



# Détection de désinformations et détection automatique d'inférences textuelles et de contradictions: nouveaux jeux de données pour le français et l'intérêt des approches hybrides et logiques



Doctorant: Maximos Skandalis

Co-encadré par: Richard Moot, Christian Retoré, Simon Robillard

12/07/2024

Journée scientifique de l'ICO



# Axes de la thèse

**Apprentissage profond et méthodes formelles pour la  
détection automatique d'énoncés contradictoires -  
application à la détection de désinformations**



# Axes de la thèse

## Apprentissage profond et méthodes formelles pour la détection automatique d'énoncés contradictoires - application à la détection de désinformations

1. Nouveaux jeux de données pour le français
2. Évaluer des modèles d'apprentissage profond
3. Obtenir la représentation logique des phrases des jeux de données
4. Utiliser un prouveur automatique pour détecter les contradictions



# Plan

- I. **Tâche et contexte**
- II. Jeux de données
- III. Évaluation
- IV. Approches hybrides/logiques

# Contexte

---

## Désinformation

- **Il existe plusieurs approches pour la lutte contre la désinformation:**
  - Détecter le langage radical ou chargé;
  - Regarder les réactions/commentaires à des tweets;
  - Caractériser les sources comme fiables / non fiables;
  - Détecter des rumeurs ou des techniques de propagande.

## Notre approche des désinformations

- **Nous nous intéressons aux désinformations en tant qu'informations contradictoires à une autre information.**
- **Nous ne nous positionnons pas par rapport à la vérité de l'un ou de l'autre énoncé comparés.**

## Contradictions

- **En TALN, la définition la plus courante est que deux énoncés sont contradictoires s'ils ne peuvent pas être vrais en même temps (dans la même situation).**
- **Tâche de détection automatique de contradictions = tâche de classification binaire de paires de phrases**

# Contexte

---

## Intérêt de la tâche d'inférence textuelle (NLI):

- Elle sert de base pour les tâches nécessitant une compréhension approfondie des différentes nuances de la langue.
- Elle peut aider à:
  - bien identifier les relations textuelles;
  - la tâche de questions-réponses;
  - l'analyse de sentiments;
  - maintenir la fidélité des traductions automatiques.
- Déterminer les relations logiques entre les phrases peut contribuer à une meilleure compréhension du contexte par les systèmes.

## Intérêt de la tâche de détection de contradictions:

- Les contradictions sont sous-représentées dans les jeux de données pour NLI.
- La relation de contradiction est symétrique.

## Motivation:

- Peu de jeux de données dédiés à ces tâches existent en français;
- La tâche d'inférence textuelle et, en particulier, celle de détection des contradictions pourraient également être utilisées pour la détection de fausses informations, ce qui présente un intérêt particulier pour notre projet de recherche.

# La tâche

---

**Sentence-pair binary or multi-class  
classification task**

# La tâche

---

**Sentence-pair** binary or multi-class  
classification task

*Sentence-pair :*

Chaque exemple du jeu de données est composé:

- D'une paire de phrases, à savoir:
  - une prémisse, et
  - une hypothèse;
- Une étiquette.



# La tâche

Sentence-pair **binary or multi-class** classification task

*Binary or multi-class :*

Entailment	Non-entailment	
Entailment	Neutral	Contradiction
Non-contradiction		Contradiction

Pour  
NLI/RTE

Pour la  
détection  
de contra-  
dictions

# La tâche

Sentence-pair binary or multi-class classification task

*Sentence-pair :*

Chaque exemple du jeu de données est composé:

- D'une paire de phrases, à savoir:
  - une prémisse, et
  - une hypothèse;
- Une étiquette.

