

Abstract

We conduct a linguistic analysis of the language usage patterns of multilingual peer supporters in two health-focused WhatsApp groups in Kenya comprising of youth living with HIV. Even though the language of communication for the group was predominantly English, we observe frequent use of Kiswahili, Sheng and code-mixing among the three languages. We present an analysis of language choice and its accommodation, different functions of code-mixing, and relationship between sentiment and code-mixing. To explore the effectiveness of off-the-shelf Language Technologies (LT) in such situations, we attempt to build a sentiment analyzer for this dataset. Our experiments demonstrate the challenges of developing LT and therefore effective interventions for such forums and languages.

Dataset

Multilingual youth in Kenya, living with HIV, actively participating in two WhatsApp based peer support forums, facilitated by a healthcare facilitator

Group – 1

1,655 messages (28 members, 14 female, 14 male, age=14-17 years)

Group – 2

4,901 messages (27 members, 21 female, 6 male, age=18-24 years)



Linguistic Analysis: Understanding the inter-personal relationships among the peer supporters

Linguistic Preferences for intent and emotions

- Role of Sheng during formal and informal conversations?
[More preferred during informal chit-chat than formal ones]
(Sheng in chitchat >> Sheng in information exchange)
- Role of each language in expressing various conversational intent?
[English: more prevalent in formal expression, Kiswahili in informal expression]

Understanding Linguistic Style Accommodation

The way by which language choice of the speakers are coordinated and how much one speaker's language choice influence the others choice too.

Groups	Swahili Acm	English Acm	Sheng Acm	Swahili/Sheng Acm
Group-1	0.099	0.054	0.079	0.207
Group-2	0.265	0.031	0.168	-

Positive accommodation in both the groups.

Kiswahili - marked code-choice for Group-2.

Average Accommodation of the speakers for different languages.

Joint accommodation by Kiswahili and Sheng.

Exploring Code-Switching Functions

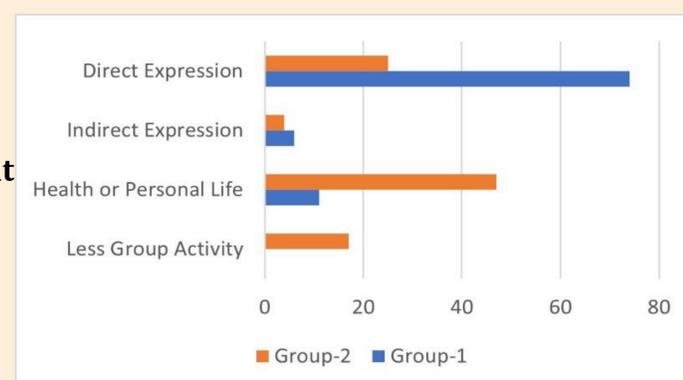
Expression of Sentiment

Which is more dominant? **Structural** (how people code-switch) patterns or **Pragmatic** patterns (why people code-switch)?.

Negative sentiment and user activity correlation?

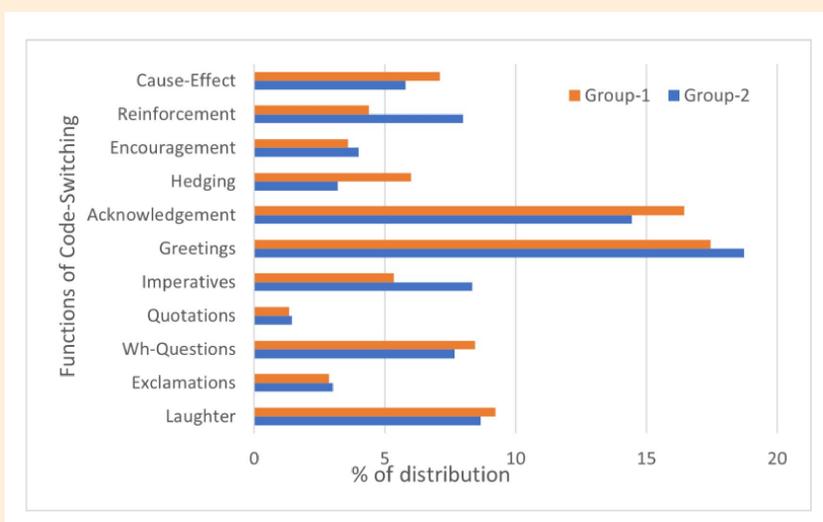
- Positive ☺

Themes of Negative sentiment



Developing Sentiment Classifier?

Tricky as we can achieve the maximum of 0.60 F1-score fine-tuned by using state-of-the-art MMLM XLM-Roberta.



Summary and Recommendations

1. Analysis and benchmarking the real-time healthcare datasets with the unique challenges of handling conversational aspects and multilingual nature should be encouraged.
2. Since Sheng plays a crucial role in effective informal support exchange, it is pivotal to infuse its vocabularies and support code-mixing in the multilingual models.
3. In order to understand the dichotomy of various demographic factors influencing behavior of the peer supporters, massive scale data collection efforts should be made.