statement)
lv6.**[07228055]**.(disclosure,revelation,revealing)
lv7.**[07228377]**.(display)
lv8.**[07228810]**.(sackcloth_and_ashes)
lv6.**[07219061]**.(denial)

lv6.**[07253354]**.(address,speech)
lv7.**[07255865]**.(oratory)
lv8.**[07256984]**.(declamation)
lv6.**[07247648]**.(denunciation,denouncement)
lv7.**[07248075]**.(tirade,philippic,broadside)
lv6.**[07245162]**.(naming)
lv7.**[07245708]**.(indication,denotation)
lv6.07226850.(informing,making_known)
lv7.**[07238811]**.(warning)
lv8.**[07240675]**.(threat)
lv5.00408356.(activity)
lv6.**[01063257]**.(demand)
lv6.**[00411260]**.(practice,pattern)
lv7.**[00419038]**.(mistreatment)
lv8.**[00419407]**.(annoyance,annoying,irritation,vexation)
lv8.**[00420657]**.(harassment,molestation)
lv9.**[00426908]**.(tease,teasing,ribbing,tantalization)
lv8.**[00420921]**.(maltreatment,ill-treatment,ill-usage,abuse)
lv6.**[00944804]**.(puncture)
lv7.**[00946605]**.(prick,pricking)
lv6.**[01059124]**.(provision,supply,supplying)
lv7.**[01059683]**.(feeding,alimentation)
lv6.00734044.(wrongdoing,wrongful_conduct,misconduct,actus_reus)
lv7.**[00772032]**.(violation,infringement)
lv7.00746303.(transgression,evildoing)

lv8.00767587.(offense,offence)
lv9.00767761.(crime,criminal_offense,criminal_offence,law-breaking)
lv10.00770190.(felony)
lv11.**[00782543]**.(larceny,theft,thievery,thieving,stealing)
lv12.**[00783339]**.(robbery)
lv10.00770581.(fraud)
lv11.**[00781784]**.(swindle,cheat,rig)
lv6.00576778.(work)
lv7.**[00635107]**.(investigation,investigating)
lv7.00656128.(care,attention,aid,tending)
lv8.**[00657767]**.(first_aid)
lv6.01223473.(behavior,behaviour,conduct,doings)
lv7.**[01226520]**.(discourtesy,offense,offence,offensive_activity)
lv8.**[01227516]**.(insult,affront)
lv6.00805278.(control)
lv7.**[00813935]**.(grasping,taking_hold,seizing,prehension)
lv8.**[00814187]**.(clasp,clench,clutch,clutches,grasp,grip,hold)
lv9.**[00418410]**.(embrace,embracing,embracement)
lv10.**[00418656]**.(cuddle,nestle,snuggle)
lv6.00819341.(protection)
lv7.**[00825193]**.(defense,defence)
lv6.01218392.(support)
lv7.**[01218681]**.(sustenance,sustentation,sustainment,maintenance,upkeep)
lv6.00583425.(occupation,business,job,line_of_work,line)
lv7.00611221.(profession)

lv8.00612720.(education)
lv9.00888759.(teaching,instruction,pedagogy)
lv10.**[00894541]**.(lecture,lecturing)
lv5.00042637.(acquiring,getting)
lv6.**[00089301]**.(capture,gaining_control,seizure)
lv6.00078239.(acquisition)
lv7.**[00090171]**.(restitution,return,restoration,regaining)
lv7.**[00079838]**.(purchase)
lv5.00875745.(judgment,judgement,assessment)
lv6.**[00876299]**.(disapproval)
lv5.01082290.(group_action)
lv6.**[01232427]**.(assembly,assemblage,gathering)
lv6.01125919.(social_control)
lv7.**[01129340]**.(enforcement)
lv8.**[01130458]**.(law_enforcement)
lv6.01108713.(transaction,dealing,dealings)
lv7.01092370.(commerce,commercialism,mercantilism)
lv8.01115160.(selling,merchandising,marketing)
lv9.**[01119289]**.(vending,peddling,hawking,vendition)
lv5.00038116.(action)
lv6.00191991.(change)
lv7.**[00400891]**.(conversion)
lv7.**[00332776]**.(motion,movement,move,motility)
lv8.**[00350195]**.(stroke)
lv9.**[00548653]**.(beat)

