

Colloque INTERLAB – 26 octobre 2022

La traduction neuronale dans une perspective interdisciplinaire

Nicolas Ballier & Maud Bénard

Cette communication présentera quelques travaux interdisciplinaires conduits au sein de l'équipe CLILLAC-ARP conduits autour de la traduction neuronale. Nous montrerons comment s'articule une collaboration de linguistes et de traductologues avec des informaticiens, des mathématiciens et des linguistes computationnels.

Nous montrerons comment nous avons essayé de nous approprier toute une partie de la chaîne de production de la traduction automatique neuronale:

a. la collecte ou la valorisation de données linguistiques

- la constitution de mémoires de traduction (projet OCTAVE)
- la valorisation de ressources terminologiques existantes (projet ARTES)
- la constitution de corpus, coeur de cible de l'activité d'une grande partie de l'équipe
- la réflexion sur la phase de sous-tokenisation : les textes subissent toute une série de transformations avant d'être "appris" par la machine. Nous explorons les conséquences de ces transformations (projet ComplexNeuroViz).

b. la production de modèles de traduction et l'analyse des traductions générées

- l'analyse des biais dans la traduction à partir de l'étude du flux informationnel (projet NeuroViz)
- l'appropriation des méthodes de production des modèles de traduction et leur spécialisation (projet SPECTRANS)
- l'acquisition d'une plate-forme pour la traduction neuronale et la production d'expériences de deep learning (projet PAPTAN)

c. la constitution d'outils et de scripts d'analyse

- outil d'aide à la décision (préselection automatique de textes médicaux à inclure pour constituer un corpus d'entraînement en fonction des textes à traduire)
- outil d'aide au calcul du score automatique (réflexion sur les différentes métriques automatiques d'évaluation de la qualité de la traduction et leur limite)
- outils de visualisation et d'extraction d'information à partir du réseaux de neurone

Nous présenterons quelques-uns des résultats obtenus, en recherche fondamentale et en recherche appliquée. Nous détaillerons quelques exemples de résultats et les perspectives qui se dégagent d'un point de vue sociétal. Au plan théorique, nous donnerons des exemples de biais architecturaux pour la traduction du genre français/anglais et les problèmes que continuent à poser pour la traduction les groupes nominaux complexes. D'un point de vue sociétal, nous montrerons comment se déploient les problématiques de cette technologie : quelle culture de la donnée développer ?