Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Chaojie Zhong (Michael) **April 7th, 2017**

A Study into Emoji Use in a Multi-Party Chat

by

Chaojie Zhong (Michael)

Dr. Hiram Maxim, Dr. Yu Li

Co-advisers

Linguistics

Dr. Hiram Maxim

Dr. Yu Li

Co-advisers

Dr. Yu Li

Adviser

Dr. Greg Orloff

Committee Member

Dr. Susan Tamasi

Committee Member

2017

A Study into Emoji Use in a Multi-Party Chat

By

Chaojie Zhong (Michael)

Dr. Hiram Maxim

Dr. Yu Li

Co-advisers

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Arts with Honors

Linguistics

2017

Abstract

A Study into Emoji Use in a Multi-Party Chat

By Chaojie Zhong (Michael)

As Computer Mediated Communication (CMC) grows, Emoji has seen a large increase in popularity in recent years, which culminated in the selection of the Face with Tears of Joy as the first ever pictograph Oxford Dictionaries Word of the Year in 2015. Although there is a plethora of folk opinions on this social sensation, academic research on emoji in use is lacking, and the existing corpora do not place an emphasis on emoji. How do people use emoji? Is it simply a transparent and intuitive way of communicating emotion? As the inventory of emoji continues to grow and people continue to send it to each other in multi-party chat (MPC), investigation into the nature and usage of emoji is warranted. This study aims to answer two questions: whether there is a consensus on the meaning of emoji, and how individuals adapt to a potentially unclear system in daily communications without much confusion. A survey was administered to find out if there was any agreement on emoji, and an existing online chat group was followed and observed to explore how emoji was used in reality with the support of earlier research on emoticon, a predecessor of emoji, and speech acts theory. The results show that emoji is interpreted differently by individuals, but the sentiment perception of emoji is highly consistent. The partially conventionalized emoji is also used mostly in sentence final position similar to punctuation, suggesting the importance of context in meaning making of emoji.

Key Words: Computer Mediated Communication, Multi-Party Chat, Emoji, Emoticon, Speech Acts, Corpus, Sentiment Perception, Conventionalization, Punctuation

A Study into Emoji Use in a Multi-Party Chat

By

Chaojie Zhong (Michael)

Dr. Hiram Maxim

Dr. Yu Li

Co-advisers

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Arts with Honors

Linguistics

2017

Table of Contents

Introduction:

A Brief History of Emoticon and Emoji

Although associated nowadays with Computer-mediated Communication (CMC), the emoticon, short for emotion icon, emerged as early as the late nineteenth century when *Puck* magazine published a set of four emoticons in 1881 (Bunner, 1881). Thirty years later Bierce (1912) suggested "an improvement in punctuation – the snigger point, or note of cachinnation: it is written thus ‿ and presents a smiling mouth. It is to be appended, with the full stop, to every jocular or ironical sentence" (p. 387). Bierce's proposal for new punctuation was a particularly notable innovation because punctuation marks became the building blocks of emoticon decades later.

Such has been the case on the contemporary online forum Reddit, where "/s" is added to online comments as a conventionalized indicator of sarcasm. Although it could be argued that "/s" represents a phrase rather than a novel punctuation mark, the Encyclopædia Britannica's definition of punctuation as "the use of spacing, conventional signs, and certain typographical devices as aids to the understanding and correct reading, both silently and aloud, of handwritten and printed texts" (Brown, 2007) fits very closely the description of "/s". Interestingly and a topic to be considered later, punctuation is purported to aid the "understanding and correct reading," a basic function that can be found in emoticon and emoji as well.

While Bierce's innovation and the "/s" on Reddit can be considered emoticons, the commonly held view of emoticons today can be traced back to the first modern emoticon for CMC, the smiley :-). Invented by Scott Fahlman in 1982, the smiley was originally proposed as a joke marker (Lovering, 2007). Fahlman's design surpassed that of Bierce in that smiley is not only a combination of punctuation marks, but an imitation of a facial expression, giving

it additional power to convey non-verbal meaning. Had the joke marker been developed in the style of "/s" for sarcasm, it would have perhaps appeared as "/j" for joke, an arguably far less effective meaning-making icon. Lo (2008) showed that emoticons perform non-verbal functions. Derks et al. (2008) further investigated emoticons and concluded the social purposes of emoticons as expressing emotion, strengthening the verbal part of a message, and expressing humor, the first two of which correlate with the functions of nonverbal emotional expression in face-to-face communication.

As for the emoji, it made its debut in 1999 (Negishi, 2014). Its creator, Japanese computer software designer Shigetaka Kurita, stated that he drew inspiration from marks used in weather forecasts and from Kanji characters in Japanese. His stated reason for the development of emoji was to create a quick and easy way for users to communicate and to clarify the meaning of the messages that were limited in character count by the email system (McCurry, 2016).

The fact that emoji developed independently from emoticons is perhaps surprising, particularly in light of the fact that some emoji providers such as Apple and Facebook are able to automatically transform a typed emoticon into emoji symbols, thereby blurring the separate development trajectories of the two. Hern (2015) notes that the two are often used interchangeably, but he also offers the following distinction: "An emoticon is a typographic display of a facial representation, used to convey emotion in a text-only medium... emoji are actual pictures… treated by the computer as letters from a non-western language." The distinction in coding difference does not indicate functional difference, but it is possible that the readily accessible emoji keyboards could render emoticon less practical to produce and reduce the emoticon usage frequency. The competition between emoticon and emoji will be discussed later, and the corpus of this study, while containing a large number of emoji symbols, have zero emoticons, which indirectly strengthens the notion that the usage data of

emoji and emoticon, although interchangeable as the descriptive terminology of two systems, do not support the functional interchangeability.

Nearly two decades after emoji was first invented, in 2015, the emoji icon known as Face with Tears of Joy was selected as the Oxford Dictionaries Word of the Year (Oxford Dictionaries, 2015). As for the future of emoji, Kurita anticipates more localization (McCurry, 2016), as public platforms and multi-party chats (MPCs) continue to grow. The latest development in different CMC platforms support this expectation. Facebook, Twitter, Google, and the prominent Chinese social media application WeChat all employ their own varieties of emoji, and have different official nomenclatures for their symbols, which complicates the research effort on emoji due to increased variation that must be addressed.

Development in Computer Mediated Communication and Multi-Party Chat

From the 2010 BBC article, "Over 5 Billion Mobile Phone Connections Worldwide," one can get a glimpse into the rapid development of mobile devices and Internet in the recent decades. In the article, Ben Wood, a mobile phone analyst at CCS Insight, remarked that in 1987 a technology insider predicted a maximum of 10,000 phones, and by the year of 2010, the annual sales of mobile phone in UK alone had reached 30 million. As the title of the article would suggest, by 2010, which is already 7 years ago, there were 5 billion connections around the world. Although a connection, an individual terminal to the mobile network, is not a unique user and the number may not reflect the actual number of users because an individual could possess more than one device, it is still indicative of the enormous and rapid growth of mobile devices that are increasingly connected to the Internet today, through which people around the world engage in communication simultaneously.

Understandably, CMC has also seen significant growth. Twitter user activity has grown from approximately five thousand tweets per day in 2007 (Weil, 2010) to 500 million tweets per day in 2013 (Krikorian, 2013). While public platforms such as Twitter have been growing, so has the multi-party chat service. A prominent Chinese social media application, WeChat, has seen an enormous increase in its number of users over the past 5 years, skyrocketing from 2.8 million users in 2011 to 846 million in 2016 (Tencent, 2016).

As the volume of CMC increases, the usage of emoji becomes more frequent as well. According to the emoji tracking website EmojiTracker, in early 2017, the Face with Tears of Joy emoji is used at least 1.5 billion times a day on Twitter alone. However, EmojiTracker only assesses emoji usage in the public domain and there is no supporting evidence that the data can represent the emoji usage in MPCs. In fact, due to the private nature of most MPCs, it is difficult to obtain a meaningful corpus and observe how emoji is really used in them. There are publically available MPC corpora such as the Ubuntu Chat Corpus by Uthus and Aha that is still updated regularly today (Uthus and Aha, 2013) but such corpora are in text form and the only non-text component is emoticon rather than emoji. The research on CMC is considerably broader and examines features such as social behaviors in CMC, including agenda control and disagreement (Shaikh et al., 2010), but the current study limits its focus to the scholarship on emoticon or emoji.

CMC Trend and Early Studies on Emoticon

To study emoji, it is helpful to understand the background trend of CMC development and examine the chronological precursors of emoji, emoticons, the research on which may be useful in shaping the current study. Despite similarities in functions, there are also distinctions between emoji and emoticons that will problematize the effort to translate earlier scholarship on emoticon to emoji.

Before either emoticon or emoji became popular in CMC, research focused on the interactive written discourse (IWD) as an emerging register. Ferrara and Whittemore (1991) found omissions of sentence elements such as pronouns, copula, and articles, and observed that IWD was a hybrid of both written and spoken language and that norms were gradually emerging. The significance of their study was that it showed the increasingly speech-like properties of CMC and the trend that novel features could be conventionalized, suggesting potential incorporation of more speech acts. Al Sa'Di and Hamdan (2005), in their study on Cyber-English and CMC, showed that emoticons were employed as substitutes for absent nonverbal cues in MPCs and they were also the only substitutes that did not reduce the speech-like quality of Cyber-English, further suggesting that there could be additional functions associated with emoticons.

Walther and D'Addario (2001) examined the functions of emoticons and found that emoticons "had few impacts on message interpretations... and when they did have an impact, they were not consistent across replications." Their discovery is significant because it reflects the purpose and function of emoticons in early usage. They emerged, in the first place, as direct correspondences to actual facial expressions. The finding that they were inconsistent and not influential suggests that they were poorly conventionalized and integrated.

