How emotions influence COVID related Code-switching and Code-mixing texts on Social Media

Xinyuan Xia xxia16@syr.edu & supervised by Professor Tej Bhatia tkbhatia@syr.edu

Syracuse University
College of Arts & Sciences

Abstract

Since the beginning of 2020, the outbreak of the new coronavirus has had a devastating impact on the global population. People's lives have been changed indefinitely, as well as the methods and means of human communication. During this quarantine time, the Internet and social media have become the most important platforms to make one's voice heard. People share their views, comments, lives, worries, or condolences of the COVID-19 and pandemic online in this must-be-isolated time to keep the essential needs for communicating and socializing. Anxiety, fear, and depression seem to be the keywords that revolves around the COVID-19, and emotions have reshaped the communication between different languages.

This paper aims to analyze the relationship between emotions and bilingualism during the COVID-19 pandemic. It firstly provides the most recent reviews on studies of COVID-19 and bilingualism and then discusses what code-switching (CS) and code-mixing (CM) is. Moreover, for data collection and analysis, it focuses on Chinese-English code-switching and code-mixing social media data from Twitter, Weibo, and YouTube to discuss the effects of emotion for language change. Even though the data is limited, it still uses qualitative analysis with supplemented quantitative data to first categorize the bilingual data into different CS/CM genres and sentiments. And then to discuss how and why bilinguals use different codes to express their sentiments and unfold the emotion and communication deprivation caused by the outbreak of coronavirus. It reveals that emotions can indeed influence the choice of language for bilinguals but with no absolute boundaries for shifting language under one particular sentiment. Different goals and purposes of using languages, people's sociolinguistic backgrounds and social contexts are all influential factors in the shifting.

Methodology & Analysis

Initial difficulties

- Twitter has a sophisticated API and it is difficult for people who are not an expert on coding to extract bilingual texts automatically.
- Twitter does not contain much Chinese-English bilingual data for COVID-19 because it is banned in China due to censorship issues .

Therefore, it eventually led to the current small amount of bilingual data from three social media platforms, Twitter, Weibo and YouTube.

Data Collection

In order to find the bilingual data for COVID-19 and emotions in Twitter and Weibo, it had to narrow down the scope manually by searching in specified keywords combinations which are comprised of English and Chinese words related to the pandemic and emotions.

Platform	keywords combinations and post number			
Twitter	#lang:zh-cn covid# 17		#lang:zh-cn covid sad# 8	
Weibo	#Covid sad# 5		#疫情 yi qing (pandemic) depress# 6	
	#疫情 yi qing (pandemic) anxious# 2		#疫情 yi qing (pandemic) depression# 5	
	#疫情 yi qing (pandemic) shit# 4		#疫情 yi qing (pandemic) mental health# 7	
	#COVID # 25		#Covid累 lei (tired)# 16	
	#疫情 yi qing (pandemic) tired# 11			

As for YouTube data, inspired by the case study of Chinese-speaking YouTube vloggers in COVID-19 (Zhang & Zhao, 2020), it finally found one channel called "See The World with DJ." This vlogger has started a new series of videos ("DJ's 90 seconds share of America's pandemic") since COVID. It aims to share pandemic information in America in both English and Chinese. The whole series are eight vlogs, and the data are extracted CS or CM sentence transcripts in the video.

Data Analysis

First, all data is transcribed and then coded into two large categories: one is for different kinds of CM/CS; the other is for sentiments like pos/neg/neu.

CM/CS	Numbers	Inter-sentential CS 20	Inter-sentential CS & Literal translation 7	
	Intra-sentential CM 63	Inter-sentential CS & Intra-sentential CM 13	
	Inter-sentential CS & Intra-sentential CM & Literal translation 2		
Sentiments | Numbers	Positive 15	Negative 63	Neutral 24
	Positive & Negative 3		

The most frequently occurred CM/CS type is intra-sentential CM, and the emotional inclination of most transcripts are towards negative. It is understandable that intra-sentential CM would appear more often because there are no restricted rules or social contexts for using any language in one sentence or post on social media.

Moreover, sentiment words/phrases/sentences in each post are also extracted and categorized into two groups by the language they use as either English sentiment words or Chinese words. And words expressing sentiments in Chinese appeared more often than the English ones.