lv9.**[01175528]**.(blow)
lv10.**[00134488]**.(smack,smacking,slap)
lv10.**[00137149]**.(kick,boot,kicking)
lv7.00199979.(change_of_state)
lv8.**[00278220]**.(wetting)
lv9.**[00279247]**.(watering)
lv10.**[00279399]**.(sprinkle,sprinkling,sparge)
lv8.00277464.(soiling,soilure,dirtying)
lv9.**[00277831]**.(contamination,pollution)
lv8.00210792.(termination,ending,conclusion)
lv9.**[00217881]**.(destruction,devastation)
lv7.00376871.(change_of_integrity)
lv8.**[00377208]**.(breakage,break,breaking)
lv8.**[00378877]**.(burning,combustion)
lv9.**[00379104]**.(arson,incendiarism,fire-raising)
lv8.**[00403900]**.(damage,harm,hurt,scathe)
lv7.00280679.(motion,movement,move)
lv8.**[00325210]**.(rise,ascent,ascension,ascending)
lv9.**[00325936]**.(climb,mount)
lv8.**[00330053]**.(flow,stream)
lv6.00162063.(choice,selection,option,pick)
lv7.**[00163453]**.(decision,determination,conclusion)
lv8.**[00166763]**.(move)
lv6.00035910.(accomplishment,achievement)
lv7.**[00037483]**.(feat,effort,exploit)

lv8.**[00043854]**.(derring-do)
lv6.00966680.(aggression,hostility)
lv7.00967515.(violence,force)
lv8.**[00967829]**.(domestic_violence)
lv4.07298313.(happening,occurrence,occurrent,natural_event)
lv5.**[07303737]**.(trouble)
lv6.**[07319470]**.(misfortune,bad_luck)
lv7.**[07329438]**.(calamity,catastrophe,disaster,tragedy,cataclysm)
lv5.07332364.(failure)
lv6.**[07334374]**.(miss,misfire)
lv5.07311046.(change,alteration,modification)
lv6.**[07334902]**.(birth,nativity,nascency,nascence)
lv6.**[07370091]**.(death,decease,expiry)
lv7.**[07347947]**.(martyrdom)

References

Agrawal, A., & An, A. (2012). Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations. In *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology.* Vol 1, 346-353.

Aristotle (1984). Metaphysics. In Barnes J. (Ed.), *The Complete Works of Aristotle: The Revised Oxford Translation*. Princeton, New Jersey: Princeton University Press. Vol 2, 1552-1728.

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics.* Vol 1, 86-90.

Balahur, A., J. M. Hermida, and A. Montoyo. (2011). Detecting Implicit Expressions of Sentiment in Text based on Commonsense Knowledge. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis.* 53-60.

Bamberg, M. (1997). Language, Concepts, and Emotions: The Role of Language in the Construction of Emotions. *Language Sciences*, 19(4), 309-40.

Bednarek, M. (2008). *Emotion Talk Across Corpora*. London: Palgrave Macmillan.

Bednarek, M. (2009). Emotion Talk and Emotional Talk: Cognitive and Discursive Perspectives. In Hanna Pishwa (Ed.), *Language and Social Cognition: Expression of the Social Mind*. Berlin; New York: Mouton de Gruyter. 395-431.

Cannon, W. B. (1927). The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *American Journal of Psychology*, 39, 106-24.

Calvo, R. A., & D' Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, 1(1), 18-37.

Chang, L. L., K.-J. Chen, & C.-R. Huang. (2000). Alternation across Semantic Field: A Study of Mandarin Verbs of Emotion. In Yung-O Biq. (Ed.) *Special Issue on Chinese Verbal Semantics. Computational Linguistics and Chinese Language Processing*, 5(1), 61-80.

Chaumartin, F. R. (2007). UPAR7: A Knowledge-based System for Headline Sentiment Tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*. 422-425.

Derks, D., Bos, A. E., & Von Grumbkow, J. (2008). Emoticons and Online Message Interpretation. *Social Science Computer Review*, 26(3), 379-388.

Descartes, R. (1649). The Passions of the Soul. In J. Cottingham, R. Stoothoff, and D. Murdoch. (Eds. & Trans.), *The Philosophical Writings of Descartes*. Cambridge: Cambridge University Press. Vol. 1, 325- 404.

Ding, H., & Riloff, E. (2016). Acquiring Knowledge of Affective Events from Blogs Using Label Propagation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*. 2935-2943.

Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S., & Weischedel, R. M. (2004). The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In *Proceedings of the fourth International Conference on Language Resources and Evaluation (LREC-2004)*.

Dowty, D. (1979). *Word Meaning in Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht; Boston: D. Reidel Publishing Company.

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, Vol. 67 (3), 547-619.

