B.Tech Project

Bliss to Text Translation using Deep Learning

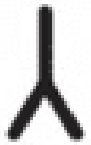
- Piyush Kumar, 19IE10022

Supervisor – Manjira Sinha

Introduction

WATER

EMOTION



MAN

- Blissymbolics were developed by **Charles K. Bliss** (1897-1985).
- **Blissymbolics** is a semantic graphical language that is currently composed of more than **5000** authorized symbols - **Bliss-characters** and **Bliss-words.** It is a **generative language** that allows its users to create new Bliss-words as needed.
- It is used by individuals with severe **speech and physical impairments** around the world, but also by others for **language learning** and support.

RAIN

HAPPINESS

WRITER

2000 +

basic symbols which can be **combined together** to create a huge variety of new symbols.

The symbols can be formed into sentences and their order is **based on English word order**

How Bliss Works ?

Bliss-character:

An individual graphic symbol (ideograph)

person earth sky

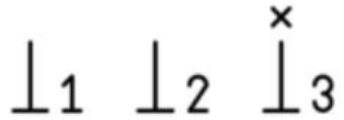
pencil mouth ear

Bliss-word:

Represents a meaning and is spelled using a sequence of one or more bliss-characters

knowledge teacher student

world language blissymbol



I you they

Indicators:

Characters used to show the part of speech of a symbol

plural action description past future

Example:

I know Blissymbolics

Problem Statement

45182

he, him, himself

17321

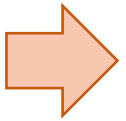
be, am, are, is

77523

to_play

98651

football

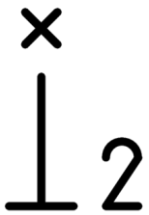


He is playing football.

34573

can, be_able_to

12573



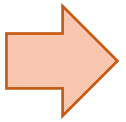
you, yourselves

76511

help, assist, aid

12986

I, me, myself



Can you help me ?

Given a **sequence of bliss symbol**, the goal is to translate the sequence into grammatically correct statement in **English** language.

Data Insights

We have used **CHILDES** dataset which is the collection and annotation of speech **to and by children.**

CHILDES (talkbank.org)

The dataset contains the transcript of communication done with children of different age group from different countries in the world. In this project dataset of North American English of age group 6-16 is used

CHILDES DATASET

MOT: see this door ?
MOT: you can open the door ?
MOT: and they can walk in the house .
CHI: 0 .
MOT: not drive in the house .
MOT: they hafta walk in the house .
MOT: are you gonna drive in the house ?
CHI: no .
MOT: it doesn't go in .
CHI: it goes in .
MOT: can you open that door ?
MOT: alright .
CHI: 0 .
MOT: are you driving in the house ?
CHI: yeah .
CHI: can't .
CHI: I can't open .
CHI: no I can't [!] .
MOT: well try again !
MOT: try again .

Methodology

Data Creation (English Corpus)

"**CHILDES**" Text Corpus → ***Manual Cleaning*** → Lemmatization using **Spacy** → Tokenization → **Cleaned "CHILDES" Data**

- Removed junk words like – "**um, hm, uh, oh, yea, huh** and **quotes**"
- Replaced apostrophe with their meaningful alternate words: **he's – he is, we're – we are, don't – do not, she'll – she will**,
- Removed chatting abbreviations like: **ty(thank you), asap, ttyl**

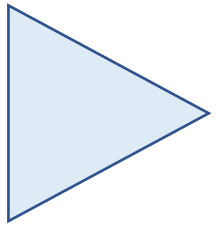
- playing/plays/played – **play**
- cars, car's, cars' – **car**
- is/are/am – **be**

```
don't ya
['do', 'not', 'ya']

it's not in my hair
['it', 'is', 'not', 'in', 'my', 'hair']

what's that
['what', 'is', 'that']

she can really really do it
['she', 'can', 'really', 'really', 'do', 'it']
```

Data Creation (Bliss Corpus)

The **official Blissymbolics dictionary** contains a unique **4-5 digit** code associated with each Bliss character or Bliss word