However, Walther and D'Addario (2001) also noted that emoticon seemed to present a negativity effect, as the addition of a negative emoticon, such as a frown or, :-(, made the text more like a combination with a negative verbal component. Although they further noted that the reverse was not true, as the addition of a positive emoticon, such as a smile or, :-), did not alter the meaning of a negative text. The discrepancy was explained in their research as an evidence of the inconsistency and the insignificance of emoticon. However, it is critical in reflecting that the functions of emoticons are not entirely random.

It must be noted that there are two flaws in the methods employed by Walther and D'Addario (2001). Firstly, their stimuli used sentences that can be interpreted as self-reinforced statements, supported by the usage of strong subjective desires and thus more resistant to modifications than shorter and more circumstantial texts would be. For example:

(1) That econ class you asked me about, it's a joy. I wish all my classes were just like it.

Instead of:

(2) That econ class you asked me about, it's a joy.

It is noticeable that (1) conveyed a pronounced desire in addition to a statement regarding the preference of the user as the second half of the message supported the notion of joy, making it difficult for a single emoticon to change the reinforced meaning significantly, while (2) the degree of joy was not specified and would possibly allow a larger degree of interpretation. In addition, the emoticons employed in the examples were limited: only frown, :-(, smile, :-), and wink, ;-) are used. Although Walther and D'Addario (2001) pointed out that wink, in particular, may denote irony, they did not include irony as a part of the stimuli and merely designated different levels of positivity or negativity. As a result, the stimuli examples were already fixated in meaning, and the function of emoticon in such a setting was only supplementary.

A similar problem can be found in Yang et al. (2007), who drew on a blog corpus to conduct one of the first CMC-based corpus-based studies. In this project the authors examined collocations with phrases in order to classify what they called emoticons, although they were more visualized and appeared to be more similar to emoji than the punctuation-based emoticon. In this approach the collocation provides context to classify the emoticons, but it also restricts the emoticons to a supplementary role. For example, in this study if an emoticon did not have a collocated phrase, such as :-) and *happy*, it was not included in the analysis. While the focus of Yang et al. (2007) is different from the current study's

examination of emoji use in MPC in that it overlooks emoticons or emoji that do not have a lexical correlation, it studies public corpora rather than private MPCs, and it does not examine the usage frequency and distribution of the symbols and the usage pattern of individual users, it nevertheless demonstrates that a corpus approach is essential to the current study because the investigation of language in use warrants a large coherent sample size that can only be found in natural corpus. Moreover, although there is a large number of variables that cannot be controlled in a natural corpus such as age, educational background, and cultural heritage, it is produced voluntarily by the users in daily life contexts, which is difficult to simulate in a controlled experiment.

Forsyth and Martell (2007) also studied emoticon with a corpus approach and used natural language processing (NLP) to analyze MPC dialogue. Although the primary purpose of their study was to construct an MPC corpus, they discovered that emoticons were used extensively, but they treated emoticons as interjections in their classification and did not examine the usage status of the emoticons. Nevertheless, it is evident that emoticons made steady progress over the roughly two-decade time frame during the development of CMC and were already highly popularized at that point to appear extensively in MPCs.

Later, Lo (2008) demonstrated that, in contrast to what Walther and D'Addario (2001) had concluded, emoticons actually allow receivers to correctly understand the level and direction of emotion, attitude, and attention expression. A reason for the remarkably different result is that Lo used short and circumstantial texts that are more open to modifications such as:

(1) Really? She didn't tell me. (Someone's response to another person).

Instead of

(2) Really? She didn't tell me. I thought she would!

It is possible that the person who produced (1) could be followed with, for example, “whatever,” and indicate indifference while (2) resembles a Walther and D’Addario stimulus and markedly shows the person’s unhappiness. Lo also included a wider range of emoticons that reflected, for example, attention and indifference, thus more comprehensively demonstrating the roles emoticons played. Lo’s (2008) finding also suggests that emoticons play a more influential role than as shown in Yang’s study that examined emoticons only in relation to their collocations.

Speech Acts and Emoticon

Providing further understanding of the use of emoticons, Dresner and Herring (2010) conducted a comprehensive analysis of the functions of emoticons by incorporating speech act theory. In his seminal work on speech act theory, Austin (1975) observed that when one produces an utterance, one typically performs three types of concomitant acts: locutionary, illocutionary, and perlocutionary, with the locutionary act being the linguistic production, the illocutionary act being the intended action performed through the locutionary act, and the perlocutionary act being “an action performed through an utterance that depends for its identity not only on the speaker’s intentions but rather also on the effect of the utterance on its audience,” such as persuasion (Austin, 1975; Dresner and Herring, 2010). Searle developed the concept of illocutionary act further and proposed to organize the illocutionary act into multiple taxonomies such as assertive illocutionary acts (e.g., statements), commissive acts (e.g., promises), directive acts (e.g., commands), expressive acts (e.g., avowals of emotion), and declarative acts (e.g., christenings) (Searle, 1969, 1985; Dresner and Herring, 2010).

Regarding whether illocutionary acts are conventional in nature, Austin and Searle agreed that they are, which was contested by Davidson (1984), who argued that what is

necessary for speech is only that a hearer be able to interpret a speaker, which can be achieved without shared conventions. Sperber and Wilson (1986) also argued that a given illocutionary act does not necessarily depend on the speaker following conventions, but rather on the hearer being able to infer the speaker's communicative intentions.

Dresner and Herring (2010) then considered an important question: how does a competent language user learn to produce speech acts with a given force and correctly ascribe illocutionary force to other people's utterances? The answer to the question fell into two camps as well, as Searle (1969) argued that ''speaking a language is engaging in a (highly complex) rule governed form of behavior," implying that one learns to use illocutionary acts in the same way one learns other social conventions, and one's having mastered them allows one to justify one's judgments with respect to illocutionary acts on the basis of one's intuitions as a language speaker. Sperber and Wilson (1986), however, considered mastery of this aspect of language use as being derived from more general human abilities to make context-dependent inferences involving interlocutors' intentions. They also noted that conventional and intentional approaches to illocutionary force were not inconsistent with each other, but rather could be combined.

Dresner and Herring (2010) proceeded to consider the connection between speech acts and sentential moods, and argued that mood would be better characterized as helping indicate the force of an utterance, while not essential to the identity of a speech act or to its success. Combining the previous considerations and exploring the other factors, they wondered whether typographic indication of force would be possible. They then observed that punctuation marks would not be nearly as numerous as types of speech acts, nor would they be correlated in a strict, rule-like fashion with the speech acts that they would indicate.

Dresner and Herring (2010) thus argued that a general function common to many emoticons is textual indication of illocutionary force. More broadly, three functions of

emoticons were identified: (a) emotion, mapped directly onto facial expression (e.g., happy or sad); (b) non-emotional meaning, mapped conventionally onto facial expression (e.g., a wink as indicating joking intent; an anxious smile); and (c) illocutionary force indicators that do not map conventionally onto facial expression (e.g., a smile as downgrading a complaint to a simple assertion). Dresner and Herring (2010) further speculated that emoticons could be viewed as an expansion of the text, similar to punctuations, a point echoed by Markman and Oshima (2007), who concluded that punctuation, in fact, would be the primary function of emoticon, based on their sequential placement at the ends of phrases, sentences, and messages.

Last but not least, Dresner and Herring (2010) also raised some questions: whether one of the aforementioned functions is more dominant, and if there are other circumstantial parameters that affect the distribution.

Emoticon vs. Emoji

To put the emoticon research into context of the emoji, it is necessary to compare the two and examine their relations with each other. Wortham (2011) has made the distinction in popular media that emoji are "the more elaborate cousins of emoticons" while emoticons are the "creative combinations of colons, parentheses and other punctuation that people use to drop a facial expression into a text message or e-mail", and emoji are a kind of "pictorial alphabet stored on a phone that can be displayed in place of the regular keyboard, making it easy to tap out a visual message." While accessibility is not directly relevant to the meaning and function of emoji or emoticons, the relatively more elaborate nature of emoji may allow them to be more adaptable to different contexts, although whether there is a correlation between the size of emoji inventory and the usage frequency remains unclear. In the same article, Sundar is quoted saying "text as a medium is particularly dull when it comes to

expressing emotions… Emoticons open the door a little, but emoji opens it even further. They play the role that nonverbal communication, like hand gestures, does in conversation but on a cellphone" (Wortham, 2011).

However, such comparison is not conclusive. In few years later, Kelly and Watts (2015) noted that emoji extended the capabilities of emoticons and found, through interviews with individuals, that emoji, similar to emoticon as Dresner and Herring discovered, demonstrated more functions than conveying emotions. The three major functions could be summarized as to maintain a conversational connection, to permit play, and to create shared and secret uniqueness.

In the same year, Pavalanathan and Eisenstein (2015) examined how emoticon could compete with emoji, the first comparative study of its kind to understand the relationship between emoticon and emoji. Their result showed that if one began to use emoji, one's emoticon usage would decrease, suggesting that emoji and emoticon do not augment each other. Rather, they seem to compete with each other and emoji appears to be preferable. This confirms that: 1. emoji share the core functions of emoticons as it can replace emoticons; 2. emoji are also more extensive and more capable in expressing meaning, which can explain why they are favorable even if their functions overlap. Whether emoji have acquired novel functions remains unclear, but the fact that they can functionally replace emoticon allows the previous study on emoticon to be translated to emoji.

Emoji as an Open System

Overall, emoji displays multiple levels of openness: the cultural environment may cause different usage pattern; the different visual variants of a single symbol allow for differential interpretations; within the same set of emoji, individual symbols can be perceived differently.

Kurita discussed the possibility of standardizing the various existing emoji systems but noted that the plethora of communities around the world may employ these icons differently, and wonders “to what degree they’re used in the same way, and to what degree there’s a local nuance” (Blagdon, 2013). SwiftKey, an adaptive keyboard provider that offers an electronic keyboard that learns from the user’s pattern and also an emoji keyboard, Swiftmoji, published an Emoji Report that affirms his interest by showing that there is a clear discrepancy in emoji usage between different countries and languages (SwiftKey, 2015).

