

# Modèle neuronal pour la résolution de la coréférence dans les dossiers médicaux électroniques

Julien Tourille<sup>1</sup>   Olivier Ferret<sup>1</sup>   Aurélie Névéol<sup>2</sup>   Xavier Tannier<sup>3</sup>

<sup>1</sup>CEA, LIST, Laboratoire Analyse Sémantique Texte et Image

<sup>2</sup>LIMSI, CNRS, Université Paris-Saclay

<sup>3</sup>Sorbonne Université, Inserm, LIMICS

TALN, 2020

1. Introduction

2. Modèle

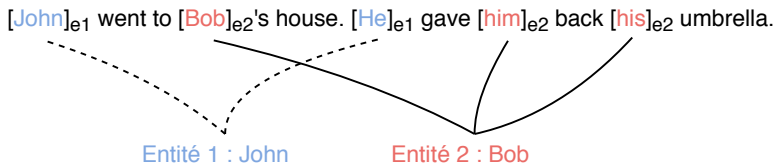
3. Données et contexte expérimental

4. Résultats et analyse

5. Conclusion et perspectives

# Résolution de la coréférence

La résolution de la coréférence consiste à identifier toutes les mentions d'entités ou d'événements et à les regrouper en classes d'équivalence (Pradhan et al. 2011)



# La coréférence dans le domaine médical

La résolution de la coréférence dans le domaine médical concerne

- les événements (e.g. procédures médicales, prises de médicaments ou tests biologiques)
- les entités (e.g. patients, médecins ou membres de la famille)

The CXR revealed **8 mm obstructing stone** ...

The **renal stone** was considered to be the cause of patient's symptoms ...

We recommended surgical procedure to remove **ureteropelvic stone** ...

Surgeon is Dr Chase ...

The patient has children but they are not nearby ...

# La coréférence dans le domaine médical

La résolution de la coréférence dans le domaine médical concerne

- les événements (e.g. procédures médicales, prises de médicaments ou tests biologiques)
- les entités (e.g. patients, médecins ou membres de la famille)

The CXR revealed **8 mm obstructing stone** ...

The **renal stone** was considered to be the cause of **patient's** symptoms ...

We recommended surgical procedure to remove **ureteropelvic stone** ...

Surgeon is Dr Chase ...

The **patient** has children but they are not nearby ...

1. Introduction

**2. Modèle**

3. Données et contexte expérimental

4. Résultats et analyse

5. Conclusion et perspectives

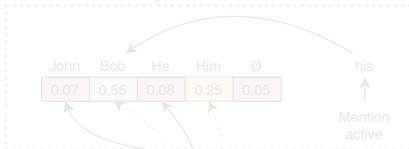


# Modèle *mention-ranking*

Modèle *mention-ranking* incorporant de l'information sur les clusters en cours de construction, inspiré par les approches récentes dans le domaine général (Lee et al. 2017 ; Wiseman et al. 2016)

[John]<sub>e1</sub> went to [Bob]<sub>e2</sub>'s house. [He]<sub>e1</sub> gave [him]<sub>e2</sub> back [his]<sub>??</sub> umbrella

Modèle mention-ranking



Lecture linéaire du texte

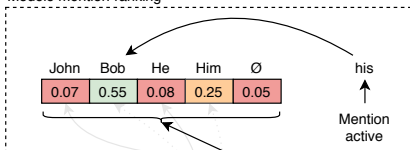


# Modèle *mention-ranking*

Modèle *mention-ranking* incorporant de l'information sur les clusters en cours de construction, inspiré par les approches récentes dans le domaine général (Lee et al. 2017 ; Wiseman et al. 2016)

[John]<sub>e1</sub> went to [Bob]<sub>e2</sub>'s house. [He]<sub>e1</sub> gave [him]<sub>e2</sub> back [his]<sub>??</sub> umbrella