Ekman, P., & Davidson, R. (Eds.). (1994). *The Nature of Emotion: Fundamental Questions (Series in Affective Science)*. New York: Oxford University Press.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal Processes in Emotion. In Davidson R. J., Scherer K. R., Goldsmith H.H., (Eds.). *Handbook of Affective Sciences*. New York: Oxford University Press. 572–595.

Frank, J. (1990). You Call that a Rhetorical Question?: Forms and Functions of Rhetorical Questions in Conversation. *Journal of Pragmatics*, 14(5), 723-738.

Fiehler, R. (2002). How to Do Emotions with Words: Emotionality in Conversations. In Fussell, S. R. (Ed.). *The Verbal Communication of Emotions*. London; New York: Routledge. 79-106.

Fillmore, C. (1982). Frame Semantics, In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm*. Seoul: Hanshin. 111-137.

Fillmore, C., & B. Atkins. (1992). Toward a Frame-based Lexicon: The Semantics of Risk and Its Neighbors. In Lehrer, A. & E. Kittay (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantics and Lexical Organization.* Hillsdale: Lawrence Erlbaum Associates. 75-102.

Foolen, A. (2004). Expressive Binominal NPs in Germanic and Romance Languages. In G. Radden & K. -U. Panther (Eds.), *Studies in Linguistic Motivation*. Berlin; New York: Mouton de Gruyter. 75–100.

Foolen, A. (2012). The Relevance of Emotion for Language and Linguistics. In Foolen, A., Lüdtke, U. M., & Racine, T. P. (Eds.), *Moving Ourselves, Moving Others: Motion and Emotion in Intersubjectivity, Consciousness and Language*. Amsterdam; Philadelphia: John Benjamins Publishing Company. 349-369.

Frijda, N. H. (1986). *The Emotions*. Cambridge: Cambridge University Press.

Frijda, N. H. (1993). The Place of Appraisal in Emotion. *Cognition and Emotion*, 7, 357-387.

Frijda, N. H., & Zeelenberg, M. (2001). What Is the Dependent? In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research*. New York: Oxford University Press. 141-155.

Gao, K., Xu, H., & Wang, J. (2015). A Rule-based Approach to Emotion Cause Detection for Chinese Micro-blogs. *Expert Systems with Applications*, 42(9),

4517-4528.

Ghazi, D., Inkpen, D., & Szpakowicz, S. (2015). Detecting Emotion Stimuli in Emotion-bearing Sentences. In Gelbukh, A. (Ed.), *Computational Linguistics and Intelligent Text Processing. CICLing 2015. Lecture Notes in Computer Science.* Cham: Springer. Vol. 9042, 152–165.

Gibbs, R. W., Leggitt, J. S., & Turner, E. A. (2002). What's Special about Figurative Language in Emotional Communication. In Fussell, S. R. (Ed.), *The Verbal Communication of Emotions: Interdisciplinary Perspectives*. Hillsdale: Lawrence Erlbaum Associates. 125-149.

Greene, S., & Resnik, P. (2009). More than Words: Syntactic Packaging and Implicit Sentiment. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.* 503-511.

Grondelaers, S., & Geeraerts, D. (1998). Vagueness as a Euphemistic Strategy. In Athanasiadou, A., & Tabakowska, E. (Eds.), *Speaking of Emotions: Conceptualisation and Expression*. Berlin; New York: Mouton de Gruyter. 357- 374.

Gui, L., Yuan, L., Xu, R., Liu, B., Lu, Q., & Zhou, Y. (2014). Emotion Cause Detection with Linguistic Construction in Chinese Weibo Text. In: Zong C., Nie J., Zhao D., Feng Y. (Eds.), *Natural Language Processing and Chinese Computing. NLPCC 2014. Communications in Computer and Information Science.* Heidelberg: Springer. Vol. 496: 457-464.

Gui, L., D. Y. Wu, R. F. Xu, Q. Lu, & Zhou, Y. (2016). Event-driven Emotion Cause Extraction with Corpus Construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016).* 1639–1649.

Gui, L., J. N. Hu, Y. L. He, R. F. Xu, Q. Lu, & Du, J. C. (2017). A Question Answering Approach for Emotion Cause Extraction. In *Proceedings of the*

2017 Conference on Empirical Methods in Natural Language Processing (ENMLP 2017). 1593–1602.

Harkins, J. & A. Wierzbicka (Eds.). (2001). Emotions in Crosslinguistic Perspective. Berlin; New York: Mouton de Gruyter.