*Binary or multi-class :*

Entailment	Non-entailment	
Entailment	Neutral	Contradiction
Non-contradiction		Contradiction

Pour NLI/RTE

Pour la détection de contradictions

# Plan

- I. Tâche et contexte
- II. Jeux de données**
- III. Évaluation
- IV. Approches hybrides/logiques

# Jeux de données

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI
- FraCaS

# Jeux de données

---

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

# Jeux de données

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

Nous avons introduit :

- DACCORD
- RTE3-FR
- GQNLI-FR

# Jeux de données

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

Nous avons introduit :

- DACCORD
- RTE3-FR
- GQNLI-FR

1034 paires de phrases extraites  
d'AFP Factuel:

- Guerre Russie-Ukraine (472)
- Covid-19 (450)
- Crise climatique (112)

# Jeux de données

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

Nous avons introduit :

- DACCORD
- RTE3-FR
- GQNLI-FR

1034 paires de phrases extraites  
d'AFP Factuel:

- 515 contradictions
- 519 non contradictions



# Le jeu de données DACCORD

- **Nous avons sélectionné les paires de phrases à la main, en relisant des articles en entier sur AFP Factuel** (structure similaire: titre, résumé, corps avec analyse).
- **Nous avons gardé les phrases qui nous ont paru d'intérêt pour la tâche étudiée.**
- **Nous avons couplé les phrases recueillies :**
  - soit entre elles;
  - soit avec des phrases construites à partir des phrases recueillies, de sorte que la paire satisfasse l'étiquette attribuée.

**303**

Articles relus

**49,81%**

Contradictions

# Jeux de données

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

Nous avons introduit :

- DACCORD
- RTE3-FR
- GQNLI-FR

800+800 paires de phrases:

- 412+410 inférences
- 299+318 cas neutres
- 89+72 contradictions

# Jeux de données

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

Nous avons introduit :

- DACCORD
- RTE3-FR
- GQNLI-FR

300 paires de phrases:

- 97 inférences
- 100 cas neutres
- 103 contradictions

# Jeux de données

---

## En anglais:

- XNLI
- MNLI
- SNLI
- FraCaS
- SICK
- RTE (1, ..., 8)
- GQNLI
- LogicNLI
- ANLI
- WANLI
  
- FOLIO
- PHEME

## En français:

- XNLI (2490 val et 5010 test)
- FraCaS (346 paires de phrases)

Nous avons introduit :

- DACCORD
- RTE3-FR
- GQNLI-FR
- Traductions automatiques de LingNLI et SICK

# Plan

I. Tâche et contexte

II. Jeux de données

**III. Évaluation**

IV. Approches hybrides/logiques

# Évaluation

- Classification à 3 étiquettes

Modèles	RTE3-EN		RTE3-FR		GQNLI		GNLI-FR	
	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1
DistilmBERT <sub>Base</sub> -cased	60,75	47,92	61,13	46,65	26,00	26,03	27,67	26,88
XLm-R <sub>Base</sub>	-	-	60,50	49,61	-	-	31,67	31,46
CamemBERT <sub>Base</sub> , 3-class	-	-	63,13	51,52	-	-	33,67	33,44
mDeBERTa-v3 <sub>Base</sub> , XNLI	67,13	56,26	67,13	55,01	28,33	27,73	28,67	27,94
mDeBERTa-v3 <sub>Base</sub> , NLI-2mil7	71,25	61,33	69,63	60,57	<b>36,67</b>	<b>37,04</b>	<b>38,33</b>	<b>38,58</b>
XLm-R <sub>Large</sub>	<b>72,88</b>	<b>63,62</b>	<b>71,25</b>	<b>62,47</b>	35,33	35,02	36,34	35,94
CamemBERT <sub>Large</sub> , 3-class	-	-	71,13	61,97	-	-	33,33	31,62

- Résultats cohérents avec les jeux de données dans leur langue d'origine

# Évaluation

- Nous avons aussi évalué la bonne prédiction de l'étiquette « Contradiction »

