

# Temporal information extraction from clinical text

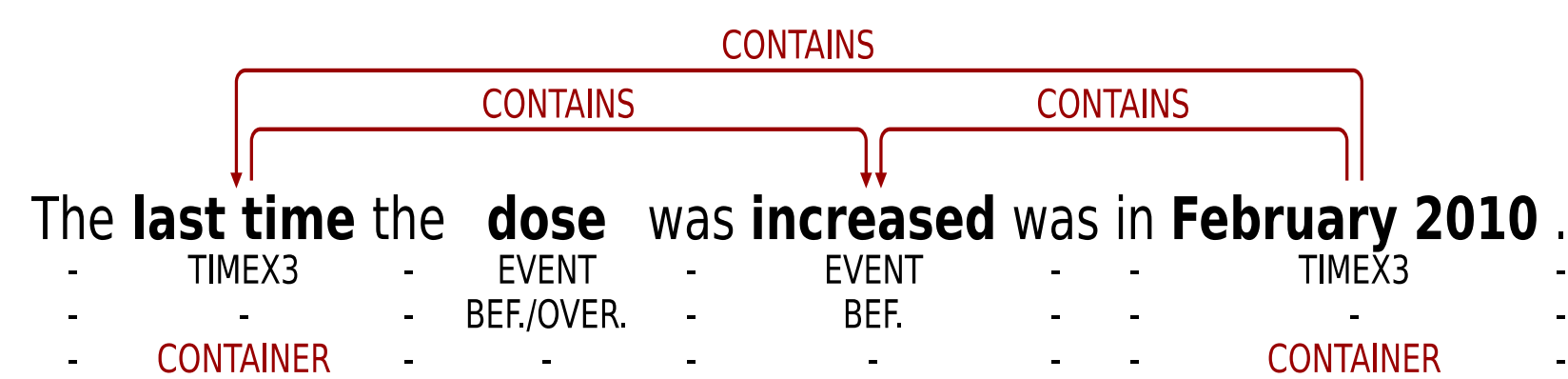
Julien Tourille<sup>1,2</sup>, Olivier Ferret<sup>3</sup>, Xavier Tannier<sup>1,2</sup>, Aurélie Névéol<sup>1</sup>

<sup>1</sup>LIMSI, CNRS, Université Paris-Saclay, F-91405, Orsay, <sup>2</sup>Université Paris-Sud, <sup>3</sup>CEA, LIST, F-91191, Gif-sur-Yvette

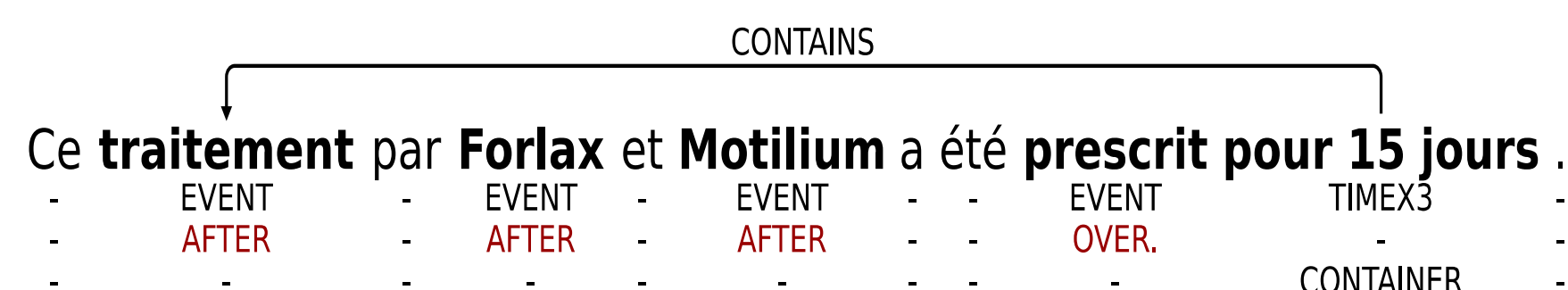
<sup>1</sup>firstname.lastname@limsi.fr; <sup>3</sup>firstname.lastname@cea.fr

## Objectives

1. **Contains relation extraction** between medical events and/or temporal expressions



2. **Document Creation Time (DCT) relation extraction** between medical events and documents



3. **Multilingual methods and models** (French and English)

## Corpora: Electronic Health Records

1. **THYME (English)**: clinical and pathological documents from the Mayo Clinic
2. **MERLOT (French)**: clinical documents from a Gastroenterology, Hepatology and Nutrition department

