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 group.
- 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 primary means of communication during remote telescope observation was 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

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
     \[ \text{Emoticon Vocabulary Size} = \frac{\# \text{Distinct Emoticons}(S_i, LE_j)}{\# \text{Distinct Emoticons in the Dataset}} \]
     \[ S_i: \text{Speaker } i \]
     \[ LE_j: \text{Language Environment } j \]
  2. Emoticon Entropy
     \[ \text{Emoticon Entropy}(t) = -\sum_{i} \frac{|D_E_i| |E_i(t)|}{\sum_j |E_j(t)|} \log_2 \left( \frac{|E_i(t)|}{|D_E_i|} \right) \]
     \[ t: \text{Time Bucket} i \]
     \[ E_i(t): \text{Emoticon } k \]
     \[ |E_i(t)|: \text{The Number of Emoticons in } E_i(t) \]
     \[ |D_E_i|: \text{The Number of Distinct Emoticons in } E_i(t) \]

Results:
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.