Traduction automatique neuronale : principes et questionnements actuels, challenges futurs.

Loïc Barrault

LIUM
Laboratoire d'Informatique
Université du Mans

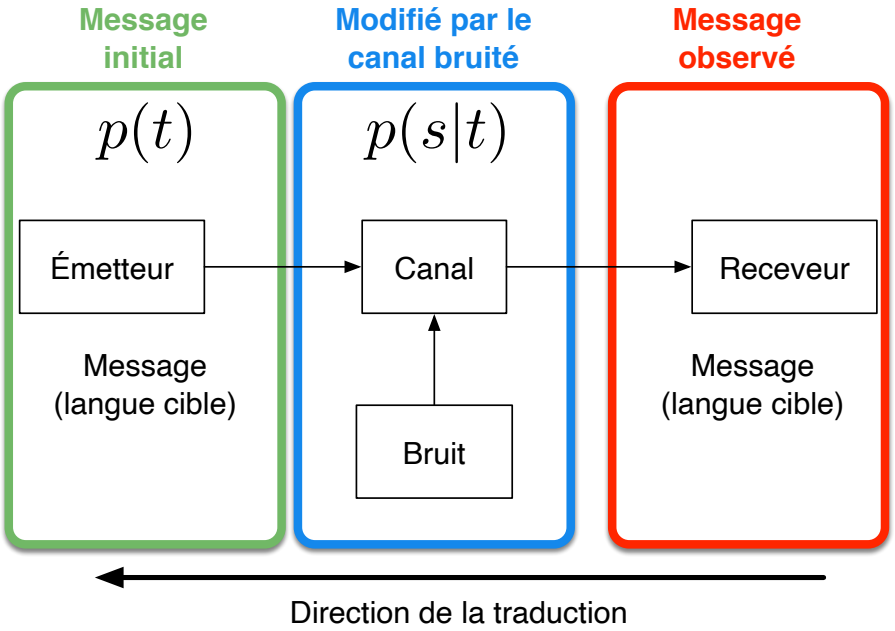
Le Mans Université

Plan

- Principes de la traduction automatique
- Principes fondamentaux des réseaux de neurones / deep learning
- Les plongements de mots ou « embeddings »
- Traduction automatique neuronale
- Recherches actuelles
- Discussion

Principes de la traduction automatique

Canal Bruité



When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Warren Weaver (1949)

Théorie de la Communication, Shannon, 1948

Traduction automatique

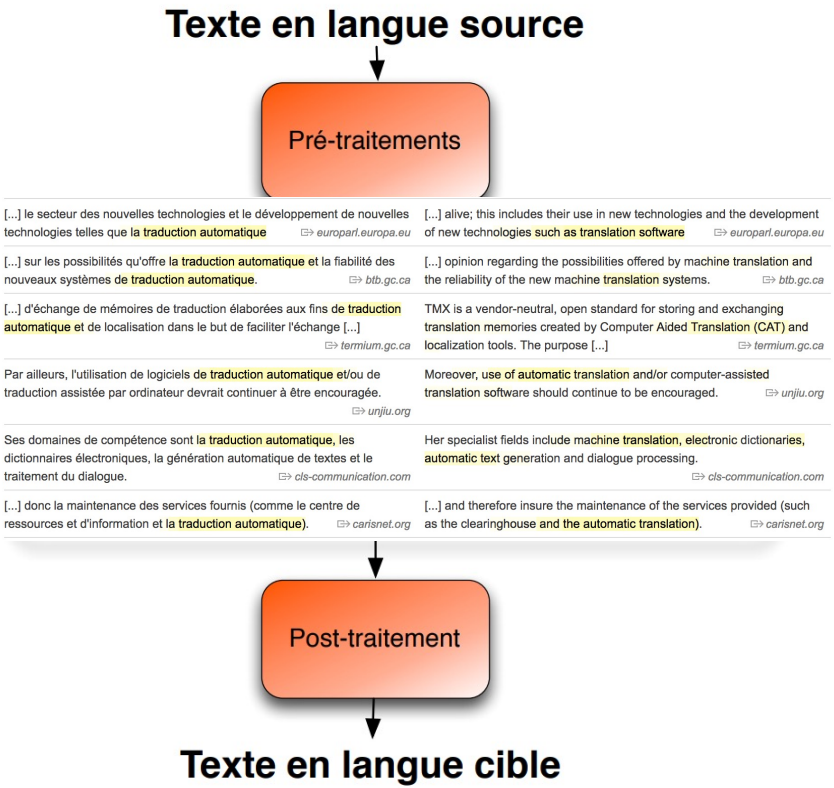
- Traduire une phrase = sélectionner parmi les **traductions possibles** celle qui est la **mieux formée**

$$t* = \underset{t}{\operatorname{argmax}} \quad p(s|t) \quad p(t)$$

- Modèle de traduction : $p(s|t)$ -> peu importe si la phrase est fluide en langue cible
- Modèle de langue : $p(t)$ -> peu importe si la phrase est une bonne traduction
- Combinaison : une traduction fluide en langue cible.

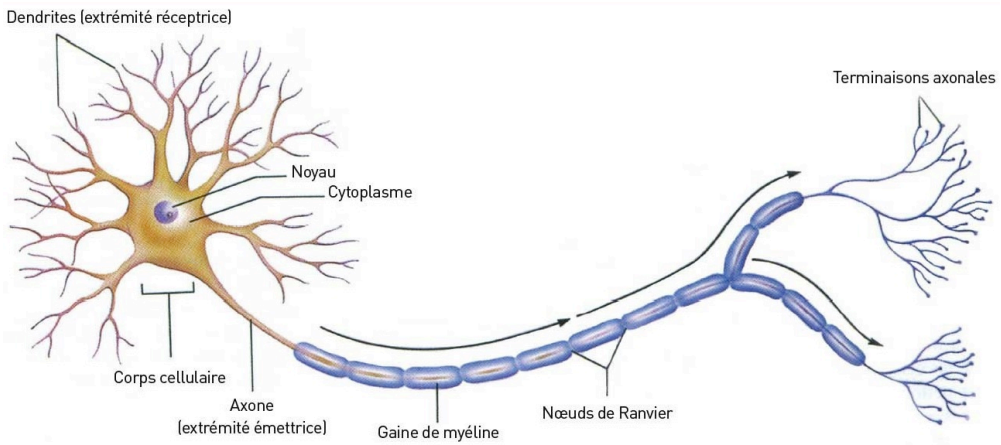
Différentes approches

- À base d'exemple
- À base de segment (phrase-based)
- À base de réseaux de neurones
  - Deep learning



Principes du deep learning

Neurone biologique



- Réception des stimulations des autres cellules via les **synapses**
- Lorsque la somme des stimulations dépasse un seuil
  - la cellule envoie une impulsion électrique le long de son axone : **dépolarisation**
- La transmission de l'impulsion électrique se fait par des **neurotransmetteurs**
  - Les bouts de l'axone touchent presque le corps ou les **dendrites** d'autres cellules

