

Le traitement automatique du langage naturel.

Samuel Cruz-Lara Sociétaire de l'ALS.

Le traitement automatique du langage naturel (TALN) ou (NLP) pour *Natural Language Processing*, est un domaine pluridisciplinaire impliquant la linguistique et l'informatique - dont l'intelligence artificielle. Son objectif est de permettre l'interaction concrète, performante et efficace entre les ordinateurs et les humains, en utilisant les langues naturelles.

Pour l'atteindre, dans un contexte désormais incontournable d'apprentissage automatique (*machine learning*), nous verrons que les ordinateurs doivent être capables de traiter de grands volumes d'information en langues naturelles, tout en tenant compte de nombreuses nuances contextuelles de celles-ci et du multilinguisme au sein d'un même groupe social ou d'un même territoire.

Je commencerai en donnant une brève description du TAL, en soulignant la problématique et les enjeux de ce domaine pluridisciplinaire, particulièrement complexe.

J'évoquerai ensuite, brièvement, l'histoire du TAL entre les années 1950 et 1990 (approche symbolique), entre les années 1990 et 2010 (approche statistique) et, à partir des années 2010, l'approche neuronale.

Je présenterai enfin, de manière très synthétique, mes activités de recherche dans le TAL au sein de l'équipe SYNALP du LORIA (UMR 7503). Ceci permettra de placer le TAL dans des contextes divers comme la normalisation, le multilinguisme, l'analyse de sentiments et la détection d'émotions, les mondes virtuels (*le metaverse*) ou encore l'analyse de textes, y compris des textes n'étant pas écrits en langage naturel.