Machine Learning, méthodes et solutions

Données à dimension variable : Traitement du langage

Traitement du langage : données

Classification :

- thème/genre (gutenberg.org : 57k livres)
- auteur (gutenberg.org : 57k livres)
- sentiment (Kaggle movie review : 222k commentaires rotten tomatoes)
- reconnaissance d'entités nommées (Kaggle Annotated Corpus for NER : 1.3M tags)
- ...

Traitement du langage : données

Compréhension :

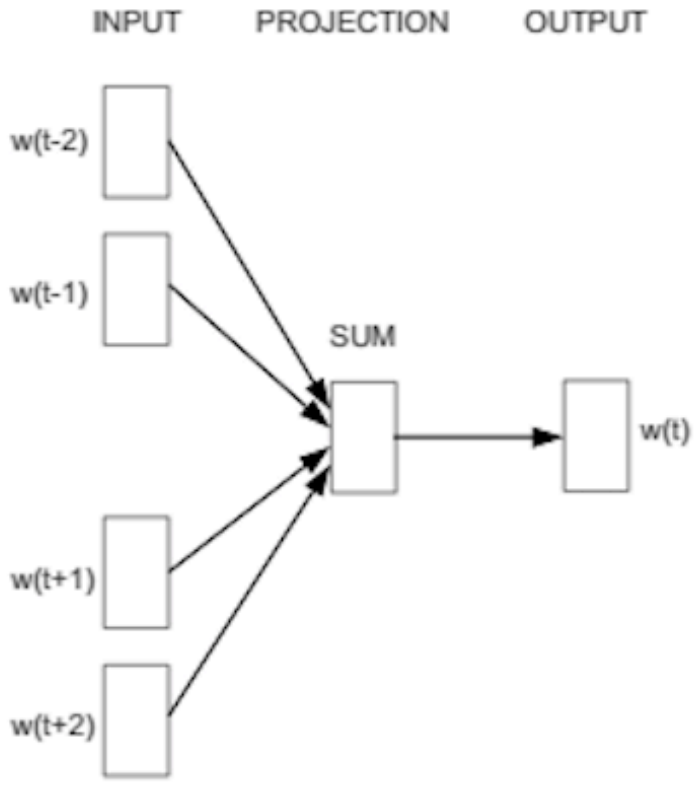
- Question/réponses (SQUAD : 150k questions)
- Traduction (europarl : 450k phrases alignées)
- ...

Traitement du langage : word embeddings

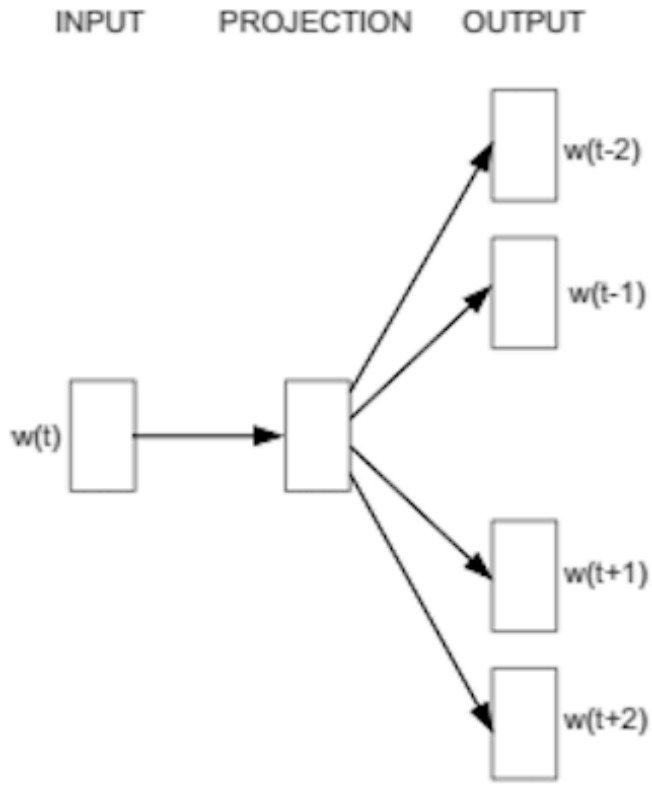
mot = indice dans un dictionnaire (dimension > 30000)
mot = vecteur “sémantique” (dimension < 1000)

- word2vec
- CBOW/Skip-Gram
- Thought vector (pour des phrases ou même des documents entiers)
- ...

