Hw11
Yao Yan
1669598

Q1

1. Apply the perl file for sentence boundary detection.
   Using commands:
   ./split-sentences.perl -l en < en-ep-99-12-17.txt > en-ep-99-12-17.tok.txt
   ./split-sentences.perl -l de < de-ep-99-12-17.txt > de-ep-99-12-17.tok.txt
2. Apply sentence_aligned.sh for alignment
   a. ./sentence_aligner.sh de-ep-99-12-17.tok.txt en-ep-99-12-17.tok.txt > de-en-aligned.txt
   b. cut -f1 de-en-aligned.txt > de-en-aligned.txt.de
   c. cut -f2 de-en-aligned.txt > de-en-aligned.txt.en

Q2

For sentence alignment comparison:
It's the same idea as hw
In my solution, the first word in the sentence is represented by "s", the others using "o". The same conversion is applied to both input le and gold le. By comparing the alignment of "s", the F1 score for sentence boundary detection can be calculated.
e.g.
"This is sentence1."
"This is sentence2."
will be converted to "soosoo"
"This is sentence1. This is sentence2."
will be converted to "sooooo"

Q3

t(maison|green) = $\frac{1}{2}$ t(vert|green)= $\frac{1}{2}$ t(la|green)= 0
t(maison|house) = $\frac{1}{2}$ t(vert|house)= $\frac{1}{4}$ t(la|house)= $\frac{1}{4}$
t(maison|the) = $\frac{1}{2}$ t(vert|the)= 0 t(la|the)= $\frac{1}{2}$