Apprentissage dans le cerveau humain

- Modification de la force des connexions, ajout ou suppression de connexions

Principe de Hebb : “***Neurons that fire together, wire together***”

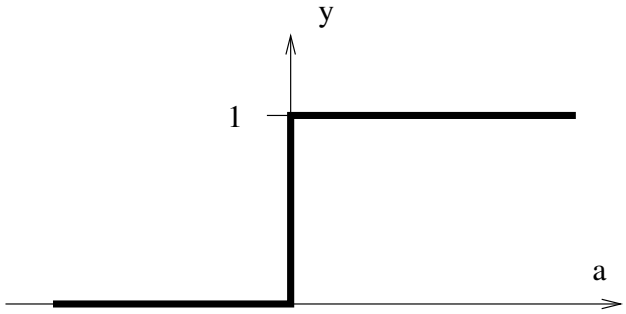
- ➢ Les neurones actifs en même temps vont avoir tendance à renforcer leurs liens
- ➢ Processus local

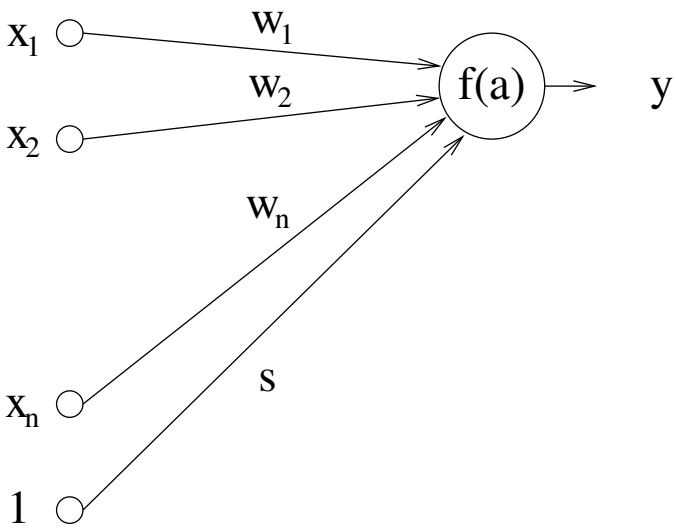
- Tout cela se fait sans aucune supervision
- Par l’expérience et retour (feedback) de l’environnement

Neurone informatique : le perceptron

- Rosenblatt (1957)
- Petite unité de calcul vaguement inspirée par le fonctionnement du cerveau humain

- f : fonction seuil

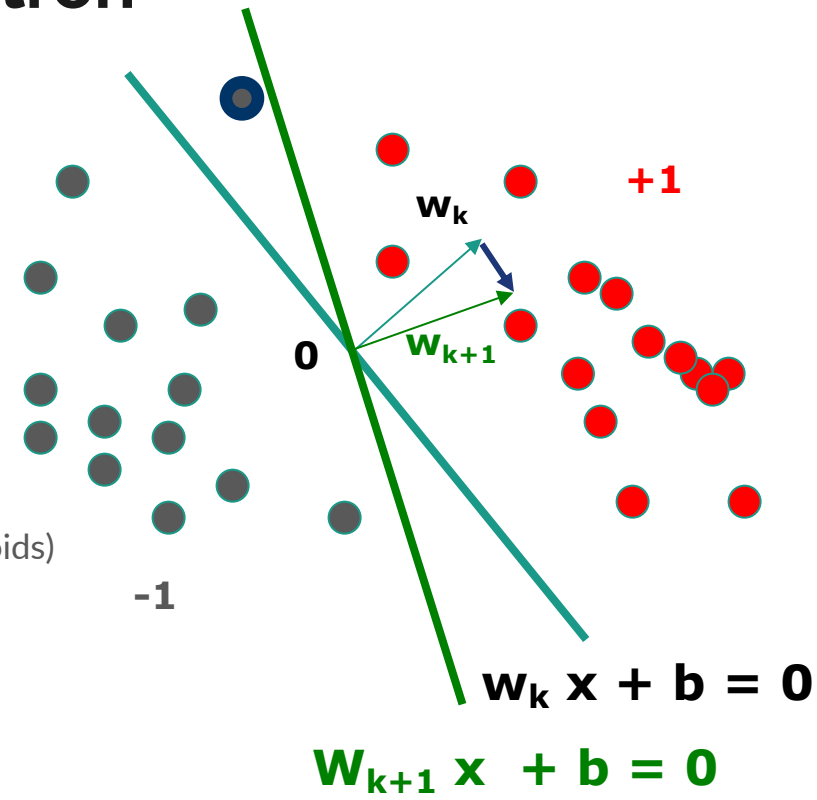




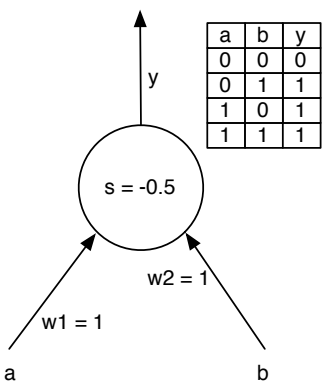
Apprentissage d'un perceptron

- Tâche de classification

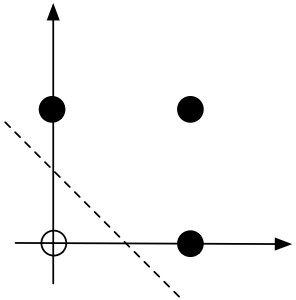
- Parcours des exemples d'apprentissage
- Calcul de la sortie pour chaque exemple
- Si la sortie est erronée, on modifie les paramètres (poids)

