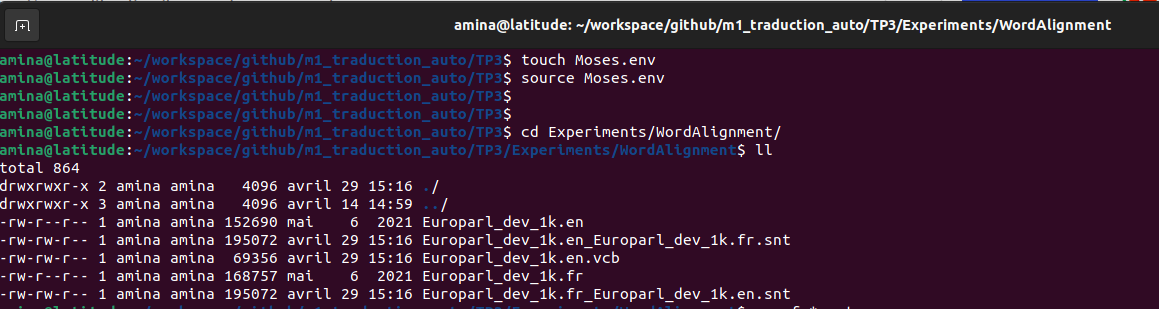
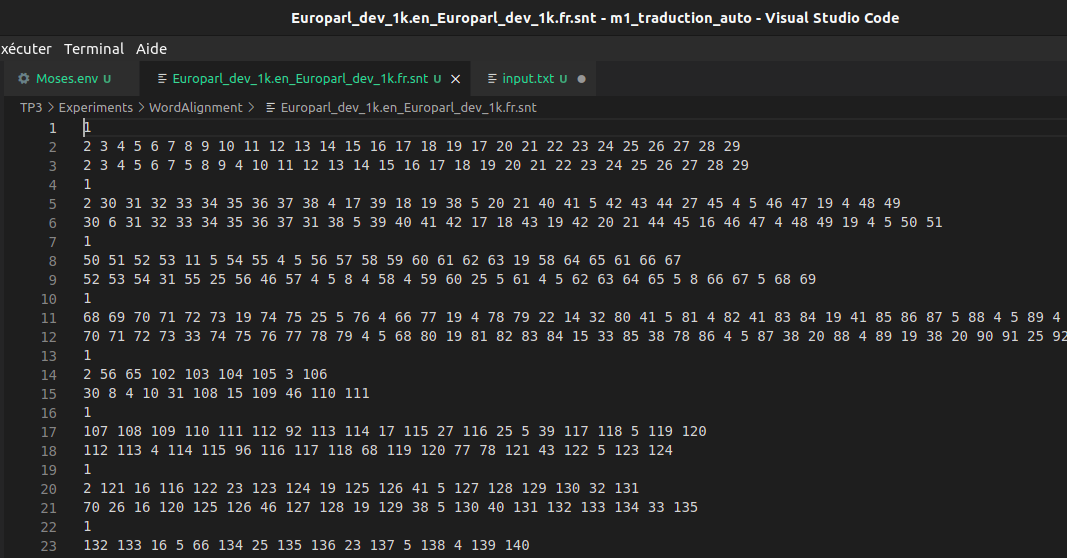
TP\_3

