

(4) Some contrastive pictures between two languages – Bangla and English from the statistics of histogram part are drawn below:

The result we found from nlp\_histogram\_2016831021.py are:

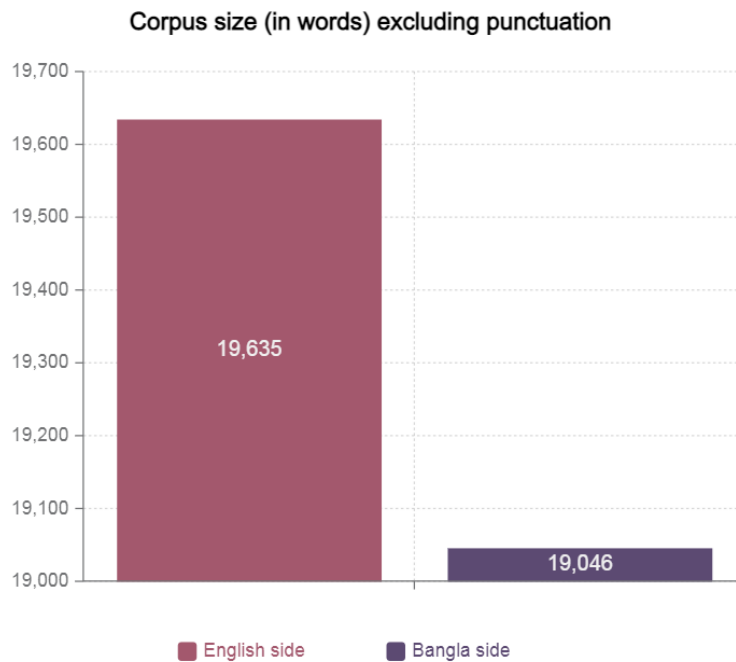
	English side	Bangla side
Corpus size (in words) excluding punctuation	19635	19046
Corpus size (in chars) excluding spaces	82972	86032
Average sentence length (in words)	4.501	4.366
Vocabulary size (no. of unique words)	2834	3514
Lexical diversity*	6.928	5.420
Corpus size (in lines)	4362	4362

\* How frequently on average each vocabulary item appears in the corpus.

And top ten frequent words in the parallel corpus are:

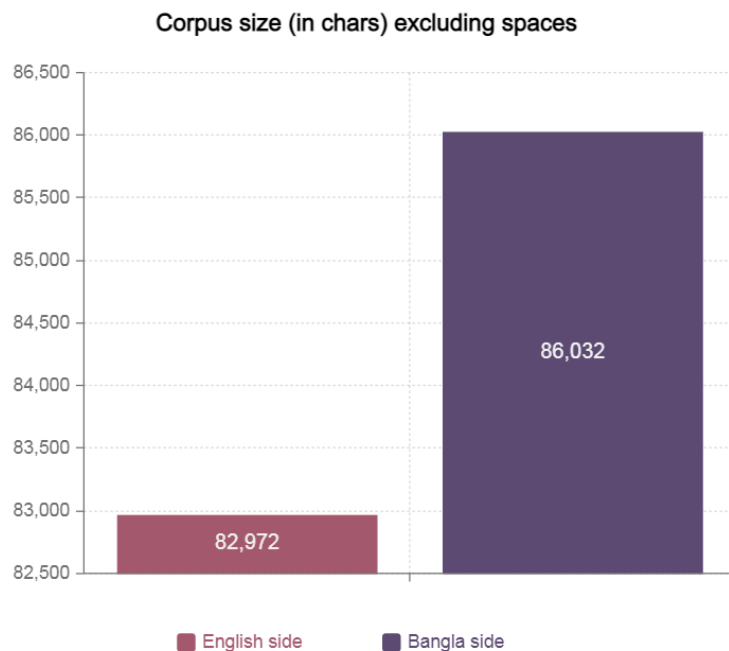
English side			Bangla side		
Words	Frequency	(%)	Words	Frequency	(%)
আমি	926	4.862	I	839	4.273
টম	584	3.066	Tom	628	3.198
আমার	448	2.352	to	586	2.984
কি	391	2.053	you	534	2.719
না	309	1.622	is	457	2.327
আপনি	225	1.181	the	370	1.884
এটা	175	0.919	a	300	1.528
তুমি	175	0.919	I'm	204	1.0389
কথা	156	0.819	your	165	0.840
আপনার	149	0.782	He	160	0.815

- Corpus size (in words) excluding punctuation:



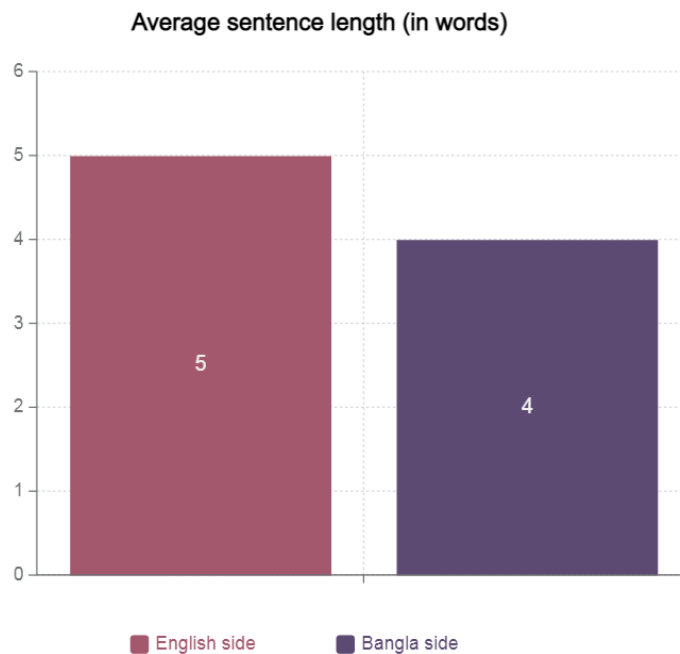
Here, Corpus size (in words) in English is more than in Bangla. In number English corpus size (in words) is 19,635 and in Bangla, it is 19,046. So Bangla has 589 fewer words in the corpus.

- Corpus size (in chars) excluding spaces:



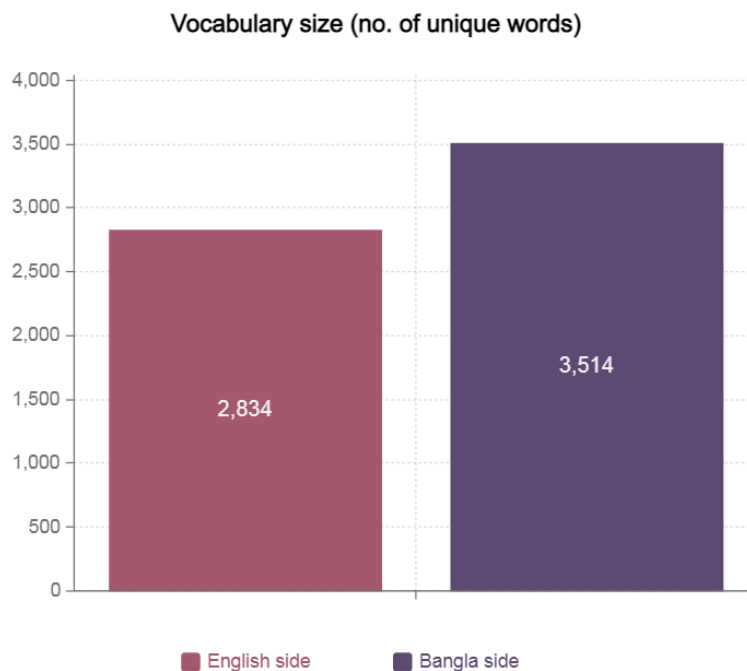
Here, the Corpus size (in chars) in English is less than Bangla. In number English corpus size (in chars) is 82,972 and in Bangla, it is 86,032. So Bangla has 3,060 more chars in the corpus.

- Average sentence length (in words):



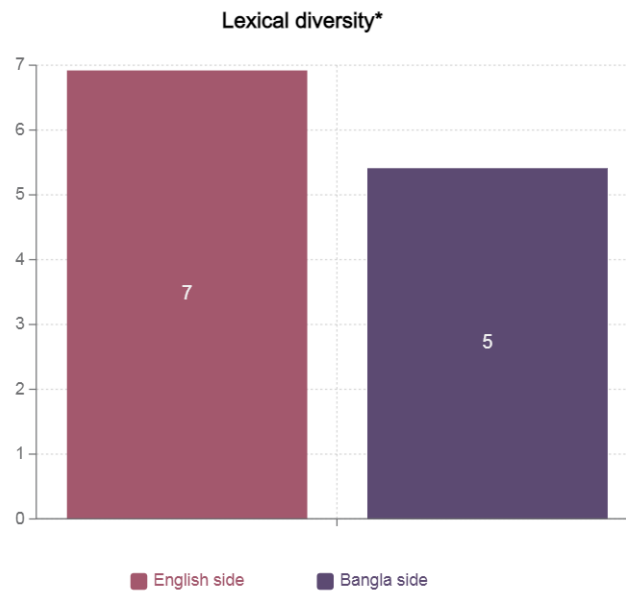
Here, the average sentence length (in words) in English is more than in Bangla. In English average length of a sentence is 5 (4.501), and In Bangla, it is 4 (4.366).