BCI reference number	character
12321	a,an,any
18228	what
16200	play-(to)
12408	approve-(to)
12342	adult,mature
18245	who,whom,that



BCI reference number	character
12321	a
12321	an
12321	any
18228	what
16200	to play
12408	to approve
12342	adult
12342	mature
18245	who
18245	whom
18245	that

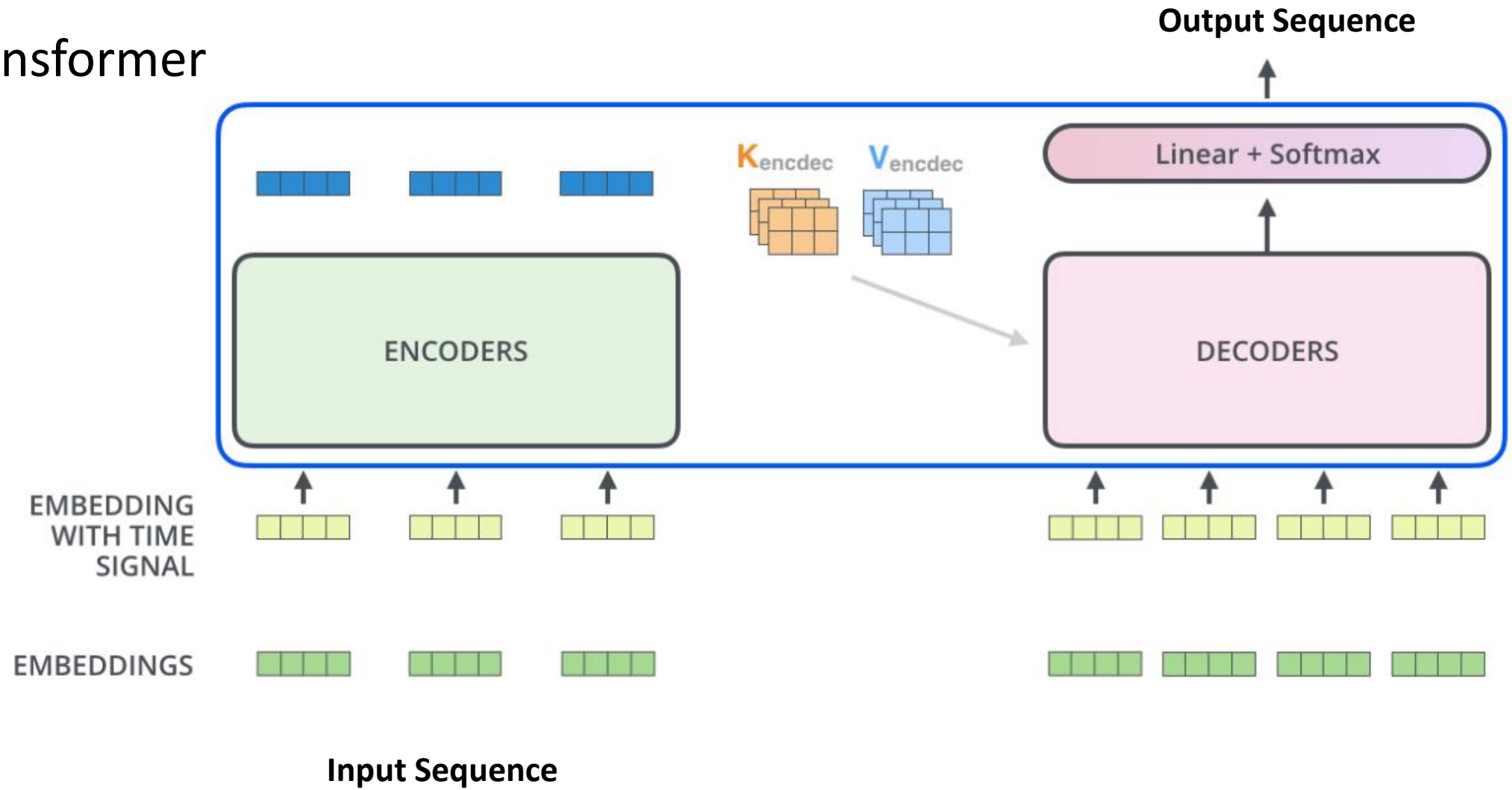
Data Creation
(Parallel Dataset)