Installtion et Utilisation de Moses

Installation de Moses



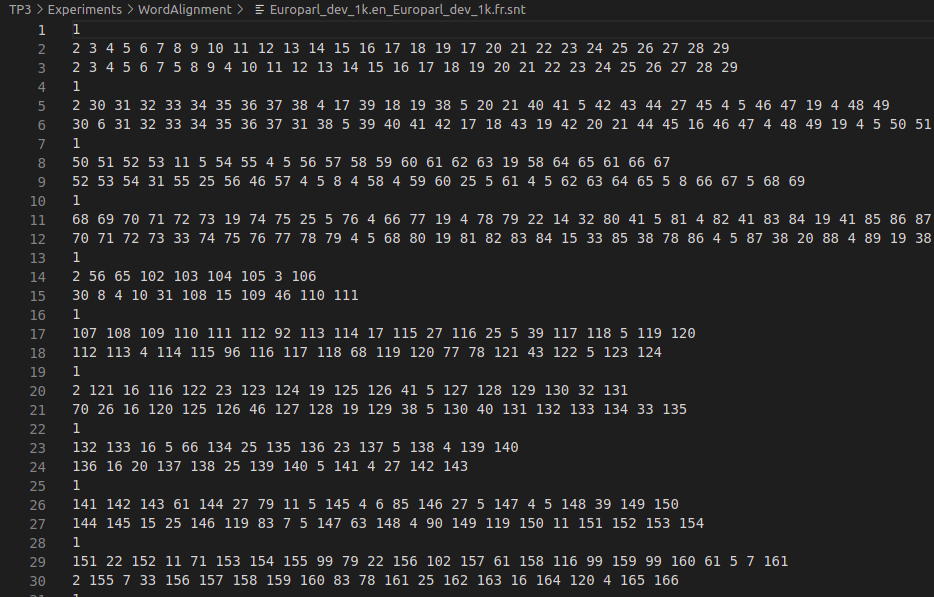


plain2snt.out ⇒ il sort 2 fichiers:

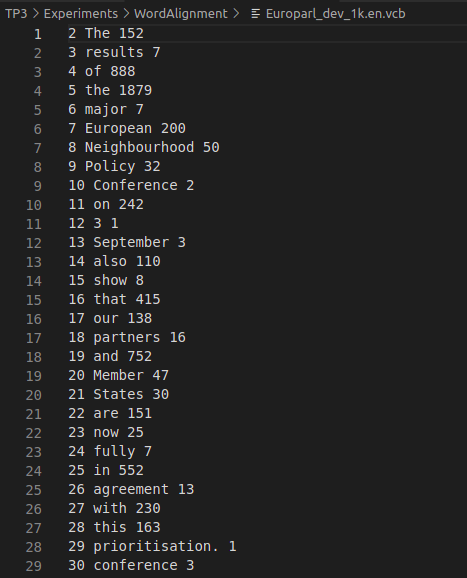


Résultat :

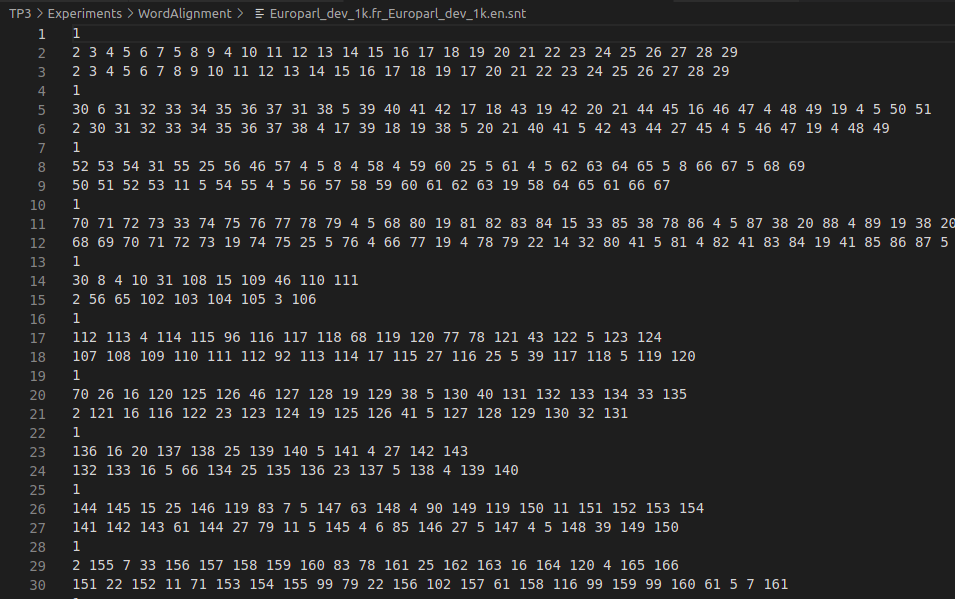
Europarl\_dev\_1k.en\_Europarl\_dev\_1k.fr.snt

