

Language Patterns and Behaviour of the Peer Supporters in Multilingual Healthcare Conversational Forums

¹Ishani Mondal, ¹Kalika Bali, ¹Mohit Jain, ¹Monojit Choudhury, ²Keshet Ronen,
³Millicent Ochieng, ³Kagonya Awori, ³Jacki O’Neill

¹Microsoft Research India, ²University of Washington, ³Microsoft Africa Research Institute, Kenya
{t-imonda, kalikab, monojitc}@microsoft.com

Abstract

In this work, we conduct a quantitative 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. We provide recommendations for language resources that should be built to address these challenges.

Keywords: Multilingual Conversations, Healthcare, Peer support groups, Code-mixing

1. Introduction

In recent years, various online chat applications such as WhatsApp and WeChat are being increasingly leveraged for establishing peer support groups beyond the formal healthcare system (Karusala et al., 2021; Bhat et al., 2021). Exchange of peer support is critical for attaining improved physical and mental well-being as it helps the individuals deal with uncontrollable and emotionally crippling life events. Prior work on peer support forums are limited to studies of social connection and engagement patterns (Sharma et al., 2020a; Sad-
eque et al., 2015), and modelling of user behavior (Hosseini and Caragea, 2021; Sharma et al., 2020b). We are not aware of work that quantitatively studies and characterizes the linguistic patterns of interaction in such forums.

In this work, we conduct analysis of the language and communication patterns in two WhatsApp based healthcare forums in Kenya, which were used by youth, living with HIV, for discussing health, lifestyle, personal and cultural issues, and connect to other individuals with similar needs and challenges. While there was a significant use of English in the conversations, the users were mostly tri-lingual, speaking English, Kiswahili and Sheng (Sheng is a Kiswahili and English-based cant). We also observe frequent code-mixing between pairs or all of the three languages. The objectives of this work are threefold.

- 1) We wish to understand the linguistic patterns of interactions in such forums that fortify peer support and create trust and bonding.
- 2) In particular, we want to study the relationship between the choice of language, and (a) linguistic accommodation, (b) conversational intent, and (c) sentiment.
- 3) We would also like to understand if we can build

positive technological interventions for such forums with the current-state-of-the-art language technology (LT) for these languages. And, if not, what language resources should be created to enable appropriate LTs. To address these questions, we begin with the linguistic annotations for this dataset created by (Mondal et al., 2021). First, we explore the patterns of language preference of peer supporters while expressing different conversational intents of peer support, such as information seeking/providing, group work and greetings (Section 4). This part of the analysis reveals that while English is mostly used for information exchange, Kiswahili and Sheng are more commonly used for greetings and informal chitchat.

Second, we adopt the framework proposed by (Bawa et al., 2018) to measure the linguistic style accommodation for code-choice. We observe that in one of the groups, Sheng and English seem to be the marked code-choices, while Kiswahili being the unmarked choice. In another group, Kiswahili and Sheng together constituted the unmarked code, and English remained the marked code-choice. This curious dichotomy is likely explainable by the difference in the demography, primarily age of the users, which has interesting linguistic repercussions that we discuss in Section 5.

Third, we study the various functions of code-mixing (Begum et al., 2016) in the conversation beyond the basic forum-specific intents. We find that structural (rather than functional) switching patterns are more dominant, which can be attributed to the short length of utterances and prevalence of interjections in the chitchat conversations. (Section 6)

Finally, we observe that, excluding the healthcare providers, most of the active members in the group express negative sentiments far more often than positive ones (Section 7). This made us believe that a poten-

tially beneficial technical intervention could be to identify the sentiment of the utterances, and nudge the users to use and/or suggest more positive ways of expressing the same. Therefore, as a case study, we take up the exercise of developing a simple triaging tool to identify negative sentiments expressed in the group. However, we found that the off-the-shelf Transformer-based Language models, XLM-R (Conneau et al., 2020) and mBERT (Devlin et al., 2019), when fine-tuned for sentiment detection, performs extremely poorly not only on Sheng, Kiswahili and code-mixed data, but also on English utterances, arguably because of their short length and heavily contextualized meanings. We conclude by providing several recommendations for developing language resources that can support LT for peer-to-peer health forums for low resource languages.

2. Background and Related Work

Kenya is amongst those regions of the world that display an extensive linguistic diversity (Spernes and Ruto-Korir, 2021; Ideh and Onu, 2017; Dwivedi, 2015). Broadly, there are three language groups in the region, namely *Bantu*, *Nilotic*, and *Cushitic*, and each group includes more than five languages, making multilinguality a norm in the country (Dwivedi, 2015). English is considered as a colonial language spoken by most of the educated people whereas Kiswahili, a *Bantu* language, is one of the official languages of the country, and being spoken by the majority of the population enjoys a near equal status with English (Dwivedi, 2015; Githiora, 2002). However, the emergence of language varieties like “Sheng” have made inroads into the lingua franca Kiswahili. An existing body of work looks into the origin, definition and evolution of Sheng; such as whether Sheng is a pidgin or creole (Rinkanya, 2015; Iraki, 2014; Ogechi, 2005; Githiora, 2002; Abdulaziz and Osinde, 1997), or a mixed-code of Kiswahili, English and other native languages. Studies show how Sheng evolved from a stigmatized ‘ghetto’ code into a prestigious code symbolizing ideological affinity, in-group identity, and linguistic innovation (Kaviti, 2015; Dizayi, 2015). Other studies have focused on the impact of Sheng’s usage patterns on African Culture (Kariuki et al., 2015; Nassenstein and Hollington, 2015; Makewa et al., 2014; Mutiga, 2013; Momanyi, 2009; Githinji, 2008; Fink, 2005), as well as on the sociolinguistic aspects of code-mixing and lexical restructuring that the language undergone (Kanana and Ny’onga, 2019; Githiora, 2018; Bosire, 2008; Githinji, 2006; Bosire, 2006). To the best of our knowledge, we are the first to linguistically analyze the conversations of peer supporters, speaking English, Kiswahili, and Sheng fluently, in online conversational forums.

