

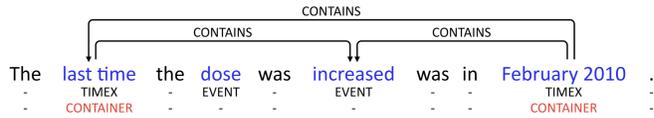
Extraction de relations temporelles dans des dossiers électroniques patient

Julien Tourille^{1,2}, Olivier Ferret³, Aurélie Névéol¹, Xavier Tannier^{1,2}

¹LIMSI, CNRS, Université Paris-Saclay, F-91405, Orsay, ²Université Paris-Sud, ³CEA, LIST, F-91191, Gif-sur-Yvette
prénom.nom@limsi.fr ; prénom.nom@cea.fr

Objectifs

1. Détection des **relations d'inclusion temporelle** entre les évènements (X contient Y)

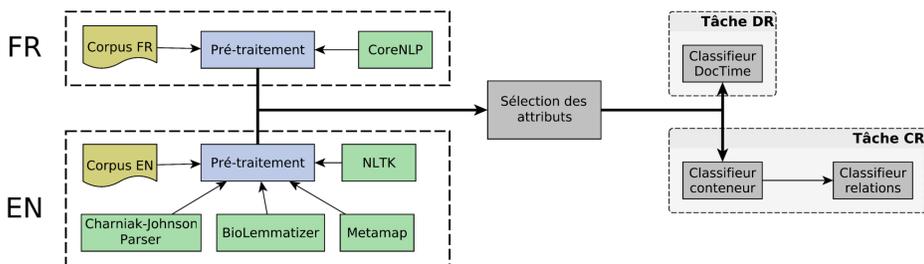


2. Détection des **relations temporelles entre les évènements et la date de création du document**



3. Développement d'un **système multilingue** (anglais et français)

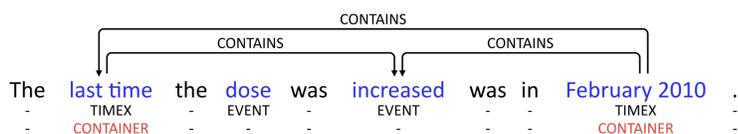
Aperçu du Système



Détection des relations d'inclusion temporelle entre les évènements

1 - Classifieur Conteneurs Narratifs

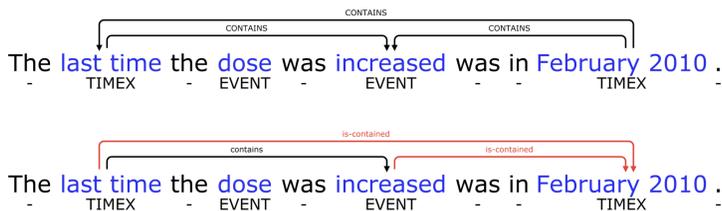
Objectif : déterminer si les entités sont susceptibles de contenir d'autres entités dans le contexte de la phrase



2 - Classifieur Relations d'Inclusion Temporelle

Objectif : classification des paires d'entités au sein des phrases

Méthode : transformation du problème à 2 catégories (*contient, pas-de-relation*) en un problème à 3 catégories (*contient, est-contenu, pas-de-relation*) afin de réduire le nombre de paires



Performance – Date Création Document

Relation	Corpus FR			Corpus EN		
	P	R	F1	P	R	F1
Bef./Over.	0,65	0,48	0,55	0,68	0,53	0,59
Before	0,85	0,44	0,58	0,87	0,83	0,85
After	0,81	0,43	0,56	0,81	0,82	0,81
Overlap	0,79	0,94	0,86	0,85	0,90	0,87
Moyenne	0,78	0,78	0,76	0,84	0,84	0,84

Corpus : Dossiers Patients

- Corpus TEMPEVAL (anglais)** : documents cliniques et pathologiques de la Mayo Clinic
- Corpus MERLOT (français)** : comptes rendus et courriers provenant d'un groupe hospitalier français

Détection des relations temporelles entre les évènements et la date de création du document (DocTime)

→ Classifieur d'évènements

Approche: classification supervisée

Classes: *before, before-overlap, overlap, after*

Stratégies

- Formes fléchies des tokens
- Représentations vectorielles des tokens calculées avec l'outil word2vec

Algorithmes

Langue	Classifieur	Algorithme	Utilisation des formes vectorielles ?
FR	CONTENEUR	SVM (Linéaire)	NON
	RELATION DocTime	SVM (Linéaire) Random Forests	OUI OUI
EN	CONTENEUR	Random Forests	NON
	RELATION DocTime	SVM (Linéaire) SVM (Linéaire)	OUI OUI

Traits d'apprentissage

Trait	DocTime	Conteneur	Relation
Type des entités	✓	✓	✓
Formes des entités	✓	✓	✓
Attributs des entités	✓	✓	✓
Positions relatives des entités dans le document	✓	✓	✓
Entités classées comme conteneurs potentiels			✓
Type du document	✓	✓	✓
Formes des entités dans les contextes gauche, centre et droit	✓	✓	✓
Types des entités dans les contextes gauche, centre et droit	✓	✓	✓
Attributs des entités dans les contextes gauche, centre et droit	✓	✓	✓
Entités marquées comme conteneur dans les contextes gauche, centre et droit			✓
Étiquettes POS des verbes de la phrase	✓	✓	
Formes des tokens contextuels (unigrammes)	✓	✓	
Étiquettes morpho-syntaxiques des tokens contextuels (unigrammes)	✓	✓	
Formes des tokens contextuels (bigrammes)	✓	✓	
Étiquettes POS des tokens contextuels (bigrammes)	✓	✓	

Performance – Inclusions Temporelles

	Corpus FR			Corpus EN		
	P	R	F1	P	R	F1
pas-de-relation	0,98	0,98	0,98	0,94	0,96	0,95
contient	0,62	0,59	0,61	0,62	0,48	0,54
Moyenne	0,96	0,96	0,96	0,90	0,91	0,91

Remerciements

Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche sous la référence CABeRnE ANR-13-JS02-0009-01 et d'un financement du Labex DigiCosme.