Regarding the nomenclature and semantics of emoji, Unicode, the character-coding provider for softwares, notes that because emoji symbols are treated as pictographs, they only receive unique identifiers and may acquire multiple meanings based on their appearances (Unicode, 2016). Most importantly, Unicode states that emoji is not universal (Unicode, 2016), as the symbols could develop new meanings due to active usage, implying that emoji is not a closed system that can be prescriptively evaluated, rather one open to individual usage and interpretation variations that can only be descriptively assessed.

Apart from the social influences, Miller et al. (2016) found that individual providers of emoji also have different visual representations of the supposedly same symbols, and the different rendering leads to variation in interpretations. In their study, survey respondents also commented that sometimes different rendering altered the meaning of a message and led to frustrations, and noted that when sending messages between different providers, emoji symbols might not be transmitted correctly. Furthermore, Miller et al. (2016) noted that even within the same set of emoji from a single provider, there proved to be differential interpretations. Nevertheless, users in MPCs can still integrate an apparently inconsistent system and communicate effectively. How MPC participants manage to incorporate emoji without confusion among fellow participants is intriguing. The answer to this question has

important implications beyond the scope of CMC, as it sheds light on how new symbols may be adopted and integrated into a linguistic system.

It leads to a critical question: if a system is open, how is it integrated into a linguistic community in a meaningful way? Kurita himself noted that “people of all ages understand that a single emoji can say more about their emotions than text. Emoji have grown because they meet a need among mobile phone users. I accept that it’s difficult to use emoji to express complicated or nuanced feelings, but they are great for getting the general message across” and emoji has “been around for more than 15 years in Japan and are just a part of life that people think of as a natural part of sending a text message” (McCurry, 2016).

The Present Study:

In light of this scholarship, two research questions come to the fore:

1. Are emoji symbols still open to interpretations, or have they become conventionalized?
2. If they are open, how are they used in MPCs even though the community might not agree on their meaning? If they are conventionalized, does their usage in MPCs reflect the newfound consensus?

The first research question is relevant because Miller et al. (2016) indicated that different interpretations were possible, but they asked the participants to interpret a sample of the most popular emoji from *different* providers, and mainly aimed to find out whether there could be any differences in interpretation of the differentially rendered emoji based on sentiment and semantics. The present study does not aim to examine participants’ interpretation of different renderings; rather, it investigates whether interpretation is coherent within the emoji set from just one provider, and thereby gain an understanding of the current emoji usage trends. Specifically, due to the sheer volume of daily usage of emoji, it could be

hypothesized that for CMC users, through their exposure to such a large amount of online discourse on a daily basis, any process of innovation and subsequent conventionalization and consensus-building regarding emoji use could take place within weeks if not days.

The second research question is the core of the investigation. It aims to examine the emoji in use and to explore if there are any consistent trends in the MPC corpus. It will reveal information that will be helpful in understanding how a meaning-making system such as emoji can be used easily and readily on a daily basis in a meaningful way, even though there would appear to be the potential for confusion among users.

Method:

This study assessed whether emoji from just one provider (WeChat), without the aforementioned rendering variations on different platforms, could elicit individual interpretation differences, and explored whether linguistic communities found common ground and consensus, an essential condition in human communication, despite the potential variance (Clark, 1996). The data for this study stemmed from:

1. a survey modeled after that introduced by Miller et. al. (2016) to investigate individual differences in interpretation of emoji symbols (Appendix 1, Section 1) ;
2. corpus analysis of a Mandarin speaking MPC on the social application WeChat.

For the survey, the emoji symbols were all from WeChat in order to avoid rendering differences. The survey was completed by forty current undergraduates of the Department of Russian and East Asian Languages and Culture (REALC) at Emory University on the main campus of the university. Both Mandarin and English native speaking students participated in the survey. The survey inquired about the linguistic background, the age group, CMC activity frequency, emoji usage frequency, source of emoji knowledge, emoji inventory status, and

whether the respondent used WeChat (Appendix 1). The WeChat usage is important because WeChat encodes its emoji with official names and users can see them when selecting from the emoji keyboard, which may facilitate conventionalization. The survey also provided 10 randomly selected WeChat emoji symbols and asked the respondent to interpret them. The purpose of the questions was to better understand the perception and use of emoji by users, to connect the survey results with the analysis results of the MPC from WeChat, and to clarify potential interfering factors such as linguistic background and WeChat usage.

The survey was constructed on Google Forms and distributed to students of the Department of Russian and East Asian Languages and Culture (REALC) by Dr. Yu Li, a faculty member of the department. REALC was selected as the sample pool because it has speakers of Mandarin, English, and other languages, thus providing sufficient linguistic diversity. Because the participants were all students between the ages of 18 and 24, they do not represent all emoji users in society at large but they are all active and regular users.

Due to developments during the course of the study, a total of two editions of the survey was administered. The first survey addressed the semantic interpretation where respondents briefly explained what an emoji meant to them, but sentiment measurement, i.e. how positive or negative an emoji appeared to be, was not included in order to keep the number of questions at a reasonable level.

A follow up survey was later administered with an added sentiment measurement component asking the respondents to evaluate the sentiment of six frequently used emoji on a numerical scale. The emoji were selected based on the findings from the corpus.

The corpus was extracted from an MPC on WeChat that was carried out from August 2015 until April 2016. The extraction was performed with the Chinese TongBu Assistant (literal meaning: sync assistant) software available at tongbu.com, which offers a range of options, but only the WeChat log extraction function was used. Fortunately, WeChat

automatically parses all emoji into the format of (official name), such as (grin), and the software compiles the chat record into a Microsoft Excel spreadsheet, expediting the process. The main challenge of obtaining a corpus is to find a group that is willing to cooperate and to extract a large enough amount of data efficiently. In the present study, this issue was solved conveniently because the researcher is a member of the family group chat and was able to distribute consent forms easily. In addition, the software involved simplified the process considerably, which would otherwise have been very tedious and laborious due to the sheer amount of messages in the corpus, roughly 3,000 at the cutoff time, April 2016, and currently exceeding 7,000.

The MPC was initiated in August 2015 as a platform for a family group chat to communicate and is still ongoing. The group is mostly bilingual in Wu and Mandarin Chinese, two varieties of Chinese that are mutually unintelligible, which means that there are dialectal elements in the MPC. There are 16 participants in the MPC, with an age distribution of 1 under 18, 3 between 18 and 24, 3 between 25 and 34, 7 between 35 and 44, and 2 between 45 and 54. The average age of the group is 34 and the median age is 37, with a standard deviation of approximately 11 years. It is a kinship group consisting of three generations, of which the researcher is a member, and the accessibility of data was an important factor in selecting this particular MPC as the source of corpus. After all identifiers and private contents were removed, pseudonyms were assigned to the members to ensure that there was no violation of privacy or breach of confidentiality. The corpus is in Mandarin Chinese and emoji and required partial translation for the current study, which was performed by the investigator, a bilingual speaker of Mandarin and English. The translation was supervised by Dr. Yu Li, a member of the Chinese program at Emory University to ensure the quality of translation.

Results

Survey Result:

Of the 40 survey participants, 39 were between 18 and 24, while one participant was between 35 and 44. Concerning the linguistic background, 44.73% of the respondents were native English speakers and 35% Mandarin Chinese, with a small proportion (between 3% and 5%) of the population speaking other languages such as Thai, Burmese, Russian, Korean, and Spanish natively. There were eleven natively bilingual individuals, of which five were bilingual in Mandarin and English. 75% of the respondents reported to be very active in CMC, 15% were somewhat active, while 10% were not very active. 25% of the respondents reported using emoji very frequently on a daily basis, 50% used emoji somewhat frequently but still on a daily basis, and 25% reported to use emoji occasionally during the week. Importantly, no respondents reported low usage.

70% of the respondents reported that they learned emoji from others as well as having tried the emoji keyboard available on smartphones, while 25% of the respondents learned emoji from online articles. More importantly, 80% of the users reported that their emoji inventory was stable, while 20% reported a more dynamic inventory.

The questions about how the participants perceived and interpreted emoji received mixed answers. Because only 47.5% of the participants were users of WeChat, there was the possibility that the participants' definition of the emoji symbols could have been influenced by the official names of the emoji provided by WeChat. To examine this possibility, the percentage of users who mentioned the exact WeChat name of an emoji in their definition was calculated; synonyms were not considered. Table 1 presents the results of this calculation. For example, in their definition of the (panic) emoji, 5% of WeChat users and 5% of non-WeChat users mentioned the word panic in their definition. Table 1 thus indicates that the percentage of respondents who mentioned the official WeChat name of the emoji does not

reveal a significant difference between the WeChat users and non-WeChat users in their use of the official WeChat name in their definition of each emoji. A t-test applied to the data Table 1 (-p-value = 0.86) confirms this absence of any statistically significant difference between the two groups of participants. Thus, although the WeChat users might have been familiar with the official WeChat names of the emoji, this knowledge did not play an important role in their definition of the emoji. Only the definition of (hug) showed a discrepancy in responses from WeChat users and non-WeChat users. A majority of the WeChat users referred to the emoji by its official WeChat name (hug) compared with none of the non-WeChat users. Thus, although the official terminology for emoji is not deterministic, it can have an influence on participants' definition of an emoji.

Table 1: Participants' Percentage Use of Official WeChat Name in Definition

Emoji	WeChat Users (%)	Non WeChat Users (%)
panic	5	5
toasted	0	0
thumbs up	32	48
hammer	5	19
sly	5	5
hug	42	0
beckon	0	14
angry	47	38

coffee	63	90

*The emoji Face with tears of joy was not included because it does not have an official name in WeChat.

To analyze further the participants' definitions of the emoji symbols, their responses were qualitatively categorized as either positive or negative evaluations. Table 2 presents the degree of consensus among the participants in their categorization of each emoji symbol as expressing either a negative or positive sentiment. For example, respondents interpreted the emoji (panic) as fear, panic, embarrassed, anxious, and irritated, all of which were categorized as negative sentiments. Their definitions were thus considered to be consistent and similar interpretations. There was also no difference between WeChat users and non-WeChat users; all interpreted (panic) as a negative emoji. For all six emoji the difference between the two groups was not significant ($p=0.95$), but the degree of their agreement about the sentiment expressed by each emoji was high.