For **processed** (after lemmatization and cleaning) of English words, we have **mapped** each of the words with their corresponding Bliss words code and formed a parallel data corpus for translation.

english	bliss
look at this	16747 12591 17720
there's a little boy	24017 12639 8521 14171 12888
shall we go eat	24261 18212 14449 13906
let's go eat	24732 12639 14449 13906
dog's gonna eat first	12380 12639 14449 17739 13906 14187
here comes the little girl	14708 13383 17700 14171 14439
what's in there	18231 12639 14932 24017

Modeling

Transformer



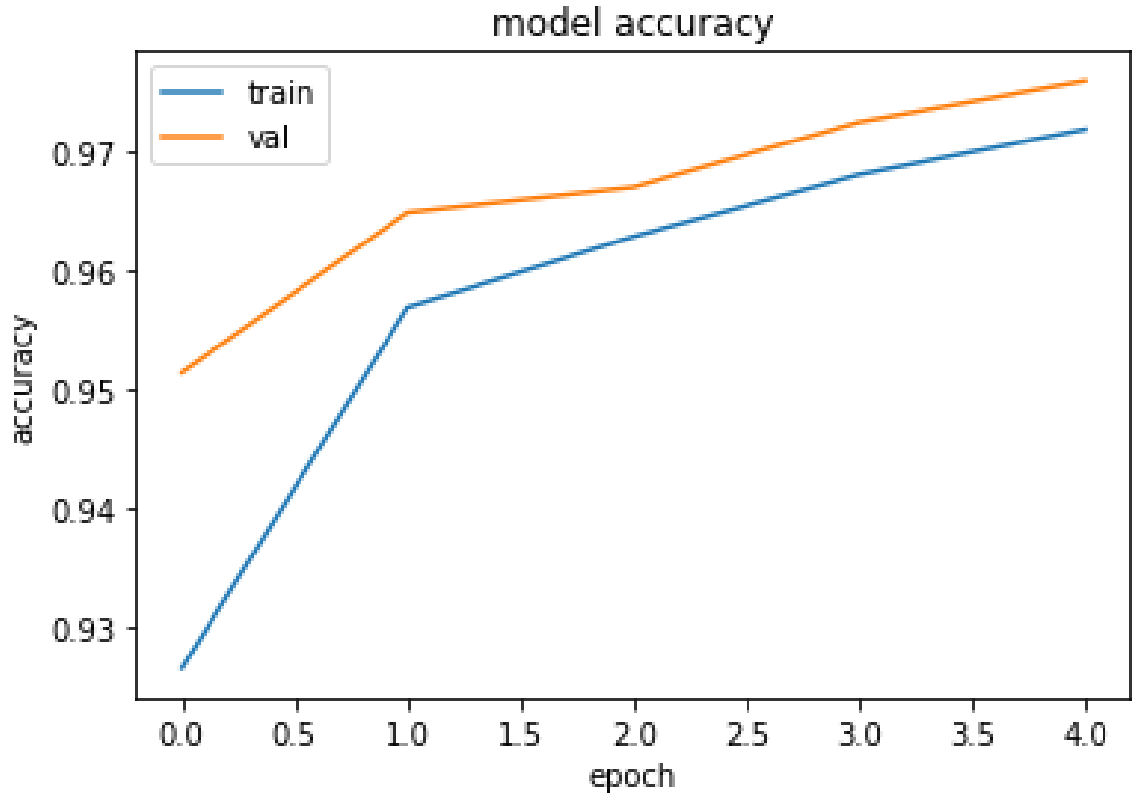
Results

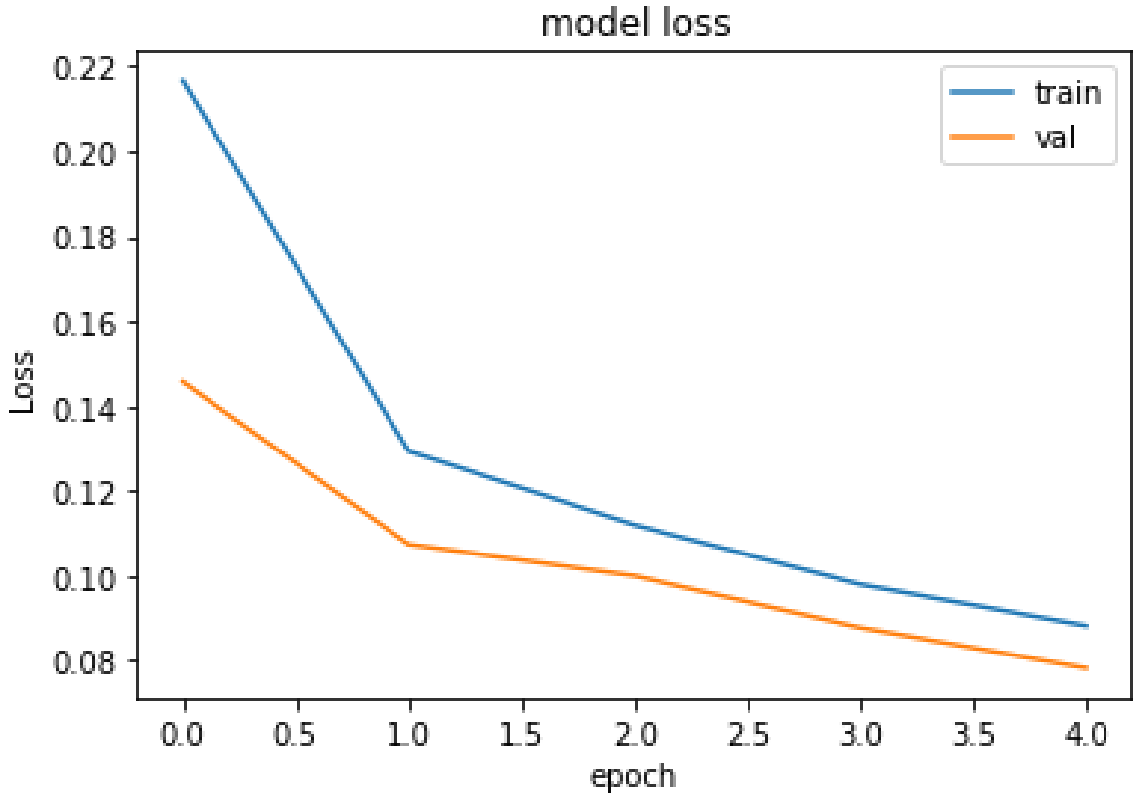
	Train Accuracy	Validation Accuracy
English to Bliss translation	**96.81**	**97.35**
Bliss to English Translation	**88.74**	**89.27**