In a multilingual context, there are more than one linguistic channel for information exchange available to the speakers interacting on social media platforms. A number of existing linguistic work seeks to understand

linguistic preferences for expressing emotion on these online social media channels such as Twitter, Facebook and Reddit (Xiang et al., 2020; Bhat et al., 2018; Rijhwani et al., 2017; Ramesh and Kumar, 2017; Pimpale and Patel, 2016; Jamatia et al., 2015; Bali et al., 2014). However, very little attention has been paid to linguistic and sociolinguistic analysis of multilingual conversation patterns of the participants in small, close-knit instant messaging applications such as WhatsApp. There has been work on the complexity in analyzing WhatsApp messages, mainly due to non-standard spellings, abbreviations, contracted forms of words, short replies, and emoticons (Makhija et al., 2020; Daniel et al., 2019; König, 2019; Sprugnoli et al., 2018; Dorantes et al., 2018). Other linguistic research focuses on dealing with adaptation of speaker’s linguistic style as a marker of coordination on Facebook and Twitter during informal conversations (Bawa et al., 2018; Danescu-Niculescu-Mizil et al., 2011; Giles et al., 2010; Sachdev and Giles, 2008), and understanding sentiment to assess emotional behavior (Nguyen and Shirai, 2015; West et al., 2014; Balahur, 2013; Arunachalam and Sarkar, 2013; Habernal et al., 2013; Abdul-Mageed et al., 2012). None of these, however, aim to understand the patterns of language preference with respect to conversational intent (such as dealing with the different informational content of the messages, showing emotional concerns towards fellow members or participating in a friendly chitchat), or the patterns of language coordination used to express different forms of peer support in instant messaging. To the best of our knowledge, we are the first to present a linguistic analysis of the messages written by youth living with HIV in a mixed-income area of Nairobi.

3. Dataset

We have used the same dataset and linguistic annotation framework mentioned in (Mondal et al., 2021). It comprises of WhatsApp chat logs from two peer-support groups for the Kenyan youth, living with HIV. Overall, there are 1,655 messages in Group-1 (28 members, 14 female, 14 male, age=14-17 years) and 4,901 messages in Group-2 (27 members, 21 female, 6 male, age=18-24 years). In order to model the conversational intent of peer supporters behind the act of sending messages, the messages are broadly classified into these major peer support categories such as *Informational*, *Emotional*, *Chitchat*, *Acknowledgement*, *Group Work*, and *Other*, which are further sub-categorized. The *Informational* support category consists of *Medical*, *Admin*, *Lifestyle*, *Personal*, and *Other*. The subcategories of *Emotional* support include *Empathy*, *Negativity*, *Happiness*, and *Hopefulness*. Moreover, the emotional behavior of peer supporters in expressing the opinions are assessed using a sentiment analysis framework comprising of *Positive*, *Negative*, *Neutral* sentiments. The morphosyntactic annotations comprise of 5 word-level language tags: *English (En)*, *Kiswahili*

(*Sw*), *Sheng* (*Sh*), *Code-Mixed Word/Phrase* (*CM*), and *Other* (*Oth*).

4. Understanding Language Preferences

Over the years, sociolinguistics are interested to understand how people speak differently in varying social contexts. A number of prior researchers (Dewaele, 2010; de Sociolinguistica, 2007; Rudra et al., 2016) have explored the language preference of users towards expressing opinions. Our work is different in the sense that, in addition to opinion, we are also interested to explore language preferences while expressing conversational intent. We ask the following questions:

1. *What is the role of Sheng during formal and informal conversations?*
2. *Which language is preferred by the multilingual youth for exchanging various forms of support?*

4.1. Definitions and Formulations

Formally, let $L = \{En, Sh, Sw\}$ be the primary set of languages in which the peer supporters exchange feelings. Different forms of peer support categories, according to the annotation schema (Mondal et al., 2021) are denoted by C which consists of $\{Info, Emo, Chchat, Ack, Gw, Oth\}$ where these denote Informational, Emotional, Chitchat, Acknowledgement, Group Work, and Other form of support respectively. Let $M = \{m_1, m_2, m_3, \dots, m_{|M|}\}$ be the set of messages from the peer supporters in WhatsApp groups. Here,

- 1) $L(M)$, $C(M)$ be the subsets of messages M that respectively contain all messages in language L , and belonging to peer support category C ,
- 2) $LC(M) = L(M) \cap C(M)$.

The preference towards a language L for expressing a type of peer support category C is given by the probability (pr) (according to Bayes' Theorem):

$$pr(L|C; M) = \frac{pr(C|L; M)pr(L; M)}{pr(C; M)} \quad (1)$$

The preference of L for expressing category C , therefore, can be quantified as:

$$pr(C|L; M) = \frac{|CL(M)|}{|L(M)|} \quad (2)$$

However, $pr(L)$, which defines the prior probability of choosing L for a message depends on a large number of socio-linguistic parameters beyond peer support categories. For instance, even though the peer supporters were not asked specifically to interact in a particular language, we found that English has been overwhelmingly more common to express any forms of support in our dataset. Thus, determining the preference of L for expressing a particular category of peer support, therefore can be quantified as the comparative measure of choosing between all the choices of languages which the users are proficient at. For instance, we consider three language choices. We can infer that language $L1$

is more preferred than $L2$ and $L3$ for expressing a category C if:

$$pr(C|L1; M) > pr(C|L2; M) > pr(C|L3; M) \quad (3)$$

The strength of the preference is directly proportional to the ratio of the probabilities: $pr(C|L1; M)/pr(C|L2; M)$ and $pr(C|L1; M)/pr(C|L3; M)$. The former ratio indicates how much likely $L1$ is preferred compared to $L2$ while expressing C ; the latter indicates the likelihood of $L1$ compared to $L3$ for expressing C .

4.2. Hypotheses

We formally define the hypotheses of language preferences by computing message-level likelihood as explained in the previous section.

Hypothesis I: We hypothesize that Kiswahili (*Sw*) is the preferred language for expression of *Chitchat* (*Chchat*) compared to English (*En*). Formally,

$$pr(Chchat|Sw; M) > pr(Chchat|En; M) \quad (4)$$

Hypothesis II: We hypothesize that English (*En*) is the preferred language for expression of *Informational* (*Info*) (Hypothesis IIa) support (similarly for the corresponding subcategories like *Medical Information* (*MInfo*) (Hypothesis IIb), *Lifestyle Information* (*SInfo*) (Hypothesis IIc), *Personal Information* (*PInfo*) (Hypothesis IId) compared to Kiswahili (*Sw*). Formally,

$$pr(Info|En; M) > pr(Info|Sw; M) \quad (5)$$

$$pr(MInfo|En; M) > pr(MInfo|Sw; M) \quad (6)$$

$$pr(SInfo|En; M) > pr(SInfo|Sw; M) \quad (7)$$

$$pr(PInfo|En; M) > pr(PInfo|Sw; M) \quad (8)$$

Hypothesis III: We hypothesize that Kiswahili (*Sw*) is the preferred language for expression of *Empathy* compared to English (*En*). Formally, we expect:

$$pr(Empathy|Sw; M) > pr(Empathy|En; M) \quad (9)$$

Hypothesis IV: Regarding expression of opinion, we hypothesize that Kiswahili (*Sw*) is the preferred language for expression of *Negative* (*Neg*) sentiment compared to English (*En*). Thus, we expect:

$$pr(Neg|Sw; M) > pr(Neg|En; M) \quad (10)$$

We can also compare it using:

$$\frac{pr(Neg|Sw; M)}{pr(Pos|Sw; M)} > \frac{pr(Neg|En; M)}{pr(Pos|En; M)} \quad (11)$$

