

# Assignment 3

Language and Society

September 24, 2021

**Deadline: 11.59 PM October 4, 2021**

## 1 Bilingualism - Field Work

Observe at least 4 bilingual (or multilingual) people around you and record how they interchangeably use the languages when following parameters vary:

1. People
2. Situation
3. Topic

For each of the candidate, mention the languages they speak and observe them for at least 2 hours. These two hours may or may not be contiguous.

## 2 Code Mixing and Code Switching

Scrap 100 tweets using the Twitter API where each tweet contains bilingual (English + Hindi) text that has at least a 60:40 ratio of the two languages. If you are not comfortable with Hindi, you can choose any other Indian Language. For each tweet, annotate the words with language i.e. each word of the tweet should be assigned a language tag (ENG or HIN). For example:

mera	dog	meri	family	ka	part	hai
HIN	ENG	HIN	ENG	HIN	ENG	HIN

Next, mark the span of the different linguistic elements like words, clauses, sentences etc. in the tweet. Analyse and explain what linguistic elements are normally mixed together.

### 2.1 Hints

1. You can use C++ or Python to scarp the tweets.
2. You need to apply for a developer account at <https://developer.twitter.com/en/apply-for-access> using your twitter account. Once you have a developer account, you will have to create a new App so that you can generate an Access token which will be used to scrap tweets
3. You can scrap tweets by making API calls using your terminal (curl command) or Postman. You can also use libraries like tweepy to scrap the tweets easily.
4. A tutorial for scraping tweets using python: Medium article

### 3 General Instructions

1. Each question carries **10** Marks
2. Plagiarism of any sort will not be tolerated and would result in strict penalties.
3. Question 1 can be submitted hand written or typed.
4. The title of your submission file should be 'Name\_RollNumber.pdf'