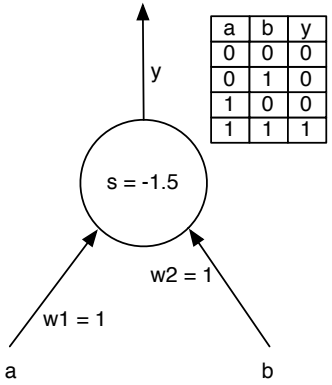


Problème linéaire vs. non-linéaire

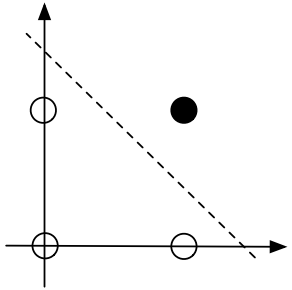


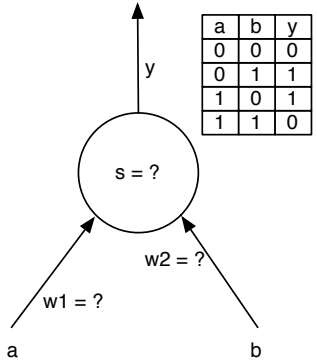
a	b	y
0	0	0
0	1	1
1	0	1
1	1	1



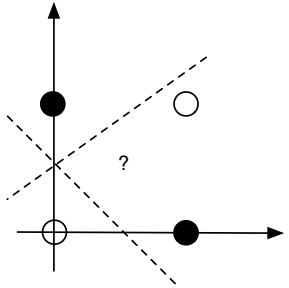


a	b	y
0	0	0
0	1	0
1	0	0
1	1	1

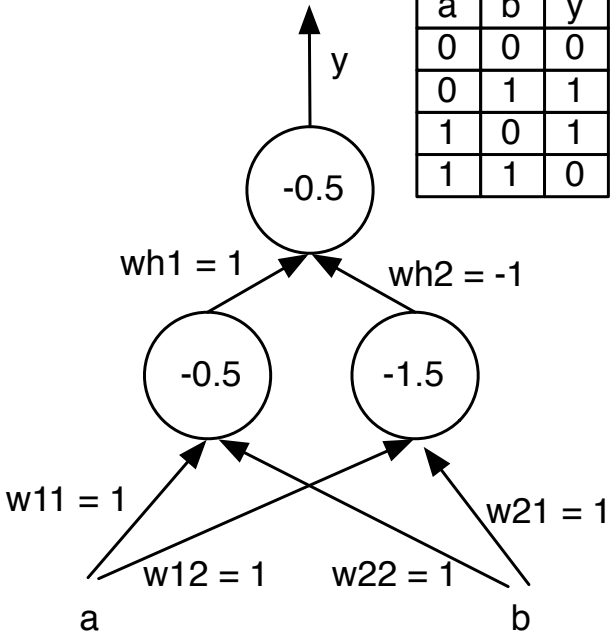




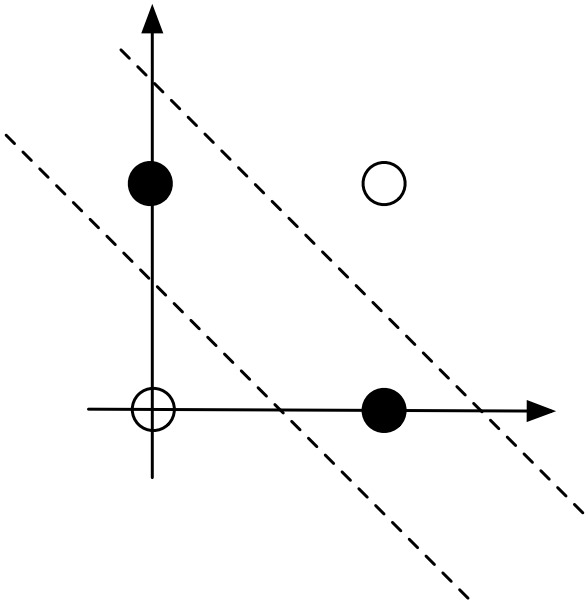
a	b	y
0	0	0
0	1	1
1	0	1
1	1	0



Perceptron multi-couche (MLP)

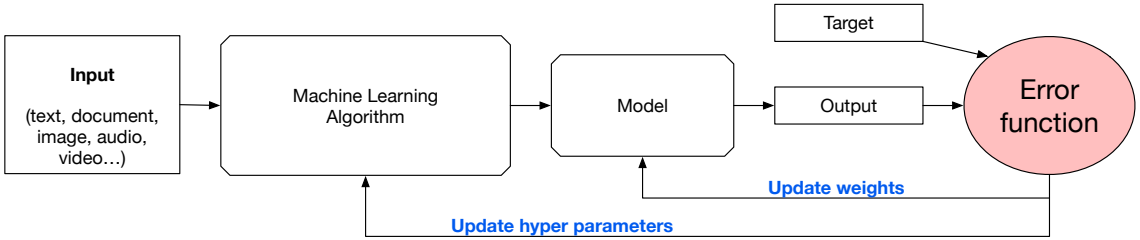


a	b	y
0	0	0
0	1	1
1	0	1
1	1	0



Apprentissage d'un MLP : rétro-propagation

- Fonction d'erreur en sortie : probabilité du mot
  - Entropie croisée (cross-entropy)
- Calcul du gradient (dérivée partielle) : $\frac{\partial E}{\partial W}$



- Mise à jour des poids :

$$W^{nouv} = W - \lambda \frac{\partial E}{\partial W}$$

- **Objectif : modifier les poids du réseau de neurones afin de minimiser l'erreur**

Représentation des mots

- vocabulaire
- 1-hot
- embeddings

Vocabulaire

- Vocabulaire : liste des mots connus par le système
  - Un mot = suite de caractères sans espace
- Quelques pré-traitements / normalisations
  - Tokenization L'école → L' école
  - Normalisation (guillemets, apostrophes, etc)

- Mot inconnu → mot spécial **<unk>**

- Techniques spécifiques pour couvrir le maximum de mots :
  - Byte Pair Encoding : les mots sont découpés (analyse statistique)
    - Ex.: maison => mais@@ on
    - Ex.: maisonnette => mais@@ on@@ nette

  - Opérer au niveau des caractères

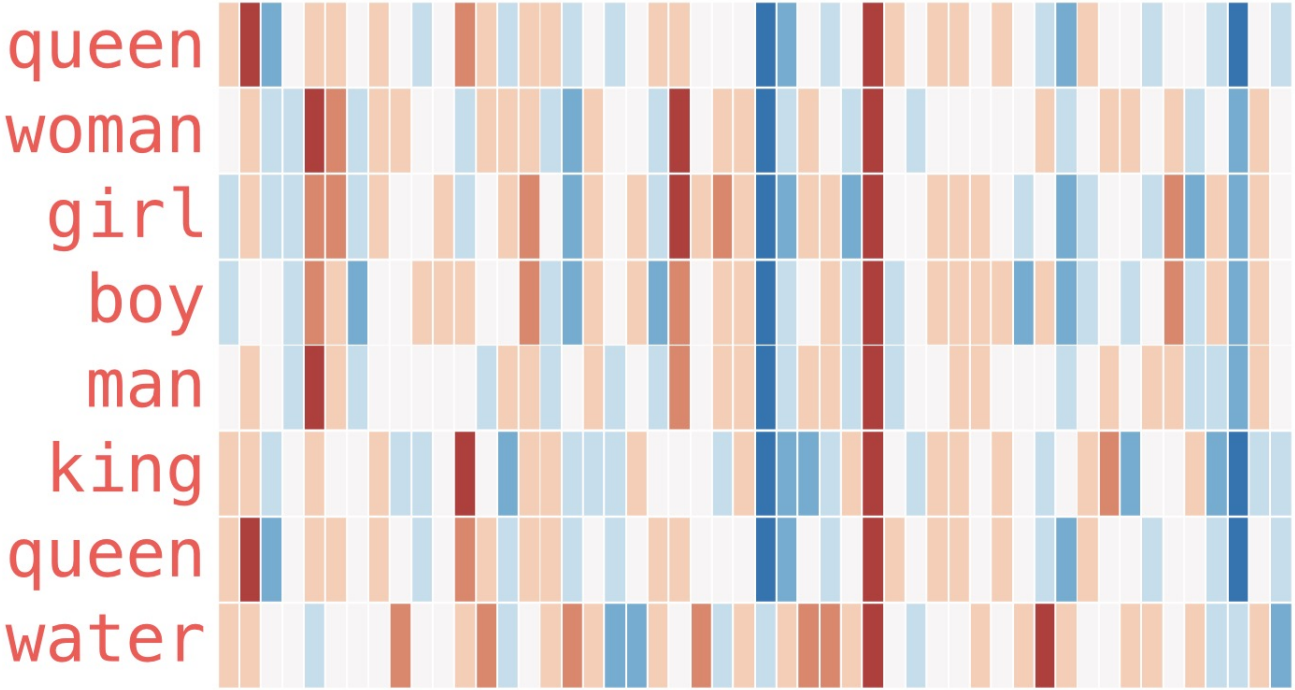
1-hot

- **Codage** spécial d'un mot : on met un 1 à l'indice réservé pour ce mot
- À chaque mot correspond un indice

la = [1,0,0,0,0,…,0]
le = [0,1,0,0,0,…,0]
… …. …
bonjour = [0,0,0,1,0,…,0]
… …. …
Zorro = [0,0,0,0,…,0,1]

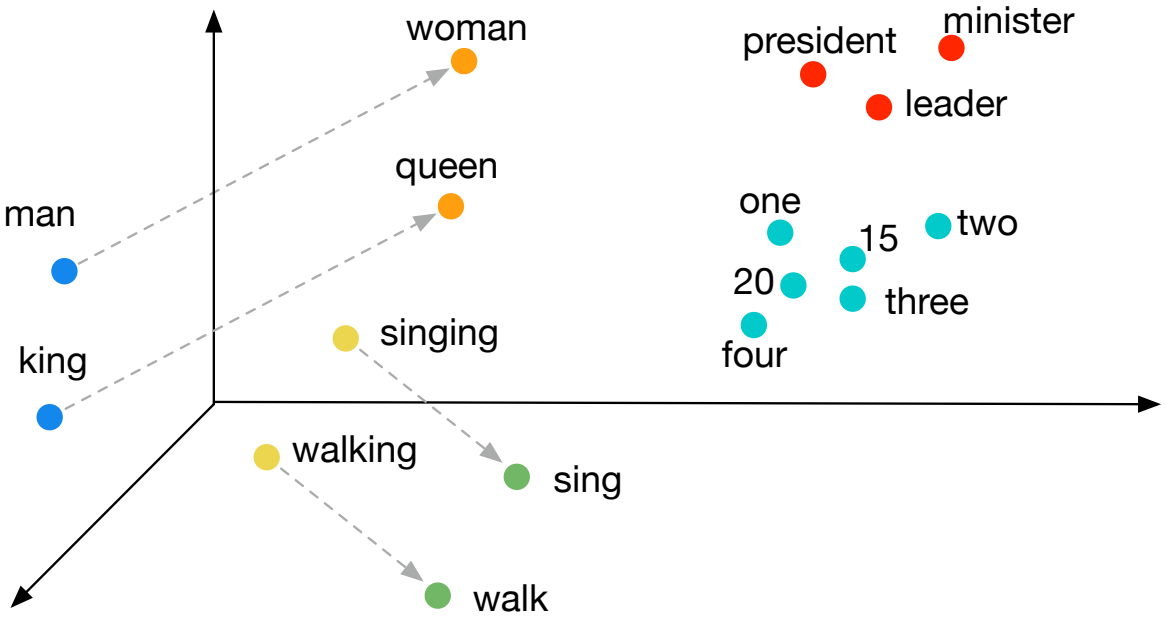
- Ne véhicule aucune sémantique
- chaque mot est à égale distance des autres

Embeddings (plongements de mots)



Source: http://jalammar.github.io/illustrated-word2vec

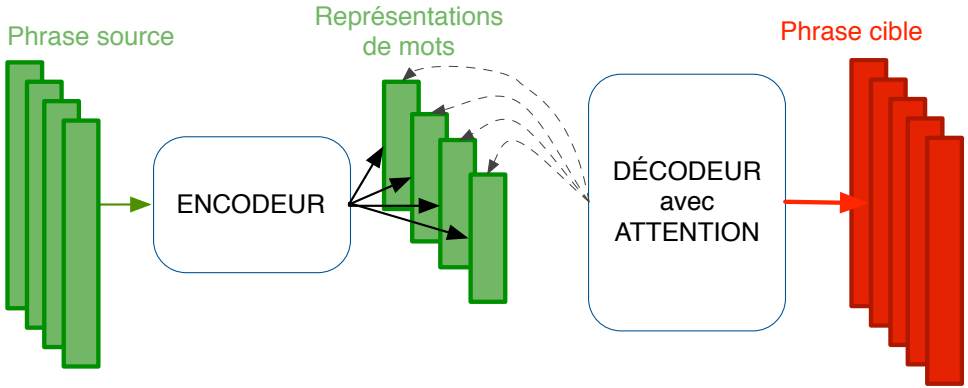
Embeddings : quelques propriétés



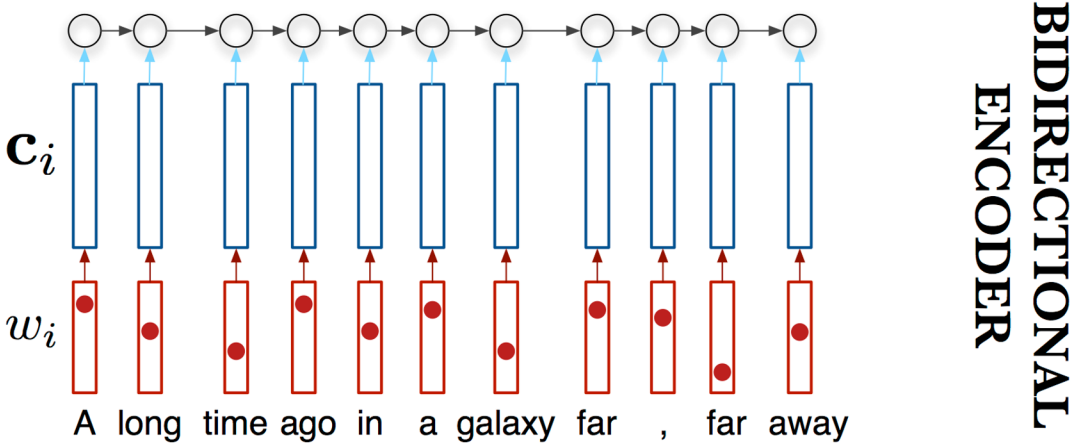
Traduction automatique neuronale
-
Neural machine translation (NMT)

