Establishing common ground in informal text communication: Emoticon use in first and second languages

Nan-Chen Chen[1], Laurie Beth Feldman[2, 3] , Judith F. Kroll[4] , Cecilia R. Aragon[1]
nanchen@uw.edu, lfeldman@albany.edu, jfk7@psu.edu, aragon@uw.edu

[1] University of Washington, [2] The University at Albany-SUNY,
[3] Haskins Laboratories, [4] Pennsylvania State University

Research Question:

Do bilingual users search for common ground with monolingual users in online text communication?

- We examined 4 years of online text communication among a stable group of adult scientists as they coordinated telescope observation via chat.
- American scientists communicated in their first language (L1) and French scientists communicated either in their L1 or in a second (L2) language.
- We analyzed emoticon vocabulary size and the emoticon entropy. The former is a measure of emoticon variety based on means; the latter is a measure of changes in emoticon production based on the distribution.

The Scientific Collaboration and Chat Dataset:

The chat dataset was produced over a four-year period by an international astrophysics collaboration consisting of about 30 members; about half of the scientists worked 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 spoke English, and English was the official language of the collaboration. Collaboration members used English in the chat whenever an English speaker was present; French speakers might revert to French whenever they were alone in the chat.

The astronomers' task is complex and required coordination on telescope observation especially when working under time pressure.

The primary means of communication during remote telescope observation were AIM (AOL Instant Messenger) chat (augmented by a virtual assistant) and VNC (virtual network computing).

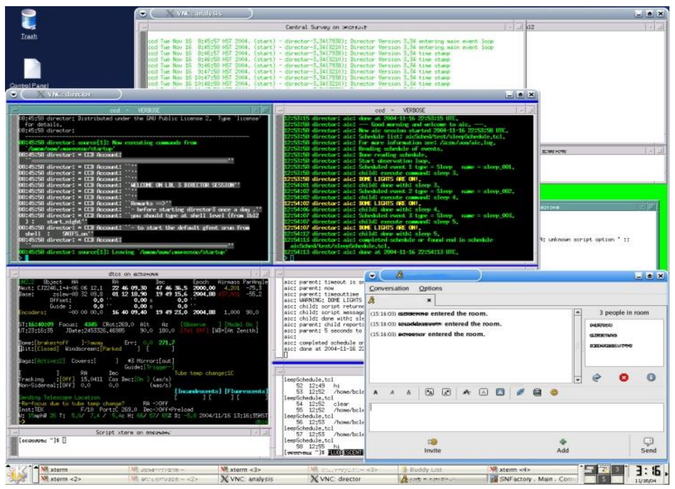


Figure 1. Scientific collaboration telescope control window with chat client

Acknowledgments:

This work 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.

For more information, contact Cecilia Aragon at aragon@uw.edu or visit http://depts.washington.edu/sccl

Methods:

- 8 native English speakers; 10 native French speakers
- each contributed 2000+ lines of chat and used 30+ emoticons
- 8926 tokens and 58 different emoticons
- 250,000+ lines of text
- Three environments: **monolingual, majority, minority**
- Two measures

1. Emoticon Vocabulary Size

$$\text{Emoticon Vocabulary Size } (S_i, \text{LE}_j) = \frac{\#\ Distinct\ \ Emoticons(S_i, LE_j)}{\#\ Distinct\ Emoticons\ in\ the\ Dataset}$$

S_i: Speaker i
LE_j: Language Environment j

2. Emoticon Entropy

$$\text{Emoticon Entropy}(t_i) = -\sum_{k}^{|DE_{t_i}|} P(E_k, t_i) \times log_2\big(P(E_k, t_i)\big)$$

$$\text{P}\big(E_{k,}\, t_i\big) = \frac{|E_k(t_i)|}{\sum_j^{|DE_{t_i}|} |E_j(t_i)|}$$

t_i: Time Bucket i,
E_k: Emoticon k
$|E_k(t_i)|$: The Number of E_k in t_i
$|DE_{t_i}|$: The Number of Distinct Emoticons in t_i
$\text{P}\big(E_{k,}\, t_i\big)$: Probablity of E_k in t_i

