Script Name	Data Level	Data Set Size	Reference
English	word	86272	
English	character	3410	
Hindi	word	15400	
Hindi	character	29970	
Bangla	word	30000	
Bangla	character	9800	
Tamil	word	8500	
Tamil	character	78000	
Telugu	character	44820	
Assamese	character	8235	