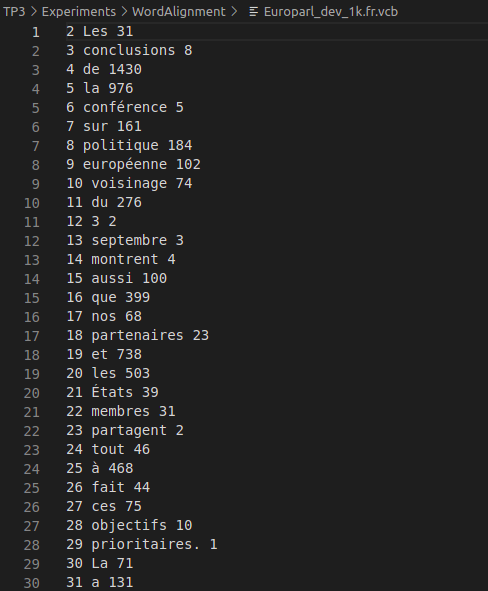


Europarl\_dev\_1k.en.vcb



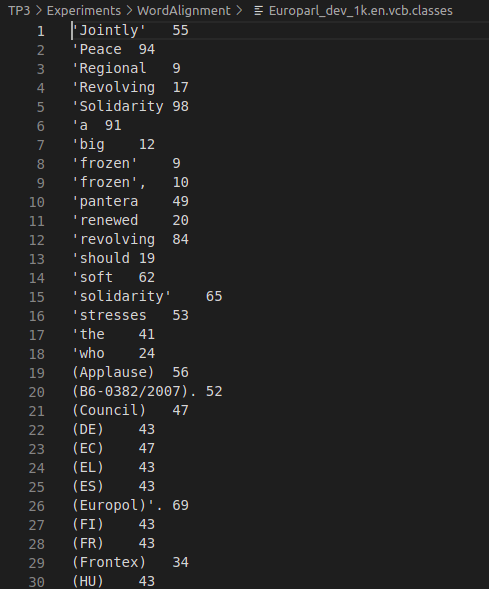
Europarl\_dev\_1k.fr\_Europarl\_dev\_1k.en.snt

Europarl\_dev\_1k.fr.vcb

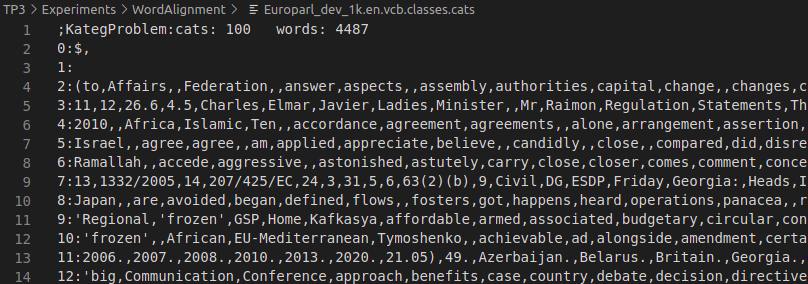


Résultat :

