

# METL: Méthodes Statistiques de Traduction Automatique

## 3C124

Paola Merlo

année académique 20019-2020

### Questions sur la Phrase-based Machine translation.

- Définition** : Un **alignement** est une fonction de la liste ordonnée des mots de la phrase cible C à la liste ordonnée des mots de la langue source S. (Une fonction est une relation binaire R entre l'ensemble A et l'ensemble B tel que pour chaque élément  $a \in A$  il y exactement un couple ordonné  $(a, b) \in R$ ,  $b \in B$ , dont le premier élément est  $a$ .)
  - Indiquez avec une liste de paires de mots et/ou syntagmes les alignements des traductions suivantes :  
(The balance was the territory of the original people ; Le reste appartenait aux autoctones).  
(The poor don't have any money ; Les pauvres sont demunis).
  - De façon plus concise, nous pouvons indiquer l'alignement par un vecteur : les positions des éléments du vecteur correspondent aux positions des mots français dans la phrase française et la valeur de chaque élément du vecteur est la position du mot anglais qui correspond au mot français indiqué par l'élément en question. Par exemple, notre alignement peut être représenté par le vecteur [1,2,3,4,5,5,5].  
Indiquez les alignements que vous avez proposés dans l'exercice précédent par des vecteurs.
  - Quel propriété doivent avoir vos alignements pour que vous puissiez les encoder avec des vecteurs ?
- Calculez l'alignement le plus probable pour la paire de phrases  $\{(the\ blue\ house, la\ maison)\}$ . D'abord énumérer tous les alignements (il y en a combien ?) et après évaluez  $P(a|e, f)$  pour chacun d'entre eux, comme vu au cours. Utilisez les probabilités de traduction suivantes :  $t(la|the) = 0.7$ ;  $t(maison|the) = 0.3$ ;  $t(la|blue) = 0.4$ ;  $t(maison|blue) = 0.6$ ;  $t(la|house) = 0.9$ ;  $t(maison|house) = 0.1$ .
- Proposez deux exemples de traduction qui requièrent des alignements de mots multiples (plusieurs mots vers plusieurs mots à la fois) pour obtenir la bonne traduction et qui donc ne seraient pas possible de représenter avec notre définition d'alignement.  
Visualisez ces traductions avec une matrice : les lignes de la matrices sont les mots de phrase source, et les colonnes sont les mots de la phrase cible. Les cases qui indiquent les alignements sont noircies.
- Construisez les alignements mot par mot qui vous permettent de compiler le dictionnaire bilingue dérivé de 12 paires de phrases donn en fin d'énoncé (page 3). Les phrases indiquées avec  $a$  dans la paire sont la source et les phrases indiquées avec  $b$  sont la cible.  
Commencez à l'aide des traductions suivantes :  
ghirok : hilat  
ok-drubel : at-drubel  
ok-voon : at-voon

ok-yurp : at-yurp  
zanzanok : zanzanat

Indiquez les alignements et le dictionnaire bilingue dérivé.

5. Ecrivez une phrase en une langue étrangère (au moins 5 mots). Reorganisez les mots qu'ils apparaissent dans un ordre français (ou anglais). Utilisez un dictionnaire bilingue pour chercher **toutes** les possibilités de traduction mot par mot. Ecrivez chacune des ces traductions dans la même colonne au dessous du mot. Effacer les mots originaux. Demandez à un ami(e) de construire une phrase en utilisant un mot de chaque colonne.

1a ok-voon ororok sprok  
1b at-voon bichat dat

2a ok-drubel ok-voon anak plok sprok  
2b at-drubel at-voon pippat rrat dat

3a erok sprok izok hihok ghrok  
3b total dat arrat vat hilat

4a ok-voon anak drok brok jok  
4b at-voon krat pippat sat lat

5a wiwok farok isok stok  
5b totat jjat quat cat

6a lalok sprok izok jok stok  
6b wat dat krat quat cat

7a lalok farok ororok lalok sprok izok enemok  
7b wat jjat bichat wat dat vat eneak

8a lalok brok anak plok nok  
8b iat lat pippat rrat nnat

9a wiwok nok isok kantok ok-yurp  
9b totat nnat quat oloat at-yurp

10a lalok mok nok yorok ghrok klok  
10b wat nnat gat mat bat hilat

11a lalok nok crrrok hihok yorok zanzanok  
11b wat nnat arrat mat zanzanat

12a lalok rarok nok izok hihok mok  
12b wat nnak forat arrat vat gat