Modèle mention-ranking



Recherche du meilleur antécédent

e<sub>1</sub> John → He  
e<sub>2</sub> Bob → him

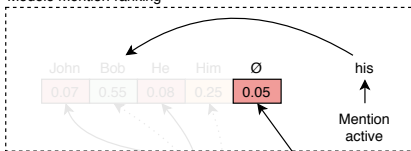
Probabilités

# Modèle *mention-ranking*

Modèle *mention-ranking* incorporant de l'information sur les clusters en cours de construction, inspiré par les approches récentes dans le domaine général (Lee et al. 2017 ; Wiseman et al. 2016)

[John]<sub>e1</sub> went to [Bob]<sub>e2</sub>'s house. [He]<sub>e1</sub> gave [him]<sub>e2</sub> back [his]<sub>??</sub> umbrella

Modèle mention-ranking



Recherche du meilleur antécédent

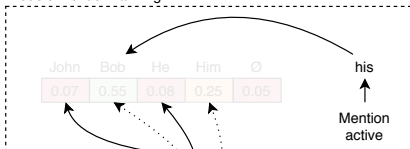
Cas sans antécédent

# Modèle *mention-ranking*

Modèle *mention-ranking* incorporant de l'information sur les clusters en cours de construction, inspiré par les approches récentes dans le domaine général (Lee et al. 2017 ; Wiseman et al. 2016)

[John]<sub>e1</sub> went to [Bob]<sub>e2</sub>'s house. [He]<sub>e1</sub> gave [him]<sub>e2</sub> back [his]<sub>??</sub> umbrella

