"Maybe It Was a Joke" - Emotion Detection in Text-Only Communication by Non-Native English Speakers

Ari Hautasaari[1], Naomi Yamashita[1], Ge Gao[1,2]

[1]NTT Communication Science Labs
2-4 Hikaridai, Seika-cho,
Soraku-gun, Kyoto, Japan
ari.hautasaari@lab.ntt.co.jp
naomiy@acm.org

[2]Department of Communication
Cornell University
Ithaca NY 14850 USA
gg365@cornell.edu

ABSTRACT

Previous studies have shown that people can effectively detect emotions in text-only messages written in their native languages. But is this the same for non-native speakers? In this paper, we conduct an experiment where native English speakers (NS) and Japanese non-native English speakers (NNS) rate the emotional valence in text-only messages written by native English-speaking authors. They also annotate all emotional cues (words, symbols and emoticons) that affected their rating. Accuracy of NS and NNS ratings and annotations are calculated by comparing their average correlations with author ratings and annotations used as a gold standard. Our results conclude that NNS are significantly less accurate at detecting the emotional valence of messages, especially when the messages include highly negative words. Although NNS are as accurate as NS at detecting emotional cues, they are not able to make use of symbols (exclamation marks) and emoticons to detect the emotional valence of text-only messages.

Author Keywords

Computer-mediated communication; text-only communication; non-native speaker; emotion;

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: Group and Organization Interfaces - Web-based interaction.

INTRODUCTION

Accurately detecting and understanding emotional states is central to all human communication. In text-only communication, however, detecting the intended emotional tone can be difficult due to the lack of non-verbal cues in text-only mediums [9, 25, 27]. Misunderstandings sometimes occur even between native speakers who try to exchange emotional cues via text-only messages [12]. This problem may be more salient in communication between native speakers (NS) and non-native speakers (NNS).

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright 2014 ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557215

Unfortunately, this is likely an issue that multilingual group members often face when contacting each other since their communication patterns are closely connected to their geographical distribution [17]. For example, when a native Japanese speaker who lives in Japan wants to contact his native English-speaking colleagues in the U.S., they will have to rely largely on text-based mediums (such as SMS or email) to overcome the time difference between their locations [14].

In addition to the limitations imposed by communication mediums used among multilingual group members, NNS may have particular difficulty in understanding the connection between the syntactic structure of a text-only message and the emotional nuances expressed by words [22] due to limited exposure to their second language. Indeed, both anecdotal and empirical evidence show that NNS generally have limited exposure to English beyond a classroom [2, 11]. Thus, it may be hard for NNS to discern the emotion expressed in a message when the wording of an English sentence does not include any direct expressions of the author's emotional state (e.g., "I am happy" vs. "I finally arrived"). To reduce such misunderstandings between NS and NNS in text-only communication, it is important to consider ways to support accurate emotion detection [12].

In this paper, we aim to uncover how native English speakers and Japanese non-native English speakers differ in detecting the emotional valence of text-only messages. Our main research questions are: Can non-native English speakers accurately detect the emotional valence of text-only messages written by native English speakers? What types of messages are problematic for non-native English speakers when detecting the emotional valence? Do native and non-native English speakers rely on the same cues to detect the emotional valence? If not, what types of cues do they rely on? Answering these questions will provide a foundation for designing text-only communication systems that support accurate emotion detection among native and non-native speakers.

In the remainder of this paper, we will discuss the related work on emotion detection in text-only communication, describe our experimental design and report our findings. Lastly, we will discuss our results in relation to previous work and draw design implications for the development of supporting tools for text-only communication between NS and NNS.

RELATED WORK

Emotion Detection in Text-Only CMC

Difficulties in emotion detection from text-only messages have been addressed by previous computer-mediated communication (CMC) theories. The Social Presence theory implies a radical claim that emotional tone is hard to detect from text-only messages. It states that problems in emotion detection are induced by the lack of non-verbal cues and intonation in text-only CMC environments. The limited range of available cues inhibits the use of full range of emotional and interpersonal information [24]. The Social Information Processing (SIP) theory, however, presents a more moderate claim. It suggests that emotional information is available in any CMC environment, including text-only communication, but it takes longer for interlocutors to detect this information [25].

Recent HCI/CSCW literature has largely supported the view of SIP. For example, Hancock et al. (2007, 2008) concluded that native English speakers are able to successfully detect positive and negative emotions in a text-only chat. Gill et al. (2008) also reported some similar findings where native English speakers were accurate in detecting more complex emotions such as joy and anger expressed in blog texts [7, 8].

Words as Emotional Cues in Text-Only CMC

It is worth noting that emotion detection is not only decided by receiver's ability to figure out what information is useful, but also by how a message is originally written by the author. In the context of text-only communication, previous research suggests that authors tend to adjust their emotional expressions to the medium to help receivers detect the intended emotions [27].