Predictions

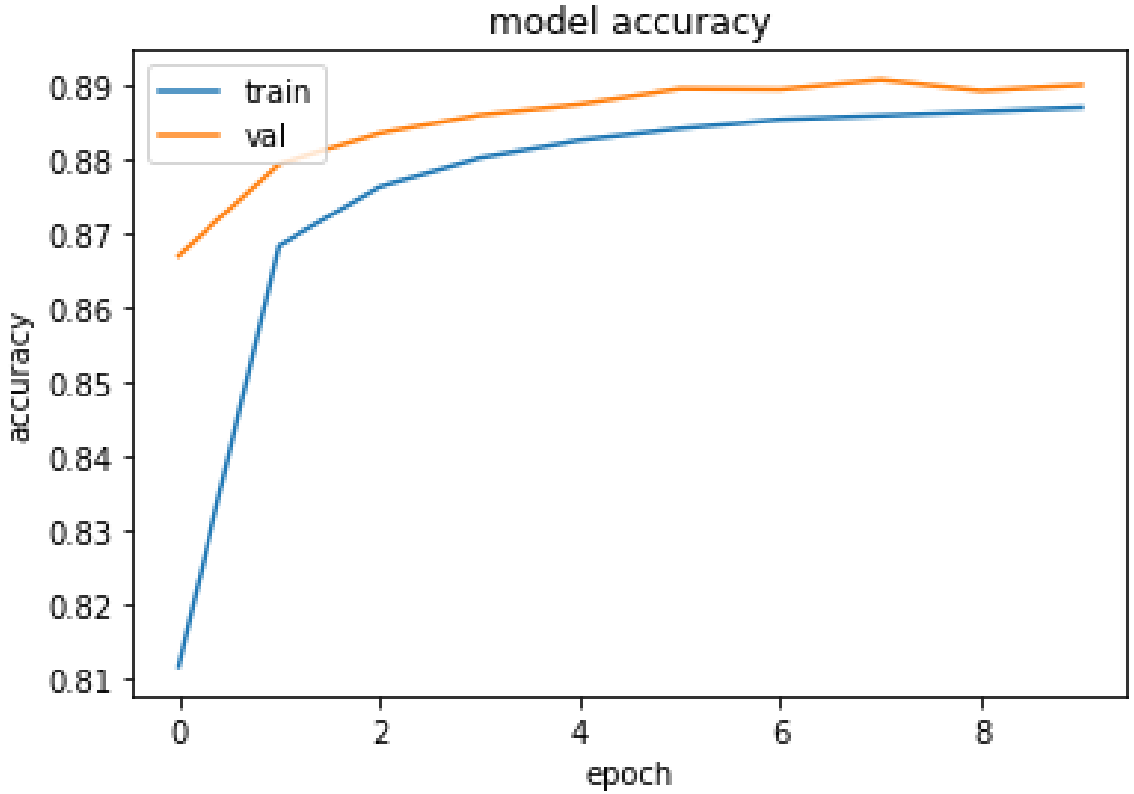
```
16747 14927 17714 14142 12602 [start] see if they fall off [end]
13114 18465 16747 17724 16482 18256 [start] can you look through that window [end]
12639 14960 25177 14675 [start] is it still fun [end]
15722 , , 14916 24261 12639 [start] no i will be [end]
17714 16747 18282 [start] they look cool [end]
18231 8521 15717 22356 [start] what a nice fish [end]
```

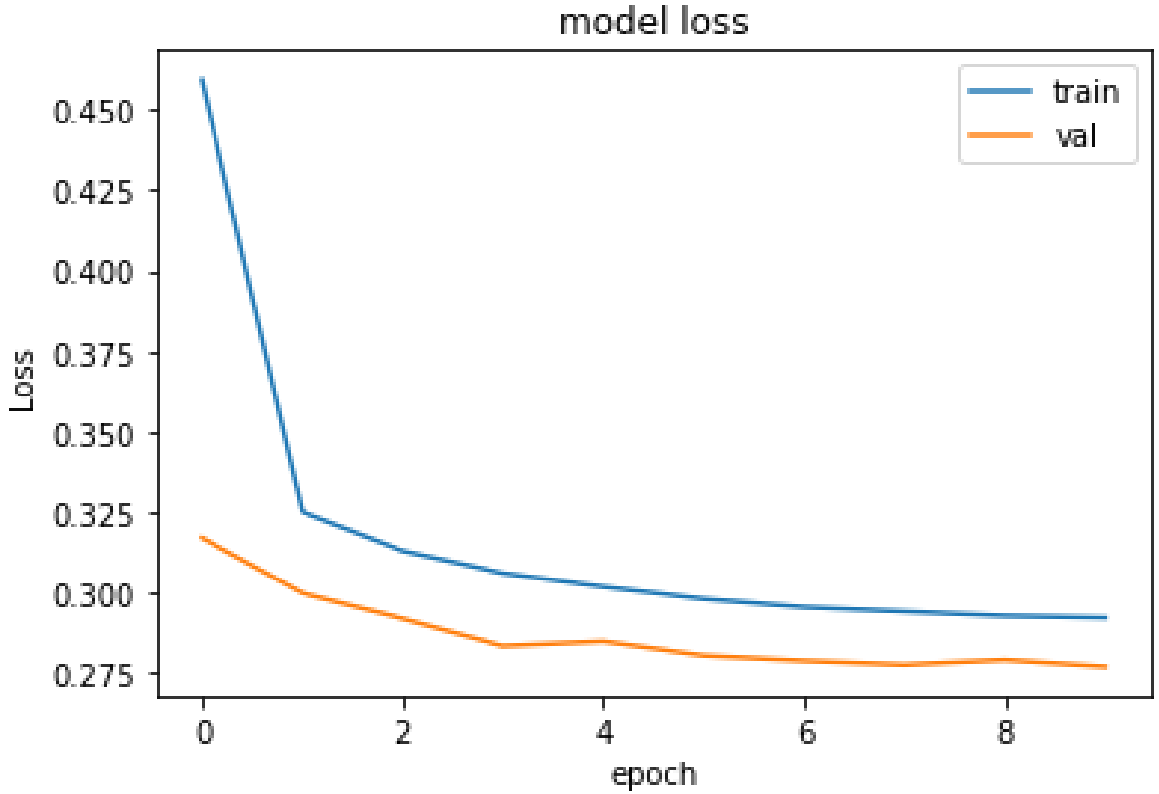
Results
(English to Bliss Translation)