Hypothesis V: Our hypothesis is Sheng (*Sh*) is more attached with *Chitchat* (*Chchat*) category of Support when compared with *Informational* (*Info*) (Hypothesis Va) and *Group Work* (*GW*) (Hypothesis Vb). Thus

$$pr(Sh|Chchat; M) > pr(Sh|Info; M) \quad (12)$$

$$pr(Sh|Chchat; M) > pr(Sh|GW; M) \quad (13)$$

Hypothesis	Probabilities	Group-1			Group-2		
		Value	Ratio	Validity	Value	Ratio	Validity
Hypothesis I	Num= $pr(Chchat Sw; M)$	0.971	1.08*	True	0.868	1.45*	True
	Den= $pr(Chchat En; M)$	0.899			0.598		
Hypothesis IIa	Num= $pr(Info En; M)$	0.200	9.52**	True	0.325	3.65**	True
	Den= $pr(Info Sw; M)$	0.021			0.089		
Hypothesis IIb	Num= $pr(MInfo En; M)$	0.054	5400	True	0.118	4.72*	True
	Den= $pr(MInfo Sw; M)$	0.00001			0.025		
Hypothesis IIc	Num= $pr(SInfo En; M)$	0.022	3.14	True	0.0209	7.21**	True
	Den= $pr(SInfo Sw; M)$	0.007			0.0029		
Hypothesis IId	Num= $pr(PInfo En; M)$	0.039	3960	True	0.00001	3.76	True
	Den= $pr(PInfo Sw; M)$	0.00001			0.025		
Hypothesis III	Num= $pr(Empathy Sw; M)$	0.019	6.33	True	0.005	1	False
	Den= $pr(Empathy En; M)$	0.003			0.005		
Hypothesis IV	Num= $pr(Neg Sw; M)$	0.174	24.86	True	0.014	1.167	True
	Den= $pr(Neg En; M)$	0.007			0.012		
Hypothesis Va	Num= $pr(Sh Chchat; M)$	0.051	12.75**	True	0.039	24.38**	True
	Den= $pr(Sh Info; M)$	0.004			0.002		
Hypothesis Vb	Num= $pr(Sh Chchat; M)$	0.051	9.107**	True	0.039	9.75**	True
	Den= $pr(Sh GW; M)$	0.006			0.004		

Table 1: Results of Message-Level Hypothesis Testing on Language Preferences. ** and * indicates strong (user-level 2-tailed t -test with p -value < 0.05) and moderate ($0.05 < p$ -value < 0.1) significance respectively, and the rest are not statistically significant. Num, Den denotes Numerator and Denominator respectively.

Note: We consider messages from both the peer supporters and facilitator, and explore the preferences of different peer support categories based on the observed statistics of only the monolingual messages (69% in Group-1 and 64% in Group-2).

4.3. Results and Observations

Table 1 shows the results for testing all the hypotheses that we wish to explore. The observed statistics are fairly consistent across the two groups in terms of message-level testing. Besides, we also determine the statistical significance of these likelihood ratio by considering user-level language preferences. This is performed because each user might have various language preferences towards expressing any sentiment or peer support conversational intent.

In order to model that, we determine the normalized probabilities of each user to determine the likelihood. Then we compute the likelihood ratio per user as explained in the previous section (Eqs 4-13). These likelihoods are then used to determine the statistical significance of each hypothesis using 2-tailed t -test with p -value. Depending on the p -value, we mark the validity of our findings as strong (p -value < 0.05), moderate ($0.05 < p$ -value < 0.1) and non-existent (p -value > 0.1). It can be observed that **Hypothesis I** and **Hypothesis IIb** hold true across both the groups with moderate statistical significance. The findings are mostly valid across both the groups, with a minor exception of **Hypothesis IIb** in Group-1.

Based on our **Hypothesis IIa**, we delve deeper to

understand what form of information exchange is most significantly expressed in English. The results, along with statistical significance is reported in Table 1, indicating that Medical and Lifestyle Information, compared to Personal Information, are significantly exchanged using English instead of Kiswahili.

Interestingly, we observe that empathetic support was not expressed significantly in Kiswahili compared to English in the forums (**Hypothesis III**). Moreover, there is a tendency to use equal amount of English and Kiswahili in Group-2 while expressing empathy. Although prior work (Rudra et al., 2016) has indicated that negative sentiment is expressed more in native language compared to English in Twitter, hereby analyzing language preference towards expressing negative sentiment (**Hypothesis IV**), it holds true but it is not statistically significant.

Hypothesis V testing shows that Sheng is more associated with informal chat (*Chitchat*) compared to formal conversations (*Informational* and *Group Work*). Existing literature (Chen and Cheng, 2020; Barz and Cohen, 2011) corroborates our findings in which Sheng has been argued to be a ‘slang’, primarily used by the Kenyan youth in informal context. Moreover, we found that the native language Kiswahili has been more frequently used in informal conversations whereas English is found more frequently in formal settings.

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	-

Table 2: Average Accommodation of the speakers in Group-1 and Group-2 for the different languages.

5. Accommodation of Language Choice

Language is inevitably at the center stage of identity construction in multilingual contexts where language choices have to be made (Giles et al., 2010). In such context, it is interesting to note the way by which the language choices of the speakers are coordinated and how much does one speaker’s choice of language affect other speakers. This phenomenon is defined as linguistic accommodation and in this section, we wish to study the linguistic patterns used by the peer supporters for accommodating each other’s language choice.

5.1. Measuring Accommodation

We adapt the mathematical formulation presented in (Bawa et al., 2018), for measuring linguistic accommodation. For each language L , F_L is true if some words from L are present in a conversation, otherwise false. F_L is said to exhibit accommodation if the likelihood of a user expressing F_L increases in an utterance (u_i) when F_L has been expressed in the previous dialog u_{i-1} . Thus, accommodation is defined as follows:

$$Acm(F_L) = P(\delta_{u_i}^{F_L} | \delta_{u_{i-1}}^{F_L}) - P(\delta_{u_i}^{F_L}) \quad (14)$$

Accommodation is hence, the difference between the observed rate of language ($P(\delta_{u_i}^{F_L} | \delta_{u_{i-1}}^{F_L})$) choice from its base rate ($P(\delta_{u_i}^{F_L})$). Instead of computing these likelihoods over the entire corpus, we compute them individually for each peer supporter since they might have different base likelihoods. Considering an utterance u from a supporter s , we redefine accommodation as the expectation of individual accommodation over all the supporters (E_s):

$$E_s(Acm(F_L)) = E_s(P(\delta_{u_i}^{\delta_{S(u)=s}} | \delta_{u_{i-1}}^{F_L}) - P(\delta_{u_i}^{\delta_{S(u)=s}})) \quad (15)$$

