Language Patterns and Behavior of the Peer Supporters
In Multilingual Healthcare Conversational Forums

Ishani Mondal, Kalika Bali, Mohit Jain, Monojit Choudhury, Keshef Ronen, Millicent Ochieng, Kagonya Awori, Jacki O'Neill

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

<table>
<thead>
<tr>
<th>Group – 1</th>
<th>Group – 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,655 messages (28 members, 14 female, 14 male, age=14-17 years)</td>
<td>4,901 messages (27 members, 21 female, 6 male, age=18-24 years)</td>
</tr>
</tbody>
</table>

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

- 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]

Exploring Code-Switching Functions
Which is more dominant? **Structural** (how people code-switch) patterns or **Pragmatic** patterns (why people code-switch)?

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.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Swahili Acen</th>
<th>English Acen</th>
<th>Sheng Acen</th>
<th>Swahili/Sheng Acen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group-1</td>
<td>0.099</td>
<td>0.054</td>
<td>0.079</td>
<td>0.207</td>
</tr>
<tr>
<td>Group-2</td>
<td>0.265</td>
<td>0.011</td>
<td>0.168</td>
<td>-</td>
</tr>
</tbody>
</table>

Positive accommodation in both the groups.

Average Accommodation of the speakers for different languages.

- **Kiswahili** - marked code-choice for Group-2.
- Joint accommodation by Kiswahili and Sheng.

Expression of Sentiment

**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.

Contact: imondal@microsoft.com, kalikab@microsoft.com