MT Assignement

Submitted by :
Shuvra Neel Roy
1915045
Date: 1/05/23

	SMT		NMT (transformer based)	
Training Size	34K		47K	
Test Size	500		2000	
Average Sentence Length	12-13 (Assamese-English)		25	
	S - Assamese T - English	S - English T - Assamese	S - Assamese T - English	S - English T - Assamese
BLEU	2.45	1.43	36.38	34.45
METEOR	0.19	0.11	0.667	0.544
TER	0.87	0.91	0.46	0.49

Transformers Results

Sentence 13	BLEU	Length ratio	Text
Human	100.00	1.00	_ In _ its _ fight _ against _ COVID - 1 9 _ since _ January _ 2 0 2 0 , _ India _ has _ crossed _ another _ peak .
Machine	47.93	0.71	_ India _ has _ achieved _ another _ milestone _ in _ its _ fight _ against _ COVID - 1 9 .

BLEU:	36.68			
Precision x brevity:	42.53 x 86.23			
Type	1-gram	2-gram	3-gram	4-gram
Individual	70.61	48.64	36.63	26.02
Cumulative	60.89	50.53	43.21	36.68
Export data	CSV			

`<<` `<` 1 `>` `>>` From 1 to 45 of all **2000** sentences.

Dataset Link :https://object.pouta.csc.fi/OPUS-Anuvaad/v1/moses/as-en.txt.zip
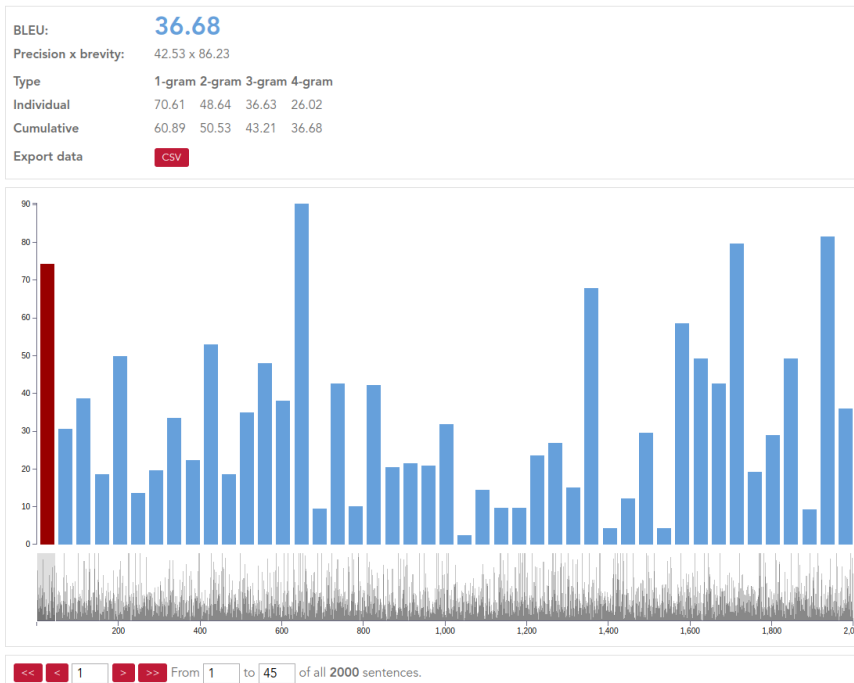
Most frequent Word:
Unfiltered:
      Assamese : আৰু : 26667
      English : 'the': 48156,
Filtered:
      Assamese : 'আৰু': 11346
      English :'the': 25854,