5.2. Experiments and Observations

As we compute the speaker-wise accommodation rates for each of the three different languages, we take into account the speakers who can speak all the three languages (base rate of all the languages is higher than 0). For our analysis, we also exclude the infrequent speakers, sharing less than 10 messages. Table 2 shows the message wise accommodation rates in Group-1 and Group-2. We found the accommodation effect to be present in both the groups. The speaker-wise accommodation results are in Appendix (Table 5 and 6). For

majority of the speakers, the rate of reciprocation (Observed) is slightly higher than the base rate (Base) for English, and in some cases the difference is not statistically significant, thereby leading to lesser accommodation rates for English. Overall, 6 users from Group-2 and 5 users from Group-1 display negative accommodative effects. Clearly, a higher base rate of expression corresponds to far less accommodation. In other words, the instances of code-choice that are uncommon and unexpected within the conversational context are likely to be accommodated for. However, looking at individual differences in these values for Kiswahili and Sheng reveals interesting patterns. Using the notion of markedness in code-choice (Myers-Scotton, 2005), in which the more salient code is more strongly accommodated for, we identify the marked language from the average accommodation scores for every conversation in both the groups.

The results show that Kiswahili clearly stands out as the marked code-choice from the accommodation scores averaged over all the speakers for Group-2. In order to test for statistical significance, we compute a pairwise t -test (Kim, 2015) between Kiswahili and English. It reveals that Kiswahili is statistically significant compared to English ($t=4.54$ with p -value=0.00063). While testing the significance of accommodation between Sheng and English pair, we found that the results are marginally significant ($t=1.88$ with p -value=0.067). Besides, we also perform a One-way ANOVA test (Ostertagova and Ostertag, 2013) for Group-2 to determine the statistical significance of accommodative effects across the three languages (Note: the t -value is 5.86 and p -value=0.0049). Whereas, in Group-1 we observe that the accommodative effect of Kiswahili is not statistically significant compared to English and Sheng.

Therefore, we computed the joint accommodative effects of Kiswahili and Sheng (Kiswahili being accommodated by Kiswahili or Sheng; Sheng being accommodated by Kiswahili or Sheng) and that of English. A significance test between the accommodation scores of these two language pairs displays a marginal statistical significance result (t -value=1.84 with p -value=0.0762). The difference in the accommodation scores of the two groups can be attributed to the variation in their age groups (adolescents in Group-1 and adults in Group-2). While Sheng is being equally used as Kiswahili by the younger Kenyans, Kiswahili is relatively more used by the older population in comparison to Sheng. Prior literature (Iraki, 2014; Mutiga, 2013; Githiora, 2002) agrees with our observation.

6. Exploring Code-Switching Functions

Code-Switching is motivated by different social, discourse, pragmatic and structural reasons (Hartmann et al., 2018; Begum et al., 2016). In our study, we explore “*Is there a pragmatic motivation (why people code-switch) or structural reason (how people code-switch) behind the phenomenon of switching languages*

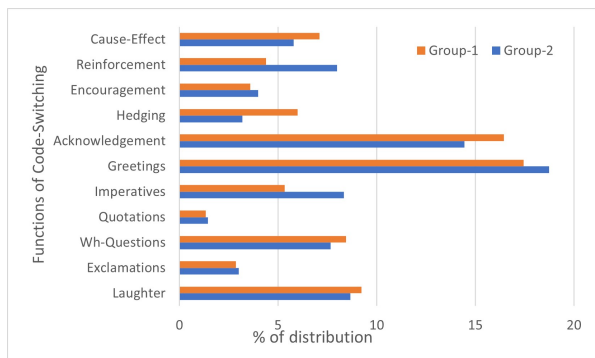


Figure 1: Distribution of the functions of Code-Switching in Group-1 and Group-2.

among the peer supporters?” We look at switching between the language pairs English-Kiswahili and English-Sheng with 99 and 331 code-switched utterances from Group 1 and Group 2, respectively. Table 3 presents the different forms of switching observed (some of which are adapted from (Begum et al., 2016)’s framework) along with examples. Further, we delve deeper to understand the distribution of different switching patterns that occur in these conversations (Figure 1).

6.1. Results and Observations

We observe that interjections or structural tag-switching is the most common CS function for the peer supporters, 60% in Group-1 and 51% in Group-2 (Figure 1). For the frequently occurring tag-switching categories such as Greetings and Acknowledgement, we attempt to understand how much language symmetry is observed at the switch points, i.e., what percentage of English to Kiswahili/Sheng and vice-versa can be observed in such conversations. We observe that English to Kiswahili switching is more common than Kiswahili to English for both the functions. 76% of Acknowledgement switching and 82% of switching in Greeting occurs from English to Kiswahili/Sheng. It is interesting to note that unlike the findings from existing studies on Twitter (Hartmann et al., 2018; Begum et al., 2016), structural tag switching patterns are more common compared to pragmatic switching phenomenon in both the peer support groups. We further observe that *Informational*, *Emotional* and *Group Work* peer support forms contain comparatively higher amount of pragmatic switching compared to those in *Chitchat* and *Acknowledgement* categories. In Group-1, 68.78% of the *Chitchat* and *Acknowledgement* tagged code-switched utterances are due to structural switching, whereas 25.13% of the *Informational*, *Emotional* and *Group Work* messages contain structural switching, majority arises due to pragmatic switching. Nearly similar patterns are observed in Group-2. 52.14% of the *Chitchat* and *Acknowledgement* tagged code-switched utterances are due to structural switching and 24.31% of the *Informational*, *Emotional* and *Group Work* mes-

sages contain structural switching. We speculate that this uneven distribution can be attributed to the need for a group identity marker to demonstrate solidarity/identify with the peer supporters. However, analysis on more conversational data would be required to confirm this.

7. Expressions of Sentiment

In order to assess the emotional behavior of peer supporters, we initially conduct a quantitative study to understand *if there is a correlation of users’ expression of a particular sentiment with their activity in the group?* (Sec 7.1). Then, we qualitatively sub-categorize the messages expressing sentiments into themes (Sec 7.2). Finally, we attempt to build a technological intervention in such scenario (Sec 7.3).

7.1. Sentiment Patterns with User Activity

We hypothesize that the most active users in the groups are more likely to express negative sentiment. Let $\{u_i, m_i\}_{1 \leq i \leq n}$ represent the sequence of users, u_i , and the corresponding number of messages delivered by them, m_i , sorted in descending order by m_i , i.e., $m_i \geq m_j$ whenever $i < j$. Let ns_i and ps_i be the count of negative and positive sentiment messages by u_i . We carry out a 2-tailed t-test to test the hypothesis that for the top- k most active users, u_1 to u_k , the fraction of negative sentiment messages is higher than the positive sentiment messages, i.e., $ns_i > ps_i$. The results are statistically significant at $k = 5$ for Group-1 ($p < 0.001$), but only mildly significant for Group-2 ($p = 0.07$). It could be possible that the users bearing negative outlook are highly active in the group for seeking support.