Modèles	DACCORD		XNLI		RTE3-FR		GNLI-FR	
	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1
DistilmBERT <sub>Base-cased</sub>	63,73	52,59	79,98	68,01	79,63	11,89	51,67	19,89
XLM-R <sub>Base</sub>	71,57	67,62	87,17	81,14	77,75	21,93	49,33	23,23
CamemBERT <sub>Base, 3-class</sub>	77,76	76,19	89,64	85,09	80,36	26,29	50,33	<b>36,05</b>
mDeBERTa-v3 <sub>Base, XNLI</sub>	80,75	78,30	90,98	86,39	85,75	30,49	52,00	20,88
mDeBERTa-v3 <sub>Base, NLI-2mil7</sub>	80,95	78,47	90,76	85,89	87,00	38,82	50,67	34,51
XLM-R <sub>Large</sub>	82,01	80,00	<b>96,49</b>	<b>94,74</b>	86,75	41,11	<b>53,33</b>	25,53
CamemBERT <sub>Large, 3-class</sub>	83,27	81,01	92,30	88,12	<b>87,63</b>	<b>41,42</b>	52,67	31,07
CamemBERT <sub>Large, 2-class</sub>	<b>84,24</b>	<b>82,49</b>	91,70	87,66	85,75	37,36	48,00	19,59

# Évaluation

- Modèles multilingues de plus en plus performants
- Les nouveaux jeux de données s'avèrent plus difficiles que XNLI pour les modèles actuels pour le français.
- Le sous-ensemble d'entraînement de XNLI est le seul disponible pour le français, ce qui affecte les performances lors des tests sur d'autres jeux de données.

Modèles	RTE3-EN		RTE3-FR		GQNLI		GNLI-FR	
	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1
DistilmBERT <sub>Base-cased</sub>	60,75	47,92	61,13	46,65	26,00	26,03	27,67	26,88
XLm-R <sub>Base</sub>	-	-	60,50	49,61	-	-	31,67	31,46
CamemBERT <sub>Base, 3-class</sub>	-	-	63,13	51,52	-	-	33,67	33,44
mDeBERTa-v3 <sub>Base, XNLI</sub>	67,13	56,26	67,13	55,01	28,33	27,73	28,67	27,94
mDeBERTa-v3 <sub>Base, NLI-2mil7</sub>	71,25	61,33	69,63	60,57	<b>36,67</b>	<b>37,04</b>	<b>38,33</b>	<b>38,58</b>
XLm-R <sub>Large</sub>	<b>72,88</b>	<b>63,62</b>	<b>71,25</b>	<b>62,47</b>	35,33	35,02	36,34	35,94
CamemBERT <sub>Large, 3-class</sub>	-	-	71,13	61,97	-	-	33,33	31,62

Modèles	DACCORD		XNLI		RTE3-FR		GNLI-FR	
	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1	Accuracy	Score F1
DistilmBERT <sub>Base-cased</sub>	63,73	52,59	79,98	68,01	79,63	11,89	51,67	19,89
XLm-R <sub>Base</sub>	71,57	67,62	87,17	81,14	77,75	21,93	49,33	23,23
CamemBERT <sub>Base, 3-class</sub>	77,76	76,19	89,64	85,09	80,36	26,29	50,33	<b>36,05</b>
mDeBERTa-v3 <sub>Base, XNLI</sub>	80,75	78,30	90,98	86,39	85,75	30,49	52,00	20,88
mDeBERTa-v3 <sub>Base, NLI-2mil7</sub>	80,95	78,47	90,76	85,89	87,00	38,82	50,67	34,51
XLm-R <sub>Large</sub>	82,01	80,00	<b>96,49</b>	<b>94,74</b>	86,75	41,11	<b>53,33</b>	25,53
CamemBERT <sub>Large, 3-class</sub>	83,27	81,01	92,30	88,12	<b>87,63</b>	<b>41,42</b>	52,67	31,07
CamemBERT <sub>Large, 2-class</sub>	<b>84,24</b>	<b>82,49</b>	91,70	87,66	85,75	37,36	48,00	19,59



# Plan







- I. Tâche et contexte
- II. Jeux de données
- III. Évaluation
- IV. Approches hybrides/logiques**