Previous works indicate a variety of cues that authors may use to exchange emotional information with receivers, including words, symbols and emoticons [1, 10, 26]. As for the use of words, authors who portray positive emotions use more words overall compared to actors portraying negative emotions [10]. Authors acting or experiencing negative emotions use fewer words [9], more affective words, words conveying negative feelings and negations [10]. Angry authors use more affective language and negative words as opposed to joyful authors who use more positive words [7].

The words used by the author have a strong impact on how NS detect the author's emotional tone. For instance, NS who follow a chat conversation rely mostly on negations and negative words to detect between author's positive and negative emotions [10].

Symbols and Emoticons as Emotional Cues in Text-Only CMC

Besides words, emotions in text-only CMC can be expressed through symbols, such as exclamation marks (!), and emoticons. Emoticons are representations of facial expressions constructed of symbols (e.g., :)), numbers and letters (e.g., 8D). Emoticons punctuate text-only messages by indicating pauses of emotional expression that would occur in spoken dialogue (e.g., laughter) [21].

Previous work has shown that native speakers rely on exclamation points, such as exclamation marks, to detect between their chat partner's positive and negative emotional states [10]. Also, the presence of emoticons strengthens the intensity of the verbal message [3], especially in the case of negative emotions [26]. In general, positive emoticons are associated with positive socio-emotional communication and negative emoticons are associated with negative socio-emotional communication [3].

These studies have shown that native speakers are able to draw accurate perceptions of an author's emotional state even when the available cues are limited. With all previous evidence showing how native speakers use different cues to detect emotions in text-only messages, we wonder whether and how the exchange of cues would work between NNS receiver and NS author. More specifically, whether non-native speakers use different types of cues than native speakers to detect emotions in text-only messages?

Emotion Detection by Non-Native Speakers

In this study, we focus on emotion detection by non-native English speakers (NNS) in complete text-only messages. Previous studies by both Russell (1983) and Romney et al. (1997) reported implications that NNS may have trouble detecting English emotion terms accurately [22, 23]. More precisely, NNS tend to recognize some, but not all nuances associated with English language emotion words even when they are fluent in the second language [22].

Besides individual emotion words, NNS may also not be familiar on how emotional tone is exchanged in symbols and emoticons in text-only messages. For example, Nishimura (2006) highlighted differences on how symbols, punctuation and spelling are used to emphasize text-only communication in English and Japanese. Further, emoticon styles, usage and variety differ widely across cultures and languages. Consequently, the context and sentiment associated with a given emoticon is dependent on the language backgrounds of the author and the receiver [18].

These studies altogether imply the possibility that NS and NNS may differ on how they detect emotions in text-only messages. Yet, little is known on what cues NNS rely on to detect emotions, what kind of messages are particularly problematic for NNS, and what causes the discrepancies between how NS and NNS perceive the same emotions. Answering these questions would inform the design of future systems for assisting NNS to accurately detect emotions in text-only CMC.

CURRENT STUDY

In the current study, we analyze how Japanese non-native English speakers detect the emotional valence of text-only messages written by native English-speaking authors.

Research Questions

In this paper, we use quantitative and qualitative data analysis to examine the following hypotheses and research questions.

Accuracy of emotional valence detection in text-only messages written by NS authors.
As mentioned in previous sections, native English speakers are able to successfully detect emotions in text-only CMC. For non-native speakers, however, emotion detection can be difficult due to their lack of fluency in the second language. Thus, we hypothesize that:

H1: Generally, NS will detect the emotional valence of a message more accurately than NNS.

In addition, we ask whether NS and NNS accuracy of detecting the emotional valence is affected by the presence or absence of highly positive or negative words in the message.

RQ1: Does the lexical sentiment of a message affect the accuracy of detecting the emotional valence for NS and NNS?

Cues for detecting emotional valence of text-only messages.
Previous works show that the emotional tone of a text-only message can be detected based on certain cues, including words, symbols and emoticons [1, 15, 26]. However, NNS may detect the emotional tone expressed in individual English words differently from NS [22, 23]. Furthermore, the use of symbols [16] and emoticons [18] in text-only communication differs between languages. Based on these, we wonder whether NNS rely on same cues as NS to detect the emotional valence of text-only messages written by native English speakers.

RQ2: Do NS and NNS make use of different types of cues (e.g., words, symbols, and emoticons) when detecting the emotional valence of text-only messages? If so, what type?

METHOD

Overview

To test our hypotheses and answer our research questions, we tested whether the language fluency (NS vs. NNS) and lexical sentiment of a message (positive vs. negative vs. objective) influenced the participants emotional valence rating and cue annotation accuracy. During the experiment, NS and Japanese NNS participants rated the emotional valence in a set of text-only messages written by native English-speaking authors. After that, we asked all participants to indicate the cues (words, symbols and emoticons) in each message that they used to resolve the emotional valence of the message.

Authors of all messages were also required to rate the emotional valence of their own messages and annotate the cues they used to generate their emotional valence rating. By doing this, we got a gold standard (i.e., author rating) for comparing the accuracy of emotional valence detection and cue annotation between NS and NNS.