7.2. Themes of Negative Sentiments

We qualitatively investigate the themes of expressing negative sentiments which include: 1) *Direct Expression* – a group member is ranting, bantering or expressing anger towards the person they are talking to; 2) *Indirect Expression* – a group member is expressing negative emotion about someone else in the group other than the addressee; 3) *Health or Personal Life* – a member is complaining about their personal problems or health issues; and 4) *Less Group Activity* – complaints regarding less interactions and non-intervention of the admin. In Group-1, 74% negative sentiments are due to direct expression of negative sentiment. Whereas in Group-2, 47%, 25% and 17% of negative sentiments are because of health/personal life related complaints, direct expression, and less group activity, respectively.

7.3. LT for Message Triaging

The peer supporters exchange thoughts related to their personal lives and also health-related struggles. An excess of messages with negative sentiment could be detrimental, not only to some users, but also to the overall trust and rapport within the peer-group leading to an ineffective and possibly harmful forum. In this context,

CS Functions	Examples	English Translation
Cause-Effect	“There are some factors that may bring this stress and even the type of your regimen, its good to talk to your doc <i>ndo waweze ku determine shidainaweza tokea wapi.</i> ”	There are some factors that may bring this. Stress and even the type of your regimen, its good to talk to your doctor so they can determine where the problem is
Reinforcement	“How about other guys. This topic am sure touches on most of us here. Please lets say something. <i>Wale wa kuchungulia tu leo tu fungue roho jameni</i> ”	How about other guys. This topic am sure touches on most of us here. Please lets say something. Those who view please let us contribute today.
Encouragement	“Hello...GrXXX here, it seems it is just the 2 of us in the group <i>watu wako wapi?</i> ”	Hello.. GrXXX here, it seems it is just the two of us in the group where are the others?
Hedging	“Guys are silent i gues <i>hawana issues nut hata kama</i> ”	Guys are silent I guess they have no issues now
Acknowledgement	“of coz <i>tunapewa 400</i> ”	Of course we will be given 400
Greetings	“Hy guys, <i>ni hizi drugs ama hii condition?</i> ”	Hey guys, is it the drugs or the condition?
Imperatives	“Please guys <i>mliamua kuwacha hii group bila notice?</i> ”	Please guys did you decided to leave the group without notifying us ?
Quotations	“One of our frnd <i>nilimgrt den akasema ”for wat</i> ”	I greeted one of our friends then they said ”for what”
Wh Questions	“What if <i>akikuambia anataka mzae utafanya aje?</i> ”	what if he says he wants you guys to have a baby,what will you do?
Exclamations	“Oooh no!! <i>reason ya kuleft ni gani?</i> ”	Oooh no!! what are the reasons of leaving?
Laughter	“Hahaha <i>kumbe nyi wawesa kwa wazazi wazuri</i> ”	Hahaha you guys can be good friends

Table 3: Different functions of code-switching annotated in the Group-1 and Group-2 conversations. Code-switching is highlighted by the change in color in the utterances. The pragmatic forms of switching are comprised of Cause-Effect, Reinforcement, Encouragement and Hedging, while the rest are parts of structural switching.

Training Paradigm	Test Set	Models used	Negative			Non-Negative			Macro-F1
			Precision	Recall	F1	Precision	Recall	F1	
Few-shot	OOD English	XLM-R	0.68	0.82	0.74	0.89	0.82	0.85	0.80
Zero-shot	Multilingual	m-BERT	0.05	0.13	0.07	0.95	0.86	0.90	0.48
Zero-shot	Multilingual	XLM-R	0.09	0.19	0.12	0.95	0.87	0.92	0.52
Few-shot	Multilingual	m-BERT	0.10	0.40	0.16	0.92	0.84	0.87	0.51
Few-shot	Multilingual	XLM-R	0.14	0.42	0.21	0.96	0.85	0.90	0.56
Zero-shot	En-Translated	m-BERT	0.12	0.17	0.14	0.96	0.87	0.91	0.56
Zero-shot	En-Translated	XLM-R	0.15	0.20	0.17	0.95	0.89	0.92	0.55
Few-shot	En-Translated	m-BERT	0.16	0.48	0.24	0.93	0.88	0.90	0.57
Few-shot	En-Translated	XLM-R	0.17	0.53	0.26	0.97	0.88	0.93	0.60

Table 4: Zero-shot and Few-shot sentiment analysis results of mBERT and XLM-R on three test sets.

LT can serve as a powerful channel to detect the mental state of the peer supporters by flagging such messages to the concerned moderator in the group, and/or by nudging the user to rephrase the message appropriately. We are interested in exploring how the state-of-the-art LT can be useful for such emergency needs.

In the recent years, the multilingual NLP community is increasingly leveraging the pre-trained Massively Multilingual Models (MMLMs) (Devlin et al., 2019; Lample and Conneau, 2019) for prediction tasks, such

as sentiment analysis on multilingual text. In order to flag the important user utterances to the moderator, we built a binary classifier for sentiment analysis (*negative* and *non-negative*) by making use of both zero-shot and few-shot classification strategies using mBERT (*bert-base-multilingual-cased*) (Devlin et al., 2019) and XLM-R (*xlm-roberta-base*) (Conneau et al., 2020). During the experiments, our train and test set consists of 50-50% split of entire dataset with nearly equal distribution of negative and non-negative classes

in the training set. In order to diversify the training samples, we also augment the training data using 1500 more examples from an out-of-domain dataset¹. We evaluate these models on three test sets: (1) Out-of-Domain (OOD) – an English test set comprising of movie reviews, (2) Multilingual – the original multilingual utterances of the peer supporters, and (3) English-Translated – the corresponding English-translated utterances of the Multilingual set. We wished to understand how difficult it is to leverage such models in this setup. The zero-shot and few-shot sentiment analysis results using class-wise Precision, Recall, F1-score, and macro-F1 are reported in Table 4. It was found that even if a model performs reasonably well on OOD, it struggles to reach a decent F1-score on the English-translated dataset of our conversation corpus. The performance drop is more acute on the multilingual counterpart, thereby pointing to the difficulties of handling low-resource languages. We analyzed and annotated the reasons of misclassification on 293 examples.

Error Analysis: Following are the main reasons:

(1) **Context** (9% of total errors): In addition to the textual content of such messages, the previous messages play a crucial role in determining the sentiment. E.g., *“All you do is walk naked at night”* (Predicted Non-Negative, annotated as Negative (sarcasm)).

(2) **Negation** (14% of total errors): When there is an inherent bias of the model in predicting the instances containing negated expressions as negative sentiment. E.g., *“I suggest that we should try and at least do this for our health let’s not think outside this box about our health but all in all we should know that it’s **not** our mistake being in this condition !”* (Predicted Negative, annotated as Non-Negative).