# Intérêt des approches logiques

Méthode:	<u>DACCORD:</u>	
	<p>P 58 caisses étaient en service dans le centre commercial Amstor le jour de l'attaque, enregistrant ce jour-là un chiffre d'affaires de 2,9 millions de hryvnia ukrainiennes, soit environ 97.000 euros. Des employés du centre commercial, blessés le 27 juin, ont témoigné auprès de l'AFP après l'attaque.</p> <p>H Amstor était fermé et vide au moment des frappes par les missiles russes.</p> <p>R Contradiction</p>	<p>P Rebekah Maciorowski, une mercenaire américaine de 28 ans, est morte au front en Ukraine, selon une publication du 11 décembre 2022 sur Facebook.</p> <p>H Si Rebekah Maciorowski est bel et bien américaine et présente sur le front en Ukraine, elle n'est ni mercenaire, ni décédée dans des combats.</p> <p>R Contradiction</p>
<b>Apprentissage profond</b>		
<b>Approche logique</b>		

Méthode:	<u>DACCORD:</u>	<u>GQNLI-FR:</u>
	<p>P Interrogée par l'AFP, l'Autorité régionale de santé (ARS) de Guadeloupe déplore une fausse information circulant et précise que ce n'est jamais elle qui passe les commandes de médicaments.</p> <p>H C'est une fausse information que ce n'est pas l'Autorité régionale de santé (ARS) de Guadeloupe qui passe les commandes des médicaments.</p> <p>R Contradiction</p>	<p>P Plus de 50% mais moins de 65% des Américains s'inquiètent du réchauffement climatique.</p> <p>H Plus de la moitié des Américains ne s'inquiètent pas du réchauffement climatique.</p> <p>R Contradiction</p>
<b>Apprentissage profond</b>		
<b>Approche logique</b>		

# Intérêt des approches logiques

Méthode:	<u>DACCORD:</u>	
	<p>P 58 caisses étaient en service dans le centre commercial Amstor le jour de l'attaque, enregistrant ce jour-là un chiffre d'affaires de 2,9 millions de hryvnia ukrainiennes, soit environ 97.000 euros. Des employés du centre commercial, blessés le 27 juin, ont témoigné auprès de l'AFP après l'attaque.</p> <p>H Amstor était fermé et vide au moment des frappes par les missiles russes.</p> <p>R Contradiction</p>	<p>P Rebekah Maciorowski, une mercenaire américaine de 28 ans, est morte au front en Ukraine, selon une publication du 11 décembre 2022 sur Facebook.</p> <p>H Si Rebekah Maciorowski est bel et bien américaine et présente sur le front en Ukraine, elle n'est ni mercenaire, ni décédée dans des combats.</p> <p>R Contradiction</p>
<b>Apprentissage profond</b>	  	  
<b>Approche logique</b>	Probablement non (trop de prémisses cachées).	Possible.

Méthode:	<u>DACCORD:</u>	<u>GQNLI-FR:</u>
	<p>P Interrogée par l'AFP, l'Autorité régionale de santé (ARS) de Guadeloupe déplore une fausse information circulant et précise que ce n'est jamais elle qui passe les commandes de médicaments.</p> <p>H C'est une fausse information que ce n'est pas l'Autorité régionale de santé (ARS) de Guadeloupe qui passe les commandes des médicaments.</p> <p>R Contradiction</p>	<p>P Plus de 50% mais moins de 65% des Américains s'inquiètent du réchauffement climatique.</p> <p>H Plus de la moitié des Américains ne s'inquiètent pas du réchauffement climatique.</p> <p>R Contradiction</p>
<b>Apprentissage profond</b>	  	  
<b>Approche logique</b>	Probablement oui.	Probablement oui.