Table 2: Participants' Percentage of Agreement in Categorizing Emoji as Negative or Positive

Emoji*	Positive (+) or Negative (-)	WeChat Users (%)	Non WeChat Users (%)	Overall (%)
panic	+	100	100	100
toasted	-	84	71	78
thumbs up	+	100	100	100
hammer	-	100	95	98

sly	+	100	95	98
angry	-	100	86	93
face with tears of joy	+	74	81	78

Although there were individual differences in defining each emoji, there was much more consensus about the sentiment conveyed by each emoji across the two groups. For example, (panic), (toasted), (hammer), and (angry) were overwhelmingly negative. (Thumbs up), (sly), and (face with tears of joy) were mostly positive, showing a high degree of agreement.

The emoji (Hug), (beckon), and (coffee) were not included in this evaluation because they do not overtly convey sentiment. (Hug) is a special case as its interpretation is influenced by its name and none of the non-WeChat users considered it to be a vehicle for sentiment.

Corpus Result:

General Description:

The corpus contains 3,256 messages, 2,468 of which are text messages that do or do not include emoji symbols, while 788 messages are either photos, animated pictures, or videos, which cannot be parsed into text and are excluded from the current study.

Of the 2,468 text messages, 806 include at least one emoji, 671 of which (83.3%) are text and emoji combinations, and the other 135 messages consist only of emoji symbols, suggesting that emoji is mostly used in textual context.

The MPC corpus presented a challenge, as there was no searchable corpus that included emoji, and the investigator had to rely on manual search algorithms to analyze the data.

Emoji Position:

Of the messages that were combinations of text and emoji, emoji was at the end of the sentence 93.6% of the time, while it headed a sentence 6.4% of the time. There were 12 messages where emoji appeared in the middle of the message, but upon close inspection, the emoji symbol can still be viewed as having been inserted at the sentence final position. For example:

1. I will head to the school for a choir rehearsal in the afternoon on Friday (shy) Come visit afterwards
2. Well played (clap) It sounds very nice

It is clear that the message fragments, although related in context before and after the emoji, are not a single coherent sentence of the same topic. In the first example the segment before the emoji is about an event at a certain time, while the segment after the emoji is an invitation. In the second example the first segment is a reaction to a person playing an instrument, while the second segment is a reaction to the music being produced. Thus, it can be concluded that within this corpus emoji never appeared in the middle of a sentence and was overwhelmingly used at the end of a sentence.

Emoji and Punctuation Distribution:

In order to inspect the aforementioned punctuation function of emoji, punctuation usage was also analyzed. Considering the fact that most emoji are used in sentence final position, only the punctuation marks that signify the end of a sentence were investigated,

namely, the period, question mark, and exclamation mark. The period is used 498 times overall, while question mark is used 200 times and exclamation mark 185 times.

To see if there was any colocalization of punctuation marks and emoji symbols, sequential arrangement of said punctuation marks and emoji symbols were searched for. A period was used 200 times before an emoji, while a question mark was never used and exclamation mark 36 times. However, of the 200 colocalization of period and emoji, 198 came from the same user, and the other two instances were part of an ellipsis and a copied and pasted post. In other words, except for one user who used periods frequently, the rest of the MPC participants did not use period before emoji at all. The only punctuation mark that colocalized with emoji, it appears, was the exclamation mark.

Emoji-only Messages:

Of the 135 messages that only contained emoji, only 43 messages were composed of multiple emoji symbols, constituting 5.33% of the emoji usage, while 92 were single emoji messages. The 43 multiple emoji messages were predominantly repetitions of the same symbols (41 or 95.3%), while only 2 were different symbols in sequence. They were (tongue)(chuckle)(clap) and (sun)(moon).

Emoji Symbol Usage and Distribution:

Of the 83 available emoji symbols on WeChat by the cutoff time stamp on the MPC, April 2016, 64 symbols are used. High frequency use of emoji was determined by those emoji symbols used 5 times or more, and 34 of those symbols qualify, which constitute roughly 41.0% of the overall emoji inventory.

In single emoji messages, only 32 different symbols are used, and 77.2% of the occurrence include only 11 different symbols, with the more frequently used ones (more than 5 times) being (grin), (nosepick), (thumbsup), (ok), and (awkward).

Concerning emoji symbol frequency and rank, the log frequency and log rank of the overall emoji usage landscape and the usage of emoji in single emoji messages were calculated and visualized in the following figures:

Figure 1: Log(Frequency) vs. Log(Rank) of All Emoji in the Corpus

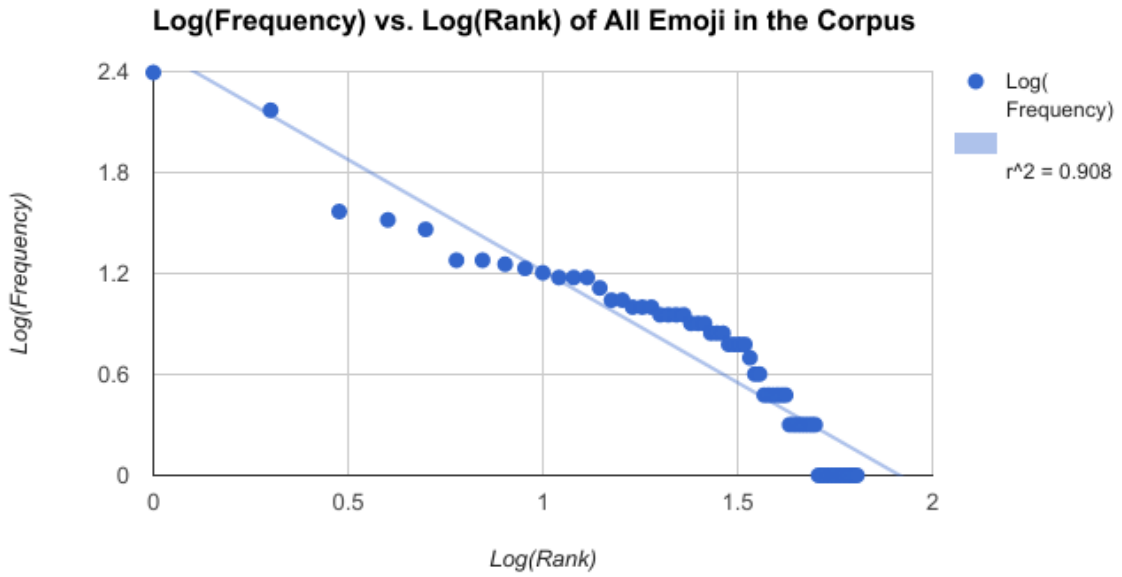
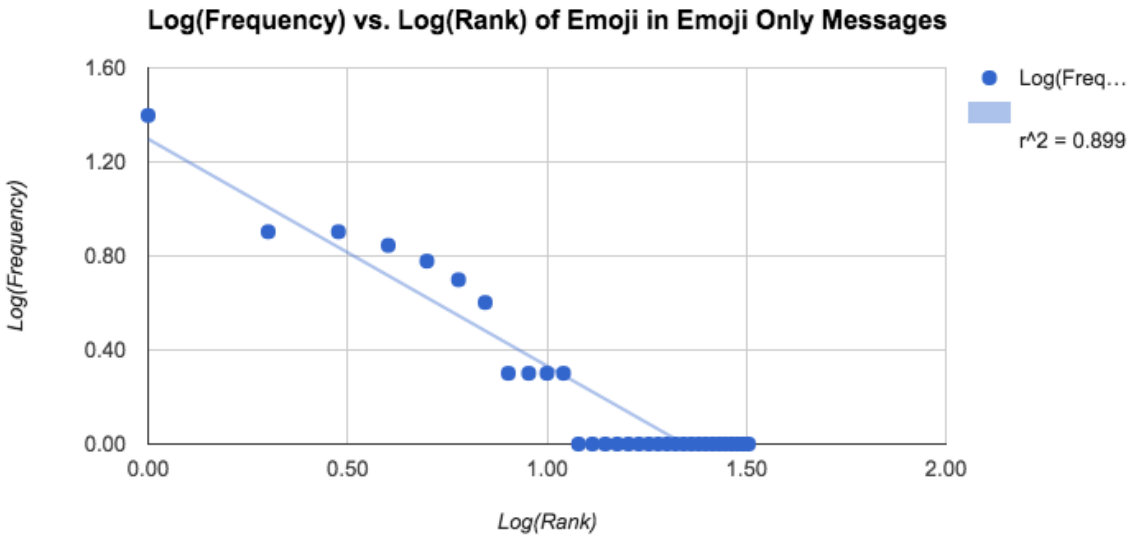


Figure 2: Log(Frequency) vs. Log(Rank) of Emoji in Emoji Only Messages

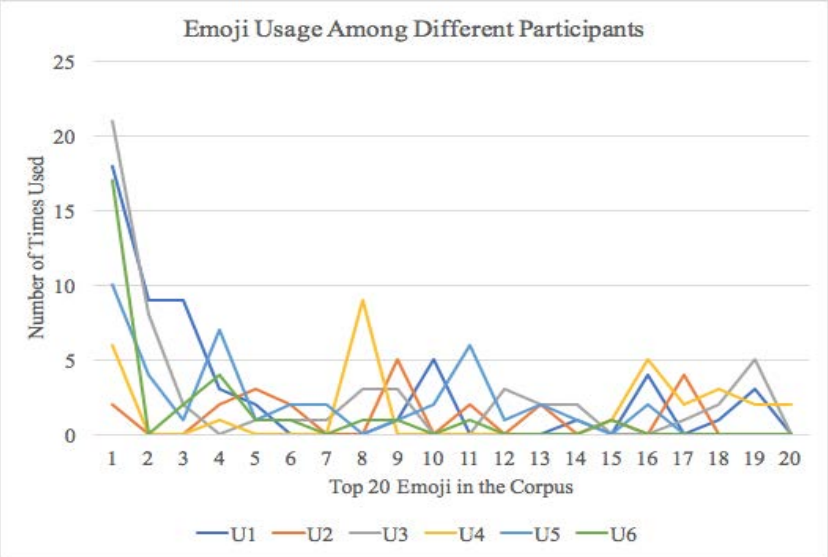


The results show a highly significant linear relationship between Log(Frequency) and Log(Rank) with an r-squared value of 0.908 and 0.899 respectively. This shows that the usage of emoji in a naturally occurring corpus follows the Zipf's Law, a statistic phenomenon where the frequency of a word and its rank on the frequency table are inversely proportional to each other.

