

Donnez le numéro de la question suivi par votre réponse.

I Partie I (10 points)

1. Expliquer la notion de fertilité utilisée en alignement lexical bilingue. (2 point)
2. Expliquer la méthode de sélection de données **mono-lingue** basée sur la différence d'entropie croisée. (4 points)
3. Expliquer le principe de fonctionnement de l'approche compositionnelle. (4 points)

II Partie II (10 points)

1. Pourquoi les réseaux de neurones récurrents semblent adaptés pour la traduction automatique ? (2 points)
2. Que représente la taille de la dernière couche d'un modèle de langage neuronal ? (2 points)
3. Quels sont les méta-paramètres d'un réseau de neurone ? (2 points)
4. Dessiner l'architecture d'un système de traduction neuronal. De quoi dépend le i-ème mot en sortie d'un système de traduction neuronal tel que présenté dans le cours ? (4 points)

III Partie III (20 points)

1. (10 points) Nous souhaiterions identifier le meilleur système de traduction FR→EN parmi les trois systèmes A, B et C. Ci-dessous la traduction de la même phrase f en utilisant les trois systèmes.

- **Source f :** Les officiers israéliens sont responsables de la sécurité de l'aéroport .
- **Sortie du système A :** Israeli officials chargeable for airport security .
- **Sortie du système B :** airport security Israeli officials are responsible for .
- **Sortie du système C :** airport security Israeli officials are responsible .
- **Traductions humaine :** Israeli officials are responsible for airport security .

- a) D'après vous, quelle est la meilleure traduction ? pourquoi ?
- b) Calculer le score BLEU (2 chiffres après la virgule) de chaque système de traduction en utilisant la traduction de référence. Quelle est la meilleure traduction selon les scores BLEU que vous avez obtenus ?

$P_2 = \frac{1}{4}$ $\frac{11111111}{11111111}$
 $P_3 = \frac{1}{4}$ $\frac{11111}{11111}$ $P_4 = \frac{1}{4}$

2. (10 points) Soit la matrice d'alignement suivante.

	Quan	tornes	tu	a	casa	?
When	x					
you		x				
are			x			
coming		x				
back		x				
home					x	
?						x

a) Quel est l'algorithme qui permet d'estimer les alignements mot à mot à partir de corpus alignés au niveau de la phrase ?

b) On souhaite utiliser la matrice d'alignement ci-dessus pour extraire les alignements au niveau groupe de mots. Trouvez les paires de segments qui ne **SERONT PAS** extraits à partir de cette matrice d'alignement :

- Quan tornes ||| When you are coming back
- tornes tu ||| you are coming back
- Quan tornes tu ||| When are you coming back
- tornes a casa ||| coming back home
- tornes a casa ?, coming back home ?
- casa ? ||| home ?
- a casa ? ||| home ?
- tu a ||| are