Traduction automatique neuronale

- Modèle séquence vers séquence de type encodeur-décodeur
- Encodeur – Décodeur
  - Encodeur : transforme la phrase en une séquence de vecteurs de faible dimension (quelques centaines)
  - Décodeur : génère la phrase de sortie mot après mot à partir de la séquence d'entrée encodée
- Équipé d'un mécanisme d'attention
  - Permet de se concentrer sur une partie de la phrase d'entrée pour générer le mot cible suivant

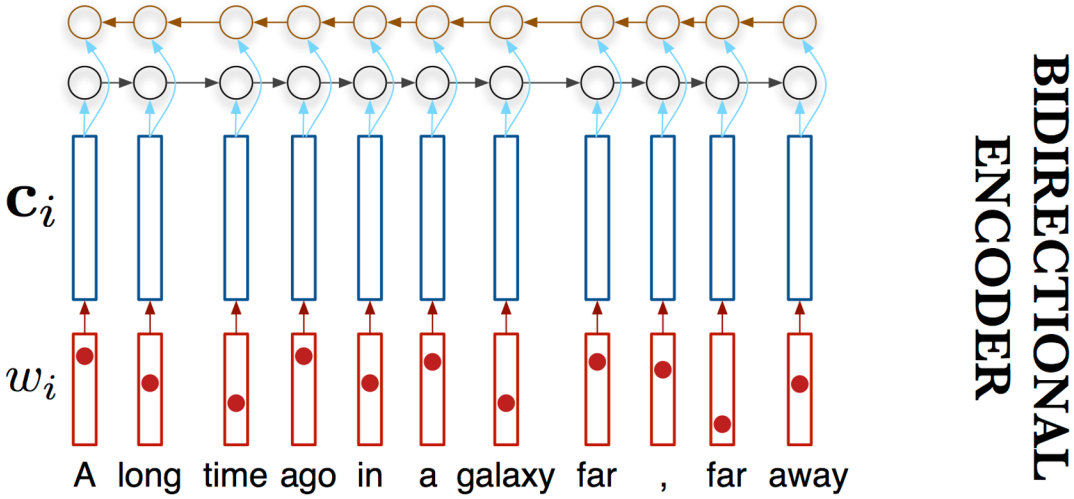


NMT : encodeur récurrent



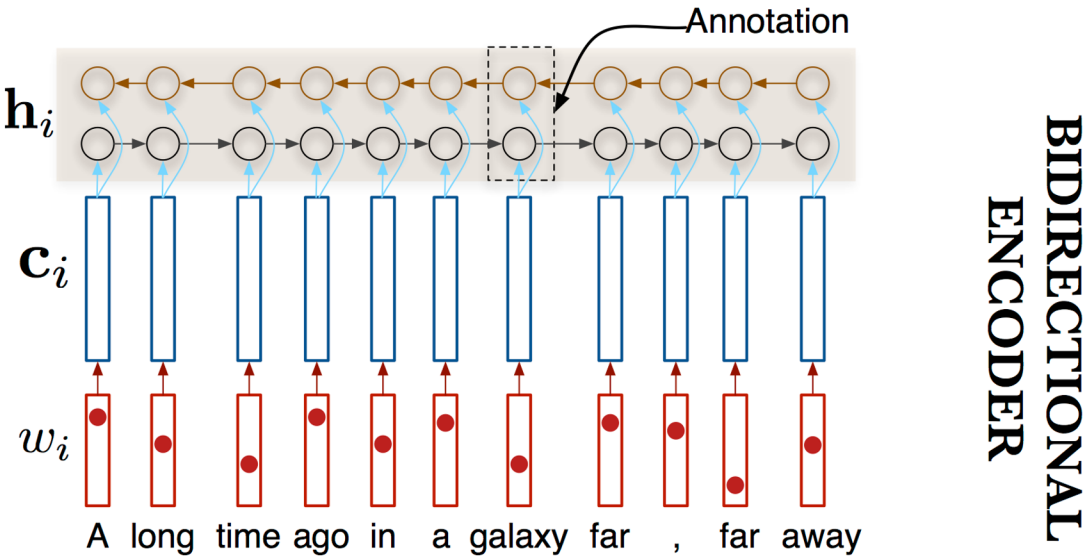
1. Encodage 1-hot + projection embedding + mise-à-jour vecteur RNN **gauche-droite**

NMT : encodeur récurrent bidirectionnel



1. Mise à jour RNN **droite-gauche**

NMT : encodeur récurrent bidirectionnel

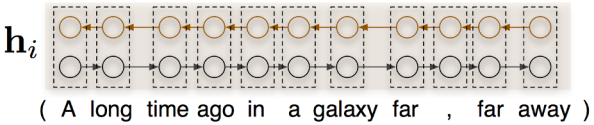


2. **Annotation** = concaténation des vecteurs **gauche-droite** et **droite-gauche**
   Chaque $\mathbf{h}_i$ encode la phrase entière avec un focus sur le i-ème mot

NMT : principe

2. Décodeur récupère les annotations et génère la traduction mot par mot.

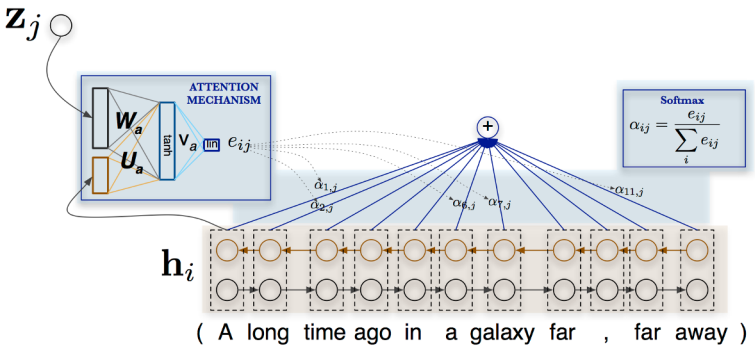
DECODER



NMT : principe

2. Décodeur récupère les annotations et génère la traduction mot par mot.
3. Poids d'attention sont calculés
   a. Petit réseau de neurones
   b. Moyenne pondérée

