

AAP AMI'24

Amélioration de la fouille d'argument par la synthétisation d'information contextuelle

Anaïs Ollagnier - Junior Fellow in AI - Équipe Wimmics

Introduction & contexte

La fouille d'arguments

Qu'est-ce que c'est ?

*“La fouille d'arguments est l'**identification automatique et l'extraction de la structure d'inférence et de raisonnement** exprimée sous forme d'arguments présentés en langage naturel.”*

Composants

- **Revendication** (“claim”): la déclaration principale ou la thèse que l'auteur de l'argument cherche à soutenir ou à défendre.
- **Prémisse** (“premise”): une déclaration ou une assertion qui est présentée comme un élément de preuve ou de soutien pour étayer la revendication principale.

Domaine d'applications: Analyse d'opinions, de controverses, de citations, explainabilité, etc.

Comment sont-ils détectés/identifiés ?

Dans la recherche académique...

Identification et/ou détection des composants d'arguments

La fouille d'arguments se décompose en plusieurs sous-tâches: détection de phrases argumentatives, détection des limites des composants d'argument ou prédiction de la structure argumentative;

La plupart des approches reposent sur l'utilisation de caractéristiques issus de l'analyse **syntactique**, **sémantique** et **discursives** ;

À chaque tâche sa méthode

La *détection de phrases argumentatives* est fondamentalement une **tâche de classification de phrases** (la prédiction de subjectivité, la classification de questions, et l'analyse de sentiment).

La détection des *limites des composants d'argument* est une **tâche d'étiquetage de séquences** (reconnaissance d'entités nommées et d'autres applications de segmentation de texte).

la prédiction de la *structure argumentative* est similaire à des **tâches de prédiction de liens** (classification de relations dans l'analyse du discours, l'estimation de similarité sémantique textuelle, et différentes applications de l'implication textuelle).

Parmi les méthodes récentes, on a recours à l'**apprentissage profond** avec des Modèles de Langage à Réseau Neuronal Récurrent.

Comment sont-ils détectés/identifiés ?

Dans la recherche académique...

Données et Annotations

Pour pouvoir mener des expériences de fouille d'arguments, l'**accès à des corpus étiquetés est essentiel** ;

Il n'existe pas de corpus de référence communément accepté pour cette tâche ; les auteurs collectent et étiquettent généralement leurs propres données ;

La plupart des études portent sur la **langue anglaise** et sont de **petite taille**;

Ce que l'on constate...

- Appréhension de la variabilité linguistique et culturelle **difficile**.
- La polyvalence des méthodes employées est **limitée**.
- Le prise en compte du contexte entourant l'argumentation est **restreinte**.

La pragmatique

“Une” définition

“Étude de l’usage du langage, en particulier au rôle du contexte”

- prise en compte du **locuteur** (plus particulièrement de ses intentions)
- au **contexte** (le sens immédiat et les inférences qui peuvent être faites lorsque le sens littéral du message est confronté à la situation – au contexte au sens large).

Qu’est ce qu’elle englobe ?

- Indices encodés grammaticalement ?
- Indices non verbaux (i.e., indices prosodiques, postures, silences, etc.) ?
- Indices dérivés des théories sur les actes de langage (intentions, croyance, désir) ?

La pragmatique en TAL

Dans la recherche académique...

Une étude publiée en 2021 révèle qu'entre 2004 et 2018 seulement 29 articles (**8,14%** des documents analysés) ont eu pour objet la modélisation de la dimension pragmatique dans ce domaine.

La plupart des approches se fonde sur les **théories de la sémantique pragmatique** dont l'objet est le sens pragmatique porté explicitement au niveau du sens littéral (ex : l'utilisation des guillemets pour véhiculer une intention sarcastique).

Cependant, l'orientation des récents travaux, publiés après 2018, se porte vers de nouveaux axes de recherche tels que l'introduction de **méthodes de probing**, la **formalisation des processus de raisonnement** ou encore **l'implémentation d'aspects se détachant de la structure linguistique** (i.e., sens commun, implicatures conversationnelles ou encore théories des actes de langage).



AAP AMI : proposition d'un postdoctorat

Contexte du Projet

Accueil

- Laboratoire I3S
- Équipe SPARKS/Wimmics

Thématiques

- **Traitement Automatique des Langues:**
 - Extraction de connaissances
 - Génération des ancres contextuelles
- **Linguistique formelle:**
 - Fondements théoriques
 - Formalisation des connaissances contextuelles

Collaborations potentielles

- Céline Poudat, linguiste et maître de conférences en analyse du discours à l'Université Côte d'Azur (Laboratoire BCL)

Projet de Recherche

Améliorer la robustesse des techniques de **fouille d'arguments** via la **synthétisation d'informations contextuelles** relevant des **aspects multimodaux** de la communication.

Projet de Recherche

Améliorer la robustesse des techniques de **fouille d'arguments** via la **synthétisation d'informations contextuelles** relevant des **aspects multimodaux** de la communication.

RQ1.

Quelles informations abstraire pour restituer la valeur pragmatique d'un énoncé ?

Projet de Recherche

Améliorer la robustesse des techniques de **fouille d'arguments** via la **synthétisation d'informations contextuelles** relevant des **aspects multimodaux** de la communication.

RQ1.

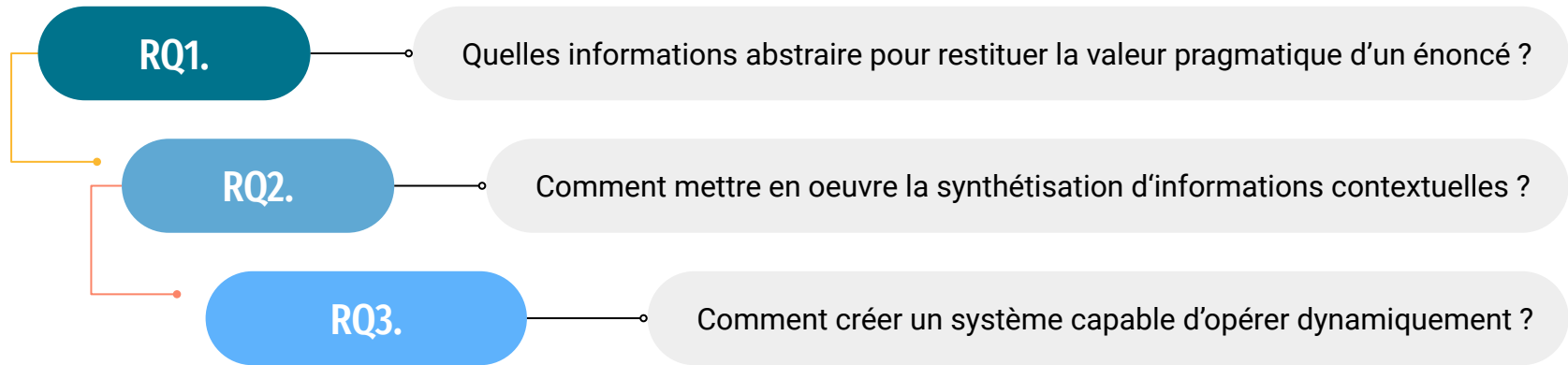
Quelles informations abstraire pour restituer la valeur pragmatique d'un énoncé ?

RQ2.

Comment mettre en oeuvre la synthétisation d'informations contextuelles ?

Projet de Recherche

Améliorer la robustesse des techniques de **fouille d'arguments** via la **synthétisation d'informations contextuelles** relevant des **aspects multimodaux** de la communication.



Merci de votre attention

Questions?

aollagnier.github.io