Hopper, P. J. & S. A. Thompson. (1980). Transitivity in Grammar and Discourse. *Language*, 56, 251-299.

Howell, S. (1981). Rules not Words. In P. Heelas & A. Lock (Eds.), *Indigenous Psychologies: The Anthropology of the Self*. San Diego, CA: Academic Press. 133–143.

Huang, C. R., Ahrens, K., Chang, L. L., Chen, K. J., Liu, M. C., & Tsai, M. C. (2000). The Module-attribute Representation of Verbal Semantics: From Semantic to Argument Structure. In *International Journal of Computational Linguistics & Chinese Language Processing*, 5(1), 19-46.

Jackendoff, R. (2007). Language, Consciousness, Culture: Essays on Mental Structure. Cambridge, MA: The MIT Press.

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

Jing-Schmidt, Z. (2008). The Manifestation of Emotion: On the Mandarin Chinese Nandao-interrogation. *Journal of Chinese Linguistics*. 36(2), 211-234.

Keltner, D., Oatley, K., and Jenkins, J. M. (2014). *Understanding Emotions*. Hoboken: Wiley.

Kozareva, Z., Navarro, B., Vázquez, S., & Montoyo, A. (2007). UA-ZBSA: A Headline Emotion Classification through Web Information. In *Proceedings of the 4th International Workshop on Semantic Evaluations*. 334-337.

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

Lau, H. Y. P. & Lee, S. Y. M. (2018). Information-seeking Questions and Rhetorical Questions in Emotion Expressions. In J. F. Hong, Q. Su, J. S. Wu (Eds.), *Chinese Lexical Semantics. CLSW2018. Lecture Notes in Computer Science*. Cham: Springer. Vol. 11173, 433-442.

Lazarus, R. S. (1966). *Psychological Stress and the Coping Process.* New York: McGraw Hill.

Lee, S. Y. M. (2010). *A Linguistic Approach to Emotion Detection and Classification*. Doctoral Dissertation. Hong Kong: The Hong Kong Polytechnic University.

Lee, S. Y. M., Y. Chen, S. S. Li, & C.-R. Huang. (2010). Emotion Cause Events: Corpus Construction and Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. 1121-1128.

Lee, S. Y. M., Zhang H., & Huang, C. -R. (2013a). An Event-based Emotion Corpus. In Liu P., Su Q. (Eds.), *Chinese Lexical Semantics. CLSW 2013. Lecture Notes in Computer Science*. Heidelberg: Springer. Vol. 8229, 635-644.

Lee, S. Y. M., Chen, Y., Huang, C. R., & Li, S. (2013b). Detecting Emotion Causes with a Linguistic Rule-based Approach. *Computational Intelligence, Special Issues on Computational Approaches to Analysis of Emotion in Text*. 29(3), 390-416.

Lee, S. Y. M., Li, S., & Huang, C. R. (2014). Annotating Events in an Emotion Corpus. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. 3511-3516.

Lee, S. Y. M. (2015). A Linguistic Analysis of Implicit Emotions. In Lu, Q. & Gao H. H. (Eds.), *Chinese Lexical Semantics. CLSW 2015. Lecture Notes in Computer Science*. Heidelberg: Springer. Vol. 9332, 185-194.

Lee, S. Y. M. (2018). Figurative Language in Emotion Expressions. In Wu Y., Hong JF., Su Q. (Eds.), *Chinese Lexical Semantics. CLSW 2017. Lecture Notes in Computer Science*. Cham: Springer. Vol. 10709, 408-419.

Lei, J., Rao, Y., Li, Q., Quan, X., & Wenyin, L. (2014). Towards Building a Social Emotion Detection System for Online News. *Future Generation Computer Systems*. 37, 438-448.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Li, W., & Xu, H. (2014). Text-based Emotion Classification Using Emotion Cause Extraction. *Expert Systems with Applications*. 41(4), 1742-1749.

Li, C. N. & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Linguistic Data Consortium (LDC). (2009). *ACE (Automatic Content Extraction) Chinese Annotation Guidelines for Events*, version 5.5.1. Retrieved from https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/chinese-events-guidelines-v5.5.1.pdf

Liu, M. C. & Hong, S. M. (2008). Mandarin Emotion Verbs: A Frame-based Analysis. *Journal of Chinese Language and Computing*. 18(3), 107-120.

Liu, M. C. (2016). Emotion in Lexicon and Grammar: Lexical-constructional Interface of Mandarin Emotional Predicates. *Lingua Sinica*. 2(1), 1-47.