Traitement du langage : word embeddings

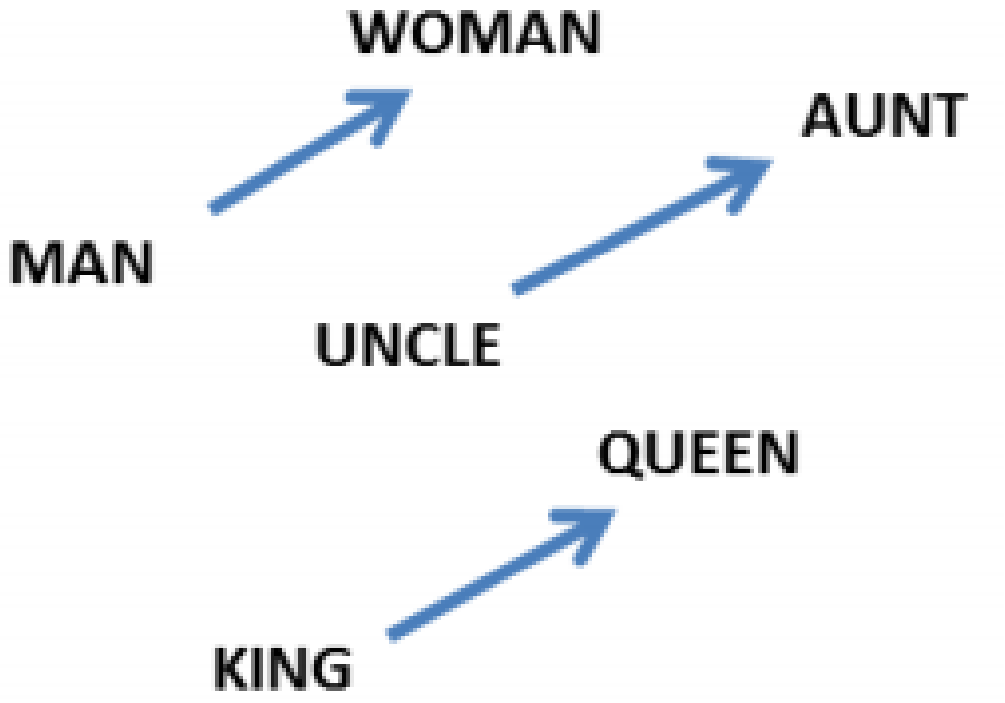


CBOW



Skip-gram

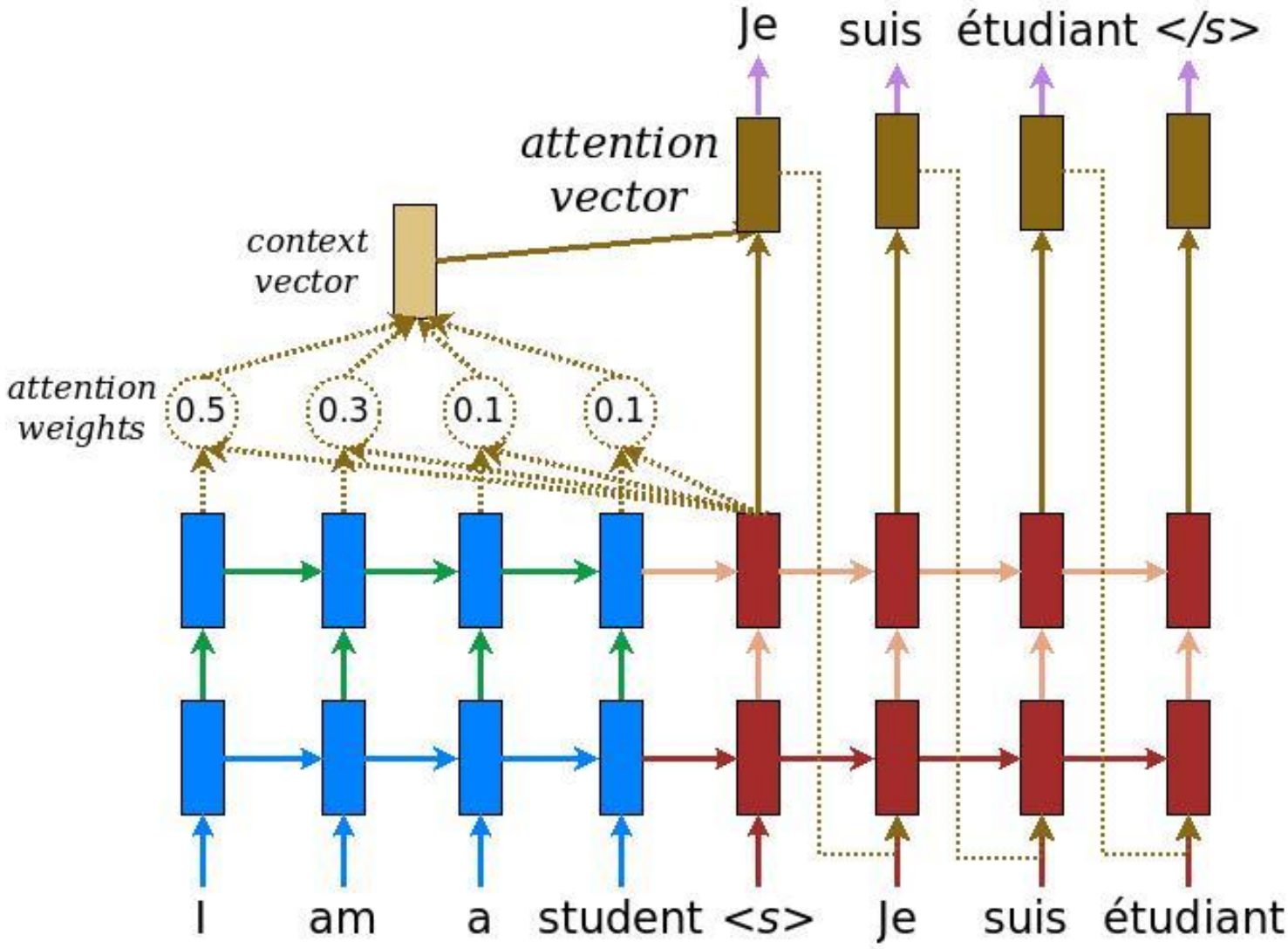
Traitement du langage : word embeddings



Démo dans l'espace word2vec

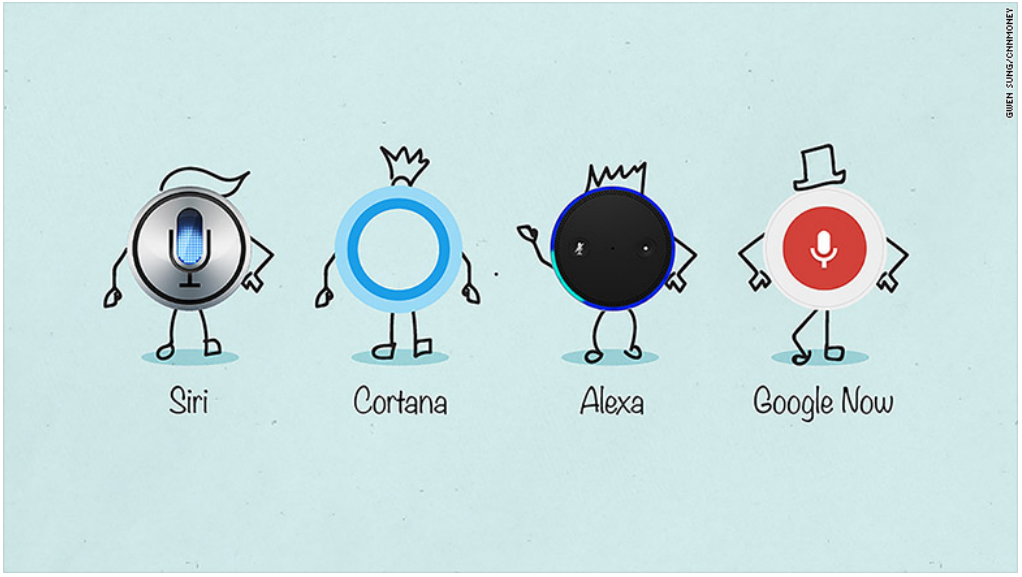
Visualisation de l'espace word2vec

Traitement du langage : Modèle à attention



Traitement du langage : Champs d'application

Transcription et synthèse de la parole

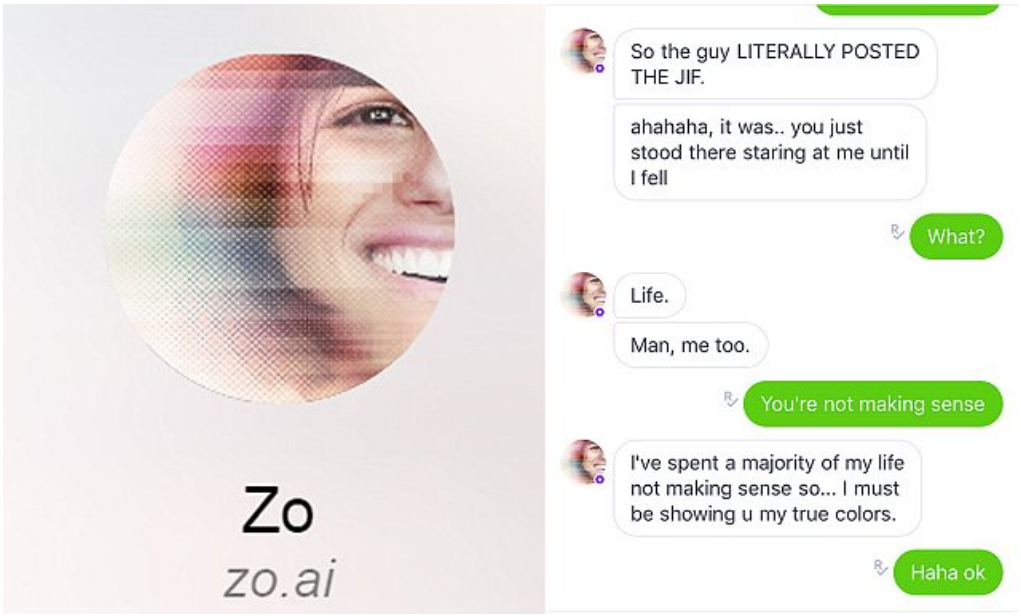


Traitement du langage : Champs d'application

Identification du locuteur

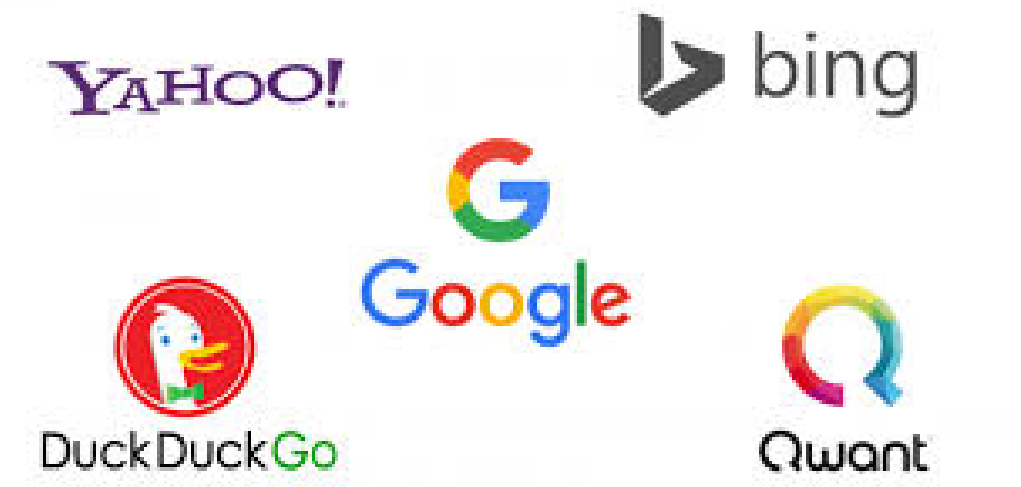
Traitement du langage : Champs d'application

Chatbot



Traitement du langage : Champs d'application

Moteur de recherche



Traitement du langage : Champs d'application

