

Emoticon and Text Production in First and Second Languages in Informal Text Communication

Cecilia R. Aragon¹, Nan-Chen Chen¹, Judith F. Kroll,² and Laurie Beth Feldman^{3,4}

¹ Human Centered Design & Engineering Dept., University of Washington, Seattle, WA, USA
{aragon, nanchen}@uw.edu

² Department of Psychology & Center for Language Science, Pennsylvania State University,
University Park, PA, USA
jfk7@psu.edu

³ Department of Psychology, The University at Albany-SUNY, Albany, USA

⁴ Haskins Laboratories, New Haven, CT, USA
lfeldman@albany.edu

Abstract. Most of the recent research on online text communication has been conducted in social contexts with diverse groups of users. Here we examine a stable group of adult scientists as they chat about their work. Some scientists communicated in their first language (L1) and others communicated either in their L1 or in a second (L2) language. We analyze the production in English of emoticons and of lines of text and compare measures in L1 and L2 speakers. L1 and L2 speakers differed significantly along multiple measures. English L1 speakers used more lines of text per message. English L2 (French L1) speakers used more emoticons per message. Patterns suggest compensatory emoticon/text productivity. In future analyses we will undertake a more fine-grained analysis of how emoticon use varies across social and linguistic settings. Computer-mediated communication is often viewed as impoverished, but even our initial research provides hints that users repurpose the technology according to social dynamics previously associated only with face-to-face communication.

Keywords: informal text communication, emoticons, bilingualism.

1 Introduction

Written communication poses special problems for understanding that do not arise when speaking and many of these problems recur in online communication. Nonetheless, advancing technology means that online communication is becoming more common not only in social but also in work domains. Many of these work contexts require people to talk in a language other than their native or first language (L1). Understanding how bilingual speakers communicate online may become crucial to productive cross-cultural team interactions and decision-making across remote workplaces. Effective social interaction encompasses understanding not only the transfer of information from others but also their emotions. That challenge is exacerbated in what has previously been considered an impoverished computer-mediated environment [1]. The insertion of emoticons is one option to convey emotion in

online text communication. Emoticons appear to function similarly to facial cues, tone of voice, and body language in face-to-face communication. We hypothesize that the use of emoticons in informal text communication (ITC) is becoming systematic and may now echo many of the patterns observed in face-to-face communication.

We focus here on emoticon use in a bilingual, cross-cultural scientific collaboration where chat serves as the primary method for coordinating scientific tasks for periods of many hours a day. The data set consists of nearly half a million lines of chat collected over a four-year collaboration and includes emoticons as well as text. The initial findings that we report here are novel in several respects. The text communication derives from a task shared by a quasi-permanent group of adult scientists. About half of them communicate in both the first (L1) and the second (L2) language and the remainder only in their L1. Therefore, we can compare L1 and L2 communication in the same social and collaborative work environment. Gesture and other nonverbal behavior enrich the text they accompany. Adopting a psycholinguistic framework, we hypothesize a parallel function for emoticons in text chat [2] as has been postulated for the role of gesture, and other nonverbal behavior in the context of face-to-face conversation. It has been asserted that gestures are less common in the presence of shared knowledge [3]. In this case, speakers may be less likely to gesture when they engage in a shared task. However, gestures may be compensatory for a lack of high proficiency in which case speakers may be less likely to gesture when proficiency is high. Our assumption is that text communication engages similar mechanisms to spoken discourse and that many of the same factors that shape spoken communication, including the use of gesture, will be present. If so, then emoticon use should be modulated by the same factors that affect displays of emotion in spoken discourse.

2 Prior Research

2.1 Bilingual Speakers

There is consensus among researchers that bilingual speakers activate both of their languages when they read, listen, or talk in either one of their languages [4]. Because both languages are active concurrently, the cognitive functioning of bilinguals is not analogous to that of monolinguals. In fact, differences between bilinguals and monolinguals are evident in many domains. For example, the tendency for a bilingual to express emotion is greater when the L2 is the ambient language. Many claim that relative to the L1, communicating in the L2 makes self-disclosure feel less threatening and allows the speaker to remain more distant [6]. Childhood memories for events from immigrants that are described in the L2 tend to have less detail and be emotionally less charged than those in the L1 [7]. Not only are bilinguals more likely to speak freely about emotionally charged topics in their L2 than L1, but at a physiological level, taboo words spoken in L2 elicit smaller electrodermal changes from baseline in bilingual speakers than do those in the L1 [8].

In the present study, we examine the consequences of bilingualism for online text communication and emoticon insertion in a professional work context with a shared task. The virtual work environment that is enabled by current technology provides a new opportunity for testing and extending models of language processing.

