**Annotation Guidelines for Project EPOCH Conversations**

**Microsoft Research Labs India Pvt. Ltd. Bangalore**

**Background:**

Patient-centered care and good communication between patients and healthcare providers has been shown to improve medical adherence. However, it is costly and difficult to provide, especially in Global Health settings where patient volumes are high and costs must be kept low. As chat tools such as WhatsApp and WeChat are becoming widely adopted, provider-patient chat offers exciting opportunities for patient outreach. In particular, we see an increase in chat groups of patient peers moderated by a medical professional as a way of extending patient support, improving medical adherence and so on. Whilst these groups are promising in terms of patient support, they can be burdensome for already overworked medical providers. Using chat to support patient care is a growing trend in Global Health (e.g. groups for HIV positive youth (Kenya), maternal and infant health (India), TB patients (Africa)). However, all parties are likely to face similar challenges. Mental health forum is another area where peer support plays a vital role.

**Project EPOCh dataset and goals:**

In Project EPOCh, we are exploring the possibility of building and deploying NLP and ML solutions to help facilitators manage chat streams, and eventually support medical adherence. In collaboration with a University of Washington team led by Keshet Ronen, we are conducting a qualitative study on two WhatsApp group chats from the Vijana-SMART study. These groups provided support for youth living with HIV in Kenya. The groups contained 27 and 28 members and lasted for 6 months. All messages from youth and their facilitator were downloaded, de-identified, translated into English, and are included in the dataset. The aim of this analysis is to categorize the types of messages, the patterns of sending and responses, and to outline which NLP and potentially other cognitive services might be appropriate to apply to this data set.

**Annotation Steps**

We would like to categorize the messages in the conversation chats from three different perspectives:

1. Annotate the categories of messages
2. Annotate the sentiment of each message
3. Annotate the language of each word in a message.

**Annotation of the categories:**

We require the annotation of the conversation chats to be carried out in two steps. Initially, we would like to categorize the different utterances into six main categories, then further subcategorize them, which is further described in Table 1.

**Step 1:**

This is the initial step of identifying the broader categories of utterances or messages. The broader categories include:

1. Informational
2. Emotional
3. Group Work
4. General chit-chat
5. Acknowledgement
6. Others

**Defining each of the categories of the utterances:**

**1. Informational:**

**Definition :**

In this category, the group members discuss and provide certain factual information, advice or warning.

There are two types of attributes for this category:

1. **Seek** : In this category, the user asks for any information or advice from their fellow group members.

**Values**: Yes / No

1. **Provide** : In this category, the user provides any information or advice to their fellow group members.

**Values**: Yes / No

**Examples:**

1. “*For any infection, a certain minimum amount of virus or bacteria is required for the infection to be transmitted. By taking ART regularly, the amount of the virus remains below that infective dose. Because of this, the virus is not transmitted. This is true for mother to child transmission, as well as sexual transmission of HIV.*” **(Attribute : Seek: No, Provide: Yes)**
2. “*TB is A potentially serious infectious bacterial disease that mainly affects the lungs.*” **(Attribute : Seek: No, Provide: Yes)**
3. “*Most people infected with the bacteria that cause tuberculosis don't have symptoms. When symptoms do occur, they usually include a cough (sometimes blood-tinged), weight loss, night sweats and fever.*” **(Attribute : Seek: No, Provide: Yes)**
4. “*HIV weakens the immune system, increasing the risk of TB in people with HIV. Infection with both HIV and TB is called HIV/TB coinfection*” **(Attribute : Seek: No, Provide: Yes)**
5. “*Hi....This is a big warning...my mother in law has just lost over Kes. 70,000 to M-Shwari and KCB-MPESA con-women...Please take note that there criminals who are going round pretending to give out free Airtel and Orange lines....in the process of registering your new free line they take all your details I.e. ID No, Date of Birth etc... After that they immediately send you some little money in your mpesa line. Thereafter they immediately contact Safaricom pretending that they have forgotten their mpesa password...Some of the questions asked by Safaricom include: your ID no, your date of birth and your last transaction-easily quoting what they have just send you...with these details Safaricom gives them a new mpesa password...and with urgency and within minutes, the crooks use your line to get maximum allowable loans from all mobile phone based creditors…*” **(Attribute : Seek: No, Provide: Yes)**
6. **“E***xplain further about circumcision***” (Attribute : Seek: Yes, Provide: No)**

**2. Emotional:**

**Definition** :

In this particular category, the group members seek emotional support or communicate empathy, love or concern towards their fellow members.

