

# Détection automatique de la corréférence événementielle dans les dossiers patients électroniques

Julien Tourille<sup>1,2,3</sup>, Olivier Ferret<sup>5</sup>, Timothy Miller<sup>4</sup>  
Aurélie Névéol<sup>1,3</sup>, Xavier Tannier<sup>6</sup>

<sup>1</sup> LIMSI, CNRS, <sup>2</sup> Univ. Paris-Sud, <sup>3</sup> Université Paris Saclay

<sup>4</sup> Boston Children's Hospital et Harvard Medical School

<sup>5</sup> CEA, LIST, Gif-sur-Yvette, F-91191 France

<sup>6</sup> Université Pierre et Marie Curie (UPMC, Paris 06), LIMICS

Séminaire OpenHealth Institute

17 novembre 2017

# Qu'est-ce que la coréférence?

Relation entre deux termes ayant **le même référent** (faisant référence à la *même entité* du monde réel)

## Exemples

- a) **Apple** est un fabricant d'ordinateur.  
**La firme à la pomme** est mondialement connue.
- b) **Bill** arrived, but nobody saw **him**.
- c) Despite **her** difficulty, **Wilma** came to understand the point.

→ **Groupes nominaux** et des **pronoms personnels**

# Qu'est-ce que la coréférence événementielle?

Relation entre deux termes faisant référence **au même événement du monde réel**

- Ancrage **spatio-temporel** (lieu et date)
- Possède des **arguments** et/ou **participants**

Exemples (Lee et al., 2012)

- a) One of the key suspected Mafia bosses arrested yesterday **has hanged** himself.  
Police said Lo Presti **had hanged** himself.  
**His suicide** appeared to be related to clan feuds.
- b) The New Orleans Saints **placed** Reggie Bush on the injured list on Wednesday.  
Saints **put** Bush on I.R

→ **Verbes et groupes nominaux**

# La coréférence dans les dossiers patients

## Phénomène hybride entre coréférence et coréférence événementielle

MMMMMMM is now approximately 5 weeks status-post **her** neurolysis for **her** plexiform neurofibromas for **pain**. **Her** **pain** is significantly improved. **She** is taking just a low-dose of Neurontin, which also seems to help her to some degree. **She** has **pain** in other distributions besides the median. **She** has only noted minimal change in **her** **hand function** following **the procedure**, and **she** states that **this** is getting better.

- 1 - MMMMMMMM – her – her – Her – She – her – She – She – her – she
- 2 - neurolysis – the procedure
- 3 - pain – pain – pain
- 4 - hand function – this

Maladies et symptômes → événements

## Contexte de recherche

Besoin d'accéder aux informations contenues dans les dossiers patients:

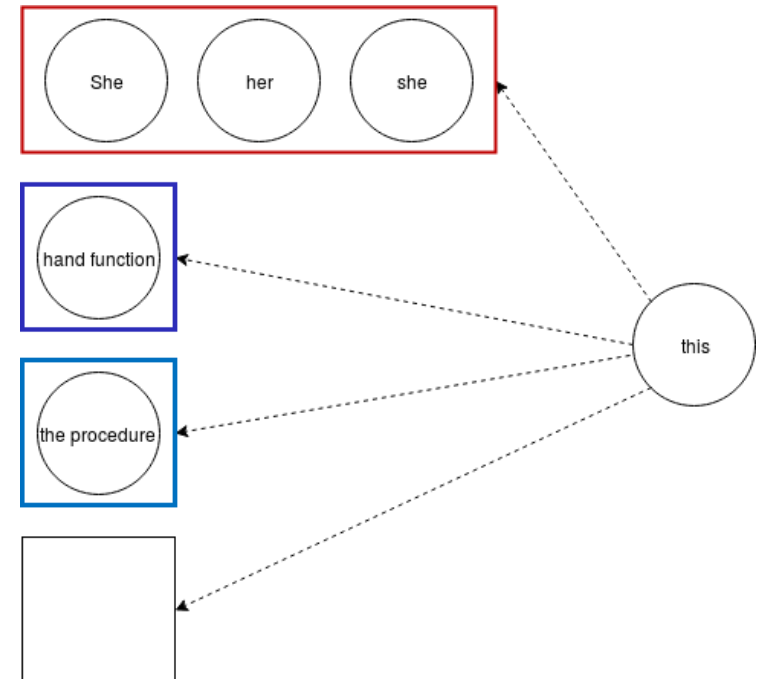
- **Réactions indésirables aux médicaments:** effets secondaires, accidents
  - **Phénotypage:** identifier les patients qui partagent les mêmes caractéristiques
  - **Aide à la décision:** apporter des informations pertinentes en temps réel aux praticiens hospitaliers
- Accès aux **chronologies des patients**
- Événements mentionnés plusieurs fois (dans le même document ou dans plusieurs documents)
- Traitement d'un **grand volume de données**
- Système automatique de résolution de la corréférence

# Définition de la tâche

**She** has only noted minimal change in **her** hand function following **the procedure**, and **she** states that **this** is getting better.

Pour chaque candidat:

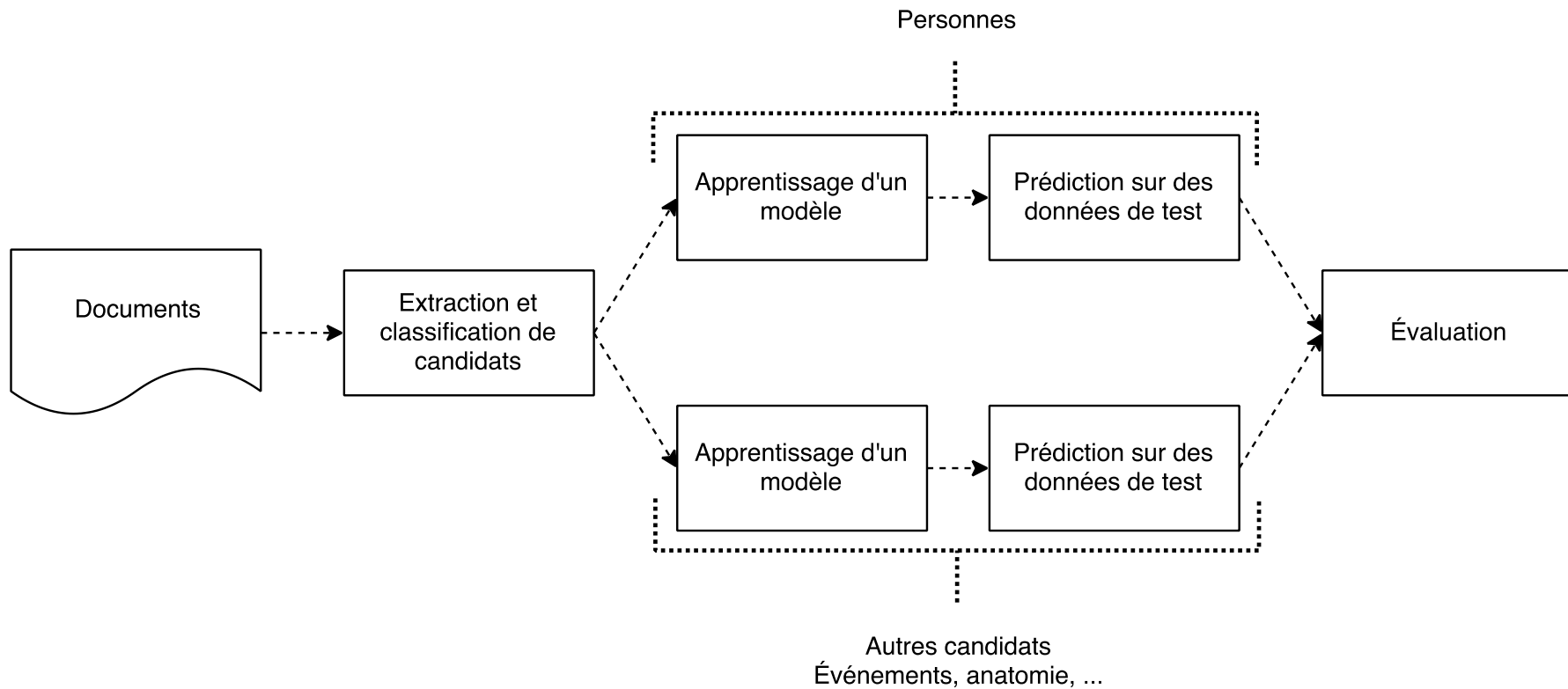
- Associer à une chaîne existante
- Démarrer une nouvelle chaîne



## Difficultés de la tâche

- Déséquilibre entre exemples positifs et négatifs: **majorité de singletons** (chaînes contenant une seule mention)
- Cas de la **cataphore**: « *Despite **her** difficulty, **Wilma** came to understand the point.* »
- Cas des **dummy pronouns**: « *I went outside and **it** was snowing.* »
- **Correspondance lexicale** erronée: Clinton → Hilary ou Bill ?
- **Évaluation**: mesurer la performance d'un système est difficile

# Résolution automatique de la coréférence





## Données

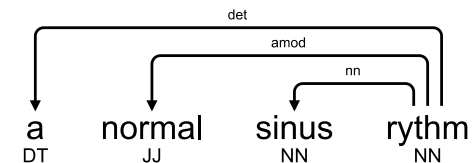
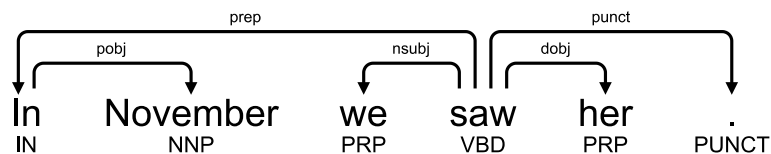
- Corpus de documents cliniques utilisé lors du 5<sup>th</sup> i2b2/VA Workshop on NLP Challenges for Clinical Records (Uzuner et al., 2012)
  - **Comptes-rendus d'hospitalisation** du *Beth Israel Deaconess Medical Center, Partners Healthcare, and University of Pittsburgh Medical*.
  - **Rapports cliniques et notes de pathologie** de la *Mayo Clinic*, et **comptes-rendus d'hospitalisation, de radiologie et de pathologie** de *University of Pittsburgh Medical*

→ 978 documents (590 pour le corpus d'entraînement, 388 pour le corpus de test)

## Modèle – Extraction de candidats

Utilisation de la sortie d'un analyseur syntaxique + règles manuelles

- Noms propres (NNP)
- Noms communs (NN) + modifieurs (a normal sinus rythm)
- Pronoms personnels (PRP)
- ...



## Modèle – Classification des candidats

➤ Chaines de coréférence impliquant des personnes:

- **très longues**
- impliquent généralement **quelques personnes**, i.e. le patient et le médecin

→ Distinction entre les **personnes physiques** et le **reste** (événements, anatomie, ...)

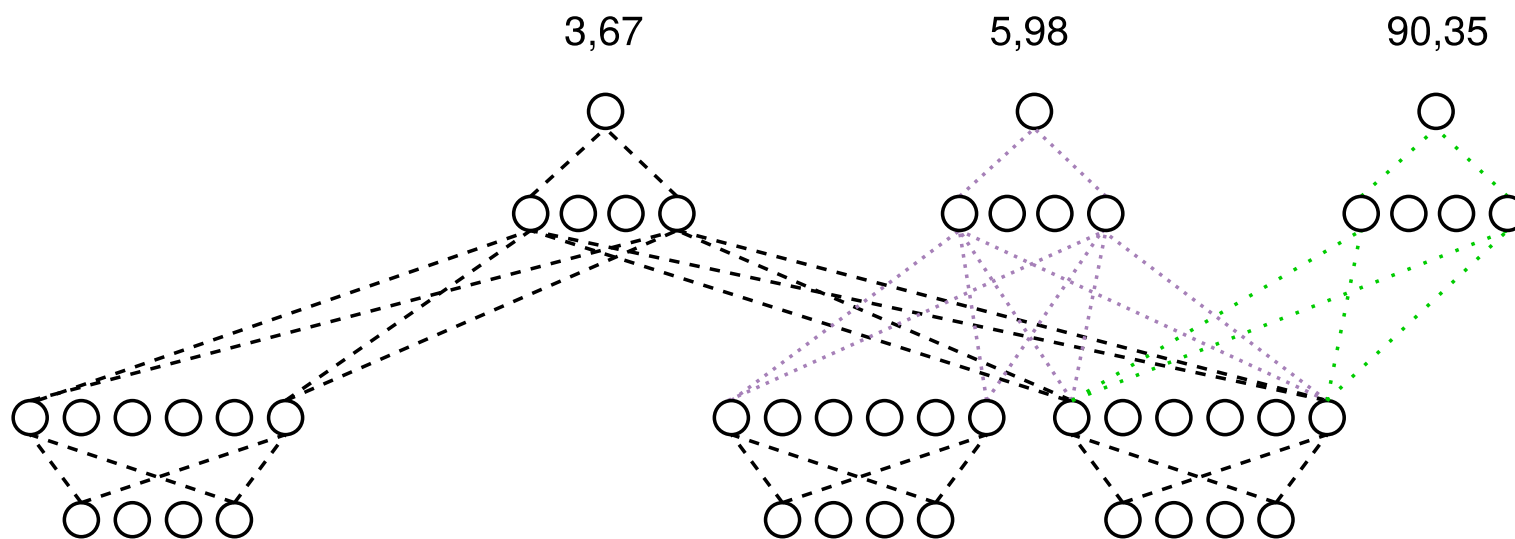
→ Construction d'un **classifieur automatique de mentions** basé sur une série de traits

- unigrammes (« *a* », « *PEG* », « *tube* »)
- bigrammes (« *a PEG* », « *PEG tube* »)
- reconnaissance de certains schémas (e.g. *Dr. XXX*)
- appartenance à des listes de vocabulaire (professions, procédures médicales, anatomie, ...)
- ...



## Modèle – Résolution de la corréférence

→ Utilisation d'un modèle **mention-ranking** basé sur des réseaux de neurones profonds

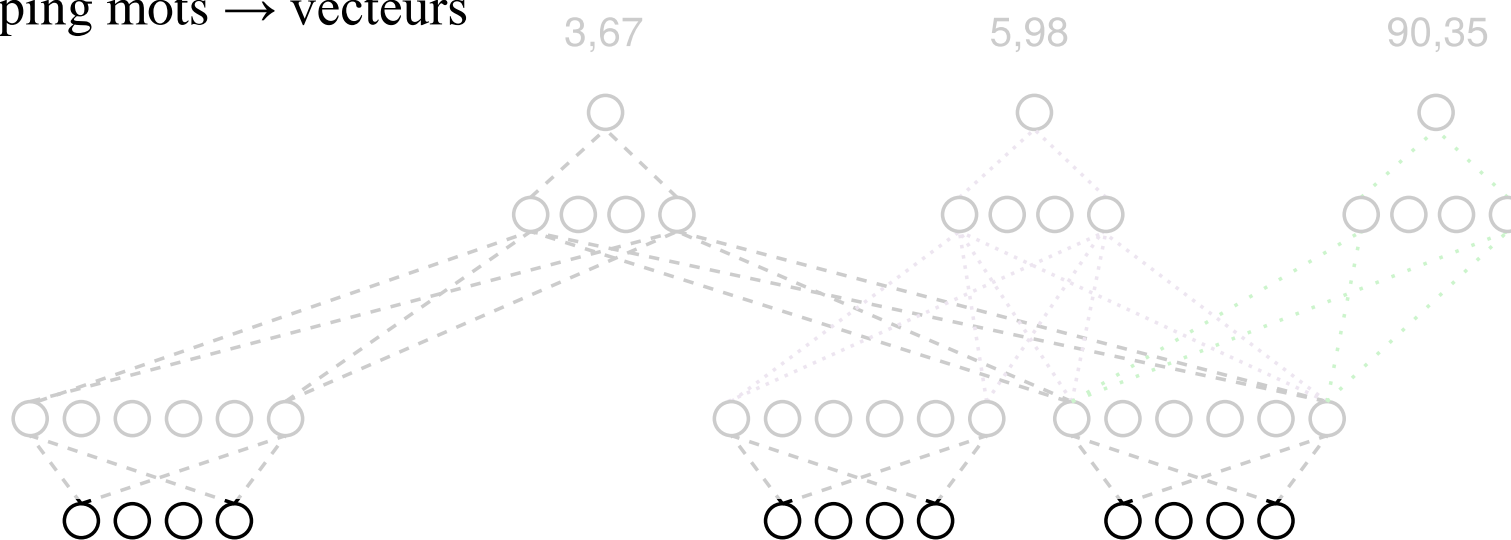


She also has [hypercalcemia](#) and has been evaluated for [this](#) without [a definite etiology](#).

# Modèle – Résolution de la corréférence

→ Utilisation d'un modèle **mention-ranking** basé sur des réseaux de neurones

Étape 1 – Mapping mots → vecteurs

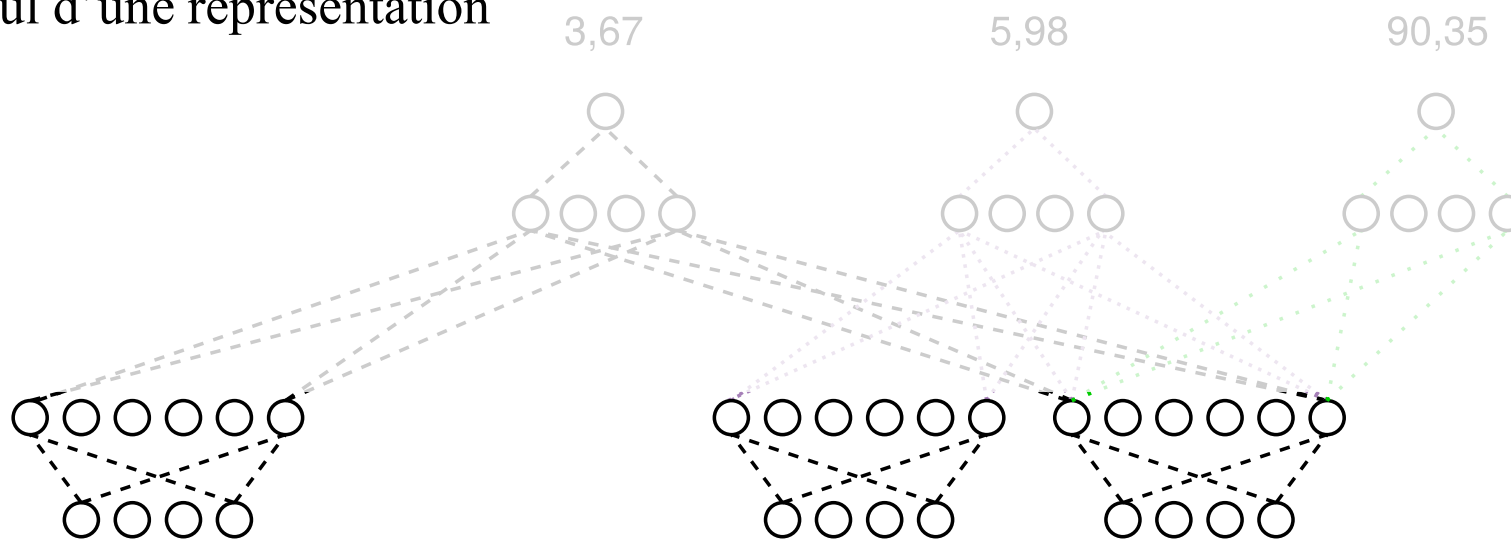


She also has [hypercalcemia](#) and has been evaluated for [this](#) without [a definite etiology](#).

# Modèle – Résolution de la corréférence

→ Utilisation d'un modèle **mention-ranking** basé sur des réseaux de neurones

**Étape 2** – Calcul d'une représentation intermédiaire

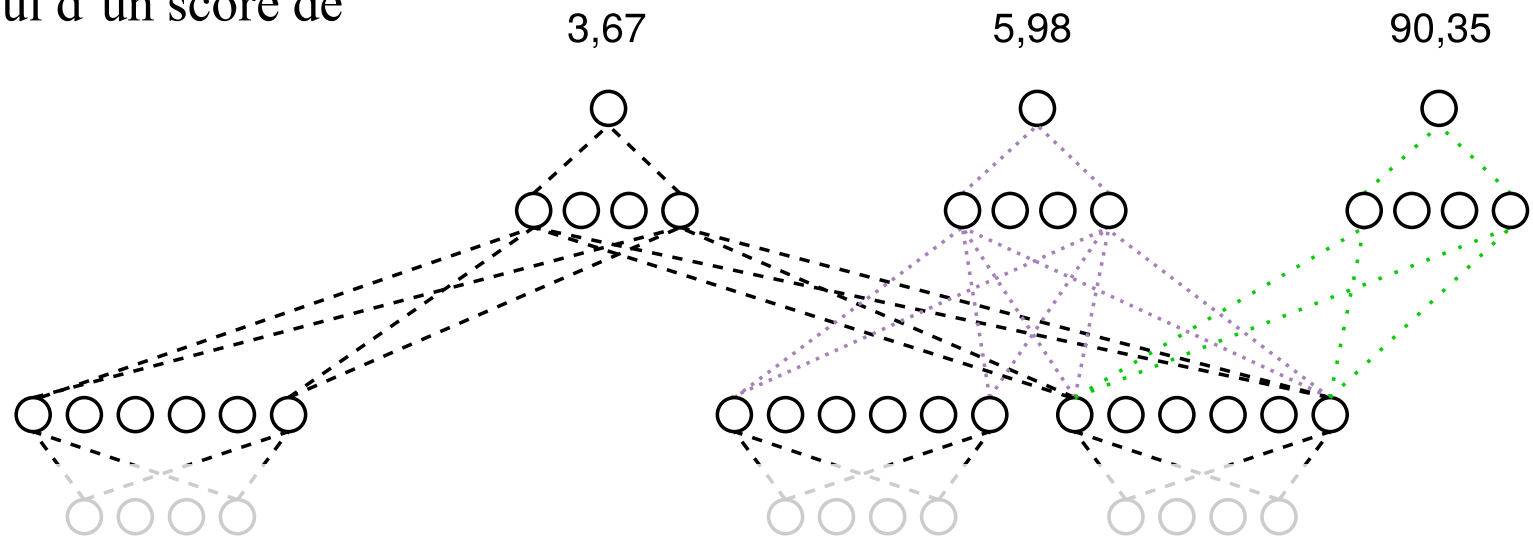


She also has hypercalcemia and has been evaluated for this without a definite etiology.