(3) **Questions** (14% of total errors): There seems to be an inherent bias of the model in predicting the instances containing questions as negative sentiment. E.g., *“Maybe share with us, **why** did you fear them??”* (Predicted Negative, annotated as Non-Negative).

(4) **Contractions** (2% of total errors): The model gets confused during prediction in presence of any contracted expression. E.g., *“**Idk**,,I guess everyone is just chilling”* (Predicted Negative, annotated as Non-Negative).

(5) **Multi-polar expressions** (5% of total errors): In this case, two different sentiments are expressed in a single utterance, but the model predicts only one. E.g., *“And whatever you guys suggest is what I will go with. Let’s continue giving our opinions. By the way I have missed you guys a lot”*.

(6) **Annotation Errors** (16% of total errors): Some of the expressions have annotation-related errors. E.g., *“You love arguing”* (Predicted negative, annotated as Non-Negative).

(7) **Language difficulty** (29% of total errors): When the corresponding English translation is predicted cor-

rectly but the code-mixed utterance is predicted incorrectly.

Overall, sentiment classification in a multi-party conversation corpus is a challenging task, since a variety of errors propagate due to the absence of contextual cues. Besides, 29% of errors are due to language related difficulties, which confirm that low-resource languages when intermingled with English poses serious threat to the predictive capabilities of the state-of-the-art models. Even though the size of the dataset was small, we still observe that the macro-F1 performance achieved by the few-shot XLM-R model on English utterances was greater (49%) than the non-English utterances (42%).

8. Conclusion and Recommendation

In this paper, we conduct linguistic analysis of the language usage patterns of multilingual youth, living with HIV, in two health-focused WhatsApp groups in Kenya. Although few researchers have focused on studying the patterns of Kiswahili usage, we are the first to study a dataset comprising of both Kiswahili and Sheng. From our experiences gathered, we arrive at the following recommendations for the community:

— Majority of the current benchmark datasets on sentiment analysis are non-conversational and monolingual in nature. A real-time healthcare dataset, like ours, is the need of the hour which comes with its unique challenges of handling conversational aspects combined with the difficulty of handling under-resourced languages. Thus, the community should focus more on building such resources and benchmark effective models on these resources for various tasks.

— Based on our observations, Sheng, a popular lingua franca amongst the urban youth in Kenya, plays a crucial role in effective informal support exchange in these peer support groups. Thus, it is pivotal to infuse its vocabularies and support code-mixing in the multilingual models that are used to develop technological interventions for this population.

— In addition to age, it would be interesting to analyze whether the expression of linguistic warmth varies with other demographic factors. Therefore, massive-scale Sheng/Kiswahili data collection efforts need to be made in order to facilitate large-scale studies.

Finally, we aggregate a set of existing language resources for Kiswahili like (De Pauw et al., 2009b; Oirere et al., 2013; Agić and Vulić, 2019; Singh et al., 2019; Piergallini et al., 2016; Masua and Masasi, 2020; De Pauw et al., 2009a; Shikali and Mokhosi, 2020; Martin et al., 2021; Adelani et al., 2021). These might motivate the community to build LT, and take up interest in carrying out linguistic analysis on such under-resourced languages. As a future work, we would like to potentially build resources for providing support to the multilingual community in developing LT for healthcare, which can also be extended to mental health related conversational forums as well.

¹<https://www.kaggle.com/c/twitter-sentiment-analysis2/data>

9. Bibliographical References

- Abdul-Mageed, M., Kuebler, S., and Diab, M. (2012). SAMAR: A system for subjectivity and sentiment analysis of Arabic social media. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 19–28, Jeju, Korea, July. Association for Computational Linguistics.
- Abdulaziz, M. and Osinde, K. (1997). Sheng and english: development of mixed codes among the urban youth in kenya.
- Arunachalam, R. and Sarkar, S. (2013). The new eye of government: Citizen sentiment analysis in social media. In *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP)*, pages 23–28, Nagoya, Japan, October. Asian Federation of Natural Language Processing.
- Balahur, A. (2013). Sentiment analysis in social media texts. In *WASSA@NAACL-HLT*.
- Bali, K., Sharma, J., Choudhury, M., and Vyas, Y. (2014). “I am borrowing ya mixing ?” an analysis of English-Hindi code mixing in Facebook. In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pages 116–126, Doha, Qatar, October. Association for Computational Linguistics.
- Barz, G. and Cohen, J. M. (2011). The culture of aids in africa : hope and healing in music and the arts.
- Bawa, A., Choudhury, M., and Bali, K. (2018). Accommodation of conversational code-choice. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 82–91, Melbourne, Australia, July. Association for Computational Linguistics.
- Begum, R., Bali, K., Choudhury, M., Rudra, K., and Ganguly, N. (2016). Functions of code-switching in tweets: An annotation framework and some initial experiments. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 1644–1650, Portorož, Slovenia, May. European Language Resources Association (ELRA).
- Bhat, I., Bhat, R. A., Shrivastava, M., and Sharma, D. (2018). Universal Dependency parsing for Hindi-English code-switching. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 987–998, New Orleans, Louisiana, June. Association for Computational Linguistics.
- Bhat, K. S., Jain, M., and Kumar, N. (2021). Infrastructuring telehealth in (in)formal patient-doctor contexts. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct.
- Bosire, M. (2006). Hybrid languages: The case of sheng.
- Bosire, B. (2008). Sheng: The phonology, morphology and social profile of an urban vernacular.
- Chen, C. and Cheng, L. (2020). African youth languages: New media, performing arts & sociolinguistic development. *Journal of African Languages and Linguistics*, 41:153 – 157.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online, July. Association for Computational Linguistics.
- Danescu-Niculescu-Mizil, C., Gamon, M., and Dumais, S. T. (2011). Mark my words! linguistic style accommodation in social media. *CoRR*, abs/1105.0673.
- Daniel, J. E., Brink, W., Eloff, R., and Copley, C. (2019). Towards automating healthcare question answering in a noisy multilingual low-resource setting. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 948–953, Florence, Italy, July. Association for Computational Linguistics.
- de Sociolinguistica, E. (2007). Blistering barnacles ! what language do multilinguals swear in ? !
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.
- Dewaele, J.-M. (2010). Emotions in multiple languages.
- Dizayi, S. (2015). The crisis of identity in postcolonial novel.
- Dorantes, A., Sierra, G., Donohue Pérez, T. Y., Bel-Enguix, G., and Jasso Rosales, M. (2018). Sociolinguistic corpus of WhatsApp chats in Spanish among college students. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 1–6, Melbourne, Australia, July. Association for Computational Linguistics.
- Dwivedi, A. (2015). Linguistic realities in kenya: A preliminary survey.
- Fink, T. K. (2005). Attitudes toward languages in nairobi.
- Giles, H., Coupland, J., and Coupland, N. (2010). Contexts of accommodation: Developments in applied sociolinguistics.
- Githinji, P. (2006). Bazes and their shibboleths: Lexical variation and sheng speakers’ identity in nairobi. *Nordic Journal of African Studies*, 15:443–472.
- Githinji, P. (2008). Sexism and (mis)representation of