**Examples:**

1. “*I know it's hard to take in the pain of losing your child 😭...but then, brave on💪🏽....This...This is not the only child you would have had 😊”*
2. “*Please never give up on yourself, I know you will survive*”.

**3. Group Work**

**Definition** :

A response that suggests that the users encourage others to remain responsive and promote universalization/togetherness for support.

**Examples:**

1. **“***Let's bring topics even on the contemporary experiences so that we can continue keeping the group more active as it should be guys!***”**
2. *“Together we shall overcome any fear”.*
3. *“It's the weekend. Let's start stories ”.*
4. **“***Guys are silent I guess they have no issues now, but we should still keep in touch”*

**4. General Chit-Chat :**

**Definition** :

In this category, the group members introduce or greet each other and promote social interactions to demonstrate friendliness.

**Examples:**

1. *“Hello, good morning”*
2. *“Happy easter guys”*
3. *“Who is calling me for easter”*
4. *“Hello, cool sunday pips”*
5. *“How was your meal guys?”*
6. *“Are you a boy or a girl”*

**5. Acknowledgement :**

**Definition** :

A response that is meant to acknowledge other's proposal or politely expressing gratitude towards someone's actions.

**Examples:**

1. **“***Thanks I'll attend the meeting*”.
2. “*Yes, I think so”*
3. *“Exactly, same here.”*
4. *“Hey, appreciate for your efforts”.*

**6. Others :**

**Definition** :

A response which does not signify any important message, might contain an emoji or any system generated message.

**Examples:**

1. “*Waiting for this message*”
2. 😂😂😂
3. 👍
4. “🍚🥦🍉🍉🍌”

**Fine-grained Sub-Categories:**

|  |  |  |
| --- | --- | --- |
| **Categories** | **Subcategories (Acronyms)** | **Attributes** |
| Informational | 1. Medical information/advice 2. Personal Life information/advice 3. Healthy Lifestyle Information/advice 4. Admin Related Information/advice 5. Other Information/advice | Seek/Provide |
| Emotional | 1. Empathy 2. Inculcating Hope 3. Expressing negative feelings 4. Expressing happiness |  |
| Group Work |  |  |
| General Chit-chat | 1. Social Interactions 2. Greetings |  |
| Acknowledgement |  |  |
| Others |  |  |

**Table 1: Categories of the Chat messages in the group and subcategories**

**The annotation scheme is described below along with the definitions of categories and subcategories with suitable examples to demonstrate:**

1. **Informational:**

**The following types of codes will fall into this category:**

1. **Medical information/advice:**

**Definition :**

This type of category deals with providing/seeking any factual information, advice or warning about some medicine, treatment or disease.

**Examples:**

1. *“Most people infected with the bacteria that cause tuberculosis don't have symptoms. When symptoms do occur, they usually include a cough (sometimes blood-tinged), weight loss, night sweats and fever.”* ***(Type: Provide)***
2. *“TB is A potentially serious infectious bacterial disease that mainly affects the lungs.”* ***(Type: Provide)***
3. *“Transmission of HIV is dependent on many factors including the viral load of the HIV positive partner, the amount of body fluid exchanged between the partners, and any injuries, sores or ulcers on the genital regions of either partner. Transmission rates for HIV are estimated to be between 0.03% to 0.3% for heterosexual vaginal intercourse.”* ***(Type: Provide)***
4. *“By taking ART regularly, the amount of the virus remains below that infective dose. Because of this, the virus is not transmitted. This is true for mother to child transmission, as well as sexual transmission of HIV.”* ***(Type: Provide)***
5. *“Peripheral Neuropathy – Pain due to nerve damage, mostly in the feet and hands. It may be described as numbness, tingling, or burning.”* ***(Type: Provide)***
6. **Personal Life information/advice:**

**Definition :**

This type of category deals with providing/seeking any factual information, advice or warning about their relationship with partner or other members of the family.