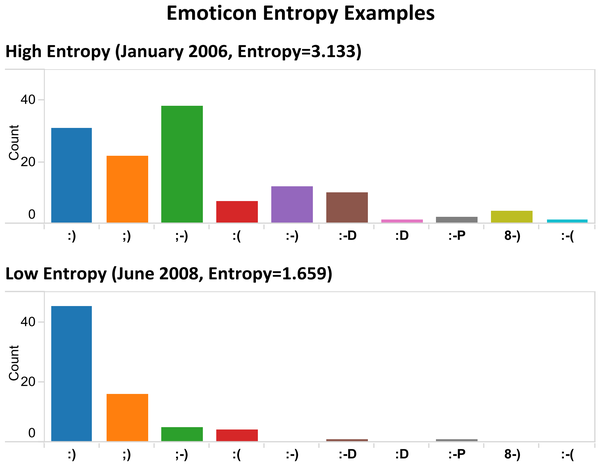


Figure 2. Months with high and low emoticon entropy: The upper chart shows the emoticon distribution in January 2006, and the lower chart the emoticon distribution in June 2008. A relatively homogeneous distribution of emoticon frequency, as appeared in January 2006, is characteristic of high entropy. Over time, collaboration members converged on a smaller set of relatively high frequency emoticons as shown in June 2008; this is characteristic of low entropy. (For illustration, we included only the top 10 emoticons. In reality calculations of emoticon entropy include all 58 emoticons.)

Results:

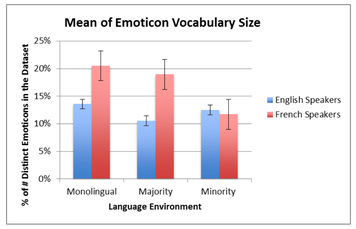
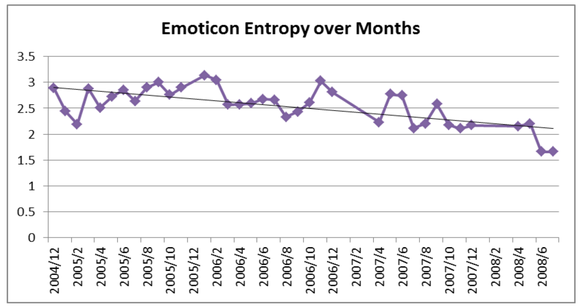


Figure 3. Mean (SE whisker) of emoticon vocabulary size (% of distinct emoticons in the dataset): A 2x3 ANOVA on language environments and language groups showed a significant interaction[F(2, 32)=6.284, p<0.005].

French speakers used a smaller emoticon vocabulary when in the minority than in the majority language environment [t(9)=3.841, p<0.004] and when in a minority relative to the monolingual environment [t(9)=4.364, p<0.002].

Figure 4. Emoticon entropy over months: The trend line shows a decreasing emoticon entropy.



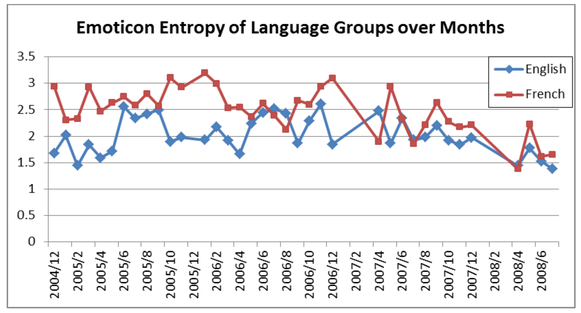


Figure 5. Emoticon entropy of the language groups (English/French speakers) over months: **The difference between the two groups decreased over time**. A paired samples t-test on the difference between the two groups in the first 10 months and the last 10 months show a significant decrease [t(9)=2.648, p<0.027].

Conclusions:

Entropy (and vocabulary size) measures show alignment in emoticon behavior over time.

- Emoticon use increased overall but differences between groups decreased later in the collaboration.
- French speakers changed more than did English speakers.