- women in sheng. *Journal of African Cultural Studies*, 20:15 – 32.
- Githiora, C. (2002). Sheng: Peer language, swahili dialect or emerging creole? *Journal of African Cultural Studies*, 15:159 – 181.
- Githiora, C. (2018). Sheng: the expanding domains of an urban youth vernacular. *Journal of African Cultural Studies*, 30:105 – 120.
- Habernal, I., Ptáček, T., and Steinberger, J. (2013). Sentiment analysis in Czech social media using supervised machine learning. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 65–74, Atlanta, Georgia, June. Association for Computational Linguistics.
- Hartmann, S., Choudhury, M., and Bali, K. (2018). An integrated representation of linguistic and social functions of code-switching. In *LREC*.
- Hosseini, M. and Caragea, C. (2021). It takes two to empathize: One to seek and one to provide. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):13018–13026, May.
- Ideh, A. and Onu, J. (2017). Multilingualism and the new language policy in south africa: Innovation and challenges.
- Iraki, F. K. (2014). Language, memory, heritage and youth: The sheng idiom.
- Jamatia, A., Gambäck, B., and Das, A. (2015). Part-of-speech tagging for code-mixed english-hindi twitter and facebook chat messages. In *RANLP*.
- Kanana, F. E. and Ny’onga, A. C. (2019). Lexical restructuring processes in sheng among the matatu crew in nakuru, kenya. *South African Journal of African Languages*, 39:42 – 55.
- Kariuki, A., Kanana, F. E., and Kebeya, H. (2015). The growth and use of sheng in advertisements in selected businesses in kenya. *Journal of African Cultural Studies*, 27:229 – 246.
- Karusala, N., Seeh, D. O., Mugo, C., Guthrie, B., Moreno, M. A., John-Stewart, G., Inwani, I., Anderson, R., and Ronen, K., (2021). “*That Courage to Encourage*”: Participation and Aspirations in Chat-Based Peer Support for Youth Living with HIV. Association for Computing Machinery, New York, NY, USA.
- Kaviti, L., (2015). *From Stigma to Status: Sheng and English in Kenya’s Linguistic and Literary Space*, volume 46, pages 223–253. 04.
- Kim, T. K. (2015). T test as a parametric statistic. *Korean journal of anesthesiology*, 68(6):540–546, Dec. 26634076[pmid].
- König, K. (2019). Sequential patterns in sms and whatsapp dialogues: Practices for coordinating actions and managing topics. *Discourse & Communication*, 13(6):612–629.
- Lample, G. and Conneau, A. (2019). Cross-lingual language model pretraining. In *NeurIPS*.
- Makewa, L., Mwitwa, M., and Ocharo, G. (2014). Student and educator perceptions of prevalence of use, attitude and views about sheng speakers. *The International Journal of Academic Research in Business and Social Sciences*, 4:265–278.
- Makhija, P., Kumar, A., and Gupta, A. (2020). hinglishnorm - a corpus of hindi-english code mixed sentences for text normalization. In *COLING*.
- Momanyi, C. (2009). The effects of ‘sheng’ in the teaching of kiswahili in kenyan schools. *The Journal of Pan-African Studies*, 2:127.
- Mondal, I., Bali, K., Jain, M., Choudhury, M., Sharma, A., Gitau, E., O’Neill, J., Awori, K., and Gitau, S. (2021). A linguistic annotation framework to study interactions in multilingual healthcare conversational forums. In *Proceedings of The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 66–77, Punta Cana, Dominican Republic, November 11. Association for Computational Linguistics.
- Mutiga, J. (2013). Effects of language spread on a people’ phenomenology: The case of sheng’ in kenya. *Journal of Language, Technology & Entrepreneurship in Africa*, 4:1–15.
- Myers-Scotton, C. (2005). *Multiple voices: An introduction to bilingualism*. Wiley-Blackwell.
- Nassenstein, N. and Hollington, A. (2015). Youth language practices in africa and beyond.
- Nguyen, T. H. and Shirai, K. (2015). Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1354–1364, Beijing, China, July. Association for Computational Linguistics.
- Ogechi, N. (2005). Trilingual codeswitching in kenya - evidence from ekegusii, kiswahili, english and sheng.
- Ostertagova, E. and Ostertag, O. (2013). Methodology and application of one-way anova. *American Journal of Mechanical Engineering*, 1:256–261, 11.
- Pimpale, P. B. and Patel, R. N. (2016). Experiments with POS tagging code-mixed indian social media text. *CoRR*, abs/1610.09799.
- Ramesh, S. H. and Kumar, R. R. (2017). A pos tagger for code mixed indian social media text - icon-2016 nlp tools contest entry from surukam. *ArXiv*, abs/1701.00066.
- Rijhwani, S., Sequiera, R., Choudhury, M., Bali, K., and Maddila, C. S. (2017). Estimating code-switching on Twitter with a novel generalized word-level language detection technique. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1971–1982, Vancouver, Canada, July. Association for Computational Linguistics.