User Emoji Profile:

Six participants were then randomly drawn from the 16-member MPC for user emoji profiling, which examined individual use of emoji and conducted comparative analysis. There were three female participants and three male participants. The overall usage comparison can be presented as the following graph:

Figure 3: Emoji Usage Profiles



Notably, there was no general consistency of regular emoji inventory between different users. Except for the top three or four emoji that were frequently used, the usage distribution of the other emoji was scattered. The size of the individual emoji inventories also varied heavily, averaging 19.0 with a standard deviation of 5.3. In addition, based on the six participants selected, there was no significant (t = 0.57) difference in total number of emoji

used between male and female, although female participants tended to use slightly more emoji (54.0) than their male counterparts (46.3).

Finally, on average, 22.9% of all possible emoji symbols were used, while only 3.0% of all emoji symbols were used highly frequently by individual participants (i.e., more than 5 times).

Follow-up Survey Result:

Because participants in the MPC used certain emoji symbols that were not included in the original survey, a second survey was conducted in order to obtain quantifiable data and to reassess the most heavily used emoji outside of the original MPC.

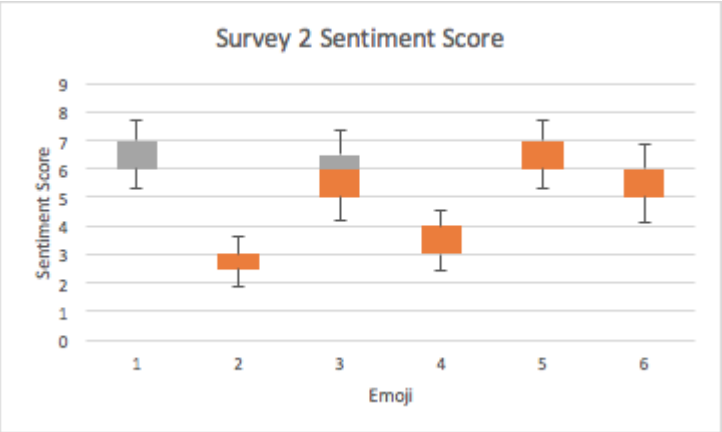
Survey participants were presented with six very frequently used emoji symbols from the MPC, (grin), (awkward), (chuckle), (shocked). (drool), and (shy) and asked two questions about the following: (1) their habitual emoji input method, i.e., whether they preferred to type emoji using an emoji keyboard, or if they typed the text equivalent of an emoji and allowed the application to automatically render the emoji; and (2) their evaluation of an emoji on a sentiment scale between 1 and 7.

Fifteen respondents from the same mail list of REALC completed the survey, although it is unknown if there were any individuals who participated in both because no identifiers were collected. All respondents reported that they input emoji from an emoji keyboard instead of typing a text equivalence and letting the electronic device translate it. They also assessed the six emoji on the scale from 1 to 7, with 7 being the most positive, and 1 being the most negative. The averages of the sentiment value can be found in the following table:

Table 3: Average Sentiment Evaluation from Survey 2

Emoji (Official Name)	Average Sentiment (Standard Deviation)
1. (grin)	6.43 (0.65)
2. (awkward)	2.93 (0.62)
3. (chuckle)	5.71 (0.83)
4. (shocked)	3.79 (0.58)
5. (drool)	6.43 (0.76)
6. (shy)	5.64 (0.93)

Figure 4: Box Plot of Sentiment Score



It appears that the frequently used emoji used in the corpus elicit very consistent sentiment evaluation from the participants, as shown above. This also confirms the preliminary conclusion from survey 1 (Table 2).

Discussion

Based on the survey results and the corpus analysis, the answers to the research questions warrant detailed explanation. Regarding the first question of whether emoji are still

open to interpretation, or if they have become conventionalized, the answer consists of two layers.

Based on the semantic descriptions in the survey results, it appears at first glance that emoji is still an open system that does not have highly conventionalized meanings. Survey respondents, when prompted to name or interpret an emoji, produced various definitions and explanations. For example, descriptions of (panic) included fear, embarrassment, anxious, etc.

However, despite the fact that participants did not arrive at a consensus in wording, emoji symbols, particularly the most frequently used ones such as (grin), elicited highly coalesced sentiment as seen in Table 2 and 3. It indicates that although emoji users did not agree on the exact term or meaning for each symbol, they were able to either infer or form a consensus on the sentiment of the emoji. Furthermore, the fact that 70% of the respondents to survey 1 named acquisition from other users as well as self-teaching as the source of their emoji inventory strongly indicates that there is possibly a process of conventionalization at play.

Therefore, emoji is a partially open system, in which the semantic descriptions of emoji vary, while the sentiment evaluations coalesce. Furthermore, based on the speech acts theory, semantic description is a locutionary act, while sentiment evaluation concerns the illocutionary. The fact that emoji proves to be highly consistent as an illocutionary item reflects the findings of Dresner and Herring (2010) on emoticons.

Consequently, answering the second research question becomes more complex. It has two parts: 1. if emoji is still an open system, how is it used in MPCs, and 2. if it has become conventionalized, whether its usage in MPCs reflects that. Based on the answer to the first research question, either part is no longer optional as both must be addressed.

Firstly, on the semantic level, emoji is not yet thoroughly conventionalized. However, it is also rarely used in isolation and mostly to accompany text and plays a peripheral role in

meaning making, as observed in the composition of the messages that include emoji, where 83.3% of the messages consist of text and emoji. Therefore, context and in this case, the texts, play an important role in helping to understand the emoji despite the potential semantic variabilities.

Secondly, the fact that frequently used emoji elicit highly coalesced sentiment evaluation and that the corpus shows a near absence of confusion can be seen as a sound evidence that a process of reaching a consensus within the linguistic community or, conventionalization, has taken place. In the entire corpus, there is only one single case of misunderstanding of the meaning of an emoji:

A: Hi all

B: (Wave)

A: You just came in and you are waving goodbye?

…

B: Isn't that just waving hand (as in greetings)?

A: That's waving goodbye, got it?

Here, user B used the emoji (wave) as a form of greeting, but user A interpreted it as a gesture of farewell, and promptly explained the nuance to user B. In this case, user A already had a clear definition of the function of a particular emoji, and was unable to correctly infer the intention of greeting.

The corpus analysis yields another interesting discovery: emoji follows the Zipfian distribution. By itself the observation does not provide any significant conclusions, but it has at least two implications. First, the use of emoji is not entirely random and that emoji symbols are not all equally deployed. In fact, emoji usage resembles that of a lexicon of a natural language. In particular, the single emoji messages also illustrate the Zipf's Law, suggesting that independently, emoji still follows a distinct distribution pattern.

Second, it is possible to construct a predictive model for a specific corpus, based on the existing information, which could lead to programs that can generate usage pattern, predict future trend, compare actual data and prediction, and infer whether the dynamics of an MPC has changed that caused a shift in usage frequency in emotionally positive or negative emoji symbols.

The observation that emoji is mostly used at the sentence final position relates to the theory proposed by Markman and Oshima (2007), which suggests that the primary function of emoji is punctuation. An observable pattern that supports this idea is that most messages that end with emoji do not have a textual component that directly corresponds to the emoji used in that context. For example:

1. The vehicle is nice, but it seems to lack a basket (grin)
2. The diet is great (commando)
3. Good cooking skills (drool)

Here, the emoji symbols (grin), (commando), and (drool) are not connected to any specific phrases in their messages. (Grin) contributes to the illocutionary act by providing an overall positive and joyful sentiment. (Commando) expresses a sense of coolness and serves to complement the statement. (Drool) is reaction to a previous photo depicting food and means the desire for the food.

Although Yang et al. (2007) analyzed collocation of emoticon and text that are directly associated with each other, such as “happy” and the smiley, it is much more difficult to identify collocation, should there be any, in the current corpus as most emoji symbols used do not directly connect with any text components of the message. This observation relates to the theory by Dresner and Herring (2010), which states that emoji functions as a textual indication of illocutionary force, and seems to be an extension of the existing punctuation

system that “aids to the understanding and correct reading… of handwritten and printed texts” (Brown, 2007).

Of course, emoji distribution alone does not convincingly argue for its function as punctuation. Another piece of evidence lies in the actual distribution of the punctuation marks themselves. The corpus shows that there is only one user who systematically uses periods, and to all other fifteen participants, the period is in fact mutually exclusive from emoji. The only sentence-final punctuation that is inclusive with emoji is the exclamation mark, which may amplify the illocutionary force of the emoji and is preserved. The mutual exclusion of the period from the emoji is unique evidence supporting the punctuation function of emoji, which may have been used to both provide the desired illocutionary force and to substitute regular periods.

A weakness of the second argument is that the disappearance of period can be seen as a result of the tendency of omission in CMC as noted by Ferrara and Whittemore (1991). Therefore, it is difficult to establish a causal relationship between emoji and punctuation use.

Limitations

With emoji seemingly being an important and frequently used component of CMC, there are several areas of research that need to be addressed in the future in order to understand their use and function.