**Examples**:

1. *“You should consult your doctor or HIV counsellor and have a detailed discussion with them regarding the same. You should also make it a point to use condoms during future sexual intercourse.”.* ***(Type: Provide)***
2. *“As we get older and have sexual relationships, it's important to protect ourselves and our partners from STIs and unwanted pregnancy. Barrier methods like condoms prevent STIs and pregnancy. There are many other methods to prevent pregnancy (they don't prevent STIs) https://www.avert.org/sites/default/files/Options%20for%20contraception%20-%20Worksheet.pdf. Also, for HIV, if we are not undetectable, our negative partner can take PrEP to protect themselves. https://twitter.com/nimejiprep?lang=en. ’’* ***(Type: Provide)***
3. **Healthy Lifestyle Information/advice :**

**Definition :**

This type of category deals with providing/seeking any factual information, advice, warning about their daily lifestyle habits.

**Examples**:

1. *“2 eggs Ksh. 20, a banana Ksh. 5, tomato Ksh. 5. Does anyone have other tips on getting enough food before taking medications? .”* ***(Type: Seek)***
2. *“You can take a lot of soup,turtle beans and fish. These work well for new moms. Don't drink coffee or tea. Use only cocoa.* ***(Type: Provide)***
3. *“I need your help , am trying to stop eating stones but I can't, at first I knew it's because of the pregnancy. But after delivery my child is almost one year and I still have stone cravings!!!”* ***(Type: Seek)***
4. *“Alcohol may lower your level of immunity making other diseases like vomiting, diarrhea etc to attack you.”* ***(Type: Provide)***
5. *“Smoking, either passively or actively is dangerous especially to us here. It raises your risk of heart disease, cancer, serious lung diseases i.e. pneumonia etc. So it's not good at all.”* ***(Type: Provide)***
6. **Admin-Related Information:**

**Definition :**

This type of message deals with providing/ seeking any factual information, advice, warning about formal meetings with the admin or purpose of the group.

**Examples**:

* 1. *a. Group is private, we shall not share our messages with anyone else not in the group*

*b. We shall respect all members and not use words like HIV, ART,CD4,*  *VL. So that people feel comfortable.*

*c. Make sure you have security code or applock in your phone*

*d. We Will be sharing challenges and good stories here*

*e. We will Be thoughtful of our messages and responses to one another*  *or pics we send*

*f. I will be sending our Weekly messages every Sunday at 10pm, but nyinyi ninaendelea to na conversation zenyu*

*g. The group proposed they will be calling me Admin. 👍🏽*

***(Type: Provide)***

1. **Other Information:**

**Definition :**

This type of category deals with providing/seeking any factual information, advice, warning about topics excluding medical, personal life, healthy lifestyle or admin related information.

**Examples:**

1. *“Hi....This is a big warning...my mother in law has just lost over Kes. 70,000 to M-Shwari and KCB-MPESA con-women...Please take note that there criminals who are going round pretending to give out free Airtel and Orange lines....in the process of registering your new free line they take all your details I.e. ID No, Date of Birth etc...After that they immediately send you some little money in your mpesa line. Thereafter they immediately contact Safaricom pretending that they have forgotten their mpesa password...Some of the questions asked by Safaricom include: your ID no, your date of birth and your last transaction-easily quoting what they have just send you...with these details Safaricom gives them a new mpesa password...and with urgency and within minutes, the crooks use your line to get maximum allowable loans from all mobile phone based creditors…***” *(Type: Provide)***
2. **Emotional:**
3. **Empathy:**

**Definition** :

A response that explicitly includes a mention of the difficulty experienced by the seeker in a way that is validating his/her distress or indicates an attempt to see things from the seeker’s perspective.

**Examples**:

* 1. *“I can relate to you at so many levels”*
  2. “*Yeah. but i understand we are dealing with the same issues in this regard*”
  3. “*Same here also and another problem is that he got me pregnan*t”

1. **Hopeful:**

**Definition** :

This category is defined when the responder encourages continued efforts/persistence and inculcates hope.

**Examples**:

* 1. *“Sorry for this dear. Like 5002 has said some things need sacrifice and time. Never give up .Just try and overcome the hurdle, all will be fine .”*
  2. *“Please try, I know you will survive”*

**3. Positivity:**

**Definition** :

In this category, the group member is narrating some good incident in a happy mode or showing some excitement towards anything.

**Examples**:

1. “*Early this month, we were blessed with a baby girl and we thank God for His mercies on us. This makes us more strong and urges us to work more on how we shall live tomorrow*”
2. “*She is turning 4 months by 1st Aug and we are so happy*”

**4. Negativity:**

**Definition** :

In this category, the group member is narrating some bad incident in a mournful mood or expressing anger/sorrow/frustration or any negative emotion towards anything.