Extraction de données

Traitement du langage : Champs d'application

Analyse de sentiments

Traitement du langage : Champs d'application

Résumé

Input: Article 1st sentence	Model-written headline
metro-goldwyn-mayer reported a third-quarter net loss of dlrs 16 million due mainly to the effect of accounting rules adopted this year	mgm reports 16 million net loss on higher revenue
starting from july 1, the island province of hainan in southern china will implement strict market access control on all incoming livestock and animal products to prevent the possible spread of epidemic diseases	hainan to curb spread of diseases
australian wine exports hit a record 52.1 million liters worth 260 million dollars (143 million us) in september, the government statistics office reported on monday	australian wine exports hit record high in september

Traitement du langage : Champs d'application

Traduction

Traduire **français** (langue identifiée)

Bonjour,

Nos filles souhaitent faire un échange ensemble dans le cadre du programme Brigitte Sauzay. Avant de passer au dossier administratif, nous souhaitions vous poser quelques questions pour s'assurer que nous envisagions cet échange entre nos filles de la même façon.

Le professeur d'allemand de Gaïa a gentiment accepté de traduire nos questions afin d'éviter des incompréhensions à cause de problèmes de traduction.

Concernant les dates, Emma nous a dit que vous souhaitiez que Gaïa soit repartie avant le 19 juin car vous partiez en vacances.