- Vocabulary size (no. of unique words):



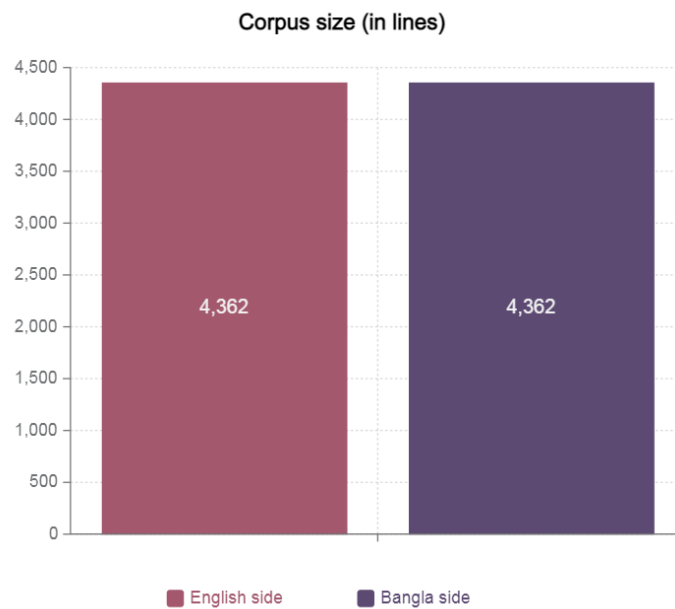
Here, The vocabulary size of English is less than in Bangla. In number, the English vocabulary size is 2,834, and the Bangla vocabulary size is 3,514. Bangla has 680 more vocabularies than English in the corpus.

- Lexical diversity:



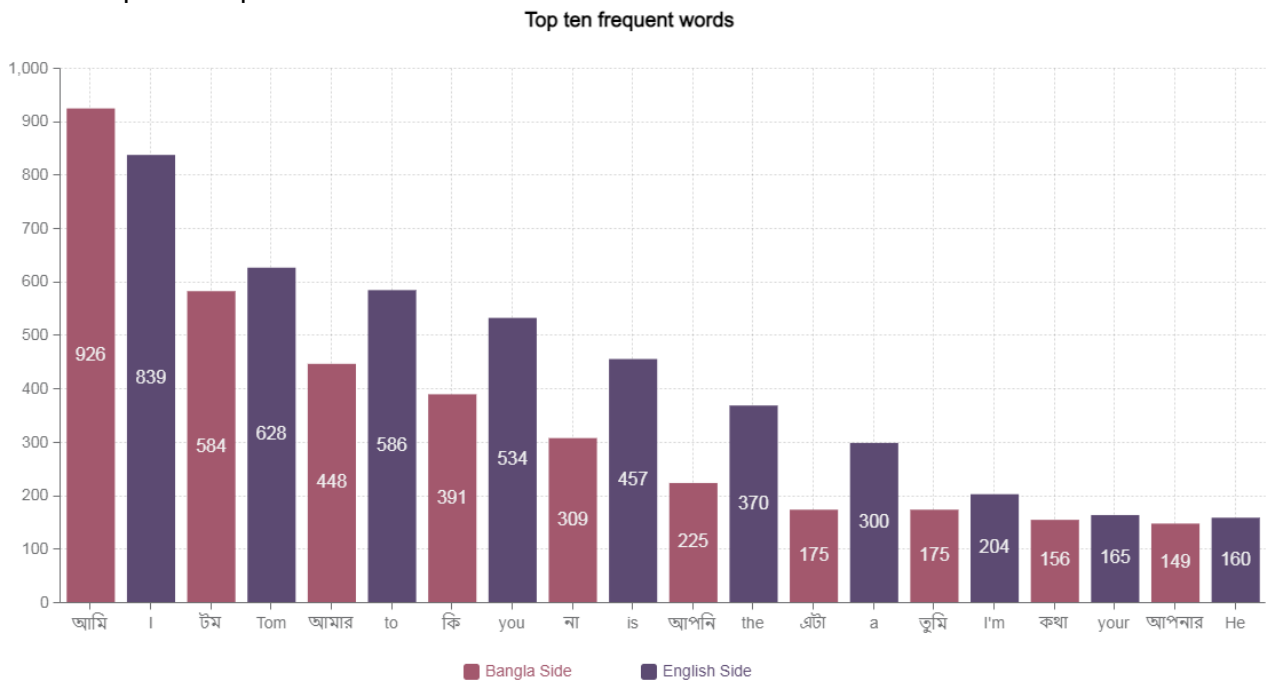
Here, Lexical diversity in English (7) is more than in Bangla (5).

- Corpus size (in lines):



Here, Both in English and Bangla, The Corpus size (in lines) is the same (4,362).

- Top ten frequent words:



Here, From the top ten frequent words, English has more reusable words than Bangla. That's why the word frequency of English words is more than Bangla words.