**Examples**:

1. “ *The water was ice cold , I felt it on my spinal cord until to the liver and twelve other organs failed.😥😤. I was scared that I found myself sitting on the toilet washing my foot and wiped the soap off my body with a towel . I cant die of cold then my body to be found naked”*
2. *“Hello guys am feeling so bad today i have a headache and i also don’t have energy please pray for me”*

**C. Group Work:**

**Definition** :

In this category, the group members promote togetherness to demonstrate team spirit, encouraging each other to be responsive and discuss regarding the official meetings with the group admin.

**Examples:**

1. “*We are stronger together.*.”
2. “*I know people are also busy with job, school etc. People are facing problems, you are not alone, We are strong together...*.”
3. *“Let's bring topics even on contemporary experiences so that we can continue keeping the group more active as it should be guys! Together we shall overcome any fear!”*

**D. Chit-chat :**

**Definition** :

In this category, the group members introduce or greet each other and talk about personal interactions to promote friendliness.

**The following types of codes will fall into this category:**

1. **Social Interactions:**

**Definition:**

In this category, the group members try to know their peers more by chatting express willingness to personally introduce themselves or arrange for personal meetings.

**Examples**:

1. *“Hey guys.. I have missed you a lot!!”*
2. *“Where do you stay?”*
3. *“I stay in Njiru”*
4. *“M from home with honey. How many are in for it??😊😊”*
5. *“Hey, welcome for lunch today”*
6. *“Come, supper is ready*😊

**2. Greetings:**

**Definition :**

It is an act of communication in which human beings intentionally make their presence known to each other, to show attention to, and to suggest a type of relationship or social status between individuals or groups of people coming in contact with each other.

**Examples:**

1. *Good morning all*
2. *Good night*
3. *Happy Easter*
4. *Good morning guys*
5. *Hello everyone*

**E. Acknowledgement :**

**Definition :** Responses used to affirm, acknowledge and/or backchannel communication, are tagged under this category.

**Examples**:

1. “Thanks..you too sleep well”.

2. “Okay, sure”

**Acronyms tables of each message categories which are needed for annotation:**

|  |  |
| --- | --- |
| **Message Categories-Subcategories** | **Acrnoyms** |
| **Informational- Medical Information/advice** | **IMI** |
| **Informational- Personal Life Information/advice** | **IPI** |
| **Informational- Healthy Lifestyle Information/advice** | **IHLI** |
| **Informational- Admin Related Information/advice** | **IAI** |
| **Informational- Other Information/advice** | **IOI** |
| **Emotional- Empathy** | **EE** |
| **Emotional- Inculcating Hope** | **EIH** |
| **Emotional- Expressing Happiness** | **EEH** |
| **Emotional- Expressing Negative Feeling** | **ENF** |
| **Group Work** | **GW** |
| **Acknowledgement** | **ACK** |
| **General ChitChat – Social Interactions** | **GCSI** |
| **General ChitChat – Greetings** | **GCG** |
| **Others** | **OTH** |

**Some Notable Pointers:**

There might be multiple categories inside a single message. For example:

1. E*xactly, and today was a good day for me. : ACK, IOI*

In this particular case, we assign two different category labels to a single response of the user. In the first span, we assign an acknowledgement and then the user provides an information in return. So the categories have been color-coded as well.

1. Usually, the emojis/emoticons are non-textual units displaying an emotion. Therefore, when we observe a response from the user, we usually categorize them based on the textual forms, and not based on the emojis.

**For example:** “*My house was flooded, I couldn’t come* 😢’

This will be categorized as an informational type of message (**IOI**), and not into (**ENF**) based on the emoji 😢

1. The annotators are required to assign one or more category labels to each single sentence spoken by the user, and also assign a remark for the same, like, why they have chosen that category over the other categories. Moreover, they might need to label the textual span as well, if there are multiple categories (as shown in example 1).
2. The annotators need to give priority to certain categories while labelling. For instance, the ‘medical information’ category needs to be necessarily labelled even if the message looks like being categorized into some other set. For example: “*Hello @5001?, I suppose you know why you are taking the drugs 🙂... They are there to help you, I guess you've got a bad impression about drugs(medicine), [Not necessarily this, but any other drugs that menu people tend to detest❌]....”* this message needs to be definitely labelled as **‘Providing medical information’ (IMI-P).**

Similarly for the message : “*Lost of appetite is not good, see a doctor or consult the admin for information or the reasons behind that*”, the category should necessarily be labelled as ‘**‘Providing medical information’ (IMI-P).**

1. Combine the incomplete sentences if coming from the same user, join them and then annotate.

1. The sentences which require more understanding of the context, the annotator might look at one or two utterances above and below to make more sense of that.