Liu, P., Du, C., Zhao, S., & Zhu, C. (2019). Emotion Action Detection and Emotion Inference: the Task and Dataset. *ArXiv, 1903.06901*.

Ma, Q. (1981). Time Object and Classification of Verbs (時間賓語和動詞的類). *Chinese Language (中國語文)*. 86-90. (In Chinese)

Mesquita, B., Frijda, N. H., & Scherer, K. R. (1997). Culture and Emotion. In Berry, J. W., Dasen, P. R., & Saraswathi T. S. (Eds.), *Handbook of Cross-Cultural Psychology, Vol 2: Basic Processes and Human Development*. Boston: Allyn and Bacon. 255-297.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*. 38(11), 39-41.

Moors, A. (2018). Appraisal Theory of Emotion. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*. New York: Springer.

Ng, B. C., Cui, C., & Cavallaro, F. (2019). The Annotated Lexicon of Chinese Emotion Words. *WORD*. 65(2), 73-92.

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of Emojis. *PloS one*, 10(12).

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

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*. (2:1-2), 1-135.

Parkinson, B. (1996). Emotions are Social. *British Journal of Psychology*. 87, 663-683.

Parkinson, B. (1997). Untangling the Appraisal-emotion Connection. *Personality and Social Psychology Review*. 1, 62-79.

Pavlenko, A. (2008). Emotion and Emotion-laden Words in the Bilingual Lexicon. *Bilingualism: Language and cognition*. 11(2), 147-164.

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.

Pustejovsky, J. (1991). The Syntax of Event Structure. *Cognition*. 41(1-3), 47-81.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: MIT Press.

Pustejovsky, J., Ingria, B., Sauri, R., Castano, J., Littman, J., Gaizauskas, R., ... & Mani, I. (2005). The Specification Language TimeML. *The Language of Time: A Reader*. 545-557.

Rashkin, H., M. Sap, E. Allaway, N. A. Smith, & Y. J. Choi. (2018.) Event2mind: Commonsense Inference on Events, Intents, and Reactions. *ArXiv:1805.06939*.

Riordan, M. A. (2017). Emojis as Tools for Emotion Work: Communicating Affect in Text Messages. *Journal of Language and Social Psychology*. 36(5), 549-567.

Roberts, R. M., & Kreuz, R. J. (1994). Why do People Use Figurative Language?. *Psychological Science*. 5(3), 159-163.

Robinson, M. D., & Clore, G. L. (2002). Belief and Feeling: Evidence for an Accessibility Model of Emotional Self-report. *Psychological Bulletin*. 128, 934-960.

Roseman, I. J. (1984). Cognitive Determinants of Emotion: A Structural Theory. In P. Shaver (Ed.), *Review of Personality and Social Psychology*. Beverly Hills, CA: Sage. Vol. 5, 11-36.

Roseman, I. J. (1991). Appraisal Determinants of Discrete Emotions. *Cognition and Emotion*. 5, 161-200.

Rosen, S. T. (1999). The Syntactic Representation of Linguistic Events. *Glot International*. 4(2), 3-11.

Sabini, J. & Silver, M. (2005). Ekman's Basic Emotions: Why Not Love and Jealousy? *Cognition and Emotion*. 19, 693-712.

Sauri, R., Goldberg, L., Verhagen, M., & Pustejovsky, J. (2009). *Annotating Events in English. TimeML Annotation Guidelines*. Brandeis University. Version TempEval-2010.

Scheff, T. (2015). Toward Defining Basic Emotions. *Qualitative Inquiry*. 21, 111-21.

Scherer, K. R. (1984a). Emotion as a Multicomponent Process: A Model and Some Cross-Cultural Data. In P. Shaver (Ed.), *Review of Personality and Social Psychology*. Beverly Hills, CA: Sage. Vol. 5, 37-63.

Scherer, K. R. (1984b). On the Nature and Function of Emotion: A Component Process Approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to Emotion*. Hillsdale: Erlbaum. 293-318.

Scherer, K. R. (1986). Vocal Affect Expression: A Review and a Model for Future Research. *Psychological Bulletin*. 99, 143-165.

Scherer, K. R. (1988). Criteria for Emotion-antecedent Appraisal: A Review. In V. Hamilton, G. H. Bower & N. H. Frijda (Eds.), *Cognitive Perspectives on Emotion and Motivation*. Dordrecht: Kluwer Academic Publishers. 89-126.