First of all, because of the survey platform used by the researcher (Google Forms), Mandarin speakers based in Mainland China, who would make valuable contribution to the survey sections of the study, were not accessible. In particular, the MPC participants were not able to complete the survey, thereby denying the study from potentially invaluable information. Although there was a number of Mandarin speakers who responded to the

surveys, they were all university students from a limited array of social backgrounds and age groups, making the survey population less representative, even though the population sizes of the two surveys are both larger than fifteen.

Secondly, the corpus was extracted from a Chinese social media application with a unique linguistic background. Most of the participants were bilingual in Mandarin Chinese and Wu Chinese and dialectal influence were found frequently, such as the use of the (cleaver) emoji to indicate the act of taking advantage of a credulous person, which correlates with the dialectal idiomatic expression of “chopping someone” as a way of indicating that someone has been duped. Reasonably, any attempt to apply the findings based on the corpus to another cultural sphere must take caution, for the cultural influence is arguably significant.

Thirdly, it is difficult to parse online chat logs into an analyzable corpus when emoji is sprinkled throughout the chat log. Although the current study benefits from the fact that WeChat has an officially designated name for each emoji and the chat log actually documents the emoji used by their names, the researcher is unfortunately not equipped with expertise in NLP to exploit the accessibility of the WeChat corpora. The manual search algorithms are both tedious and error prone, and they are not able to yield statistically significant evidence that will be more concrete.

Related to the technological limitations, at the onset of the current study, the researcher did not expect the corpus to be readily accessible. Prior to adopting the extraction software, it was estimated that the attempt to translocate the chat history to computer platform would be much too tedious and impractical. The software was discovered and used late into the project and the unexpected success of both extracting the entire MPC history and parsing the emoji symbols by their official names presented the researcher with an overwhelming amount of data. Consequently, the current study has yet to statistically scrutinize the corpus, and all qualitative and quantitative findings are preliminary.

Fourthly, implementing one comprehensive survey instead of two would allow for a cleaner comparison of the sentiment evaluation with the semantic evaluation of emoji. In its current design, participants completed the two evaluations separately and at different times.

Last but not least, within the current research design it remains to be seen how convincingly one can argue that conventionalization underlies successful emoji use. For example, while the absence of confusion and clarification requests in the MPC could indicate that the emoji used have been conventionalized, one needs to consider the chat behavior and history of each MPC group member. In other words, although the members of the MPC had been WeChat users for some time prior to the establishment of the group, they were not in direct and equal access communication with each other. Member A might be actively chatting with member B and C, while having little to no contact with member D, who could have a private chat with C but not B. As one can observe, the web of dual participant chats indicates that any process of conventionalization and consensus building can only be applied to no more than a limited number of participants. However, the MPC at hand is much larger and has sixteen members. Nevertheless, that there was only one clarification request in the entire corpus indicates that the process of conventionalization and consensus formation is more or less complete. In fact, knowing that all MPC group members had been active emoji users before the MPC began in August 2015 it is possible that the participants in the MPC already understood what the emoji symbols meant before they even joined the MPC group. In other words, if conventionalization of emoji is needed for successful communication, then that process occurred and was completed before the MPC commenced.

Also, the end result of conventionalization would be that all participants adopt the same conventions regarding how and when to use certain emoji, which will lead to a shared inventory. However, this does not necessarily mean that the inventories displayed in the corpus must be similar to each other. At the same time, because the inventories not displayed

cannot be observed in a corpus, it is unknown to what extent those emoji are conventionalized. In fact, different users favor very different sets of emoji and one participant may actively use an emoji that another participant never uses. For example, one participant favors (nosepick) and used it 5 times, while never using (shy). Another participant shows the exact opposite regarding the two emoji, and uses (shy) 5 times while never using (nosepick). This indicates that individual usage of emoji is driven by need and personal preference, and that even though a convention may be in place, some participants simply never find any use for that emoji. However, if there is no use for an emoji, why would it be conventionalized and stored in the first place?

In conclusion, although data suggest that emoji are heavily conventionalized, the detailed mechanism of conventionalization and how individuals participate in such a process remains unclear.

Future Directions

The next immediate potential project following the current study would be to interview the MPC participants and inquire their interpretation and use of emoji, which will reveal how exactly they define emoji and understand them. The results can then be compared to those of the current study. It is also desirable to administer a new survey that includes emoji in context, which can be compared to the existing survey results to tease out the significance of context.

Furthermore, although in the current study individual usage patterns of six participants are compiled, the statistical comparison of them and the graphical presentation of the finding remain limited. Ideally, all sixteen participants will be individually assessed and compared to each other in order to provide a complete overview of the emoji usage landscape

of the MPC. Currently, due to the limitations on sample size and available statistical tools, the findings are restricted in their implications.

The current study also stands to benefit from NLP tools to process the corpus data more efficiently. In fact, the researcher tested the Stanford coreNLP, and it is able to analyze the texts in Mandarin. However, coreNLP is based on dependency theory that the researcher is unfamiliar with, and it does not include Wu Chinese, which, although similar to Mandarin in writing, may contain transliteration of Wu phrases that will appear unintelligible to a Mandarin speaker. The unique multi-dialect mixture and emoji parsed as alphabetic words will require modifications to it.

Finally, to this date, the MPC is still active and will remain so for the foreseeable future. Since the cutoff date in April 2016, thousands more new messages have be produced. As such, it provides a great source of emoji in use and a growing contemporary corpus that would benefit from further analysis with the aid of NLP and more sophisticated statistical tools.

Appendix

Survey Forms:

First Survey

Background Information

No identifier is collected. Please answer as accurately as possible.

Thank you!

What is your native language(s)?

What is your age group?

18-24

25-34

35-44

45-54

55+

Do you actively participate in casual electronic communication, e.g. Email, text, Facebook, Twitter, chat software, with friends?

Very actively

Somewhat actively

Not very actively

How often do you use Emoji in those contexts?

All the time (very frequently everyday)

Often (not as frequent, but still on a daily basis)

Occasionally (Once or more every week)

Rarely (Less than once per week)

Never (What is Emoji?)

Other:

How did you learn to use Emoji? Feel free to check all that applies: (Multiple choices allowed)

From others directly (My friends kept sending those weird things)

From electronic readings (Saw this cool article once)

From meddling with smartphone keyboards (Oooh what is this? O.o)

Other:

How dynamic is your Emoji inventory? Do you stick to a collection of preferred symbols, or do you frequently adopt new symbols and integrate them into your inventory? No need to write a long answer, but details will be appreciated.

Do you use the social app WeChat?

Yes

No, but I have heard of it or seen it before.

Nope, never heard of it either.

Emoji: Name and Meaning

Please name the following Emoji symbols, briefly explain what they mean to you, and how you would use them in context (e.g. to indicate

satire, to emphasize a phrase, to indicate mood, etc. feel free to describe the circumstances in detail. Also, more than one usage may apply.)

1

(panic)

2

(toasted)

3

(thumbsup)

4

(hammer)

5

(sly)

6

(hug)

7

(beckon)

8

(angry)

9

(coffee)

10

(face with tears of joy)

Follow Up Survey

Background Information

No identifier is collected. Please answer as accurately as possible.

Thank you!

What is your native language(s)?

What is your age group?

18-24

25-34

35-44

45-54

55+

Do you actively participate in casual electronic communication, e.g. Email, text, Facebook, Twitter, chat software, with friends?

Very actively

Somewhat actively

Not very actively

How often do you use Emoji in those contexts?

All the time (very frequently everyday)

Often (not as frequent, but still on a daily basis)

Occasionally (Once or more every week)

Rarely (Less than once per week)

Never (What is Emoji?)

Other:

How do you input Emoji?

Select from an Emoji keyboard

Type in the text equivalent and let the application match Emoji

Other:

How did you learn to use Emoji? Feel free to check all that applies:

From others directly

From electronic readings

From meddling with smartphone keyboards

Other:

How dynamic is your Emoji inventory? Do you stick to a collection of preferred symbols, or do you frequently adopt new symbols and integrate them into your inventory? No need to write a long answer, but details will be appreciated.

Do you use the social app WeChat?

Yes

No, but I have heard of it or seen it before.

Nope, never heard of it either.

Emoji: Name and Meaning

Please name the following Emoji symbols, briefly explain what they mean to you, and rate how positive or negative the emotion that is conveyed seems to you.

1

(grin)

2
(awkward)

3
(chuckle)

4
(shocked)

5
(drool)

6
(shy)

All of the emoji questions also include an emotion scale in the following format
Very Negative 1 2 3 4 5 6 7 Very Positive

Figures and Tables:

Table 1

Participants' Percentage Use of Official WeChat Name in Definition

Emoji	WeChat Users (%)	Non WeChat Users (%)
panic	5	5
toasted	0	0
thumbs up	32	48
hammer	5	19
sly	5	5

hug	42	0
beckon	0	14
angry	47	38
coffee	63	90

Table 2

Participants' Percentage of Agreement in Categorizing Emoji as Negative or Positive

Emoji*	Positive (+) or Negative (-)	WeChat Users (%)	Non WeChat Users (%)	Overall (%)
panic	+	100	100	100
toasted	-	84	71	78
thumbs up	+	100	100	100
hammer	-	100	95	98
sly	+	100	95	98
angry	-	100	86	93
face with tears of joy	+	74	81	78

Table 3

Average Sentiment Evaluation from Survey 2

Emoji (Official Name)	Average Sentiment (Standard Deviation)
1. (grin)	6.43 (0.65)
2. (awkward)	2.93 (0.62)
3. (chuckle)	5.71 (0.83)
4. (shocked)	3.79 (0.58)
5. (drool)	6.43 (0.76)
6. (shy)	5.64 (0.93)