2.2 Emoticon Use in Informal Text Communication

Text-based computer-mediated communication that is spontaneously generated without editing or rewriting constitutes what we call "informal text communication" (ITC). It is more similar to spontaneous speech than to formal writing. Examples include chat, instant messaging, microblogs such as Twitter, and certain forum posts/message boards. Note that e-mail, online profiles, product reviews, blog posts, and webpages are excluded because they resemble written discourse. ITC may be synchronous or asynchronous; essential is that it serves to communicate with other humans without substantial reflection, similar to spontaneous speech.

Even a perusal of the literature [9, 10] indicates that the options when communicating in text-based chat environments are becoming more elaborated and less prescriptive, and that humans are adapting available technologies to meet social needs such as to convey emotion in communication. We hypothesize that the use of emoticons in ITC parallel many of the practices observed in face-to-face communication.

3 Background on the Scientific Collaboration and Chat Dataset

The chat dataset was produced by an international astrophysics collaboration consisting of about 30 members; about half of the scientists work at several different locations in the U.S. and the other half in three research institutes and universities in France. All the French scientists also speak English, and English is the official language of the collaboration. Collaboration members use English in the chat whenever an English speaker is present; French speakers may revert to French whenever they are alone in the chat.

The astronomers' task is complex and requires coordination on telescope observation especially when working under time pressure. The primary means of communication during remote telescope observation are AIM (AOL Instant Messenger) chat (augmented by a virtual assistant) and VNC (virtual network computing).

4 Analysis and Results

4.1 Data

Description of the Chat Corpus. The corpus consists of a total of 485,045 chat messages. The logs include 1,319 days (nearly four years from 2004 to 2008), and cover approximately 12-hour sessions during which observations from a remote telescope were coordinated. A line of chat refers to a single message. Messages are posted as soon as the user hits return in the chat client.

The chat data included only participants who contributed over 2000 lines of chat and who used more than 30 emoticons over the period of data collection. The resulting corpus included the productions of 8 native English speakers and 10 native French speakers, with a total of 259,256 lines of chat and 3343 emoticons.

Participants. A total of 18 astrophysicists in the United States and France who were members of the collaboration formed the sample. There were 8 American scientists and 10 French scientists. Typically 5-6 people were present in the chat at any one time.

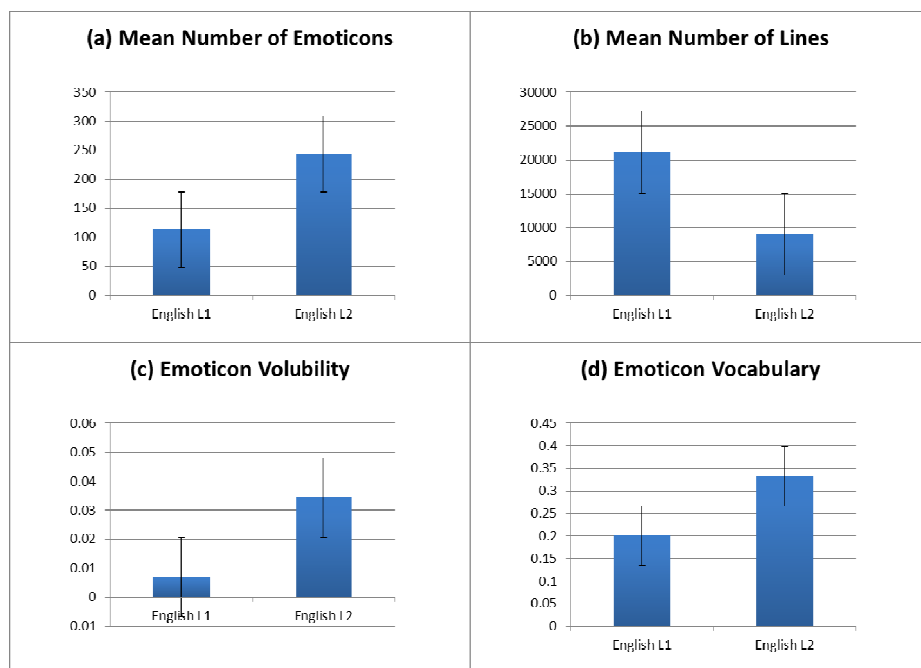


Fig. 1. (a) English L1 speakers (American) tended to insert more emoticons into their text than English L2 speakers did. (b) English L1 speakers (American) on average produced more lines of text than did English L2 speakers (French). (c) English L2 speakers (French) used more emoticons per text message (higher emoticon volubility) than did English L1 speakers (American) did. (d) English L2 speakers (French) had larger emoticon vocabularies size did English L1 speakers (American).

4.2 Results

Token Based Measures of Emoticon Use. Results of an ANOVA contrasting L1 and L2 speakers showed that English L2 speakers tended to insert more emoticons into their text messages than did English L1 speakers [$F(1, 16) = 2.859, p < .11$]. This is depicted in Fig. 1a. Results are consistent with the claim that displays of emotion are more frequent and easier to produce in one's L2. Even if reliable, an alternative

account based on overall proficiency with emoticons cannot be dismissed from these data alone. Conversely, English L1 speakers tended to write longer text messages than did English L2 speakers [$F(1, 16) = 3.354, p < .086$]. This is depicted in Fig. 1b. Most importantly, English L2 speakers used more emoticons per text message than did English L1 speakers [$F(1, 16) = 7.319, p < .016$]. We define production of emoticons per lines of text as *volubility*. It is depicted in Fig. 1c. Ongoing analyses further examine the potential compensatory relation between emoticon and text usage.