■EN-SEN ■CN-SEN
55% 45%
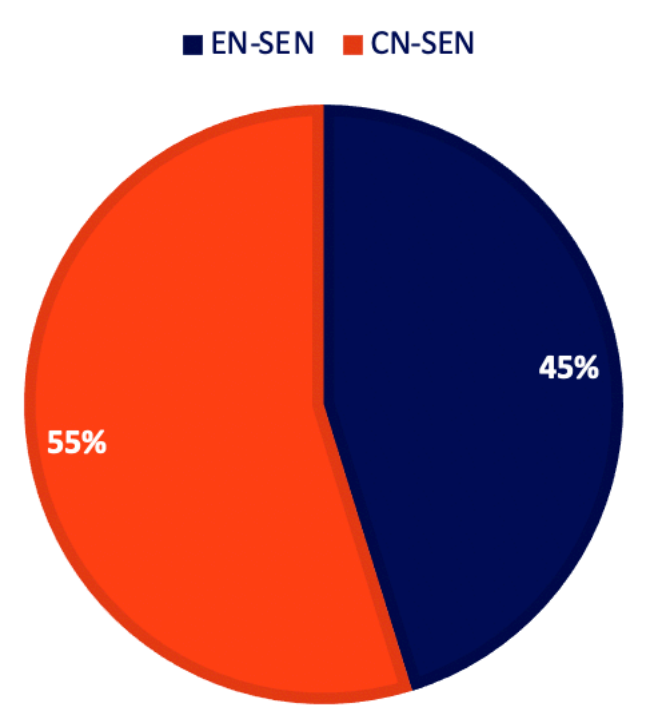
Data Example

English—*Chinese*

It's sad to say that the world's number one country is actually governed by the President Trump who is an internet troll. *hao cheng shi jie di yi de guo jia, jing ran shi ge pen zi zai li zheng zhi guo, zhen shi bei ai a!*

It's sad to say that the world's number one country is actually governed by the President Trump who is an internet troll. 号称世界第一的国家竟然是个喷子在理政治国，真是悲哀啊!

[T: It's sad to say that the world's number one country is actually governed by the President Trump who is an Internet troll. It is really pathetic that the so-called world NO.1 country is actually governed by an internet troll.]

(Inter-sentential CS/literal translation)

Negative: the word "sad" and phrase "*bei ai*" --- "pathetic"; it expresses the feeling that Trump is a terrible president just knows to tweet.

Findings

- From linguistic perspective
  - Majority of data: matrix language is Chinese
  - Mix English sentences for convenience and informality
  - Sociolinguistic & Cultural Background
- From emotional perspective
  - L1 (Chinese) is the most frequent choice in terms of negativity; choose L1 to express their deeply anxious nightmares, rough days or overwhelming feelings towards the pandemic. Use the L2 to gain distance, exercise self-control. (Pavlenko, 2008)
  - L2 is not for detachment but for intimacy (Dewaele, 2010)
  - L1 to express strong emotions like swearing, but also use L2 to swear (Pavlenko, 2008; Dewaele, 2010)
- For YouTube data

All data are inter-sentential CS and the dominant sentiment of vlogs is positive.

CS & emotions happened mainly because:

  - The YouTuber is an Asian-American, use English to express emotional scenarios would be more comfortable.
  - Socio-cultural background of the content

Conclusions

- People use social media as emotional outlets to let off steam and stay connected with the society.
- It has provided an irresistible chance for bilinguals or multilinguals to share opinions in different languages.
- The choice of language for expressing polarity sentiments is either bidirectional which L1 and L2 could both be shifted or unilateral that only use L1 to express a certain feeling.
- The sociolinguistic backgrounds and objectives pay an irreplaceable role in language shifting and emotions.
- Therefore, it means that for analyzing emotions and CS/CM, the most significant parts are identifying the social contexts underneath the conversation and understanding the social and linguistic purposes behind the texts.

Improvements and future works

- It would be great if it could get the sociolinguistic information of the bilingual speakers online either through individual interviews or surveys that could gather their opinions on shifting languages.
- The methods of collecting large amount of bilingual data on social media are too sophisticated. If there are more convenient and easier way to collect abundant bilingual data, the results would be more detailed and convincing.

References

Dewaele, Jean. 2010. *Emotions in multiple languages*. New York: Palgrave.

Pavlenko, Aneta. 2008. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, 11(2), 147-164. doi:10.1017/S1366728908003283

Zhang, Leticia-Tian, and Sumin Zhao. 2020. Diaspora micro-influencers and COVID-19 communication on social media: The case of Chinese-speaking YouTube vloggers". *Multilingua* 39(5). 553-563. https://doi.org/10.1515/multi-2020-0099 Web.