$$\tilde{\mathbf{h}}_\mathbf{j} = \sum_i \alpha_{ij} \mathbf{h}_i$$

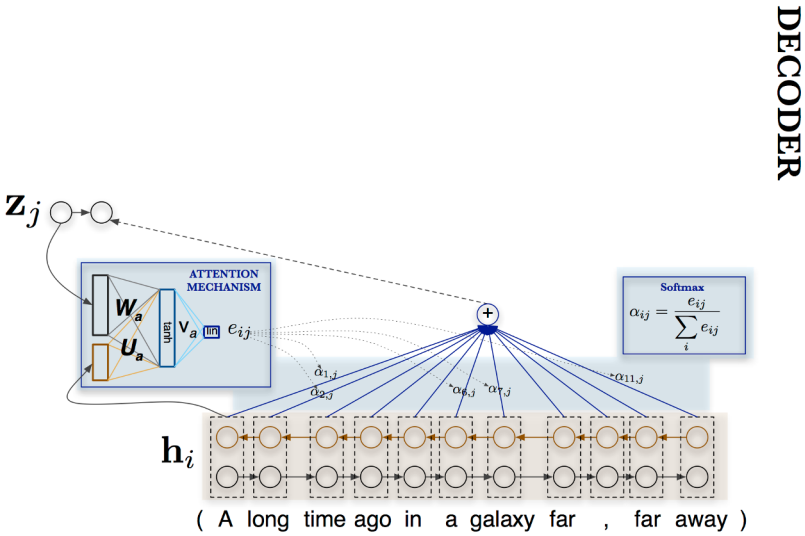


NMT : principe

2. Décodeur récupère les annotations et génère la traduction mot par mot.
3. Poids d'attention sont calculés
   a. Petit réseau de neurones
   b. Moyenne pondérée

$$\tilde{\mathbf{h}}_j = \sum_i \alpha_{ij} \mathbf{h}_i$$

4. Mise à jour RNN du décodeur

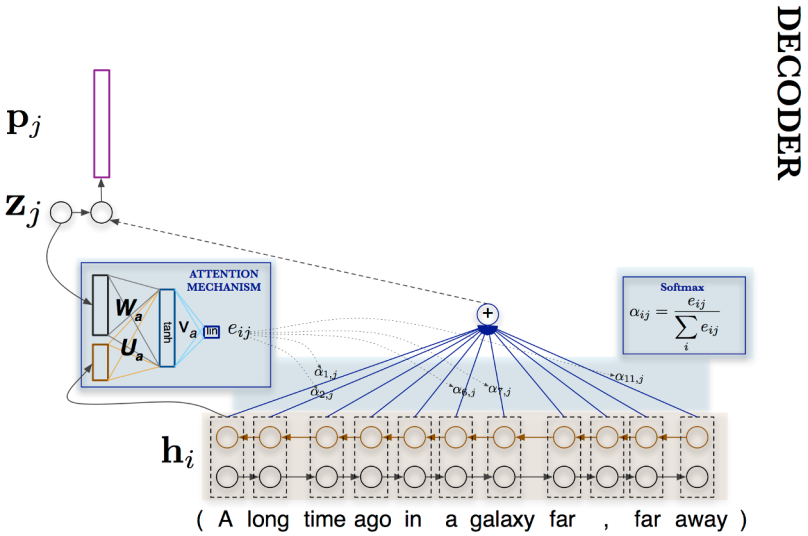


NMT : principe

2. Décodeur récupère les annotations et génère la traduction mot par mot.
3. Poids d'attention sont calculés
   a. Petit réseau de neurones
   b. Moyenne pondérée

$$\tilde{\mathbf{h}}_\mathbf{j} = \sum_i \alpha_{ij} \mathbf{h}_i$$

4. Mise à jour RNN du décodeur
5. Distribution de probabilité pour tous les mots
6. Génération du mot suivant
   a. Le plus probable
   b. Ou recherche par faisceau (beam)

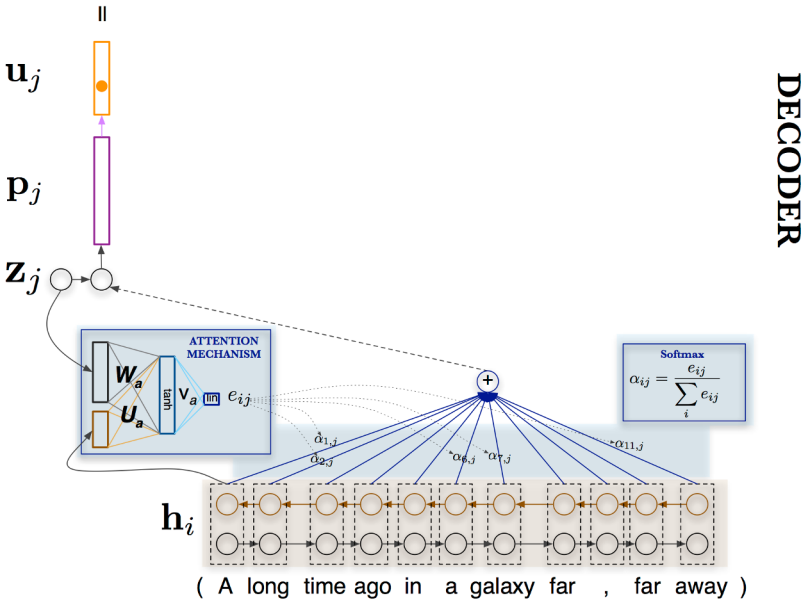


NMT : principe

2. Décodeur récupère les annotations et génère la traduction mot par mot.
3. Poids d'attention sont calculés
   a. Petit réseau de neurones
   b. Moyenne pondérée

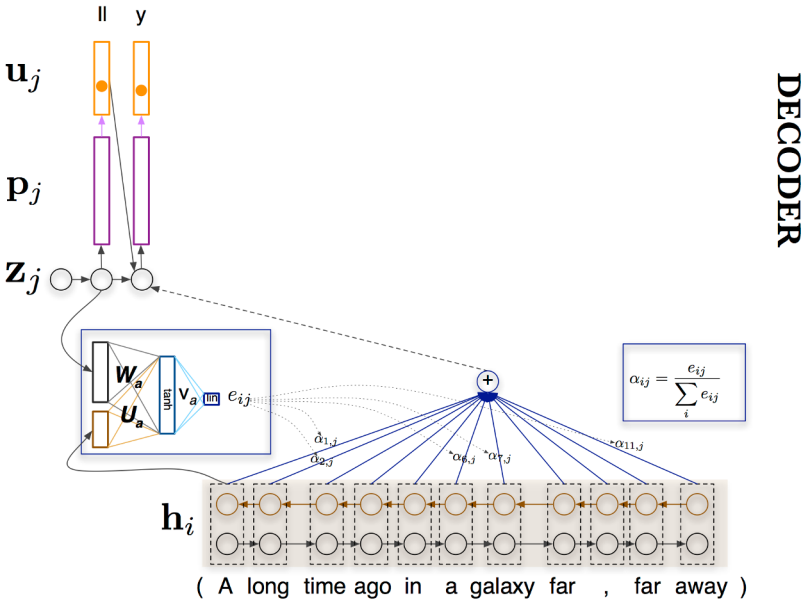
$$\tilde{\mathbf{h}}_j = \sum_i \alpha_{ij} \mathbf{h}_i$$

4. Mise à jour RNN du décodeur
5. Distribution de probabilité pour tous les mots
6. Génération du mot suivant
   a. Le plus probable
   b. Ou recherche par faisceau (beam)