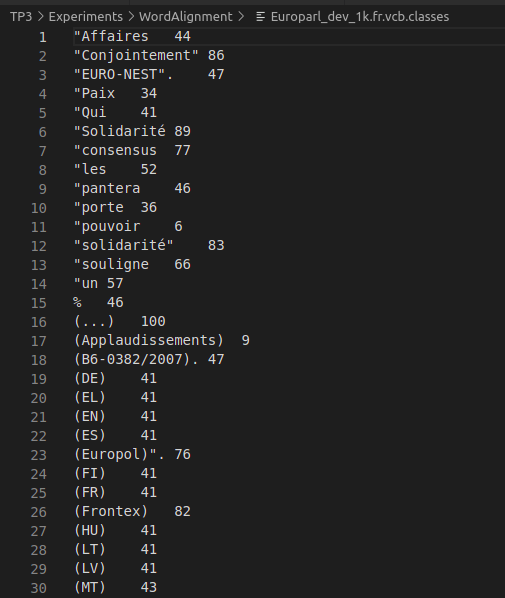
Europarl\_dev\_1k.en.vcb.classes



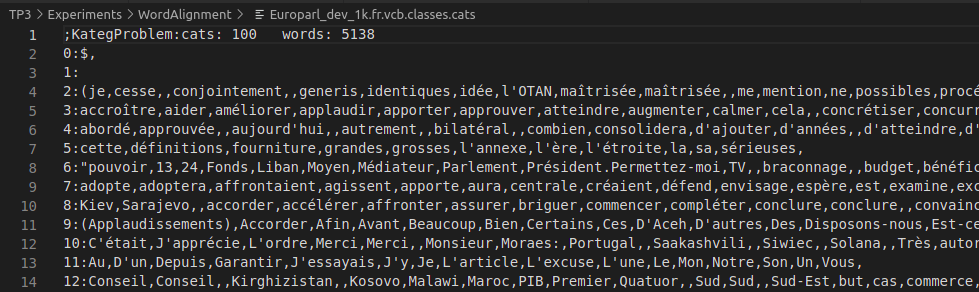
Europarl\_dev\_1k.en.vcb.classes.cats

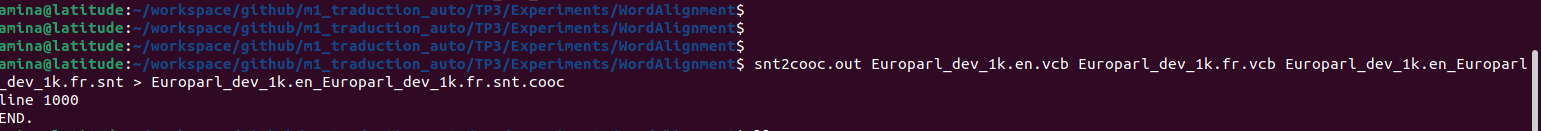


Europarl\_dev\_1k.fr.vcb.classes

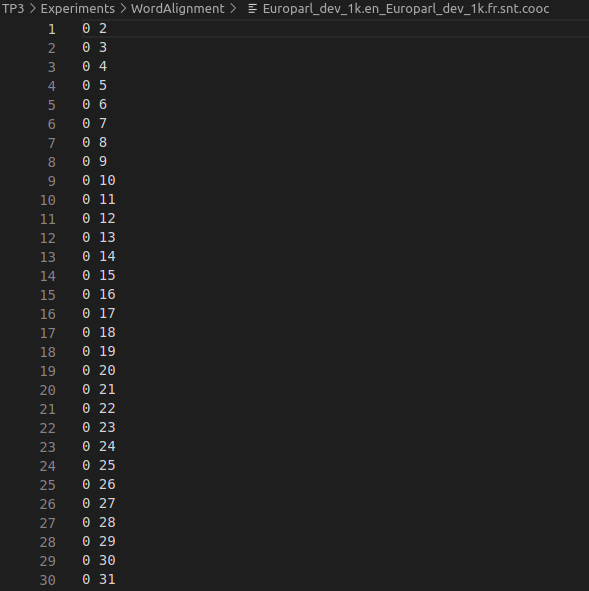


Europarl\_dev\_1k.fr.vcb.classes.cats

4.

Résultat :

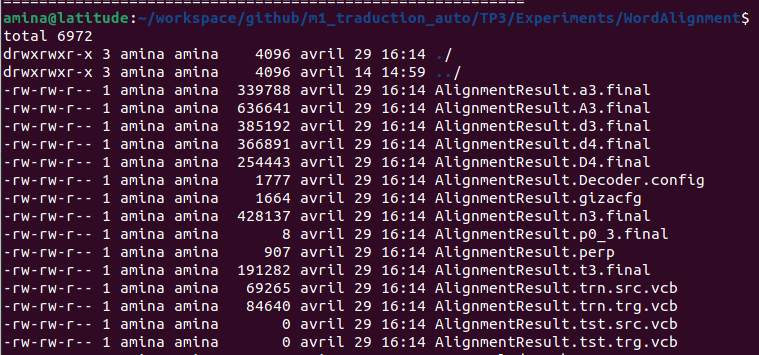
Europarl\_dev\_1k.en\_Europarl\_dev\_1k.fr.snt.cooc