Figure 1

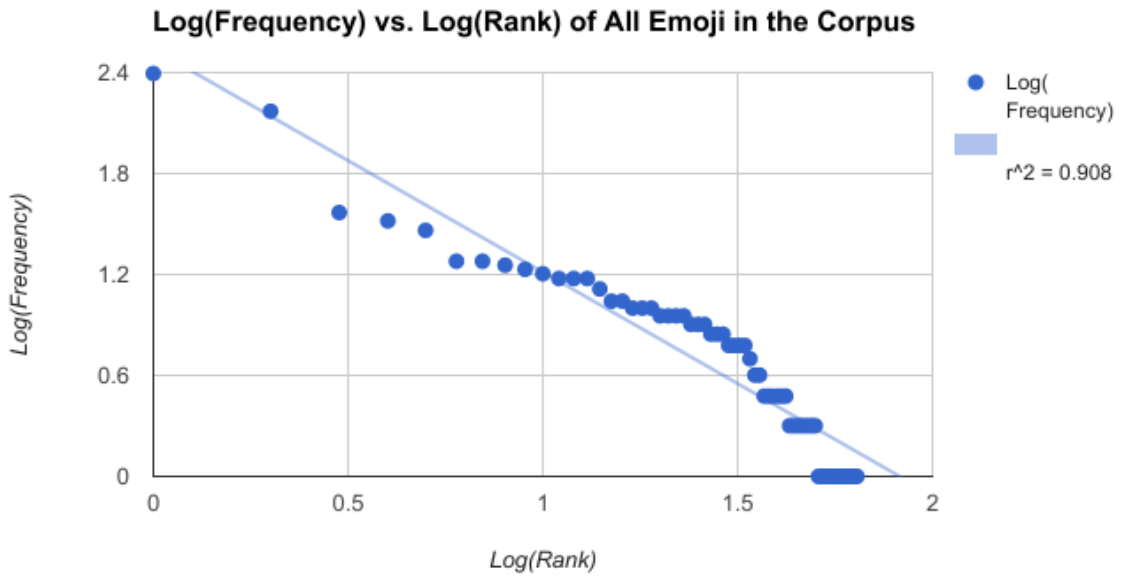


Figure 2

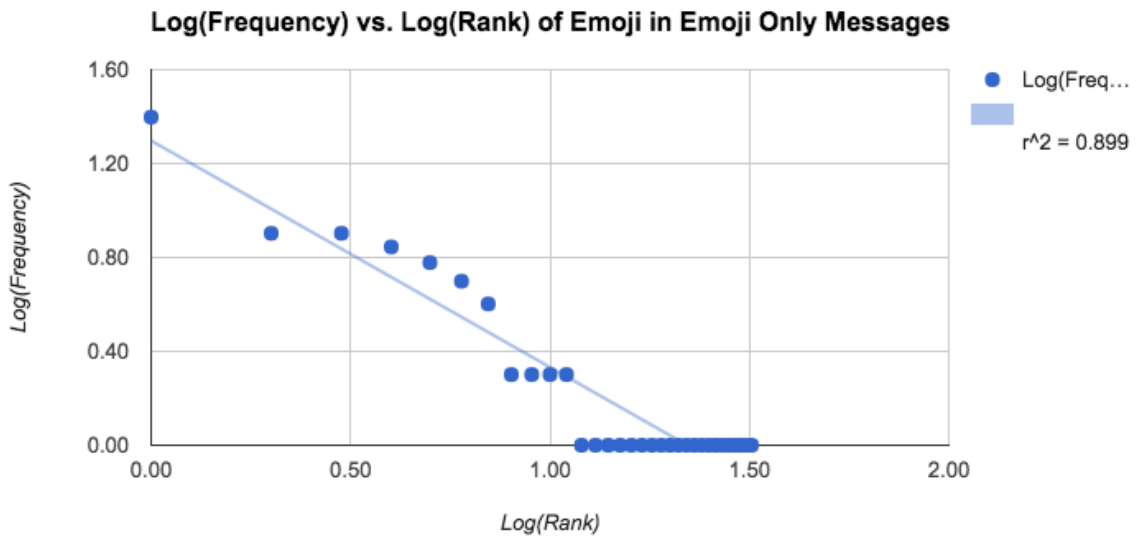


Figure 3

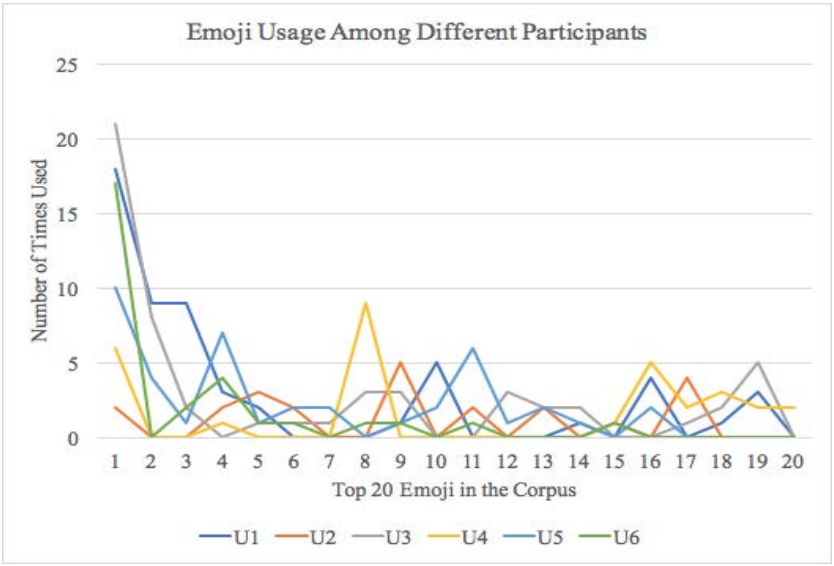
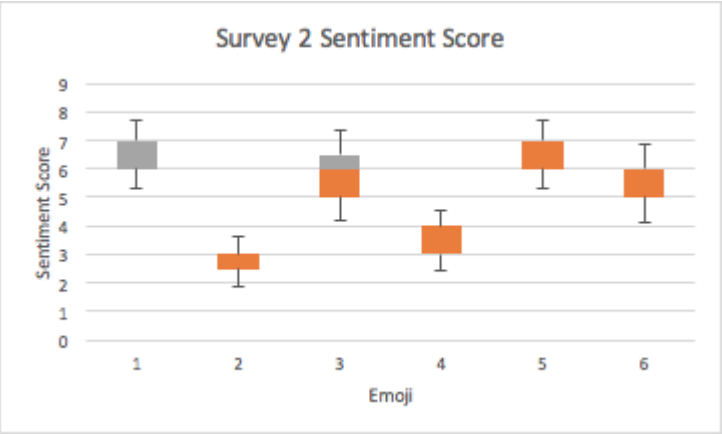


Figure 4



IRB

Bilingual Consent Form

Emory University
埃默里大学
Consent to be a Research Subject
研究知情同意书

Title: The Interpretation and Usage of Emoji
题目：绘文字的理解与使用

Principal Investigator: Dr. Hiram Maxim
主要研究者：Hiram Maxim博士

Funding Source: None
资金来源：无

Introduction
介绍：

You are being asked to be in a research study. This form is designed to tell you everything you need to think about before you decide to consent (agree) to be in the study or not to be in the study. **It is entirely your choice. If you decide to take part, you can change your mind later on and withdraw from the research study. You can skip any questions that you do not wish to answer.**
您被邀请参与一项研究。本同意书旨在告知您在决定同意或者不同意参加本研究项目之前所需考虑的一切。**参与与否完全是您的选择。如果您决定参与，之后您依然可以改变主意并退出本研究项目。您可以跳过任何您不想回答的问题。**

Before making your decision:
在您作出决定之前：

- Please carefully read this form or have it read to you
- 请仔细阅读本同意书，或让人读给您听
- Please ask questions about anything that is not clear
- 请就任何不清楚的部分提问

You can take a copy of this consent form, to keep. Feel free to take your time thinking about whether you would like to participate. By signing this form, you will not give up any legal rights.
您可以保留本同意书的备份。请耐心考虑您是否愿意参加本研究项目。签署本同意书不会使您放弃任何法律权力。

Study Overview
项目概述：

The purpose of this study is to find out whether there is a consensus on the naming and definition of Emoji characters, and how a linguistic community forms a communal standard, should there be none to begin with.

本项目的用意在于研究对绘文字的命名和定义是否有共识，以及如果最初没有共识的话，一个语言社群是如何就其达成共识的。

Procedures

步骤：

Participants of section 1 will be asked to complete an online survey, which will ask them to provide a concise linguistic background and to name and describe a list of twelve Emoji characters selected at random.

第一部分的参与者会被要求完成一份在线调查。该调查将请求参与者提供简要的语言背景，并命名和描述十二个随机择取的绘文字。

Participants of section 2 have already produced online chat sessions, which with their consent will serve as the data source for the analysis. They will be asked to authorize the researchers to access the chat records.

第二部分的参与者已经完成了在线聊天部分，聊天内容将作为数据源以供研究人员分析。研究人员将请求他们授权查看聊天记录。

No identifiers will be collected, and all participants of section 2 will be assigned pseudonyms prior to the study. Data from section 1 will be compiled, and data from section 2 will be filtered to remove private information.

本研究项目不会收集任何身份信息，第二部分的参与者将会在研究项目开始之前获得化名。第一部分的数据将会被汇总，第二部分的数据将会被滤去私人信息。

Risks and Discomforts

风险与不适：

There are no reasonably foreseeable discomforts associated with this study. The risk of breach of confidentiality is minimal as no identifiers will be collected, and data will be aggregated.

本研究项目没有任何可以预见到的不适。泄漏参与者真实身份的风险因为不收集身份信息以及数据汇总而最低化。

New Information

新信息：

It is possible that the researchers will learn something new during the study about the risks of being in it. If this happens, they will tell you about it. Then you can decide if you want to continue to be in this study or not. You may be asked to sign a new consent form that includes the new information if you decide to stay in the study.

在研究过程中，研究人员有可能发现新的参与风险。如果有此情况，他们会告知您。您可以决定是否继续参与本研究项目。他们可能会请您签署一份新的涵盖新信息的同意书。

Benefits

利益：

This study is not designed to benefit you directly. This study is designed to learn more about the interpretation and usage of Emoji. The study results may be used to help others in the future.

本研究项目并非为您获取直接收益而设计。本研究项目旨在了解绘文字的理解与使用。本研究项目的成果也许会在将来对他人有益。

Compensation

补偿：

You will not be offered payment for being in this study.
您不会因为参加本研究项目而获得报酬。

Confidentiality

保密协定：

Certain offices and people other than the researchers may look at study records. Government agencies and Emory employees overseeing proper study conduct may look at your study records. These offices include the Office for Human Research Protections, the funder(s), the Emory Institutional Review Board, the Emory Office of Research Compliance. Emory will keep any research records we create private to the extent we are required to do so by law. A study number rather than your name will be used on study records wherever possible. Your name and other facts that might point to you will not appear when we present this study or publish its results.
除研究人员之外的某些机构与个人可能会查看研究记录。监督研究行为正当与否的政府机关和埃默里员工可能会查看您的研究记录。这些机构包括：人类研究保护办公室、资助者、埃默里校级审查委员会以及埃默里研究合规办公室。埃默里将在法律要求的范围内保留被创建的任何私人研究记录。研究人员在实验记录中将尽可能使用研究编号而非您的姓名。在我们介绍本研究项目或发布其结果时，您的姓名和其它可能指向您的细节将不会出现。

Study records can be opened by court order. They may also be produced in response to a subpoena or a request for production of documents.
研究记录可以在法院指令下开封。它们亦可应传票或提供文件的指示而提供给法院。

Voluntary Participation and Withdrawal from the Study

自愿参与及退出本研究项目：

You have the right to leave a study at any time without penalty. You may refuse to do any procedures you do not feel comfortable with, or answer any questions that you do not wish to answer.
您有权随时退出本研究项目并且不受任何处罚。您可以拒绝完成任何您感到不适的步骤，或不回答任何您不想回答的问题。

The researchers also have the right to stop your participation in this study without your consent if:
如果有下列事项，研究人员有权在未经您同意的情况下终止您的参与：

- They believe it is in your best interest;
- 他们认为这符合您的最佳利益；
- You were to object to any future changes that may be made in the study plan;
- 您会反对将来研究计划中可能作出的任何改动；
- or for any other reason.
- 或任何其它原因。

Contact Information

联系信息：

Contact Chaojie Zhong (Michael) at 402-671-8151 (US) or 138-0167-2065 (China):
请通过 402-671-8151（美国）或 138-0167-2065（中国）联系钟超杰：

- if you have any questions about this study or your part in it, or

- 如果您有任何有关本研究项目或您的参与的问题，或
- if you have questions, concerns or complaints about the research
- 如果您对于本研究项目有问题、担忧或抱怨。

Contact the Emory Institutional Review Board at 404-712-0720 or 877-503-9797 or irb@emory.edu:
请通过 404-712-0720（美国）或 877-503-9797（美国）或 irb@emory.edu 联系埃默里校级审查委员会：

- if you have questions about your rights as a research participant.
- 如果您对于您作为研究参与者的权利有疑问。
- if you have questions, concerns or complaints about the research.
- 如果您对于本研究项目有问题、担忧或抱怨。
- You may also let the IRB know about your experience as a research participant through our Research Participant Survey at http://www.surveymonkey.com/s/6ZDMW75.
- 您亦可通过以上调查链接让实验审查委员会了解您作为研究参与者的体验。

Consent
同意

Please, print your name and sign below if you agree to be in this study. By signing this consent form, you will not give up any of your legal rights. We will give you a copy of the signed consent, to keep.
如果您同意参与本研究项目，请在以下空间填写并签署您的姓名。通过签署本同意书，您不会放弃您的任何法律权利。我们会提供给您一份本同意书的副本以供保留。

_____
Name of Subject
参与者姓名

_____ _____
Signature of Subject
参与者签名 日期 时间

_____ _____
Signature of Person Conducting Informed Consent Discussion
进行知情同意谈话的人签名 日期 时间

Works cited

Al-Sa'Di, R. A., & Hamdan, J. M. (2005). Synchronous online chat□□ English: Computer-mediated communication. World Englishes, 24(4), 409-424. doi:10.1111/j.0883-2919.2005.00423.x

Austin, J. L. (1975). How to do things with words. Oxford university press.

Bierce, A. (1912). The collected works of Ambrose Bierce ...: Antepenultimata (p. 387). New York &: The Neale publishing company.

Blagdon, J. (2013, March 04). How emoji conquered the world. Retrieved September 28, 2016, from http://www.theverge.com/2013/3/4/3966140/how-emoji-conquered-the-world

Brown, T. (2007, September 17). Punctuation. In Encyclopædia Britannica. Retrieved February 21, 2017, from https://www.britannica.com/topic/punctuation

Bunner, H. C. (Ed.). (1881, March 30). Typographical Art. Puck, 9(212), 65-65. Retrieved February 17, 2017, from https://babel.hathitrust.org/cgi/pt?id=osu.32435081696189;view=1up;seq=67.

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Davidson, D. (1984). Communication and convention. In D. Davidson, Inquiries into truth and interpretation (pp. 265–280). Oxford, UK: Clarendon Press.

Derks, D., Bos, A. E., & Grumbkow, J. V. (2008). Emoticons in Computer-Mediated Communication: Social Motives and Social Context. CyberPsychology & Behavior, 11(1), 99-101. doi:10.1089/cpb.2007.9926

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. Communication theory, 20(3), 249-268.

Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive Written Discourse as an Emergent Register. Written Communication, 8(1), 8-34. doi:10.1177/0741088391008001002

Forsyth, E. N., & Martell, C. H. (2007). Lexical and Discourse Analysis of Online Chat Dialog. International Conference on Semantic Computing (ICSC 2007). doi:10.1109/icsc.2007.55

Hern, A. (2015, February 06). Don't know the difference between emoji and emoticons? Let me explain. Retrieved March 15, 2017, from https://www.theguardian.com/technology/2015/feb/06/difference-between-emoji-and-emoticons-explained

Kelly, R., & Watts, L. (2015). Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design.

Krikorian, R. (2013, August 16). New Tweets per second record, and how! [Web log post]. Retrieved February 7, 2017, from https://blog.twitter.com/2013/new-tweets-per-second-record-and-how

Lo, S. (2008). The Nonverbal Communication Functions of Emoticons in Computer-Mediated Communication. CyberPsychology & Behavior, 11(5), 595-597. doi:10.1089/cpb.2007.0132

Lovering, D. (2007, September 18). Smile: The :-) turns 25. Retrieved February 21, 2017, from http://usatoday30.usatoday.com/tech/news/2007-09-18-smiley-emoticon-anniversary_N.htm

Markman, K. M., & Oshima, S. (2007, October). Pragmatic play? Some possible functions of English emoticons and Japanese kaomoji in computer-mediated discourse. In Association of Internet Researchers Annual Conference (Vol. 8).

McCurry, J. (2016, October 27). The inventor of emoji on his famous creations – and his all-time favorite. Retrieved February 21, 2017, from https://www.theguardian.com/technology/2016/oct/27/emoji-inventor-shigetaka-kurita-moma-new-york-text

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). "Blissfully happy" or "ready to fight": Varying Interpretations of Emoji.

Negishi, M. (2014, 3/26/2014). Meet Shigetaka Kurita, the Father of Emoji. Retrieved from http://blogs.wsj.com/japanrealtime/2014/03/26/meet-shigetaka-kurita- the-father- of-emoji/

Oxford Dictionaries, (2016, June 22). Oxford Dictionaries Word of the Year 2015 is… OxfordWords blog. Retrieved September 28, 2016, from http://blog.oxforddictionaries.com/2015/11/word-of- the-year- 2015-emoji/

Pavalanathan, U., & Eisenstein, J. (2015). Emoticons vs. emojis on Twitter: A causal inference approach. arXiv preprint arXiv:1510.08480.

Searle, J. R. (1969). Speech acts: An essay in the philosophy of language (Vol. 626). Cambridge university press.

Searle, J. R. (1985). Expression and meaning: Studies in the theory of speech acts. Cambridge University Press.

Shaikh, S., Strzalkowski, T., Broadwell, G. A., Stromer-Galley, J., Taylor, S. M., & Webb, N. (2010). MPC: A Multi-Party Chat Corpus for Modeling Social Phenomena in Discourse. In LREC.

Sperber, D., & Wilson, D. (1986). Relevance. Oxford: Basil Blackwell.

SwiftKey. (2015, November 25). Emoji meanings: Report reveals most-used emoji list. Retrieved September 28, 2016, from https://blog.swiftkey.com/americans-love-skulls-brazilians-love- cats-swiftkey- emoji-meanings- report/

Tencent, & China Internet Watch. (n.d.). Number of monthly active WeChat users from 2nd quarter 2010 to 3rd quarter 2016 (in millions). In Statista - The Statistics Portal. Retrieved February 7, 2017, from https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/

Unicode. (2016, July 6). Emoji and Dingbats. Retrieved September 28, 2016, from http://unicode.org/faq/emoji_dingbats.html

Uthus, D. C., & Aha, D. W. (2013, March). The Ubuntu Chat Corpus for Multiparticipant Chat Analysis. In AAAI Spring Symposium: Analyzing Microtext (Vol. 13, p. 01).

Walther, J. B., & D'addario, K. P. (2001). The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. Social Science Computer Review,19(3), 324-347. doi:10.1177/089443930101900307

Weil, K. (2010, February 22). Measuring Tweets [Web log post]. Retrieved February 7, 2017, from https://blog.twitter.com/2010/measuring-tweets

Wortham, J. (2011, December 06). Whimsical Texting Icons Get a Shot at Success. Retrieved March 19, 2017, from http://www.nytimes.com/2011/12/07/technology/emoji-in-iphones-signals-a-shot-at-mainstream-success.html

Yang, C., Lin, K. H., & Chen, H. (2007). Building emotion lexicon from weblog corpora.Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions - ACL '07. doi:10.3115/1557769.1557809