TALN 2ème partie

Résumé

Le **Traitement Automatique du Langage Naturel (TALN)** est une discipline de l'intelligence artificielle qui permet aux machines de comprendre, analyser et générer du texte. Pour traiter des séquences de mots, les premiers modèles développés étaient les **Réseaux de Neurones Récurrents (RNN)**. Bien que ces modèles soient adaptés aux données séquentielles, ils présentent des limites, notamment une perte d'information sur les longues phrases.

Pour pallier ce problème, les **Long Short-Term Memory (LSTM)** ont été introduits. Ces réseaux améliorés sont capables de retenir des informations sur de plus longues distances dans une phrase, ce qui les rend plus performants pour des tâches comme la **traduction automatique** et la **modélisation du langage**. Toutefois, ces modèles restent complexes à entraîner et lents à exécuter.

Une avancée majeure a été l'introduction du **modèle Encodeur-Décodeur (Seq2Seq)**, qui transforme une phrase source en une représentation numérique avant de générer sa traduction. Pour améliorer encore la qualité des traductions, le **mécanisme d'attention** a été intégré, permettant au modèle de se concentrer sur les parties les plus importantes d'une phrase, plutôt que de traiter toute la séquence de manière uniforme.

Les **Transformers** ont ensuite révolutionné le domaine en éliminant les contraintes des RNN et LSTM. Grâce au **mécanisme d'auto-attention**, ils analysent simultanément l'ensemble du texte sans suivre un ordre strict. Cette approche améliore à la fois la rapidité et la précision des modèles. Des technologies avancées comme **GPT** (utilisé pour générer du texte) et **BERT** (conçu pour comprendre le contexte des mots) sont basées sur cette architecture.

Aujourd'hui, le TALN est utilisé dans de nombreuses applications : **traduction automatique**, **chatbots**, **résumé automatique de texte** et **correction grammaticale**. Malgré leur efficacité, les modèles récents comme les Transformers nécessitent une grande puissance de calcul et d'importants volumes de données pour fonctionner de manière optimale.

Terme	Définition
TALN (Traitement Automatique du Langage Naturel)	Domaine de l'IA qui permet aux machines de comprendre et générer du texte.
Tokenisation	Processus de découpage d'un texte en unités plus petites (mots, phrases).
Word Embedding	Technique qui transforme les mots en nombres pour capturer leur signification et leurs relations.
RNN (Réseau de Neurones Récurrent)	Modèle d'IA qui analyse le texte séquentiellement mais qui oublie rapidement les premières informations d'une phrase.
LSTM (Long Short-Term Memory)	Version améliorée des RNN qui conserve mieux les informations sur le long terme.
Seq2Seq (Encodeur-Décodeur)	Modèle qui transforme une phrase en représentation numérique avant de générer une traduction.
Mécanisme d'Attention	Méthode permettant à un modèle de se concentrer sur les parties les plus importantes d'un texte.
Transformer	Modèle avancé qui remplace les RNN et LSTM, offrant un traitement plus rapide et précis du texte.
Auto-Attention	Technique qui permet à un modèle de comprendre les relations entre tous les mots d'une phrase simultanément.
GPT (Generative Pre-trained Transformer)	Modèle d'IA basé sur les Transformers, utilisé pour générer du texte de manière autonome.
BERT (Bidirectional Encoder Representations from Transformers)	Modèle basé sur les Transformers, conçu pour analyser le sens des mots en tenant compte de leur contexte.