Traduire en **anglais**

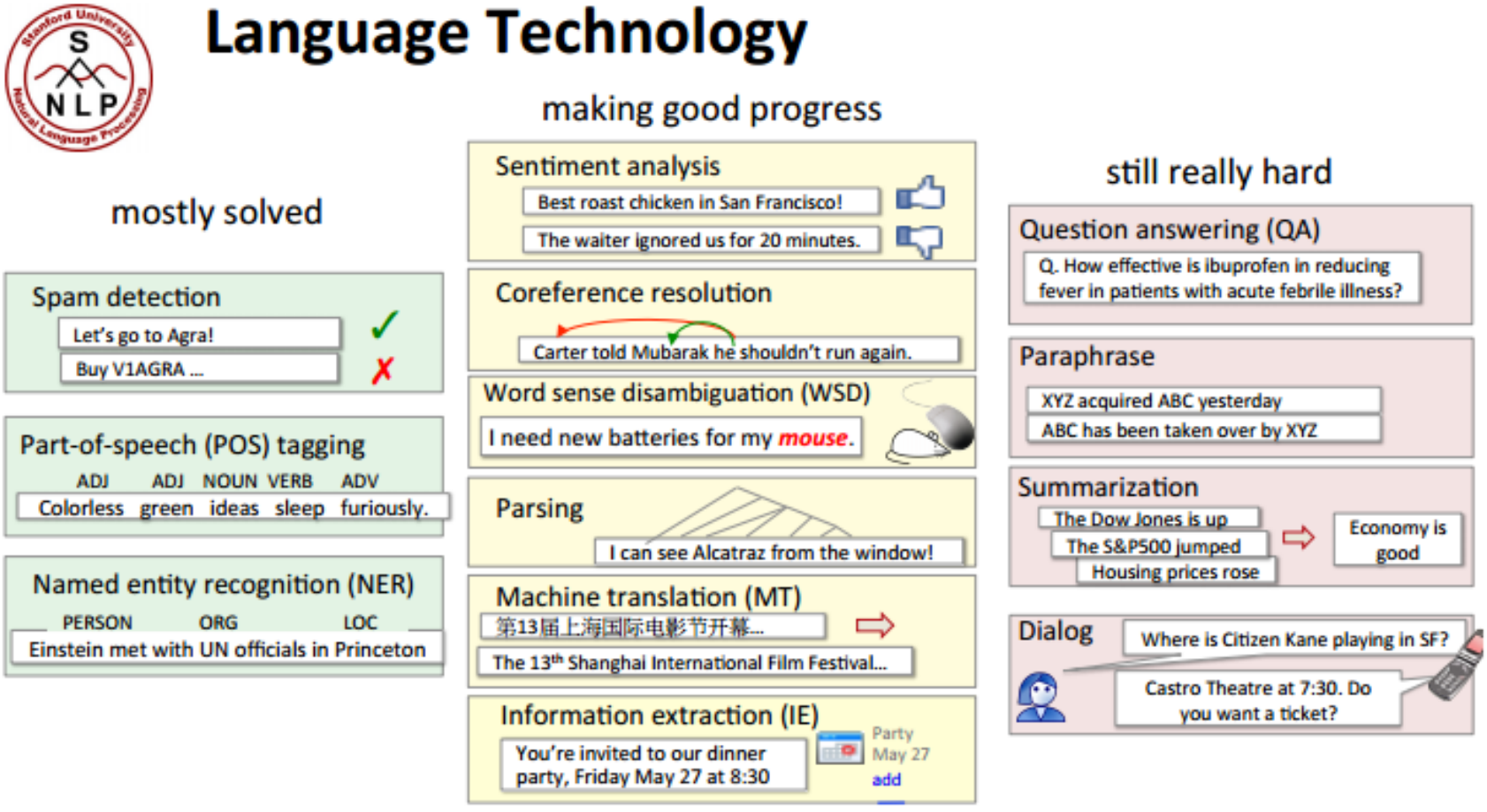
Hello,

Our daughters would like to do an exchange together as part of the Brigitte Sauzay program. Before moving on to the administrative file, we wanted to ask you a few questions to ensure that we consider this exchange between our daughters in the same way.

Gaïa's German teacher kindly agreed to translate our questions in order to avoid misunderstandings due to translation problems.

Regarding the dates, Emma told us that you wanted Gaïa to leave before June 19 because you were going on vacation.

Traitement du langage : Champs d'application



$\approx$ 2015

Traitement du langage : Champs d'application