Participants

There were 20 native English speakers (NS) and 20 non-native English speakers (NNS) that participated in this study. All NS participants had received their primary education (from the age of 6 to 18, elementary school to high school) in English speaking countries. All NS participants reported English as their only native language.

The NNS participants in this study were all Japanese native speakers. They all had spent less than two years in English speaking countries. They speak English as their second language, but not fluently. We required a minimum score of 750 in the TOEIC[1] English proficiency test for all NNS participants.

Materials for Participants

Categorization of Messages Based on Lexical Sentiment
We extracted a total number of 98 messages from Facebook with the authors' consent, which formed the initial message pool for our experiment materials. All messages were public status updates written by native English-speaking authors. None of the messages included photos, videos, hyperlinks, or any other additional content besides text.

In order to get an equal distribution of messages with different lexical sentiment (positive, negative and objective), we categorized all messages using a software called SentiWordNet[2]. We used SentiWordNet to identify and count the proportion of emotional words in each message (Table 1). SentiWordNet includes a human rated synset classification of words with three numerical scores in the positive, negative and objective scales [6]. We input all words in each message to SentiWordNet and identified the sentiment score of each individual word.

The lexical sentiment of each message was calculated as the ratio of highly positive words to highly negative words (rated over 0.5 in a 0-1.0 scale)[3]. In total, 33.67% of messages were categorized as positive, 32.65% were categorized as negative and 33.67% were categorized as

[1] Test of English for International Communication

[2] http://sentiwordnet.isti.cnr.it/

[3] SentiWordNet allows user feedback on the values assigned to each individual word. In this paper, we used the values as of April 15th, 2013.

	Rater	Sampling message with cue annotations	Rating scale and results	Emotional valence according to the rating
Lexical sentiment	SentiWordNet	Holy moly that flight was freakin' scary :O The WIND was nuts :O They almost diverted us to another Tokyo airport. Hats off to the pilots :) OMG :O	#Positive words = 0 #Negative words = 1	Negative
Author rated valence & cue annotation	NS author of the message	Holy moly that flight was freakin' scary :O The WIND was nuts :O They almost diverted us to another Tokyo airport. Hats off to the pilots :) OMG :O	4 on a 7-point scale	Neutral
Participant rated valence & cue annotation	NS participants	Holy moly that flight was freakin' scary :O The WIND was nuts :O They almost diverted us to another Tokyo airport. Hats off to the pilots :) OMG :O	5 on a 7-point scale	Slightly positive
	NNS participants	Holy moly that flight was freakin' scary :O The WIND was nuts :O They almost diverted us to another Tokyo airport. Hats off to the pilots :) OMG :O	2 on a 7-point scale	Negative

Table 1. Example of a message with lexical sentiment calculated by SentiWordNet, gold standard rating and cue annotation by NS author, and participants' ratings and cue annotations by NS and NNS (overlap between author's and participants' annotation is indicated in green, and additional annotations are indicated in yellow).

objective (i.e., messages that had neither positive or negative lexical sentiment). Table 1 includes an example of a message with a negative lexical sentiment (negative words indicated in red).

All messages were fully anonymized, excluding all person names, profile photos, and other personal information. We combined these messages in to one dataset and randomized their order.

Author Rating and Cue Annotation

To compare the accuracy of participants' emotional valence detection and cue annotation on each message, we use the authors' emotional valence ratings and cue annotations as the gold standard. We asked all authors to rate each of their own messages for the emotional valence in a 7-point Likert scale (1 = very negative, 4 = neutral, 7 = very positive). Secondly, we asked them to annotate the words, symbols and emoticons in each message that reflected their rating. The author annotations were used to calculate the cue selection accuracy of NS and NNS. Table 1 includes an example of author rating and cue annotation.

Procedures

In the experiment, we asked both NS and NNS participants to read a set of messages (N=98). In the first step, they rated the emotional valence of each message. After that, they annotated all cues in the messages (i.e., words, symbols and emoticons) that they were using to determine their emotional valence rating in the first step (Table 1). The task instructions for NNS participants were given in Japanese.

NNS participants were allowed to use English to Japanese bilingual dictionaries during the experiment, so that their emotional valence rating and cue annotation would not be heavily influenced by their individual English language abilities (grammar, vocabulary, etc.).

Measures

Participants' emotional valence detection in the message: Both NS and NNS participants rated their evaluation of the emotional valence of each given message on a 7-point Likert scale (1 = very negative, 4 = neutral, 7 = very positive). The rating score reflects the participants' emotional valence detection in the message. The consistency of emotion ratings given by NS (Krippendorff's α = .68) and NNS (Krippendorff's α = .64) formed a reliable scale.

Participants' cue annotation in the message: Both NS and NNS participants annotated three types of cues from the messages (words, symbols and emoticons), that influenced their emotional valence rating on the given message.

RESULTS

We conducted Mann-Whitney U tests to explore if NS and NNS detect the emotional valence in each type of messages differently. Non-parametric test was used because the data was not normally distributed.

Emotional Valence Rating Accuracy

Our *H1* and *RQ1* discussed the differences in NS and NNS ability to detect the emotional valence of NS authors' messages. To verify our *H1*, we firstly calculated participants' emotional valence rating accuracy. Participants' rating accuracy is calculated as a correlation (Spearman's correlation coefficient) with the author ratings. We then compared the average correlation of NS and NNS.

Accuracy of Emotional Valence Detection between NS and NNS

To explore our *H1*, we conducted a Mann-Whitney U test on the effect of language proficiency (NS vs. NNS) to emotional valence rating accuracy (N=98). The result fully supported our *H1*. A Mann-Whitney U test showed a significantly higher correlation between NS and author ratings compared to the correlation between NNS and author ratings (r_{NS}=0.51, r_{NNS}=0.41; Z=4.26, p<.05). This indicated that NS participants could detect the emotional valence of NS authors' messages more accurately than NNS participants.

Accuracy of Emotional Valence Detection in Messages with Different Lexical Sentiment

In our *RQ1* we asked whether the lexical sentiment of a message would affect the accuracy of emotion detection for NS and NNS. To answer this research question, we conducted a Mann-Whitney U test on the effect of lexical sentiment of a message (positive vs. negative vs. objective) on the emotional valence rating accuracy for both NS and NNS.

The results indicated that the NS and NNS participants' accuracy of detecting the emotional valence followed a similar pattern (Figure 1). More specifically, both NS and NNS were most accurate at rating the emotional valence of messages with positive lexical sentiment, and least accurate at rating messages with negative lexical sentiment. This effect was significant with both NS (r_{POS}=0.57, r_{NEG}=0.45; Z=2.23, p<.05)[†] and NNS (r_{POS}=0.46, r_{NEG}=0.35; Z=2.53, p<.05)[°]. Similarly, the rating accuracy for messages with positive lexical sentiment was higher than rating accuracy for messages with objective lexical sentiment for both NS and NNS, but this result was significant only for NS (r_{POS}=0.57, $r_{OBJECTIVE}$=0.47; Z=3.45, p<.05)[‡].

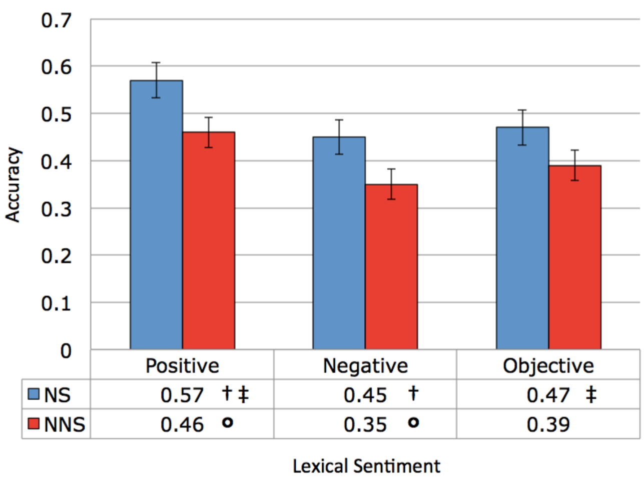

Figure 1. NS and NNS emotional valence rating accuracy for messages with positive (N=33), negative (N=32) or objective (N=33) lexical sentiment.

These results answered our *H1* and *RQ1*. They indicated that the participants' language background and the lexical sentiment of a message had an effect on the participants' accuracy of detecting the emotional valence of the text-only messages.

Cues for Assessing Emotional Valence

Our *RQ2* discussed the differences in NS and NNS ability to annotate proper emotional cues in NS authors' messages. To answer this research question, we firstly calculated the participants' cue annotation accuracy. NS and NNS cue annotation accuracy is indicated by Spearman's correlation coefficient with author cue annotations (match or no match). Empty author annotations (i.e., no cues selected by author) are handled as either 1 (no cues selected also by participant) or 0 (cues selected by participant).

Accuracy of Emotional Cue Annotation between NS and NNS

In our *RQ2* we asked whether NS and NNS participants would use different types of cues (e.g., words, symbols and emoticons) to generate their emotional valence rating. To answer this research question, we first used the authors' annotations as a gold standard to compare whether there are differences in the accuracy of cue annotation between NS and NNS.

The results given by Mann-Whitney U test did not show any significant difference in the cue annotation accuracy between NS and NNS (r_{NS}=0.62, r_{NNS}=0.66; Z=-1.28, p=*n.s.*). However, after splitting all messages based on their lexical sentiment (Figure 2), our data showed that the cue annotation accuracy was significantly higher in messages with positive lexical sentiment compared to negative lexical sentiment for both NS (r_{POS}=0.69, r_{NEG}=0.47; Z=2.88, p<.05)[†] and NNS (r_{POS}=0.74, r_{NEG}=0.60; Z=3.15, p<.05)[°]. Similarly, the cue annotation accuracy was significantly higher in messages with positive lexical sentiment compared to objective lexical sentiment for both NS (r_{POS}=0.69, $r_{OBJECTIVE}$=0.54; Z=2.53, p<.05)[‡] and NNS (r_{POS}=0.74, $r_{OBJECTIVE}$=0.61; Z=2.23, p<.05)[**].

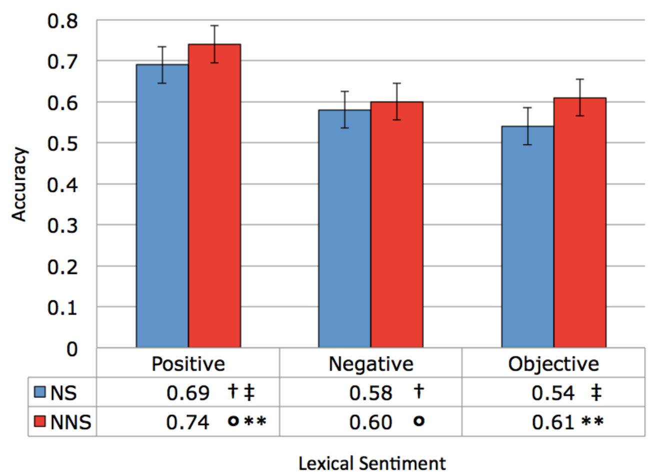

Figure 2. NS and NNS cue annotation accuracy in messages with positive (N=33), negative (N=32) or objective (N=33) lexical sentiment.

These results answered the first part of our *RQ2.* They indicated that participants' accuracy on annotating proper emotional cues varied with the lexical sentiment of a message.

NS and NNS Annotation of Words, Symbols and Emoticons as Emotional Cues

As shown in Table 1, the participants also annotated words, symbols and emoticons that were not included in the author annotations. To explore the second part of our *RQ2* and answer what different types of cues NS and NNS make use of, we examined the discrepancies in word, symbol and emoticon annotations between NS and NNS in all messages (N=98).

First, we calculated the ratio of additional words selected by NS and NNS. Secondly, we calculated the ratio of symbols and emoticons selected by NS and NNS from all messages (N=98). For symbols, we focused on exclamation marks since previous work has indicated that they are used as one NS author strategy to express emotions in text-only messages [10].

We did not find any significant difference in the ratio of words selected as emotional cues between NS (M=0.31, SD=0.17) and NNS (M=0.30, SD=0.16) from all messages (N=98). Furthermore, we tested the word annotation accuracy based on the gold standard author annotations separately, and found no significant difference between NS and NNS (r_{NS}=0.62, r_{NNS}=0.65; Z=-0.97, p=*n.s.*). However, our results indicated that NS chose exclamation marks and emoticons significantly more frequently than NNS (Table 2).

	Exclamation Marks	Emoticons
NS	0.57	0.32
NNS	0.29	0.07
	χ^2 = 14.87, p<.05	χ^2 = 18.35, p<.05

Table 2: Ratio of exclamation marks and emoticons selected by NS and NNS.

These results partially answered the second part of our *RQ2*. They indicated that NS annotated significantly more symbols (exclamation marks) and emoticons as emotional cues from messages than NNS. Variance on word selection turned out to be the same between NS and NNS.

Emotional Valence Rating Accuracy between NS and NNS in Messages Including Symbols and Emoticons as Emotional Cues

A Mann-Whitney U test indicated that NS were significantly more accurate in their emotional valence rating for messages that included exclamation marks or emoticons than NNS (r_{NS}=0.78, r_{NNS}=0.69; Z=3.39, p<.05). Furthermore, a Mann-Whitney U test showed that NS were significantly more accurate than NNS at rating the emotional valence of messages where the lexical sentiment was positive (r_{NS}=0.67, r_{NNS}=0.57; Z=2.02, p<.05) or negative (r_{NS}=0.70, r_{NNS}=0.49; Z=3.48, p<.05) and the messages included exclamation marks or emoticons

The emotional valence rating accuracy of NS did not vary with the lexical sentiment when messages included exclamation marks and emoticons. However, NNS were significantly less accurate at detecting the emotional valence when a message included exclamation marks or emoticons and its lexical sentiment was either positive (r_{POS}=0.57, $r_{OBJECTIVE}$=0.69; Z=-2.34, p<.05) or negative (r_{NEG}=0.49, $r_{OBJECTIVE}$=0.69; Z=-3.23, p<.05).

These results answered the second part of our *RQ2*. Altogether, our results indicated that NS and NNS use similar words to detect the emotional valence of text-only messages. However, NS make use of symbols (exclamation marks) and emoticons significantly more than NNS. Furthermore, our results showed that the lexical sentiment of a message (positive and negative vs. objective) had an effect on NNS rating accuracy, but not on NS rating accuracy when the messages included exclamation marks and emoticons.

DISCUSSION

In summary, we found the following results on NNS detection of emotional valence in text-only messages written by native English-speaking authors:

- NNS were significantly less accurate than NS at detecting the emotional valence of English messages written by native English speakers.
- NNS were least accurate at detecting the emotional valence of messages where the lexical sentiment was negative.
- NNS were as accurate as NS at detecting the proper emotional cues in messages.
- NNS were significantly less accurate than NS at detecting the emotional valence of messages that included symbols (exclamation marks) and emoticons.

While the differences may seem relatively small, the effect size (Cliff's delta ranging from δ=0.32 to δ=0.67) indicates the robustness of our results. In the next sections, we discuss our findings in relation to previous work in more detail.

Explanation of the Findings

Why are NNS less accurate than NS at detecting the emotional valence of text-only messages?

We hypothesized that NS are more accurate at rating the emotional valence of text-only messages than NNS (*H1*), and our results fully supported this. In our related work section, we discussed some previous works that gave an indication on why this might be the case: NNS may have trouble detecting the nuances in single emotion words even when they are advanced second language learners [22].

What makes our results unique is that while previous work has focused on the semantic similarity of individual emotion words in multiple languages, we analyzed emotional valence detection in full text-only messages. Furthermore, we calculated the accuracy of NS and NNS emotional valence detection based on author ratings for each message. Our results indicated that even when more contextual information is available (i.e., complete messages), NNS have trouble detecting the emotional valence accurately.

Previous research suggests that emotional responses to emotional expressions may differ between first and second languages depending on the first language background [4]. Thus, first language background may explain some difficulties that the Japanese NNS in this study faced when rating the emotional valence of text-only messages.

How did the lexical sentiment of a message affect NNS emotional valence rating?

In our *RQ1*, we asked whether the lexical sentiment of a message has an effect on the emotional valence rating accuracy of NS and NNS. Our results indicated that both NS and NNS were most accurate at detecting the emotional valence of messages that included highly positive words. This result is also related to the cue selection accuracy (*RQ2*), where both NS and NNS were most accurate at detecting emotional cues in messages with positive lexical sentiment.

Previous works suggested that NS make use of negative words and negations to detect between author's positive and negative emotional states [10]. NNS had particular problems with detecting the emotional valence of messages that included highly negative words as opposed to highly positive words. Furthermore, as depicted in Figure 2, NNS were least accurate at detecting the emotional cues in messages with negative lexical sentiment.

One likely explanation for this discrepancy is that while NNS can successfully detect the important emotion words, they may not be familiar with the nuances of negative English words when used as part of sarcastic and cynical messages. Indeed, some NNS participants routinely wrote "*maybe it was a joke*" as an additional explanation for their emotional valence rating in negative messages if they were not confident that they understood the emotional nuance correctly.

How come NNS were less accurate in their emotional valence ratings but this did not hold true for emotional cues?

Our second research question (*RQ2*) asked whether NS and NNS rely on same cues when detecting the emotional valence of a text-only message. Despite the fact that NNS were significantly less accurate at detecting the emotional valence in each message (Figure 1), our results showed that NS and NNS were as accurate at detecting proper emotional cues from the messages (Figure 2).

Permitting the use of bilingual dictionaries for NNS might have helped the NNS participants rate each message more accurately. Interestingly, however, the use of bilingual Japanese to English dictionaries did not solve the problem for NNS to accurately detect the emotional valence of the text-only messages. One possible explanation is that dictionaries may have helped the NNS detect the proper emotion words, but did not help in perceiving them properly. In case the NNS would have had a longer tenure in an English speaking country, their emotional valence rating accuracy would likely be closer to that of NS.

Why NNS did not rely on symbols and emoticons when detecting the emotional valence of text-only messages?

As suggested in previous work, native English speakers often rely on symbols (exclamation marks) and emoticons to detect emotions in text-only messages [1, 9, 15, 26]. However, our results concluded that NNS make use of exclamation marks and emoticons significantly less than NS when detecting the emotional valence of messages (Table 2).

Previous research has highlighted the differences in the use of symbols [16] and emoticons [18] between different languages. The pragmatic use of emoticons may also differ in the native language of the NNS. For example, Japanese emoticons may not convey emotions same as Western emoticons [26], but may serve a different, more complex purpose in socio-emotional communication [13, 28].

One possible explanation for our results is that the Japanese NNS may not be familiar with the Western emoticons (e.g., :O or :P) and their relation to the verbal content of a message. Another possible reason for the Japanese NNS not to rely on the available emoticons is the lower expressiveness of Western emoticons compared to Japanese emoticons (e.g., ＿|￣|○ meaning: disappointment) [28]. In short, NNS may be unable to detect the connection between the verbal emotional content and the symbols and emoticons depending on how they are used in the NNS native language.

Design Implications

In this paper, we concluded that NNS were significantly less accurate at detecting the emotional valence of text-only messages compared to NS. However, our results also implied that NNS did not have trouble detecting the relevant words used as emotional cues in each message.

Secondly, we highlighted a discrepancy between NS and NNS annotation of symbols (exclamation marks) and emoticons as emotional cues in the messages. In short, NNS did not make use of the available emoticons when detecting the emotional valence of a text-only message written by NS authors.

Based on these results, we propose two design implications for a system to bridge the gap between NNS receiver's and NS author's understanding of the emotional tone conveyed in a text-only message.

1) *System for classifying the lexical sentiment in emotion words based on non-native speaker's ratings.*

Existing software for automatic analysis of emotions conveyed in words and sentences are tuned based on native English-speakers' evaluations (e.g., SentiWordNet [6], LIWC [19]). For example, the system used in this paper, SentiWordNet, includes a numerical score for positive-negative polarity of subjective terms based on human ratings [6]. Based on results in this paper, the existing systems may not adequately predict NNS understanding of the emotional nuances in individual words and messages.

A system that would gather human annotations of emotion words specifically by non-native speakers would be one solution to alleviate the problem highlighted in this paper. For example, synset classification of emotion terms rated by NNS could be compared to a classification rated by NS (e.g., SentiWordNet). In case of discrepancies in the ratings, the system could let the NS author know about the possible misunderstanding the word choice may cause for NNS receivers.

A less transparent approach would be to automatically replace the emotion words with disparate ratings in NS and NNS synsets to the closest matching synonym only for NNS receivers. While this approach would be less transparent for the NS authors, it would impose a lower burden on them, as well as on other NS who receive the same message (e.g., one-to-many messages).

2) *System for translating emotional cues in symbols and emoticons for non-native speakers.*

Previous research has implied that symbols (exclamation marks) and emoticons help NS to determine between positive and negative emotions [10] and the strength of the emotion [3, 26]. However, our results indicated that not all available cues are useful for NNS (i.e., exclamation marks and emoticons).

Due to the variety of emoticons in different languages, NNS may have difficulties connecting the verbal emotional content in their second language to the symbols and emoticons. Thus, simply replacing the emoticons with the closest equivalent in the NNS native language may not be helpful. However, since the NNS in our study relied mostly on words to detect the emotional valence of a message, a solution that would translate the emotional tone of the emoticon verbally to NNS receivers would likely be helpful. For example, the basic human emotions (anger, disgust, fear, sadness, surprise, anticipation, trust, joy/happiness [5, 20]) could be used to automatically translate the emotional tone of symbols and emoticons to NNS receivers.

Informing the NS author about low expressivity of symbols and emoticons as emotional cues for the NNS may help prevent misunderstandings between NS and NNS. A transparent approach would be a system that would assist NS authors to translate their emotional expressions in symbols and emoticons. This would not lower the amount of cues available for NS receivers, but rather give NNS more verbal cues to detect the emotions conveyed in text-only message.

Future Directions

For future studies, we are interested in expanding this study to authors and participants with different language backgrounds. For example, NNS whose native language is closer to the grammatical and conceptual structure of English, or have had a longer tenure in an English speaking country might perceive emotions in English text-only messages differently. Thus, adjusting the native language of the authors and/or the language background of the participants may yield different results. We are also interested in expanding this study to include multiple emotional dimensions (e.g., joy, anger, sadness, etc.). Further, we are interested in how our results may apply to various conversational contexts, such as when the author and receiver are familiar with each other.

CONCLUSIONS

Accurately detecting emotions in text-only CMC is difficult even for native speakers. Authors need to employ alternative strategies for expressing their emotional states as opposed to when using rich communication channels [9, 15, 25]. Yet, little is known on how non-native speaking receivers detect the emotions in text-only messages written by native speaking authors based on these cues.

By looking at how accurately Japanese non-native English speakers were able to detect the emotional valence in text-only messages, we have been able to identify key problems in multilingual distance communication. When detecting the emotional valence of messages, non-native English speakers are not able to reach the accuracy of native English speakers. Furthermore, non-native English speakers have particular problems detecting the emotional valence when text-only messages include highly negative words as opposed to highly positive words.

Despite their problems in detecting the emotional valence of a message, Japanese non-native English speakers are able to accurately detect the proper emotional cues, particularly words, from text-only messages. However, while native speakers rely on symbols (exclamation marks) and emoticons to detect the emotional valence, non-native speakers are unable to make use of these cues. In other words, even when native English-speaking authors include emotional cues in their messages that both native and non-native speakers are able to accurately detect, non-native speakers detect the emotion expressed with these cues differently.

This study highlights challenges for socio-emotional communication in text-only mediums between native English-speaking authors and non-native English-speaking receivers. Especially when the majority of communication is done over text-only mediums (such as email or instant messaging), misunderstandings in emotion detection between native and non-native speakers can have detrimental effects in interpersonal and work relationships. We hope that this work illustrates the discrepancies between how non-native speakers detect emotions in text-only communication, and how native English speakers express them as message authors.

ACKNOWLEDGMENTS

We would like to thank all Facebook message authors who agreed to let us use their public status updates as part of our dataset. We would also like to extend our gratitude to the anonymous reviewers for their insightful comments and suggestions for improving this paper.

REFERENCES

1. Aman, S. and Szpakowicz, S. Identifying expressions of emotion in text. In *Proc. TSD 2007* (*LNCS 4629*), Springer (2007), 196-205.
2. Benson, M.J. Attitudes and motivation towards English: A survey of Japanese freshmen. *RELC Journal 22, 1* (1991), 34-48.
3. Derks, D., Bos, A.E.R. and von Grumbkow, J. Emoticons and online message interpretation. *Social Science Computer Review, 26, 3* (2008), 379-388.
4. Dewaele, J-M. The emotional weight of *I love you* in multilinguals' languages. *Journal of Pragmatics 40, 10* (2008), 1753-1780.
5. Ekman, P. An argument for basic emotions. *Cognition and Emotion 6* (1992), 169–200.
6. Esuli, A. and Sebastiani, F. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proc. LREC 2006*, 2006, 417-422.
7. Gill, A.J., French, R.M., Gergle, D. and Oberlander, J. The language of emotion in short blog texts. In *Proc. CSCW 2008*, ACM (2008), 299-302.
8. Gill, A.J., Gergle, D., French, R.M. and Oberlander, J. Emotion rating from short blog texts. In *Proc. CHI 2008*, ACM (2008), 1121-1124.
9. Hancock, J.T., Gee, K., Ciaccio, K. and Lin, J.M-H. I'm sad you're sad: emotional contagion in CMC. In *Proc. CSCW 2008*, ACM (2008), 295-298.
10. Hancock, J.T., Landrigan, C. and Silver, C. Expressing emotion in text-only communication. In *Proc. CHI 2007*, ACM (2007), 929-923.
11. Hyland, F. Learning autonomously: Contextualising out-of-class English language learning. *Language Awareness 13, 3* (2004), 180-202.
12. Kato, Y., Kato, S. and Akahori, K. Effects of emotional cues transmitted in e-mail communication on the emotions experienced by senders and receivers. *Computers in Human Behavior, 23, 4* (2007), 1894-1905.
13. Miyake, K. How young Japanese express their emotions visually in mobile phone messages: A sociolinguistic analysis. *Japanese Studies 27, 1* (2007), 53-72.
14. Mok, D., Carrasco, J.A. and Wellman, B. Does Distance Still Matter in the Age of the Internet?. *Urban Studies 46, 13* (2009), 2747-2783.
15. Murphy, K. L. and Collins, M. P. Development of communication conventions in instructional electronic chats. *Journal of Distance Education 12, 1–2* (1997), 177-200.
16. Nishimura, Y. Linguistic innovations and interactional features of casual online communication in Japanese. *Journal of Computer-Mediated Communication 9, 1* (2006) http://onlinelibrary.wiley.com/doi/10.1111 /j.1083 - 6101.2003.tb00356.x/full (accessed: 15.9.2013).
17. Olson, G.M. and Olson, J.S. Distance matters. *Human-Computer Interaction 15, 2* (2000), 139-178.
18. Park. J., Barash, V., Fink, C. and Cha. M. Emoticon style: Interpreting differences in emoticons across cultures. In *Proc. ICWSM 2013*, AAAI Press (2013), 466–475.
19. Pennebaker, J. W., Francis, M. E. and Booth, R. J. Linguistic inquiry and word count. Mahwah, NJ: Erlbaum Publishers (2001).
20. Plutchik, R. *The Psychology and Biology of Emotion.* New York: HarperCollins (1994).
21. Provine, R.R., Spencer, R.J. and Mandell, D.L. Emotional expression online: Emoticons punctuate website text messages. *Journal of Language and Social Psychology 26, 3* (2007), 299-307.
22. Romney, A.K., Moore, C.C. and Rusch, C.D. Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. In *Proc. National Academy of Sciences 94, 10* (1997), 5489-5494.
23. Russell, J.A. Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology 45, 6* (1983), 1281-1288.
24. Short, J., Williams, E. and Christie, B. *The Social Psychology of Telecommunications*. New York: Wiley (1976).
25. Walther, J. Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research 19* (1992), 52-90.
26. Walther, J. and D'Addario, K.P. The impacts of emoticons on message interpretation in computer–mediated communication*. Social Science Computer Review 19* (2001), 324-347.
27. Walther, J.B., Loh, T. and Granka, L. Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of Language and Social Psychology 24, 1* (2005), 36-65.
28. Yuasa, M., Saito, K. and Mukawa, N. Emoticons convey emotions without cognition of faces: An fMRI study. In *Proc. CHI 2006*, ACM (2006), 1565-1570.