Scherer, K. R. (1993). Studying the Emotion-antecedent Appraisal Process: An Expert System Approach. *Cognition and Emotion*. 7(3-4), 325-355.

Scherer, K. R. (1999). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion*. Chichester: John Wiley & Sons. 637-663.

Scherer, K. R. (2009). The Dynamic Architecture of Emotion: Evidence for the Component Process Model. *Cognition and Emotion*. 23, 1307-1351.

Skovholt, K., Grønning, A., & Kankaanranta, A. (2014). The Communicative Functions of Emoticons in Workplace E-mails. *Journal of Computer-Mediated Communication*. 19(4), 780-797.

Smith, C. S. (1990). Event Types in Mandarin. *Linguistics*. 28, 309-336.

Smith, C. S. (1991). The Parameter of Aspect. *Studies in Linguistics and Philosophy*. 43, 27-64.

Smith C. A. & Ellsworth, P. C. (1985). Patterns of Cognitive Appraisal in Emotion. *Journal of Personality and Social Psychology*. 48, 813-838.

Smith, C. A., & Lazarus, R. S. (1990). Emotion and Adaptation. In Pervin, L. A., & John, O. P. (Eds.), *Handbook of Personality: Theory and Research*. New York: Guilford. 609-637.

Smith, C. A., & Pope, L. K. (1992). Appraisal and Emotion: The Interactional Contributions of Dispositional and Situational Factors. In M. S. Clark (Ed.), *Review of Personality and Social Psychology, Vol. 14. Emotion and Social Behavior*. Thousand Oaks, CA: Sage Publications. 32-62.

Talmy, L. (2000). *Toward a Cognitive Semantics, Vol. 2. Typology and Process in Concept Structuring*. Cambridge: MIT Press.

Tai, J. H. Y. (1982). Additional Discussion of *"rere de he yi wan cha"* (也谈"热热的喝一碗茶"). *Journal of Chinese Linguistics*. 10(1), 81-85. (In Chinese)

Tai, J. H. Y. (1984). Verb and Times in Chinese: Vendler's Four Categories. In Testen, D., Mishra, V., & Drogo, J. (Eds.), *Papers from the Parasession on Lexical Semantics*. Chicago: Chicago Linguistic Society. 289-298.

Teng, S. H. (1986). Temporal Structure of Chinese Verbs (汉语动词的时间结构). In *Proceedings of the 1st International Symposium on Chinese Teaching*. Beijing: Beijing Languages Institute Press. (In Chinese)

Tian, Y., Galery, T., Dulcinati, G., Molimpakis, E., & Sun, C. (2017). Facebook Sentiment: Reactions and Emojis. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. 11-16.

Tsai, M. C., Huang, C. R., Chen, K. J., & Ahrens, K. (1998). Towards a Representation of Verbal Semantics -- An Approach Based on Near Synonyms. *Computational Linguistics and Chinese Language Processing, 3*(1), 61-74.

Tsai, M. C., Huang, C. R., & Chen, K. J. (1999). From Near-synonyms to the Interaction between Syntax and Semantics. In Yin, Y. M., Yang, Y. L., Chan, H. C., (Eds.), *Chinese Languages and Linguistics, Vol. 5: Interactions in Language*. Taipei: Academia Sinica. 439-459.

Tokuhisa, R., Inui, K., & Matsumoto, Y. (2008). Emotion Classification Using Massive Examples Extracted from the Web. In *Proceedings of the 22nd International Conference on Computational Linguistics.* Vol. 1, 881-888.

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

Vendler, Z. (1957). Verbs and Times. *The Philosophical Review*. 66(2), 143-160.

Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca: Cornell University Press.

Verstraete, J. C. (2001). Subjective and Objective Modality: Interpersonal and Ideational Functions in the English Modal Auxiliary System. *Journal of Pragmatics*. 33(10), 1505-1528.

Vu, H. T., Neubig, G., Sakti, S., Toda, T., & Nakamura, S. (2014). Acquiring a Dictionary of Emotion-provoking Events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Vol. 2: Short Papers.* 128-132.

Wallace, A. & Carson, M. (1973). Sharing and Diversity in Emotion Terminology. *Ethos*. 1(1), 1–29.

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

Yan, J., Bracewell, D., Ren, F., & Kuroiwa, S. (2006). A Semantic Analyzer for Aiding Emotion Recognition in Chinese. In Huang D. S., Li K., & Irwin G.W. (Eds.), *Computational Intelligence. ICIC 2006. Lecture Notes in Computer Science*. Berlin: Springer. Vol. 4114, 893-901.