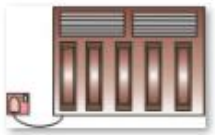
Results
(Bliss to English Translation)





Mean Meteor Score: 0.6504

Experiments
(Pictograms Addition)


Air conditioner


Almirah

Baking powder

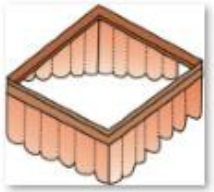

Baking tray


Baseball


Bed bug


Bee


Been

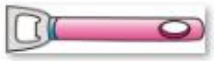

Biscuit cutter

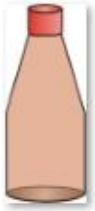

BJP Party


Blouse


Bone


Boric powder


Bottle opener


Bottle

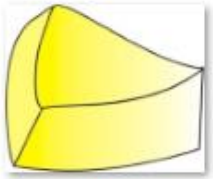

Broom


Cake mould


Carrom board


Cello


Chauki


Cheese

Chemist shop

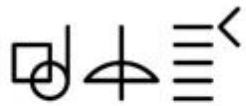

Chess

Chilli powder

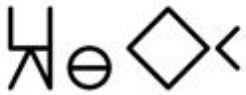
Experiments
(Pictograms Addition)

Volcano

Violin



Baseball



Globe

Experiments
(Changing Tenses)

```
change_tense("I am going to play with friends in the ground", "future")
```

```
'I will be going to play with friends in the ground'
```

```
change_tense("They will come tomorrow", "past")
```

```
'They came tomorrow'
```

```
change_tense("There had been some confusion with that", "present")
```

```
'There is some confusion with that'
```

```
change_tense("Hey, did you finish that project we were working on last week?", "present")
```

```
'Hey, do you finish that project we am working on last week?'
```

```
change_tense("The sun is shining brightly outside", "future")
```

```
'The sun will be shining brightly outside'
```

Experiments
(Changing Tenses)