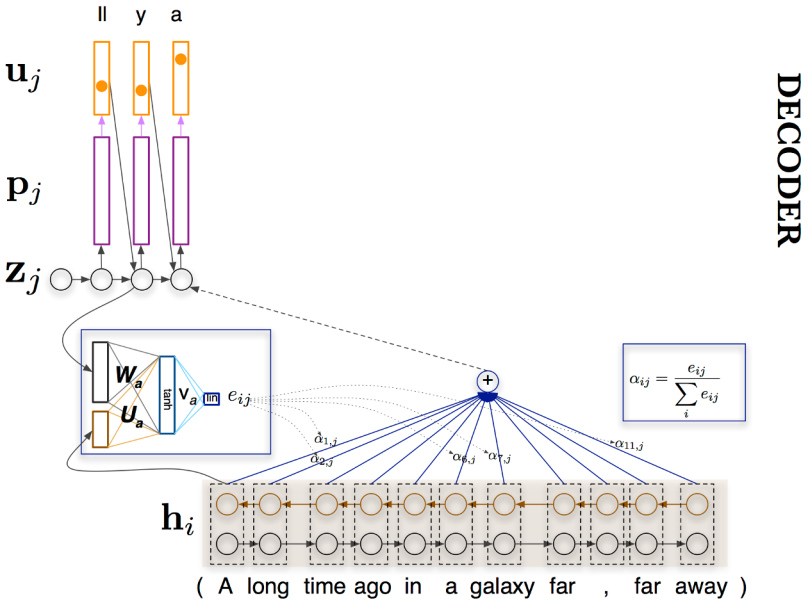
NMT : principe

- Le RNN décodeur utilise le context source et l'embedding du mot généré à l'étape précédente pour générer le mot suivant



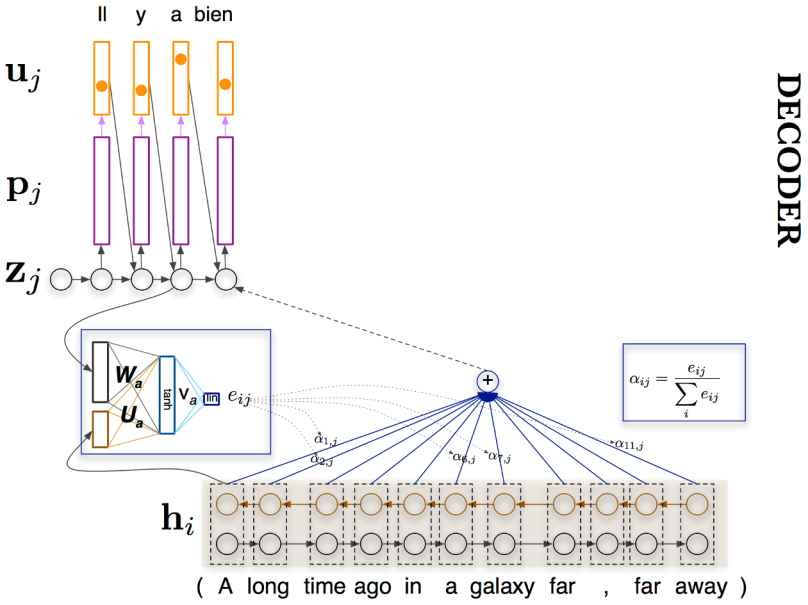
NMT : principe

- À chaque étape, un nouvel ensemble de poids d'attention est calculé
- **Les annotations ne changent pas !**
- **La mémoire du RNN décodeur a changé !**



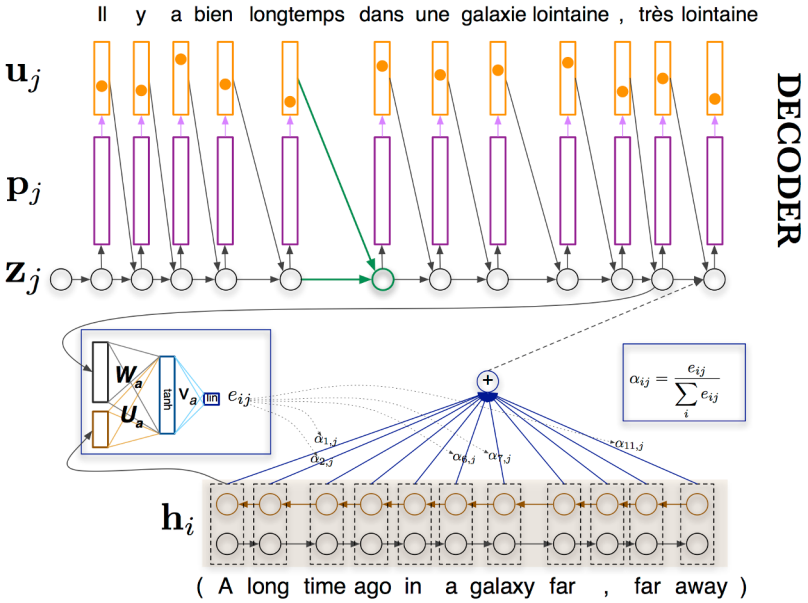
NMT : principe

- À chaque étape, un nouvel ensemble de poids d'attention est calculé
- **Les annotations ne changent pas !**
- **La mémoire du RNN décodeur a changé !**
- **Le décodeur est un modèle de langue conditionnel**



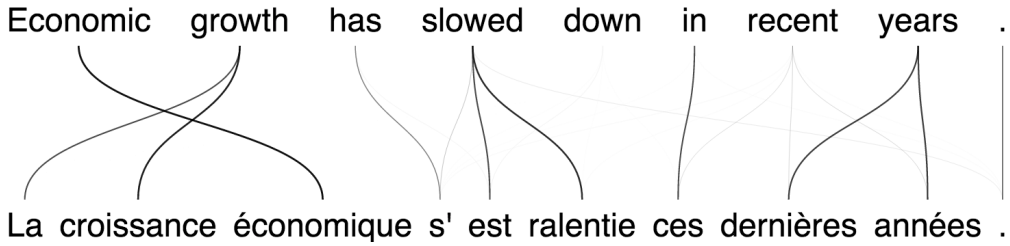
NMT : principe

- À chaque étape, un nouvel ensemble de poids d'attention est calculé
- **Les annotations ne changent pas !**
- **La mémoire du RNN décodeur a changé !**
- **Le décodeur est un modèle de langue conditionnel**
- Et ainsi de suite…
- … jusqu'à ce que le mot EOS (fin de séquence) soit généré.