# Modèle – Résolution de la coréférence

→ Utilisation d'un modèle **mention-ranking** basé sur des réseaux de neurones

**Étape 3** – Calcul d'un score de coréférence



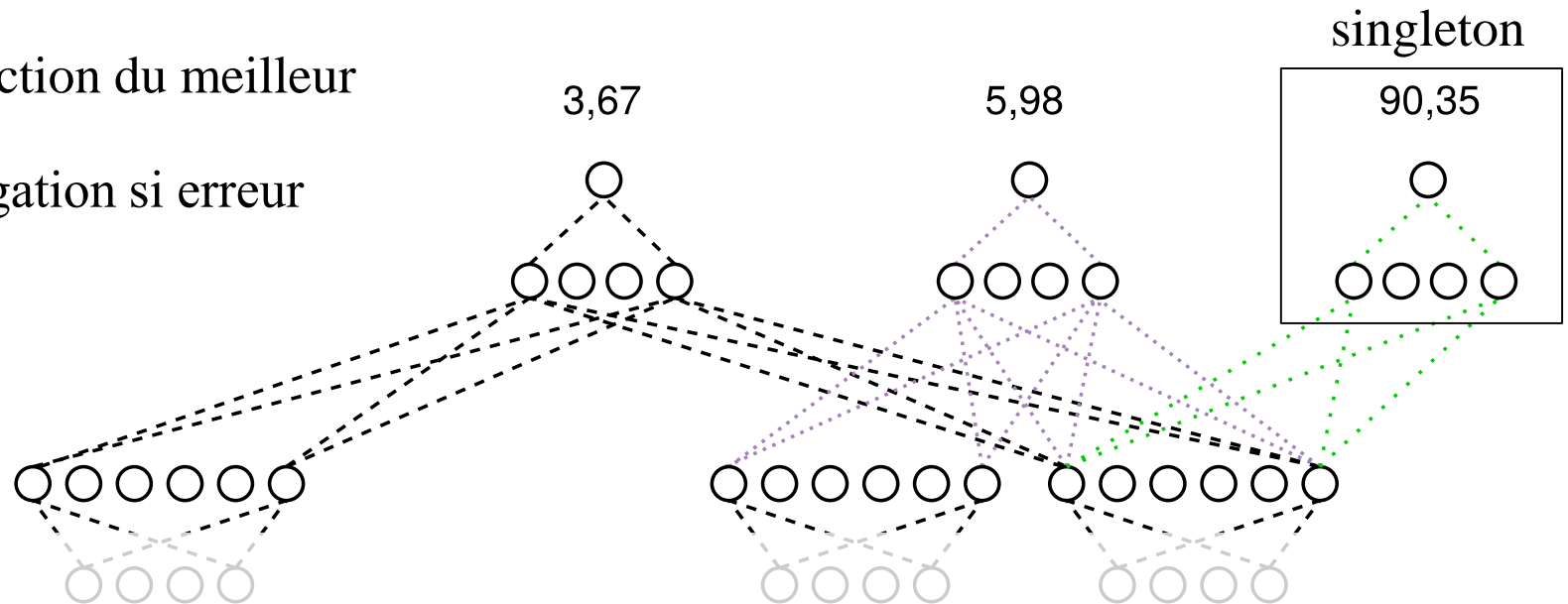
She also has hypercalcemia and has been evaluated for this without a definite etiology.

# Modèle – Résolution de la corréférence

→ Utilisation d'un modèle **mention-ranking** basé sur des réseaux de neurones

**Étape 4** – Sélection du meilleur antécédent

→ Rétropropagation si erreur



She also has hypercalcemia and has been evaluated for this without a definite etiology.



# Performance et agenda de travail

- Expériences toujours en cours
- **Résultats préliminaires:**
  - Score de 0,64 (un seul modèle pour les deux types de chaînes)
  - Score dans le **même ordre de grandeur** que les meilleurs modèles présentés dans la littérature
- **Agenda de travail:**
  - **Prendre en considération l'état des chaînes** lors du calcul du score de coréférence
  - **Prendre en considération d'autres éléments** lors du calcul tels que les **relations temporelles** qui existent entre les mentions candidates

# Questions ?

Détection automatique de la corréférence  
événementielle dans les dossiers patients  
électroniques

Julien Tourille, Olivier Ferret, Timothy Miller  
Aurélie Névéol, Xavier Tannier