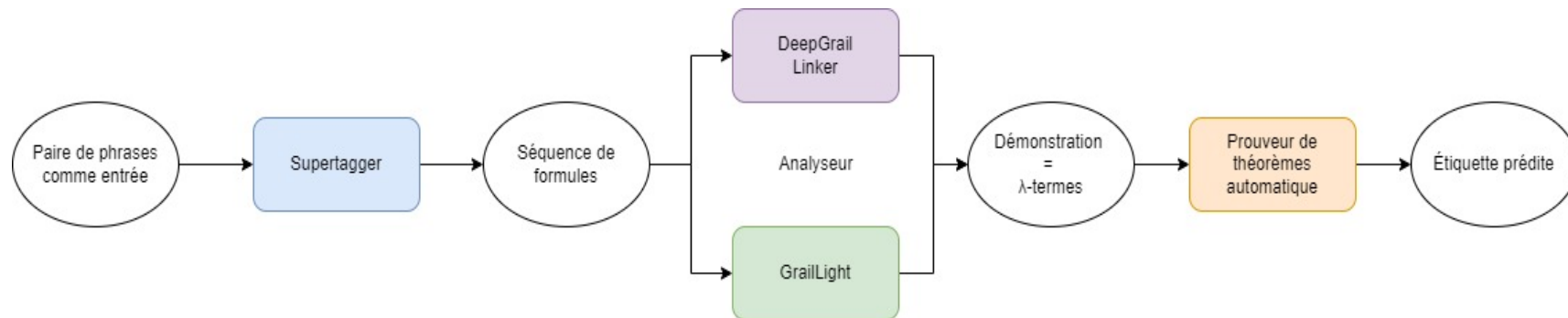
# Intérêt des approches logiques

- **Explicabilité:**
  - Nous disposons d'une démonstration.
- **Meilleur traitement :**
  - de la négation,
  - de la quantification, et
  - d'autres phénomènes sémantiques.
- **Vers une complémentarité des approches**

# Prouveurs automatiques pour le langage naturel

Système	Stratégie de preuve	Logique	Prouveur	Parseur sémantique	Abduction	Arithmétique
Mineshima et al. 2015	Tactiques ad hoc	HOL	Coq	CCG Parser (C&C)		
Abzianidze 2015, 2016	Tableau	Logique naturelle / HOL	NLogPro	C&C et EasyCCG puis LLFgen	✓	
Martínez-Gómez et al. 2017	Tactiques ad hoc	FOL	Coq	C&C et EasyCCG	✓	
Chatzikyriakidis et al. 2019, Bernardy et al. 2021	Tactiques ad hoc	HOL	Coq	Grammatical Framework		✓
Haruta et al. 2022	Résolution	FOL typée	Vampire	C&C, EasyCCG et depccg	✓ (WordNet et VerbOcean)	✓
LINC (Olausson et al. 2023)	Résolution/model building	FOL	Prover9/Mace4	LLM (StarCoder+, GPT 3.5, GPT 4)		

# Notre pipeline



## ▪ Choix:

- FOL ou HOL.
- Nous voulons un démonstrateur le plus automatisé possible.

# En résumé

- 1. Nous avons introduit certains nouveaux jeux de données pour les tâches d'inférence textuelle et de détection automatique de contradictions:**
  - a. DACCORD, un nouveau jeu de données pour la détection de contradictions;
  - b. Des traductions de RTE-3 et GQNLI de l'anglais au français.
- 2. Nous avons réalisé une évaluation de récents LLMs sur ces nouveaux jeux de données:**
  - a. Ces nouveaux jeux de données s'avèrent plus difficiles que XNLI pour les modèles actuels pour le français.
  - b. Le sous-ensemble d'entraînement de XNLI est le seul disponible pour le français, ce qui affecte les performances lors des tests sur d'autres jeux de données.
- 3. Nous avons présenté un panorama des approches logiques pour ces tâches et leur intérêt par rapport aux approches neuronales**
  - a. Explicabilité et meilleur traitement de certains phénomènes linguistiques vs possibilité de prédire un plus grand nombre de cas.



# Merci pour votre attention!

Jeux de données:



Dernier papier:

