1. Download and install crf++ from <https://taku910.github.io/crfpp/>

2) format your data into 7 columns

WORD ROOT prefix infix suffix reduplication POSTAG..

Since you don’t have the POStag of your data put XXX instead.

Below is the sample format.

isang isa \_ \_ ng \_ JJ-JN  
umaga umaga \_ \_ \_ \_ NN-NNC   
ng ng \_ \_ \_ \_ PP-OM  
disyembre disyembre \_ \_ \_ \_ NN-NND  
, , \_ \_ \_ \_ PM-PMC   
ang ang \_ \_ \_ \_ DT-CDSG  
bapor bapor \_ \_ \_ \_ NN-NNC   
tabo tabo \_ \_ \_ \_ NN-NNP   
naghahatid hatid nag \_ \_ ha VB-IMAF   
ng ng \_ \_ \_ \_ PP-OM  
maraming dami ma \_ ng \_ JJ-JCJD   
sakay sakay \_ \_ \_ \_ NN-NNC   
sa sa \_ \_ \_ \_ PP-DM  
laguna laguna \_ \_ \_ \_ NN-NNP   
, , \_ \_ \_ \_ PM-PMC   
ay ay \_ \_ \_ \_ LM-LM   
hirap hirap \_ \_ \_ \_ JJ-JCCS   
na na \_ \_ \_ \_ CC-CCP   
hirap hirap \_ \_ \_ \_ JJ-JCCS   
sa sa \_ \_ \_ \_ PP-DM  
pagsalunga salunga pag \_ \_ \_ NN-NNC   
sa sa \_ \_ \_ \_ PP-DM  
mabilis bilis ma \_ \_ \_ JJ-JCJD   
na na \_ \_ \_ \_ CC-CCP   
agos agos \_ \_ \_ \_ NN-NNC   
ng ng \_ \_ \_ \_ PP-OM  
ilog-Pasig ilog-Pasig \_ \_ \_ \_ NN-NNP   
. . \_ \_ \_ \_ PM-PMP

You can use your stemmer to produce the affixes…

2) Sentences must end with proper punctuation. And sentences are separated by a single space.

3) Run the command below:

$ crf\_test -m bigram.model test.data