- Rinkanya, A. (2015). Code switching in Kenyan women's literature after 2000 // *matatu* 46.
- Rudra, K., Rijhwani, S., Begum, R., Bali, K., Choudhury, M., and Ganguly, N. (2016). Understanding language preference for expression of opinion and sentiment: What do Hindi-English speakers do on Twitter? In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1131–1141, Austin, Texas, November. Association for Computational Linguistics.
- Sachdev, I. and Giles, H., (2008). *Bilingual Accommodation*, pages 353 – 378. 01.
- Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P., and Bethard, S. (2015). Predicting continued participation in online health forums. In *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, pages 12–20, Lisbon, Portugal, September. Association for Computational Linguistics.
- Sharma, A., Choudhury, M., Althoff, T., and Sharma, A. (2020a). Engagement patterns of peer-to-peer interactions on mental health platforms. In *ICWSM*.
- Sharma, A., Miner, A., Atkins, D., and Althoff, T. (2020b). A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276, Online, November. Association for Computational Linguistics.
- Spernes, K. I. and Ruto-Korir, R. (2021). Multilingualism and curriculum: A study of how multilingual learners in rural Kenya use their languages to express emotions. *International Journal of Educational Development*, 81:102328.
- Sprugnoli, R., Menini, S., Tonelli, S., Oncini, F., and Piras, E. (2018). Creating a WhatsApp dataset to study pre-teen cyberbullying. In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, pages 51–59, Brussels, Belgium, October. Association for Computational Linguistics.
- West, R., Paskov, H. S., Leskovec, J., and Potts, C. (2014). Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2:297–310.
- Xiang, R., Wan, M., Su, Q., Huang, C.-R., and Lu, Q. (2020). Sina Mandarin alphabetical words: a web-driven code-mixing lexical resource. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 833–842, Suzhou, China, December. Association for Computational Linguistics.
- Constantine Lignos and Chester Palen-Michel and Happy Buzaaba and Shruti Rijhwani and Sebastian Ruder and Stephen Mayhew and Israel Abebe Azime and Shamsuddeen Muhammad and Chris Chinenye Emezue and Joyce Nakatumba-Nabende and Perez Ogayo and Anuoluwapo Aremu and Catherine Gitau and Derguene Mbaye and Jesujoba Alabi and Seid Muhie Yimam and Tajuddeen Gwadabe and Ignatius Ezeani and Rubungo Andre Niyongabo and Jonathan Mukiibi and Verrah Otiende and Irero Orife and Davis David and Samba Ngom and Tosin Adewumi and Paul Rayson and Mofetoluwa Adeyemi and Gerald Muriuki and Emmanuel Anebi and Chiamaka Chukwunke and Nkiruka Odu and Eric Peter Wairagala and Samuel Oyerinde and Clemencia Siro and Tobius Saul Bateesa and Temilola Oloyede and Yvonne Wambui and Victor Akinode and Deborah Nabagereka and Maurice Katusiime and Ayodele Awokoya and Mouhamadane MBOUP and Dibora Gebreyohannes and Henok Tilaye and Kelechi Nwaike and Degaga Wolde and Abdoulaye Faye and Blessing Sibanda and Orevaoghene Ahia and Bonaventure F. P. Dossou and Kelechi Ogueji and Thierno Ibrahima DIOP and Abdoulaye Diallo and Adewale Akinfaderin and Tendai Marengereke and Salomey Osei. (2021). *MasakhaNER: Named Entity Recognition for African Languages*.
- Agić, Željko and Vulić, Ivan. (2019). *JW300: A Wide-Coverage Parallel Corpus for Low-Resource Languages*. Association for Computational Linguistics.
- De Pauw, Guy and de Schryver, G and Wagacha, P. (2009a). *A corpus-based survey of four electronic Swahili-English bilingual dictionaries*.
- De Pauw, Guy and Wagacha, Peter W and de Schryver, Gilles-Maurice. (2009b). *The SAWA corpus: a parallel corpus English-Swahili*.
- Gati L. Martin and Medard E. Mswahili and Young-Seob Jeong. (2021). *Sentiment Classification in Swahili Language Using Multilingual BERT*.
- Masua, Bernard and Masasi, Noel. (2020). *Enhancing text pre-processing for Swahili language: Datasets for common Swahili stop-words, slangs and typos with equivalent proper words*. Elsevier.
- Aaron Mogeni Oirere and Ratnadeep R. Deshmukh and Pukhraj Shrishrimal. (2013). *Development of Isolated Numeric Speech Corpus for Swahili Language for Development of Automatic Speech Recognition System*.
- Piergallini, Mario and Shirvani, Rouzbeh and S. Gautam, Gauri and Chouikha, Mohamed. (2016). *Word-Level Language Identification and Predicting Codeswitching Points in Swahili-English Language Data*. Association for Computational Linguistics.
- Shikali, Casper S and Mokhosi, Refuoe. (2020). *Enhancing African low-resource languages: Swahili data for language modelling*. Elsevier.
- Jasdeep Singh and Bryan McCann and Nitish Shirish Keskar and Caiming Xiong and Richard Socher.

10. Language Resource References

David Ifeoluwa Adelani and Jade Abbott and Graham Neubig and Daniel D'souza and Julia Kreutzer and

(2019). *XLDA: Cross-Lingual Data Augmentation for Natural Language Inference and Question Answering*.

Appendix

User	Swahili/Sheng			English		
	Obs	Base	Acm	Obs	Base	Acm
U1'	0.000	0.144	-0.145	0.643	0.816	-0.173
U2'	0.666	0.297	0.369	0.656	0.640	0.015
U3'	0.294	0.175	0.119	0.737	0.859	-0.123
U4'	0.200	0.256	-0.055	0.750	0.884	-0.134
U5'	1.000	0.166	0.833	1.000	0.667	0.333
U6'	0.633	0.338	0.295	0.761	0.486	0.275
U7'	0.242	0.168	0.074	0.792	0.758	0.033
U8'	0.444	0.131	0.313	0.741	0.833	-0.092
U9'	0.423	0.109	0.315	0.838	0.868	-0.030
U10'	0.285	0.218	0.066	0.852	0.666	0.185
U11'	0.36	0.295	0.064	0.619	0.522	0.096
U12'	0.666	0.214	0.452	1.000	0.571	0.428
U13'	0.613	0.439	0.1737	0.75	0.398	0.352
Adm	0.245	0.126	0.119	0.637	0.809	-0.172
Avg			0.207			0.054

Table 5: Group-1 Speaker-Level Accommodation.

User ID	Swahili			English			Sheng		
	Observed	Base	Acm	Observed	Base	Acm	Observed	Base	Acm
U1	0.737	0.468	0.269	0.755	0.708	0.047	0.316	0.234	0.082
U2	0.564	0.506	0.058	0.728	0.646	0.082	0.144	0.120	0.024
U3	0.681	0.498	0.183	0.774	0.555	0.219	0.384	0.078	0.306
U4	0.711	0.484	0.227	0.734	0.578	0.156	0.275	0.108	0.167
U5	0.621	0.643	-0.022	0.763	0.667	0.096	0.250	0.297	-0.047
U6	0.625	0.097	0.528	0.782	0.851	-0.069	0.800	0.029	0.771
U7	0.578	0.480	0.098	0.787	0.572	0.215	0.261	0.197	0.064
U8	0.733	0.101	0.632	0.770	0.804	-0.034	0.286	0.054	0.232
U9	0.606	0.452	0.154	0.708	0.667	0.041	0.273	0.143	0.130
U10	0.583	0.325	0.258	0.723	0.724	-0.001	0.133	0.122	0.011
U11	0.569	0.482	0.087	0.730	0.850	-0.120	0.173	0.202	-0.029
U12	0.542	0.408	0.134	0.750	0.693	0.057	0.053	0.214	-0.161
U13	0.666	0.326	0.340	0.894	0.750	0.144	0.500	0.184	0.316
U14	0.796	0.190	0.606	0.733	0.612	0.121	1.000	0.064	0.936
U15	0.599	0.200	0.399	0.800	0.857	-0.057	0.000	0.028	-0.028
U16	0.750	0.555	0.195	0.750	0.889	-0.139	0.166	0.315	-0.149
U17	0.529	0.351	0.178	0.750	0.631	0.119	0.000	0.070	-0.070
Admin	0.510	0.16	0.35	0.689	0.835	-0.146	0.400	0.084	0.316
Average			0.265			0.031			0.168

Table 6: Group-2 Speaker-Level Accommodation.