Modèle mention-ranking



e<sub>1</sub> John → He

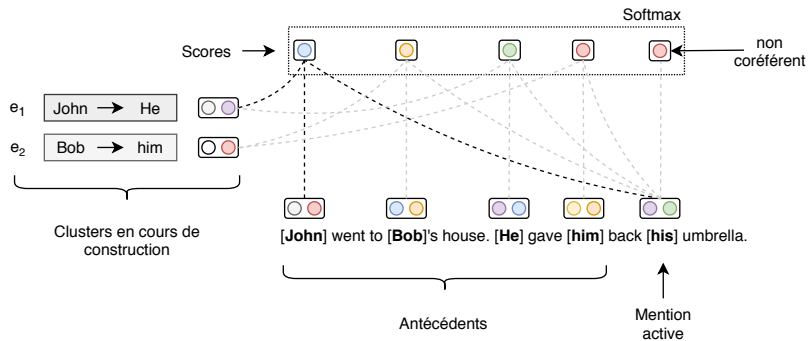
e<sub>2</sub> Bob → him



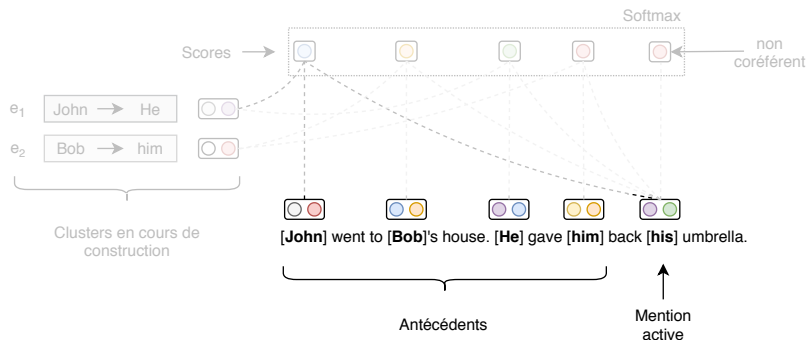
Clusters en cours de construction

Recherche du meilleur antécédent

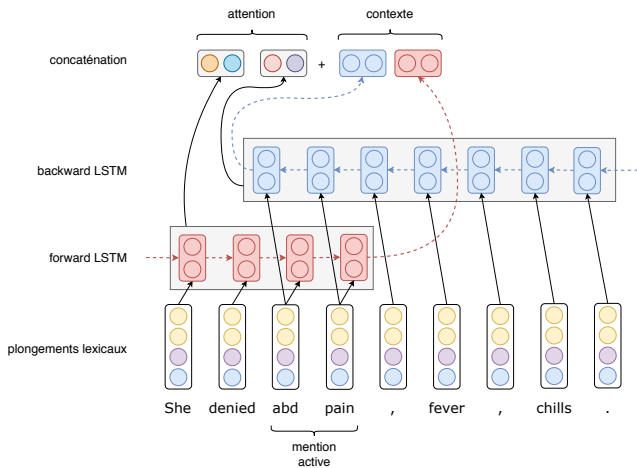
## Présentation détaillée du modèle



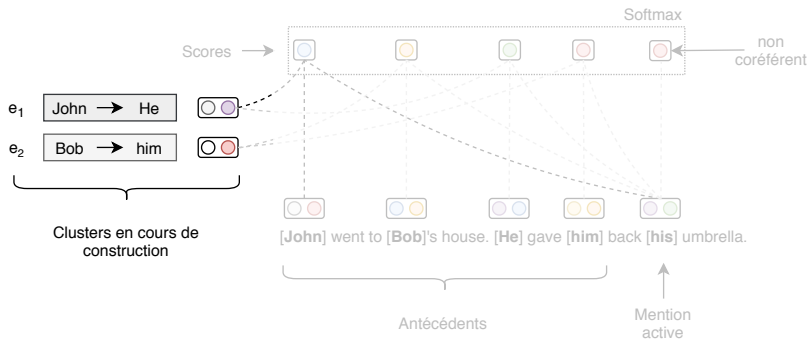
# Représentation des mentions



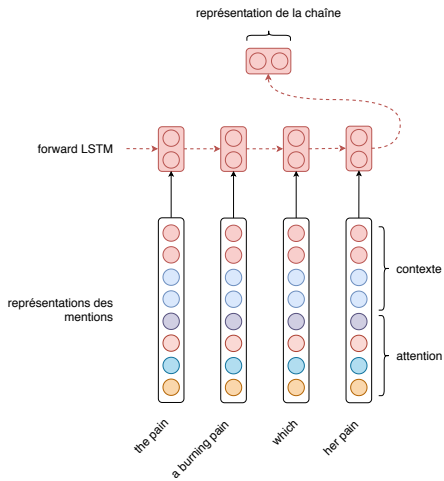
# Représentation des mentions



# Représentation des clusters

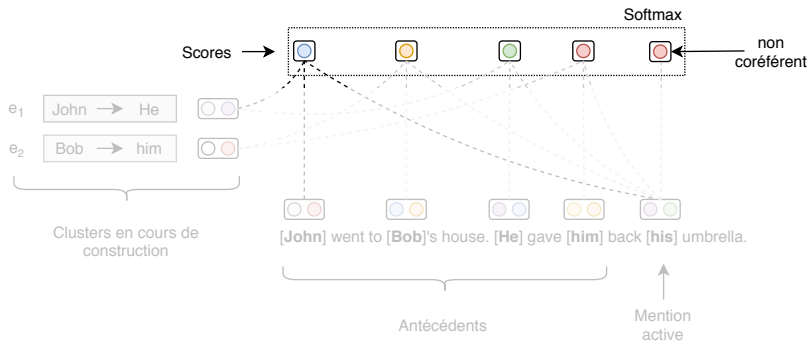


# Représentation des clusters

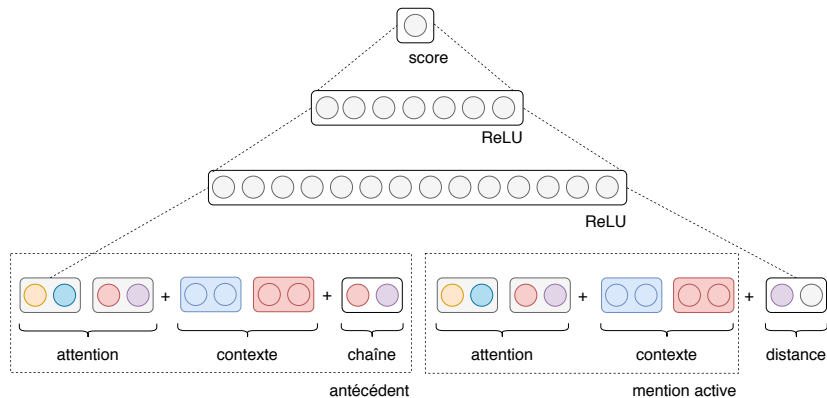




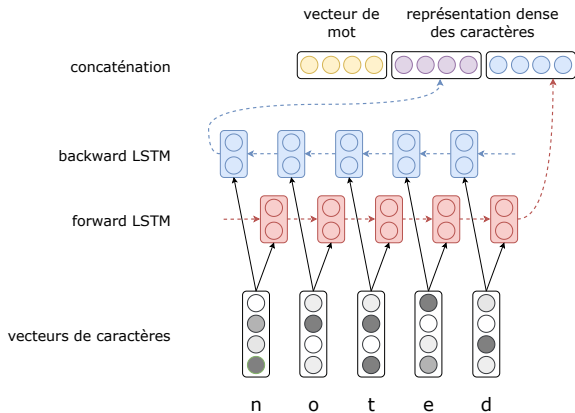
# Classifieur de paires



# Classifieur de paires



# Plongements lexicaux



1. Introduction

2. Modèle

**3. Données et contexte expérimental**

4. Résultats et analyse

5. Conclusion et perspectives

# Corpus

## Corpus i2b2 tâche 1C (Uzuner et al. 2012)

- 451 documents annotés avec 5 événements et entités : *person*, *problem*, *test*, *treatment* et *pronoun*

Institution	Train	Test	Total
BETH	115	79	194
PARTNERS	136	94	230
Combinés	251	173	424

Statistiques concernant le corpus i2b2/VA tâche 1C

Institution	Nombre chaînes	Long. moy.	Long. max.
BETH	1 816	4,2	122
PARTNERS	1 395	4,4	105

Statistiques concernant le nombre et la longueur des chaînes de coréférence du corpus i2b2/VA tâche 1C

# Contexte expérimental (1/2)

- Utilisation des mentions *gold* → pas de détection préalable des mentions
- Apprentissage de deux modèles : personnes vs. événements
- Configurations structurelles
  - Modèle nu
  - + Mécanisme d'attention
  - + Représentation des caractères
  - + Représentation des clusters en cours de construction
- Stratégies d'apprentissage
  - + Pré-entraînement sur les mentions coréférentes (Clark et Manning 2015; Wiseman et al. 2015)
  - + Filtrage des antécédents
  - + Spécialisation du classifieur de paires selon le type

## Contexte expérimental (2/2)

- Taille des couches cachées des LSTMs : 100 pour le principal, 25 pour la représentation des caractères
- Mini-lots de 1 document
- Plongement lexicaux pré-entraînés sur MIMIC-III ([Johnson et al. 2016](#))
- Dropout sur les couches cachées : 0,2
- Dropout sur les plongements en entrée : 0,5
- Apprentissage des états initiaux ([Gers et al. 2002](#))
- Prise en compte de l'aspect non-déterministe : 10x entraînements par expérience ([Reimers et Gurevych 2017](#))

1. Introduction

2. Modèle

3. Données et contexte expérimental

**4. Résultats et analyse**

5. Conclusion et perspectives



# Résultats (1/2)

condition	P	R	F1	P	R	F1	P	R	F1
	Personnes			Événements			Combinés		
baseline	87,77	82,82	85,17	65,15	54,62	59,30	76,89	67,94	72,02
	(± 0,71)	(± 0,99)	(± 0,46)	(± 1,81)	(± 1,35)	(± 0,45)	(± 1,31)	(± 0,89)	(± 0,38)
attention	88,08	83,42	85,64 ↑	66,39	54,18	59,56 ↑	77,77	67,93	72,41 ↑
	(± 0,82)	(± 0,55)	(± 0,38)	(± 1,45)	(± 2,08)	(± 0,86)	(± 1,01)	(± 1,27)	(± 0,40)
caractères	87,87	82,55	85,08 ↓	66,39	53,26	59,02 ↓	77,78	67,13	71,96 ↓
	(± 0,98)	(± 0,69)	(± 0,21)	(± 1,57)	(± 1,76)	(± 0,92)	(± 1,14)	(± 1,01)	(± 0,46)
chaîne	87,48	82,36	84,71 ↓	65,86	52,12	58,11 ↓	77,29	66,55	71,37 ↓
	(± 0,77)	(± 0,72)	(± 0,49)	(± 1,03)	(± 1,10)	(± 0,58)	(± 0,69)	(± 0,64)	(± 0,37)
filtrage	87,60	82,82	85,08 ↓	65,42	54,37	59,30 =	76,98	67,85	72,03 ↑
	(± 1,48)	(± 0,89)	(± 0,39)	(± 1,31)	(± 1,37)	(± 0,56)	(± 1,00)	(± 0,90)	(± 0,29)
spécialisation	88,52	82,34	85,25 ↑	63,66	49,62	55,68 ↓	76,64	64,92	70,17 ↓
	(± 0,76)	(± 0,86)	(± 0,34)	(± 1,12)	(± 1,60)	(± 0,72)	(± 0,84)	(± 1,00)	(± 0,38)
pré-entraînement	88,93	82,60	85,60 ↑	68,63	56,70	61,99 ↑	79,09	69,27	73,79 ↑
	(± 0,30)	(± 0,79)	(± 0,41)	(± 2,09)	(± 2,77)	(± 1,11)	(± 1,43)	(± 1,53)	(± 0,53)
optimal	88,65	82,91	85,62	67,72	57,51	62,16	78,36	69,86	73,82
	(± 1,22)	(± 0,69)	(± 0,61)	(± 1,34)	(± 1,61)	(± 0,78)	(± 1,04)	(± 1,10)	(± 0,47)

## Résultats (2/2)

	# i2b2	P	R	F1
Xu et al. (2011)	1	82,38	78,13	80,20
Cai et al. (2011)	5	75,27	73,96	74,60
notre modèle		78,36	69,86	73,82
Jindal et Roth (2011)	9	65,53	83,48	73,41
Dai et al. (2011)	8	76,08	65,65	70,48
Anick et al. (2011)	7	79,61	61,67	69,41

Comparaison de notre système à ceux de la campagne i2b2. Les scores sont obtenus *via* le script officiel CoNLL et calculés en excluant les singletons. Nous rapportons les scores des systèmes pour lesquels une conversion entièrement automatique du format i2b2 vers CoNLL a été possible.

1. Introduction

2. Modèle

3. Données et contexte expérimental

4. Résultats et analyse

5. Conclusion et perspectives

# Conclusion et perspectives

## Principales contributions

- Modèle neuronal bout en bout pour la résolution de la coréférence dans le domaine médical
- Tests d'ablation sur les différents modules du modèle
- Script de conversion du format i2b2 vers le format CoNLL  
<https://github.com/jtourille/i2b2-coref-task1c-converter>

## Limites et perspectives

- Performances encore en dessous des meilleurs systèmes i2b2
- Détection parfaite des mentions
- Non prise en compte des aspects temporels
- Pas d'utilisation des modèles récents fondés sur le *transformer*

# Remerciements

- Labex DigiCosme, projet CÔT
- Agence Nationale de la Recherche, projet CABeRneT  
ANR-13-JS02-0009-01

Merci pour votre attention

- Anick, P., P. Hong, N. Xue et al. (oct. 2011). « Coreference resolution for electronic medical records ». In : *Proceedings of the 2011 i2b2/VA/Cincinnati Workshop on Challenges in Natural Language Processing for Clinical Data* (Boston, Massachusetts, USA). i2b2.
- Cai, J., E. Mujdicza, Y. Hou et al. (oct. 2011). « Weakly supervised graph-based coreference resolution for clinical texts ». In : *Proceedings of the 2011 i2b2/VA/Cincinnati Workshop on Challenges in Natural Language Processing for Clinical Data* (Boston, Massachusetts, USA). i2b2.
- Clark, Kevin et Christopher D. Manning (juil. 2015). « Entity-Centric Coreference Resolution with Model Stacking ». In : *Proceedings of the Joint Conference of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing* (Beijing, China, 26-31 juil. 2015). Association for Computational Linguistics, p. 1405-1415.
- Dai, H. J., C. Y. Wu, C. Y. Chen et al. (oct. 2011). « Co-reference resolution of the medical concepts in the patient discharge summaries ». In : *Proceedings of the 2011 i2b2/VA/Cincinnati Workshop on Challenges in Natural Language Processing for Clinical Data* (Boston, Massachusetts, USA). i2b2.

- Gers, Felix A., Nicol N. Schraudolph et Jürgen Schmidhuber (2002). « Learning Precise Timing with LSTM Recurrent Networks ». In : *Journal of Machine Learning Research* 3, p. 115-143.
- Jindal, P. et D. Roth (oct. 2011). « Using domain knowledge and domain-inspired discourse model for coreference resolution for clinical narratives ». In : *Proceedings of the 2011 i2b2/VA/Cincinnati Workshop on Challenges in Natural Language Processing for Clinical Data* (Boston, Massachusetts, USA). i2b2.
- Johnson, Alistair E. W., Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo A. Celi et Roger G. Mark (2016). « MIMIC-III, a Freely Accessible Critical Care Database ». In : *Scientific Data* 3.
- Lee, Kenton, Luheng He, Mike Lewis et Luke Zettlemoyer (sept. 2017). « End-to-end Neural Coreference Resolution ». In : *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Copenhagen, Denmark, 7-11 sept. 2017). Association for Computational Linguistics, p. 188-197.



Pradhan, Sameer, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel et Nianwen Xue (juin 2011). « CoNLL-2011 Shared Task : Modeling Unrestricted Coreference in OntoNotes ». In : *Proceedings of the 15th Conference on Computational Natural Language Learning* (Portland, Oregon, USA, 23-24 juin 2011). Association for Computational Linguistics, p. 1-27.

Reimers, Nils et Iryna Gurevych (sept. 2017). « Reporting Score Distributions Makes a Difference : Performance Study of LSTM-networks for Sequence Tagging ». In : *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Copenhagen, Denmark, 7-11 sept. 2017). Association for Computational Linguistics, p. 338-348.

Uzuner, Özlem, Andreea Bodnari, Shuying Shen, Tyler Forbush, John Pestian et Brett R. South (2012). « Evaluating the State of the Art in Coreference Resolution for Electronic Medical Records ». In : *Journal of the American Medical Informatics Association* 19.5, p. 786-791.

Wiseman, Sam, Alexander M. Rush et Stuart M. Shieber (juin 2016). « Learning Global Features for Coreference Resolution ». In : *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies* (San Diego, California, USA, 12-17 juin 2016). Association for Computational Linguistics, p. 994-1004.

Wiseman, Sam, Alexander M. Rush, Stuart Shieber et Jason Weston (juil. 2015). « Learning Anaphoricity and Antecedent Ranking Features for Coreference Resolution ». In : *Proceedings of the Joint Conference of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing* (Beijing, China, 26-31 juil. 2015). Association for Computational Linguistics, p. 1416-1426.

Xu, Y., J. Liu, J. Wu et al. (oct. 2011). « EHUATUO : a mention-pair coreference system by exploiting document intrinsic latent structures and world knowledge in discharge summaries :2011 i2b2 challenge. ». In : *Proceedings of the 2011 i2b2/VA/Cincinnati Workshop on Challenges in Natural Language Processing for Clinical Data* (Boston, Massachusetts, USA). i2b2.