

Language and Society

Assignment 5

Name: Harinie Sivaramasethu

Roll number: 2021114008

Language: Tamil

Abbreviations used:

Matrix Language -> ML

Embedded Language -> EL

Code-switching -> CS

Dataset:

I have used code-mixed sentences of languages Tamil-English from datasets given in the below link. It has taken comments of youtube videos of popular Tamil movie trailers.

<https://github.com/bharathichezhiyan/DravidianCodeMix-Dataset>

In particular, you can find the specific dataset I've used by accessing:

DravidianCodeMix-2020.zip > tamil_sentiment_full_train.csv

Code-switching:

Code-switching is defined as the alternation of two languages within a single discourse, sentence, or constituent (Grosjean, 1982; Poplack, 2000; Clyne, 2000). Grosjean (1995) defines CS as shifting completely (emphasis added) to the other language for a word, a phrase, a sentence, etc.

Matrix/ Embedded language:

	Matrix	Embedded
1	Dominant language	Inserted elements/ contributing language
2	Its grammar sets the morphosyntactic frame	
3	First language of the speaker or the language in which the morphemes or words are more frequently used in speech	
4	Determines the syntax of a code-switching instance and its presence is obligatory in the instance.	EL in the CS instance can be one or more and is the code of a lesser degree of contribution in the CS instance.

50 Code-mixed Sentences:

We shall take sentences that are already romanized. We will also be ignoring the end tags (positive, negative, not-tamil, unknown_state) while analysing.

I shall use red to denote English and blue for Tamil.

1. **First like button vijay setupati fans.**

ML: Tamil

EL: English

CS percent: 66.67%

2. **Jayam Ravi thalaivar kitta bagiranga mannippu kekkalanna indha padatha purakkanippom Twitter hashtag pathingala thalaivar fans power Ra Boycot comali**

ML: Tamil

EL: English

CS percent: 26.31%

3. **Semma trailer super surya fans like**

ML: Tamil

EL: English

CS percent: 66.67%

4. **mammokka hero da always will be a hero**
ML: English EL: Tamil CS percent: 25%
5. **mokka bhairava teaser ela but naa suriya fan taan**
ML: Tamil EL: English CS percent: 55.5%
6. **Surya Ku love seen irukadhu polaiyee**
ML: Tamil EL: English CS percent: 33.33%
7. **Marana Massssss..... Love you Thalaiva**
ML: English EL: Tamil CS percent: 60%
8. **Tevidiya pasanga naxalite ku support panranunga**
ML: Tamil EL: English CS percent: 16.66%
9. **Anbu ezhuchi ithula thodanguthu da Marana waiting**
ML: Tamil EL: English CS percent: 14.28%
10. **0.42 bgm vera level gvp.....Vetri maaran no words**
ML: Tamil EL: English CS percent: 55.5%
11. **ajith style ke and avaroda hard work ke 5 times paakalam**
ML: Tamil EL: English CS percent: 54.54%
12. **Vera vera vera level thala top takker**
ML: Tamil EL: English CS percent: 28.57%
13. **Twitter pathutu nera inga vanthavangala like my comment**
ML: Tamil EL: English CS percent: 50%
14. **k harsha sry but antha rajesh punda avanku ena payriya pundanu naynapa thayoli**
ML: Tamil EL: English CS percent: 15.38%
15. **Yaru Yaru September la parka varinga oru like podunga**
ML: Tamil EL: English CS percent: 22.22%
16. **Thalaivaaaaaaaaa yangal thanga thalaiva thala I love you**
ML: English EL: Tamil CS percent: 62.5%
17. **Eppa ethuku evulo dislike ? Poramaiya ?**
ML: Tamil EL: English CS percent: 20%
18. **Jayam ravi epdi ji ongaluku mattum ipdilam story amayuthu.. Padam vetri pera vaalthukal..
To director also.**
ML: Tamil EL: English CS percent: 25%
19. **frst like thalaivaaaa i love u surya anna**
ML: Tamil EL: English CS percent: 62.5%

20. **Supar dupar hero am big fan Vikram sir**
ML: English EL: Tamil CS percent: 23.57%
21. **Ennaya Beethoven music** uh apudiye aataya pottu vachi irukaanga! Aanmai prachanayo
ML: Tamil EL: English CS percent: 18.18%
22. **Kalkitru selvaraghavan and suriya political story super**
ML: Tamil EL: English CS percent: 57.14%
23. **50 times ku mela paththutten. Vera level trailer... Waiting Fans hit likes.....**
ML: Tamil EL: English CS percent: 66.67%
24. **Nalla irukku...Aarambam maari aagama irundha It will be a blockbuster**
ML: Tamil EL: English CS percent: 45.45%
25. **Last dance semma iruku na like poduga**
ML: Tamil EL: English CS percent: 57.14%
26. **Andha last 2:20 sec le vara kurumba pakaveee waiting Thalaivaaaaa**
ML: Tamil EL: English CS percent: 40%
27. **Jayam Ravi. Verum Hollywood remake layae nadichittu irukkaru. Idhu endha padathoda remake?**
ML: Tamil EL: English CS percent: 25%
28. **bro pongalukunu solitu 10th relese pannuranga but pongal vanthu 15th**
ML: Tamil EL: English CS percent: 50%
29. **Erukku Vera leval fun erukku**
ML: Tamil EL: English CS percent: 40%
30. **It's seems like kekka bekka comedy short film**
ML: English EL: Tamil CS percent: 25%
31. **Rajini tha mass vj ajith lam tusu**
ML: Tamil EL: English CS percent: 14.28%
32. **sweet saapda porom scene la Ironman feel varude pa.. yenakku mattum daana**
ML: Tamil EL: English CS percent: 33.33%
33. **Dislike pannavan yara irupan sollunga friends**
ML: Tamil EL: English CS percent: 33.33%
34. **Kavunda kannada nee hip hop tamizhan illai nee hip hop kannadan**
ML: Tamil EL: English CS percent: 36.36%