**New Topic Identification:**

There will be a checkbox in the annotation tool to opt for in order to signify that the group is talking in a different topic than those of the previous conversations, more formally topic-switching. For instance, if the group was initially talking about something related to admin meetings, and then if someone starts a topic on seeking health-related information, then that could signify a topic change.

**Example 1:**

1.Initially:

|  |  |  |  |
| --- | --- | --- | --- |
| 03/04/2019 | 22:24 | 5067 | mmelala |
| 03/04/2019 | 22:25 | 5003 | Same |
| 03/04/2019 | 22:28 | 5020 | Mimi bado kuna mazuri😉😉 |
| 03/04/2019 | 22:28 | 5067 | mapema aje |
| 03/04/2019 | 22:30 | 5067 | akuna labda wewe useme mazuri |
| 03/04/2019 | 22:30 | 5020 | Karibu sapa |
| 03/04/2019 | 22:31 | 5067 | This message was deleted |
| 03/04/2019 | 22:32 | 5067 | santi mh nko full |
| 03/04/2019 | 22:32 | 5020 | Sawa mresh |
| 03/04/2019 | 22:33 | 5067 | ok |
| 04/04/2019 | 0:04 | 5042 | Ndio nalala |
| 04/04/2019 | 0:05 | 5039 | Ooh Gud9t |
| 04/04/2019 | 0:12 | 5042 | Sem |
| 04/04/2019 | 8:07 | 5067 | morning |
| 04/04/2019 | 8:14 | 5017 | Morning ✌🏽 |
| 04/04/2019 | 8:27 | 5041 | How are you guys?? |
| 04/04/2019 | 8:33 | 5019 | Good |
| 04/04/2019 | 8:34 | 5017 | Awesome💯 |
| 04/04/2019 | 8:35 | 5019 | Yas |
| 04/04/2019 | 8:53 | 5067 | good |
| 04/04/2019 | 8:55 | 5067 | mmenyamaza aje c mlete story |
| 04/04/2019 | 8:56 | 5019 | Story mapema hii |
| 04/04/2019 | 9:00 | 5067 | Kwani story inaletwa sangapi |
| 04/04/2019 | 9:09 | 5019 | Ooh my goodness! |
| 04/04/2019 | 9:11 | 5041 | Jessos |
| 04/04/2019 | 12:34 | 5003 | Great |
| 04/04/2019 | 12:34 | 5067 | good |
| 04/04/2019 | 12:35 | 5003 | Yeah am good |

**Target Message:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 04/04/2019 | 12:48 | 5003 | Am name is MaXXX am 19yrs and right now am stressed out I don't even understand what the right thing to do. I lost my son this year on feb13 until now I feel like I killed my own baby. I see my baby smile everyday I think I'm going crazy.maybe it was my fault or my family's fault I don't know.I was born positive but nobody cared to tell me about it..i came to know about it when I was 6mnts pregnant it was too late so I gave birth to a positive baby I failed to give him medication on time.since day one my baby was always sick but looking really cute and healthy little did I know that he was just passing by. |  |

Previously, there was mostly greetings and other kinds of chit-chat among the peer-supporters of the group until a person starts speaking about his/her difficulties surviving the situation, this kind of notable topic change should be marked, otherwise greetings and other types of chit-chat need not be flagged.