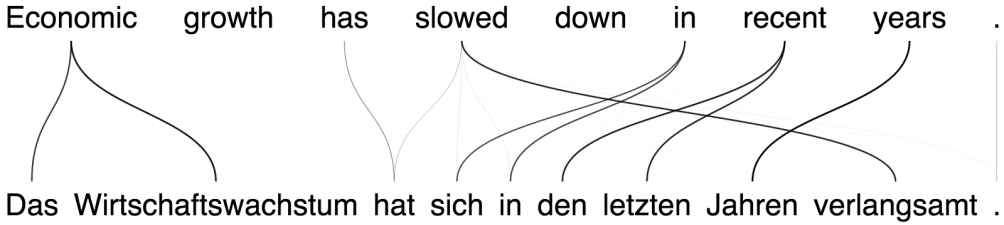


Attention => alignement lexical

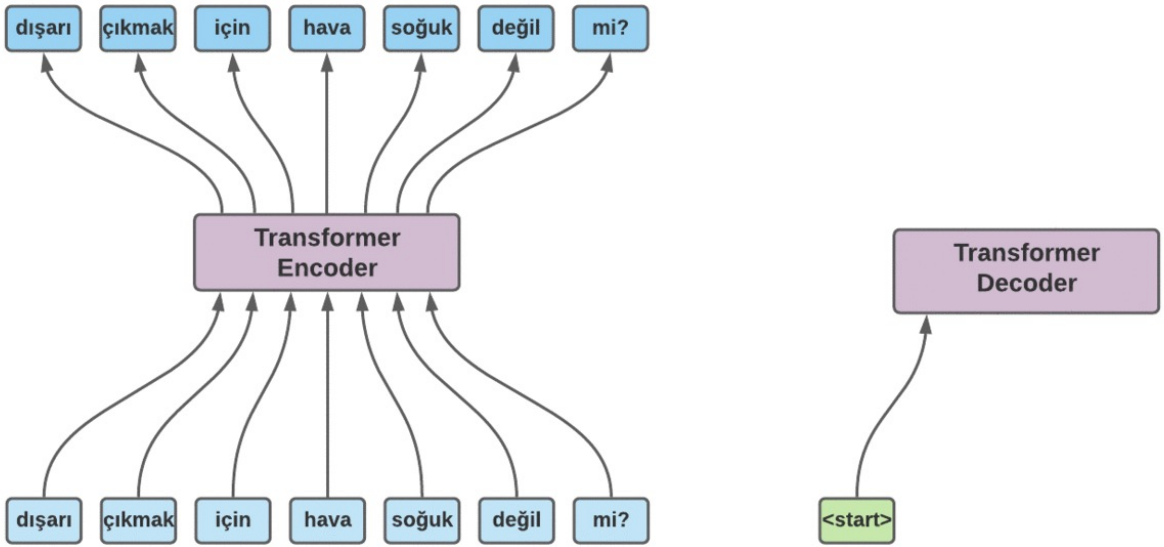
English-French



English-German



Modèle à base de transformers



Source : https://towardsdatascience.com/neural-machine-translation-inner-workings-seq2seq-and-transformers-229faff5895b

Recherches actuelles

Traduction automatique

- Les systèmes actuels fonctionnent au niveau de la phrase
  - Un **contexte** plus étendu est nécessaire
  - Contexte intra-textuel (document, anaphores, etc…)
  - Contexte extra-textuel (par ex. interlocuteurs pour le sous-titrage multilingue)

- Intégration de bases de connaissances
  - Est-il possible d'en extraire le **sens commun** ?

- Modèles multilingues
  - Langues peu dotées

- Modèle exploitant des informations linguistiques
  - Ex. : grammaire, morphologie

Traduction multimodale

- Objectif : exploiter plusieurs modalités
- Ex.: texte et vision
- A baseball player in a black shirt just tagged a player in a white shirt.
- Un joueur de baseball en maillot noir vient de toucher un joueur en maillot blanc.
- Une joueuse de baseball en maillot noir vient de toucher une joueuse en maillot blanc.

Traduction de la parole

- Systèmes en cascade
- Système neuronal direct “ end-to-end ”

Traducteurs et traduction automatique

Des questions se posent !

Courrier international

EN CE MOMENT GILETS JAUNES PRÉSIDENTIELLE RDC SYRIE GRAND DÉBAT HOROSCOPE

Langues. Les traducteurs automatiques finiront-ils par remplacer les professionnels ?

CULTURE 〉 LE TEMPS - LAUSANNE

Publié le 16/02/2017 - 13:05

Veille CFTTR

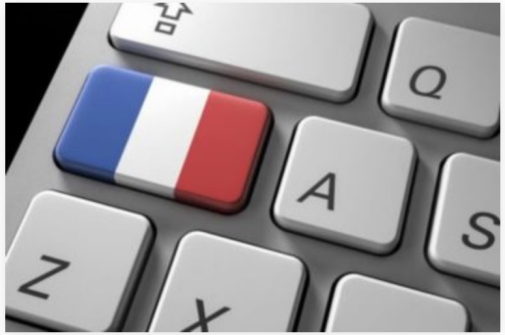
Le site de veille du Centre de Formation des Traducteurs, Terminologues et Rédacteurs techniques

Le métier de traducteur : condamné à disparaître ?

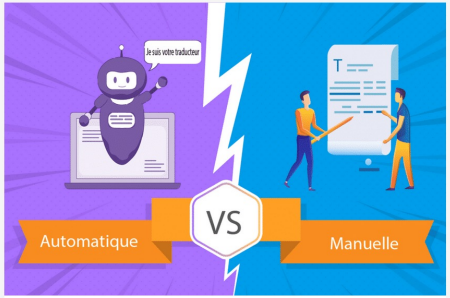
beelingwa SERVICES

Traducteurs et interprètes : tous chômeurs en 2025 ?

Après la Traduction Assistée par Ordinateur (TAO).....place à la Traduction Automatique (TA) !



Le traducteur « humain » a-t-il encore un avenir ?



TRADUCTION AUTOMATIQUE : POURQUOI LES MACHINES NE REMPLACERONT-ELLES JAMAIS LES TRADUCTEURS HUMAINS ?

Quelles sont donc les complications et limites des systèmes de traduction automatique ? Ces systèmes peuvent-ils vraiment remplacer les traducteurs humains ? Découvrez les réponses à ces questions et bien d'autres encore dans cet article.

Traducteurs vs. Post-éditeurs

TradOnline
More than translations

Tout sur la traduction automatique et la post-édition !

par Tradonline | Mar 20, 2018 | Nos services, Technologies | 0 commentaires

Traduire

Revue française de la traduction

222 | 2010 : Traduire pour le théâtre

Hors cahier thématique

La post-édition : l'avenir incontournable du traducteur ?

La post-édition :
l'avenir incontournable du traducteur ?

Anne-Marie Robert

SFT (Société française des traducteurs)

Ma vision

- La traduction automatique reste un **outil**
  - à **maîtriser** et utiliser à bon escient
  - permet d'améliorer la productivité (éventuellement)

- Traducteur = post-éditeur ?
  - la traduction automatique reste une suggestion

- MT: bonnes performances sur des phrases simples, ne nécessitant pas ou peu de **contexte**
  - moins performante pour des textes plus complexes (poésie, littérature,…)

- [À l'heure actuelle,] les ordinateurs ne **comprennent** rien !
  - Que veut dire comprendre … ?

Discussion / questions ?