Type Based Measures of Emoticon Use. In a second set of analyses we examined *vocabulary size*, or the number of different emoticons that L1 and L2 speakers used. Results are expressed in terms of number of different emoticons relative to the total number of emoticon types documented in the corpus (total=64). Here the effect of L1 [$F(1, 16) = 3.95, p < .06$] was marginally significant. It suggested that English L2 speakers are using more different emoticons over all. This is depicted in Fig. 1d.

5 Discussion and Conclusions

This work contrasts emoticon production in ITC in a work setting by American speakers of English as the L1 and French speakers of English as the L2. We introduced four measures of production. First we examined mean number of emoticons and mean number of lines of text in each message. Then we examined the mean number of emoticons per lines of chat. We defined this token-based measure of emoticon production as *volubility*. Most important is that the *production rate* based on number of emoticons per lines of chat was higher for participants communicating in an L2 than in an L1 English. Finally, we looked at emoticon *vocabulary size*, a type measure of emoticon production that reflects the mean number of unique emoticons produced by an individual divided by the number of unique emoticons present in the entire chat (64). Vocabulary size was higher in L2 than L1 English.

Most novel in our findings is the greater volubility and vocabulary of emoticons produced by scientists communicating in their L2 than in their L1. Results are consistent with the claim that communicating in the L2 makes it easier to convey emotion because it introduces emotional distance and reduces any sense of vulnerability relative to communicating in the L1 [6, 8].

Not only do L2 chatters have higher volubility than L1 chatters, they also produce fewer lines of text. Consistent with the claim that emoticons in chat serve a purpose similar to gesture and facial expressions in face-to-face communication the increased emoticon volubility among scientists communicating in their L2 could reflect a type of compensatory behavior. Accordingly, the reduced accessibility of words to express an intended meaning in L2 could be causally related to the increased production of non-verbal elaborations such as emoticons. This pattern warrants further investigation.

6 Limitations and Future Work

We have only begun to examine chat behavior and how it varies for communication in an L1 as compared to an L2 and our sample consists of a relatively limited number of individuals. However, number of observations per person over a four-year period is disproportionately huge, adding reliability to our analyses.

More important is that status of English as either a first or second language could not be manipulated experimentally. Therefore, it remains possible that the difference between groups (emoticons per lines of text; number of different emoticons) reflects skill differences in emoticon use rather than constraints on communicating in one's L1 or L2. To attenuate the contribution of uncontrolled differences across groups, subsequent analyses will examine interactions between L1/L2 speakers and various language environments defined by the composition of L1 and L2 speakers. We ask whether productions in L2 vary more with language environment than do those in L1. It is unlikely that these interactions reflect group differences.

Acknowledgments. The writing of this paper was supported in part by NIH Grant HD053146 and NSF Grants BCS-0955090 and OISE-0968369 to J.F. Kroll; by NIH Grant HD053146 to Haskins Laboratories.

References

1. Gill, A.J., French, R.M., Gergle, D., Oberlander, J.: The language of emotion in short blog texts. In: CSCW, vol. 8, pp. 299–302 (2008)
2. Gullberg, M., McCafferty, S.G.: Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition* 30, 133 (2008)
3. Holler, J., Wilkin, K.: Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes* 24, 267–289 (2009)
4. Kroll, J.F., Bobb, S.C., Wodniecka, Z.: Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition* 9, 119–135 (2006)
5. Zhang, S., Morris, M.W., Cheng, C.-Y., Yap, A.J.: Heritage-culture images disrupt immigrants' second-language processing through triggering first-language interference. *Proceedings of the National Academy of Sciences* (2013)
6. Pavlenko, A.: *Emotions and Multilingualism*. Cambridge University Press (2007)
7. Schrauf, R.W., Hoffman, L.: The effects of revisionism on remembered emotion: the valence of older, voluntary immigrants' pre - migration autobiographical memories. *Applied Cognitive Psychology* 21, 895–913 (2007)
8. Harris, C.L., Aycicegi, A., Gleason, J.B.: Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics* 24, 561–579 (2003)
9. Brooks, M., Kuksenok, K., Torkildson, M.K., Perry, D., Robinson, J.J., Scott, T.J., Anicello, O., Zukowski, A., Harris, P., Aragon, C.R.: Statistical affect detection in collaborative chat. In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pp. 317–328. ACM (2013)
10. Park, J., Barash, V., Fink, C., Cha, M.: Emoticon Style: Interpreting Differences in Emoticons Across Cultures. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media* (2013)