**Example 2:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 11:23 | Vijana-SMART 1 | Hey GXXX, did you get help?? |  | Hey GXXX, did you get help?? |
| 19:39 | 5061 changed their phone number to a new number. Tap to message or add the new number. |  |  |  |
| 16:42 | 5030 | Yes. I am well 💊💊💊💊 |  | Yes. I am well 💊💊💊💊 |
| 16:59 | Vijana-SMART 1 | That's good. Thank you |  | That's good. Thank you |
| 13:25 | 5036 | How are you |  | How are you |
| 14:53 | Vijana-SMART 1 | How are guys... |  | How are guys... |
| 17:57 | 5032 | How is ur mid break? |  | How is ur mid break? |
| 18:57 | 5030 | Hi this is GrXXX, I 've missed all of you. I have alot of homework but we will going for our midterm soon 🙋🏽‍♂👍 |  | Hi this is GrXXX, I 've missed all of you. I have alot of homework but we will going for our midterm soon 🙋🏽‍♂👍 |
| 19:13 | Vijana-SMART 1 | Nycc |  | Nice |
| 19:41 | 5036 | That's good |  | That's good |

**Target Message:**

|  |  |
| --- | --- |
| 5055 | Hello good pple, kindly be alert, if you are using the year of your birth as M- pesa pin kindly change it ,my friend has being conned 45,598/= this evening. There are people out here pretending to be calling from Huduma Number Headquarters (Nairobi) that your Huduma Number is out, they call you by your full Names and ask to confirm if you are the legitimate owner of the phone. Later they ask for the year of birth!! they use an Airtel line to call as someone Next to him operates the safaricom phone line,... Please don't be conned.. share widely with others,. Thanks and Have ablessed Sunday |

**Sentiment Analysis Annotation Framework:**

Sentiment is a view or opinion that is held or expressed by an individual. In this setup, we need to understand the sentiment of each message in the chat to get a deeper sense of understanding the patterns by which the peer supporters interact with each other in the HIV positive groups in Africa.

There are three possible sentiments that we would like to annotate:

**1. Negative Sentiment**

**2. Neutral Sentiment**

**3. Positive Sentiment**

**A . Negative Sentiment** :

It expresses some sort of negative feeling or view or opinion about someone or something.

**Example**:

**a**. “*You have a mental problem it's no secret.”*

This expresses some sort of anger/rudeness against someone, which clearly expresses a negative view or opinion.

**b. “***You should try going offline”*

This also expresses some sort of anger/rudeness against someone, asking him/her to not disturb anymore, which clearly expresses a negative view or opinion.

**c. “***Hi admin, for the past few days have been feeling low and oppressed I want to quit the drugs because they do me harm than good, am loosing appetite, I don't feel salt anymore. My nails are changing to white. I am tired* !!!”

This expresses a sad feeling of the speaker displaying unwellness, therefore needs to be classified as “negative” sentiment.

**B. Neutral Sentiment:**

It neither expresses a positive nor a negative sentiment of the speaker, basically a neutral feeling, it could be a general comment, acknowledgement, chitchat or any factual advice or a simple greeting.

**Example**:

**a**. “*VaXXX said that then went missing...But she said it's her friends experience..”*

This expresses a neutral comment.

**b. “***When are you coming?”*

This also expresses some neutral comment/question.

**c.**” *We are good*”

This also expresses some neutral comment

d.” *Good morning👋🏽👋🏽*”

Though this expresses somewhat positive attitude of the speaker, but since this is simply a greeting, it should fall under the category of neutral sentiment.

e. ” *You can take a lot of soup, turtle beans and fish. These work well for new moms. Don't drink coffee or tea. Use only cocoa”*

This is simply an advice which does not bear any sentiment of user.

**C. Positive Sentiment:**

The sentiment needs to be classified as positive if the speaker feels strong and positive at any particular utterance, except the normal aspects such as any form of greetings.

**Examples**:

1. “*I'm much happy to interact and share with you guys!”*

*2. “Oh same here its really healthy and healthy”*

*3. “I have been great am just chilling..wassup”*

*4. “Hi guys I used to fear bt now days I don't fear my mum now ,DADY and one friend*"

**Language ID Annotation Framework:**

To understand the interactions from the perspective of code-mixing, we need to annotate the language used by the group members on a word-level basis. For each one of the messages, we would like to annotate the language of each word (word-level) language id annotation. For example:

**Original Message**: Mimi nilichukua my siz tukalost

**Language** (SW) (SW) (En) (En) (CM)

**Translation**: I took my sister and we got lost

We are considering five types of language identification codes:

1) **En**: Stands for English

2) **Sw** : Stands for Swahili

3) **Sh**: Stands for Sheng

4) **Code-mixed** : For code-mixed languages.

5) **Oth**: For other languages.

For example: if you are not sure or you don’t know what language vocabulary does the current word fall into, or the word is some named entity or there is some other reason why the word doesn’t really fall into these 4 main categories.