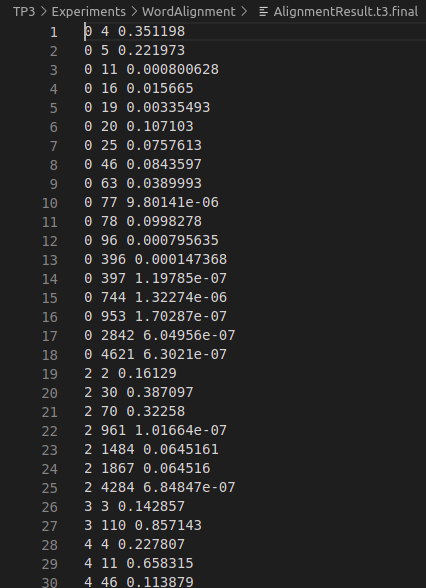
5.

GIZA++ -S Europarl\_dev\_1k.en.vcb –T Europarl\_dev\_1k.fr.vcb –C Europarl\_dev\_1k.en\_Europarl\_dev\_1k.fr.snt -CoocurrenceFile Europarl\_dev\_1k.en\_Europarl\_dev\_1k.fr.snt.cooc -o AlignmentResult

Résultat :



- La table de traduction : "AlignmentResult.t3.final"



5.

mgiza -S Europarl\_1000\_en.txt.vcb –T Europarl\_1000\_fr.txt.vcb –C Europarl\_1000\_en.txt\_Europarl\_1000\_fr.txt.snt -CoocurrenceFile Europarl\_1000\_en.txt\_Europarl\_1000\_fr.txt.cooc -o AlignmentResult

II.Utilisation:

1. Récupération de corpus
2. Tokenisation du corpus Anglais-Français

$MOSES\_HOME/mosesdecoder/scripts/tokenizer/tokenizer.perl -l en < Europarl\_train\_10k.en > Europarl\_train\_10k.en.tok

$MOSES\_HOME/mosesdecoder/scripts/tokenizer/tokenizer.perl -l fr < Europarl\_train\_10k.fr > Europarl\_train\_10k.fr.tok

Résultat:

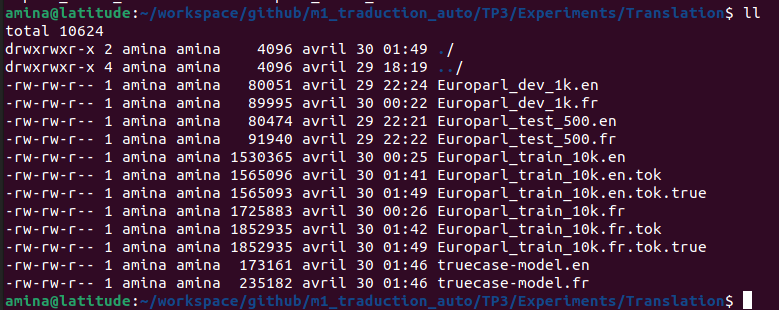
Europarl\_train\_10k.en.tok

Europarl\_train\_10k.en.tok

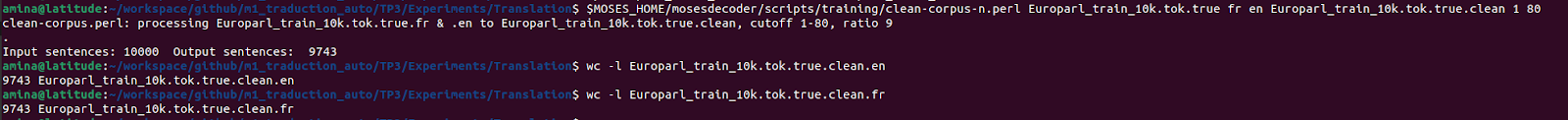
c. Changement des majuscules en minuscules du corpus Anglais-Français

1. Apprentissage du modèle de transformation

2. Transformation:

￼d. Nettoyage en limitant la longueur des phrases à 80 caractères

Résultats :

2) Construction du modèle de langage pour le corpus Anglais: