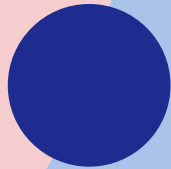


# **PRÉSENTATION DE L'ARTICLE**



# L'ARTICLE

2

Titre : **A deep neural network model for speakers coreference resolution in legal texts**

Auteurs : Donghong Ji , Jun Gao , Hao Fei , Chong Teng , Yafeng Ren

Publié dans : Information Processing and Management 57 (2020) [[lien](#)]

Langue : Anglais

Langue des données : Chinois

Keywords :

- Legal text mining
- Coreference resolution
- Court record document
- Neural networks
- Attention mechanism

- Extraction d'information dans des textes juridiques
  - Résolution de corréférence
    - **Résolution de corréférence des mentions des intervenants du procès**
- Problématique : peu de recherche existante sur cette tâche, et plus particulièrement dans les domaines professionnels spécialisés
- Obstacle : difficulté de mettre en place des chaînes de traitement NLP, notamment en aval
- Proposition : Modèle neuronal innovant, spécialisé sur les comptes rendus d'audience

# LES DONNÉES

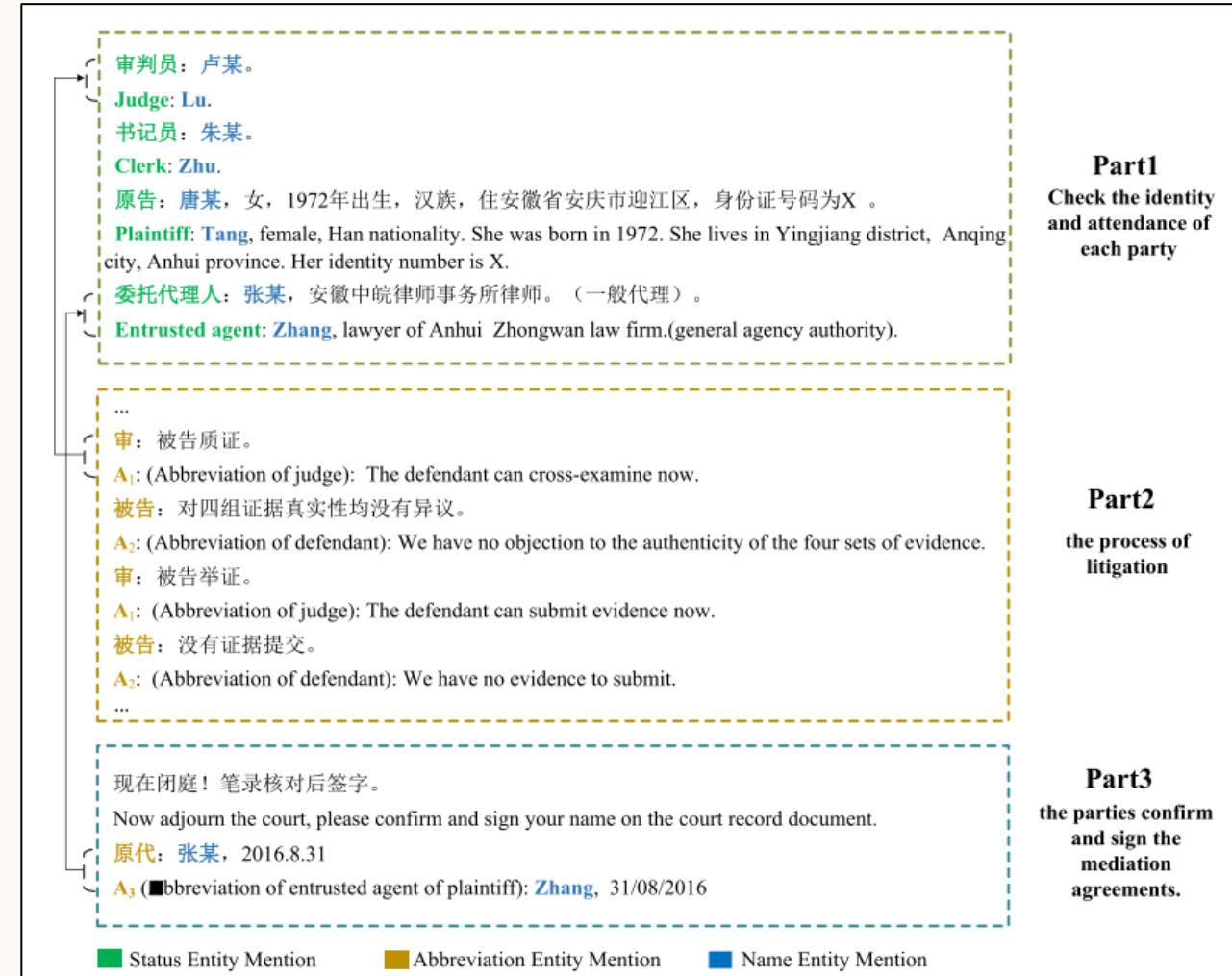
Origine : collecte de données d'affaires civiles réelles ayant eu lieu entre 2012 et 2019

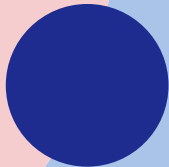
Annotation : Les entités cibles y sont annotées et liées, seulement quand au moins 3 d'entre elles sont coréférentes.

Quantité : 1289 documents, 9 464 annotations de coréférence (avant filtrage)

Répartition : Train/Dev/Test ration = 7:1:2.

Category	Train	Dev.	Test	Total
Document	903	129	257	1289
Average sentence length	3416	3415	3490	-
Max sentence length	14,488	9492	11,301	-
Average sentence length(filtered)	379	415	374	-
Max sentence length(filtered)	1510	1379	1186	-
Entity Type	Train	Dev.	Test	Total
Abbreviation entity	586	171	223	768
Name entity	3439	681	1127	4636
Status entity	14	11	13	14





# MÉTHODOLOGIE

Tâche : résolution de la coréférence des intervenants / acteurs du procès.  
Objectif : relier les abréviations (A) aux noms (N), puis les abréviations (A) aux statuts (S).  
Transformation du document en une séquence de mots (W) complétée par 3 jeux distincts d'entités : A / N / S

Proposition de modèle neuronal avec 3 modules majoritaires :

- module d'encodage des informations contextuelles autour des entités (span representations)
- module neuronal de représentation graphique des relations construites entre les entités (GCN)
- mécanisme multi notations (Biaffine + FFNN) pour déterminer les scores de coréférence entre les paires.

Les méthodes sont très bien décrites dans l'article et reproductibles. Les auteurs discutent du fait que les méthodes à base de règles peuvent bien fonctionner et sont faciles à utiliser pour la résolution de coréférence mais ne sont pas du tout robustes face à divers scenarios (variété de texte)

Abbreviation	Name	Status
审	卢某	审判员
Abbreviation of judge	Lu	Judge
答	张某	被告
Abbreviation of Reply	Zhang	Defendant
二被代	吴某	委托诉讼代理人
Abbreviation of entrusted agent of the second defendant	Wu	Entrusted agent