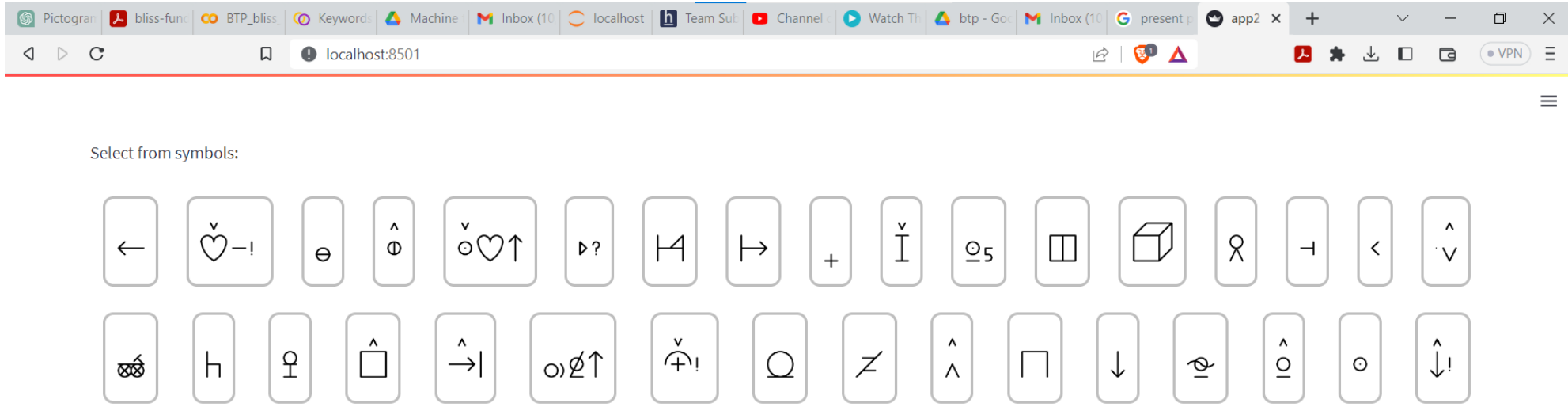
```
[20] change_tense("I am typing on my computer, using the keyboard to input letters and
```

```
'I was typing on my computer, using the keyboard to input letters and symbols ont
o the screen. The room was quiet except for the sound of my fingers tapping the k
eys. Outside, the sun was shining and birds were chirping in the trees. I was fee
ling focused and productive as I worked on this task. Occasionally, I paused to t
ake a sip of water and stretched my arms and neck to prevent any strain. As I con
tinued typing, I was aware of the progress I was making and the time I was spendi
ng on this project. Despite any challenges that may came up along the way, I was
confident that I 'll was able to complete it successfully.'
```

```
ter that evening, I relaxed at home and watched a movie before heading to bed, feeling content a
```

```
'I will wake up early and will go for a run around the park. The air will be crisp and refreshin
g, and I will enjoy the feeling of the sun on my skin. After my run, I will shower and will get
ready for work. I will spend the morning attending meetings and working on various projects, mak
ing steady progress throughout the day. In the afternoon, I will meet up with a friend for lunch
and we will catch up on each other's lives. Later that evening, I will relax at home and will wa
tch a movie before heading to bed, feeling content and grateful for a productive and enjoyable d
ay.'
```

Experiments
(Web Service)



Conclusion

- The aim of our work was to create a computer-based tool for educational and communicational purposes for people with severe speech impairments.
- The machine learning method using transformer performed well on both side translation that is bliss-to-English and English-to-bliss

- Project Link - https://github.com/mepiyush2000/BTP_bliss_symbol_project
- Future work will look into generation-heavy and transfer approaches for Bliss-to-Text translation considering blissymbolics as bag of words
- Add more number of pictograms for training and language model text generation using bliss symbols
- Improve website UI and make it more easy to use by dividing the symbols intp section and sub sections.

References

- Electronic usage of BLISS symbols https://www.researchgate.net/publication/238474375_Electronic_usage_of_BLISS_symbols
- Application of Blissymbolics to the non-vocal communicatively handicapped https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=2595&context=etd
- Web-service for translation of pictogram messages into Russian coherent text https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8633677
- Natural Language Generation from Pictographs https://aclanthology.org/W15-4711.pdf
- From Bliss Symbols to Grammatically Correct Voice Output: A Communication Tool for People with Disabilities https://link.springer.com/content/pdf/10.1023/A:1013682632553.pdf
- A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation Syst https://aclanthology.org/2022.amta-research.19.pdf
- Blissymbolics Communication International https://www.blissymbolics.org/
- CHILDES Dataset https://psyling.talkbank.org/years/1985/jcl-childes.pdf

Thank You