35. Na rajini agave 10 time mela pathutta ah
ML: Tamil EL: English CS percent: 25%
36. Selvaragavan movei waiting Vera level & Surya mass hero marana waiting
ML: Tamil EL: English CS percent: 54.54%
37. Enna padaippu epperpatta karpana thalaivaa your great
ML: Tamil EL: English CS percent: 28.57%
38. Kanna moochi ray ray kandu pudi yaare effect
ML: Tamil EL: English CS percent: 12.5%
39. Ilayaraja Is Back To Form
ML: English Tamil: Tamil CS percent: 20%
40. Indha padamavadhu vijayku hit padama maranumpa
ML: Tamil EL: English CS percent: 16.66%
41. low caste pasangala pathina padama Avlova hit agathu
ML: Tamil EL: English CS percent: 37.5%
42. Padam vera level Thala fans hit
ML: Tamil EL: English CS percent: 50%
43. Lighta adanga maaru bgm maathiri theriyuthu
ML: Tamil EL: English CS percent: 16.66%
44. Trailer Elam Nallathan Iruku ana MuGamoodi Padam Mari Agama irundha Serithan!!
ML: Tamil EL: English CS percent: 9.09%
45. Mohan Anna mobile number anupunga pls
ML: Tamil EL: English CS percent: 50%
46. Yethukku last scenela blur effectu nallave illa
ML: Tamil EL: English CS percent: 57.14%
47. Jothika Yudham solra scene padathula enga dawwww?
ML: Tamil EL: English CS percent: 14.28%
48. Trailer late ah parthavanga like podunga.
ML: Tamil EL: English CS percent: 50%
49. Omg .. use head phones. Enna bgm da saami
ML: English EL: Tamil CS percent: 37.5%
50. I think sivakarthickku hero getup set aagala
ML: English EL: Tamil CS percent: 28.57%

General observations and Present Conjectures:

- Code mixing occurs at the lexical and phonological level
- At the lexical level, code mixing generally occurs when people do not remember the Tamil equivalent of that particular lexical item in English.
- The reason could be that the present-day youth have been learning some words mostly in English. It could be that they have been listening to their family members using those words only. Words like 'comment, like, share, fans, super, late, wait' have become a part of the language repertoire of everybody.
- Nouns and verbs are preferred more while code mixing. Adjectives, adverbs and conjunctions are also used by the youth when they code mix. (Here, the reason why we are looking at the youth is because the dataset is from youtube comments, where one can reasonably assume is mostly the youth or middle-aged people)
- 1) Rabbit tree keela thungiduchu. 2) Tired aaicha so thungichu.
The above examples are from speech samples of students. This shows that students tend to use Tamil as the matrix language and English words embedded into it. Though English and Tamil have different word order (Tamil SOV and English SVO), the code mixing and insertion of English words is neither artificial nor deliberate.
- The mixing of foreign words always entails their phonetic modification. The English words take some phonological features of Tamil. Eg) ball+u=baalu, Car+u=carru, door+u=dooru. The /u/ insertion is a part of Tamil phonology. English words are pronounced with the phonology of Tamil.

Present Conjectures behind Code Mixing:

- Reasons for code mixing usually tends to be
 1. Don't know Tamil equivalents
 2. Easy to express in English than in Tamil
 3. Lexical inadequacy in Tamil
 4. Some English words are easy to pronounce
 5. Easily understandable in English
 6. Prestigious to use English words
 7. Habitual use of English lexical items

General norms of politeness for Tamil:

1. Elders are addressed politely as mama or mami (uncle or aunt). The respectful term in Tamil neengal (a respectful 'you') rather than nee (which is 'you' when referring to peers) is used when addressing them.
2. Adding "Thiru" for men and 'Thirumathi' for married woman and "Selvi" for unmarried woman in front of the name, for showing respect. This is similar to the use of Mr., Mrs., Miss in English.
3. "avargal—அவர்கள்", The morpheme 'gal' is added to address a person with the expression of respect.
4. The first most common greeting to a visitor will be Vaanga, Vanakkam, Nalla irukeengala, Saapiteengala? thanni kudinga. Please come, greetings, how are you? Have you had your food? Please drink water. So, this question itself tells the Tamils culture. It is normal Tamil tradition that we first ensure that the visitor had his food and not hungry at the time he visits our home. After this, one is usually welcomed with a glass of water before even speaking other things.

5. Vanakkam (Hello)- This word is usually accompanied with the gesture of Joining hands in a praying posture and smiling gently.
6. 'Thazhmaiudan' translates to 'with sincerity' or crudely translates to 'lower than you in status'. This is used as closing phrases in letters to depict politeness.
7. Reverence to elders is shown by seeking their blessings by prostrating oneself at their feet.