# LE MODÈLE

6

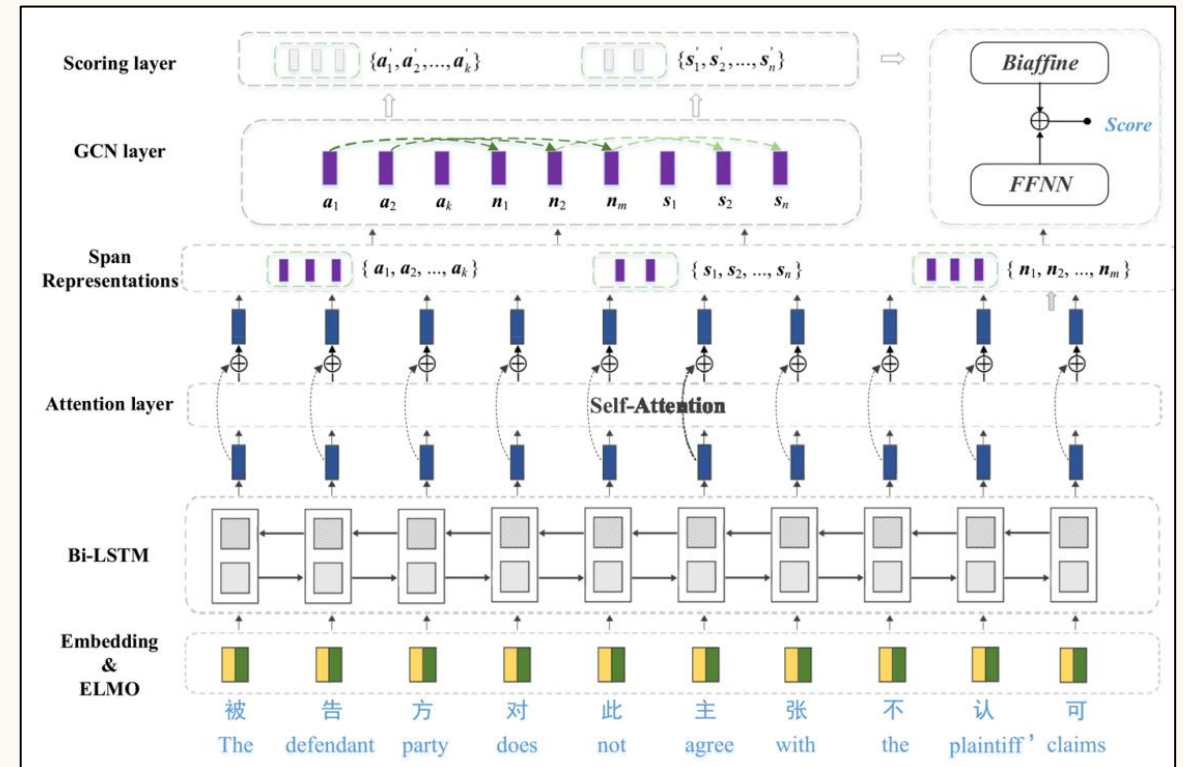
Results of different models.

Model	Precision(%)	Recall(%)	F1(%)
Sentence-classification (BERT)	58.73	76.41	66.41
Lee et al. (2018) (A-N)	89.08	78.10	83.22
Zhang et al. (2019a) (A-N)	70.86	94.68	81.61
our model (A-N)	82.34	84.75	<b>83.53</b>
Rule-based (A-S)	95.21	72.48	82.31
Lee et al. (2018) (A-S)	88.64	77.85	82.87
Zhang et al. (2019a)(A-S)	82.44	88.75	85.48
our model (A-S)	85.31	89.87	<b>87.53</b>

Approche à base de règle pour N-S (F1 score)

N-S mapping (A-N-S) = 96.22%

S-N mapping(A-S-N) = 96.46%



Note : ablation des composants pour vérifier leur contribution au modèle (section 6)

Points forts :

- Démonstration d'un modèle permettant d'obtenir des résultats satisfaisant sur la résolution de coréférence des acteurs dans un compte rendu de procès
- Amélioration vis-à-vis de méthodes à base de règles
- Reproductibilité de l'expérience, clarté et explications détaillées, illustrées
- Explication des erreurs relevées
- Capacité **potentielle** de généralisation du modèle proposé, sur des textes différents et des langues variées

La conclusion soutient l'obtention de résultats significatifs par une « large marge », ce qui est vrai par rapport au modèle de classification de phrases mais pas vraiment par rapport aux autres modèles des travaux cités où la marge est plus fine (cf. diapo 7)

Les auteurs mentionnent leur volonté d'exploiter les connaissances spécifiques au domaine pour **enrichir les significations sémantiques des abréviations** et **d'intégrer les informations structurelles** du document dans le but d'obtenir une meilleure résolution de la coréférence.

**MERCI**