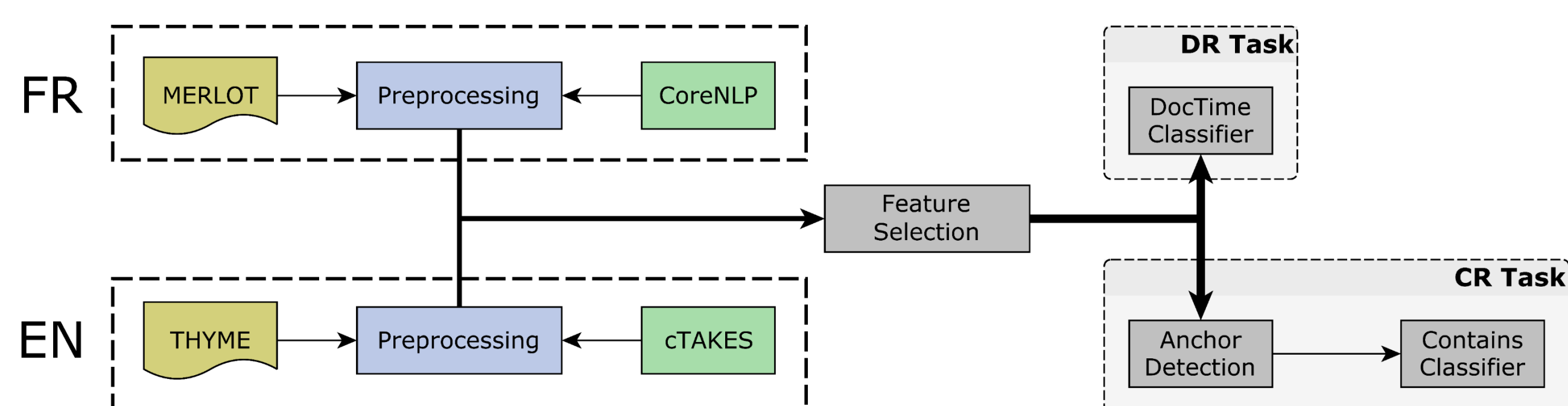
## Temporal relation extraction between medical events and Document Creation Time (DCT)

### → Event classification

**Method:** supervised classification

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

## System Overview



## Strategies

1. Plain lexical forms
2. Word embeddings computed with word2vec

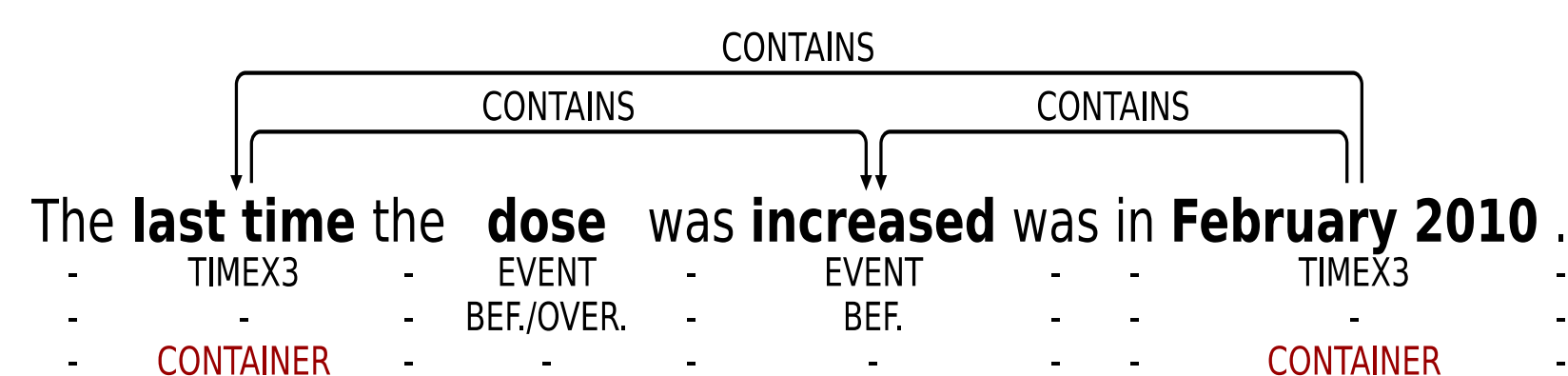
## Best Algorithms and Strategies

Language	Classifier	Algorithm	Word Embeddings ?
	IS_CONTAINER	SVM (Linear)	NO
FR	CONTAINS_REL	SVM (Linear)	NO
	DocTime	SVM (Linear)	NO
EN	IS_CONTAINER	SVM (Linear)	NO
	CONTAINS_REL	SVM (Linear)	NO
	DocTime	SVM (Linear)	NO

## Contains relation extraction

### 1 – Container anchor identification

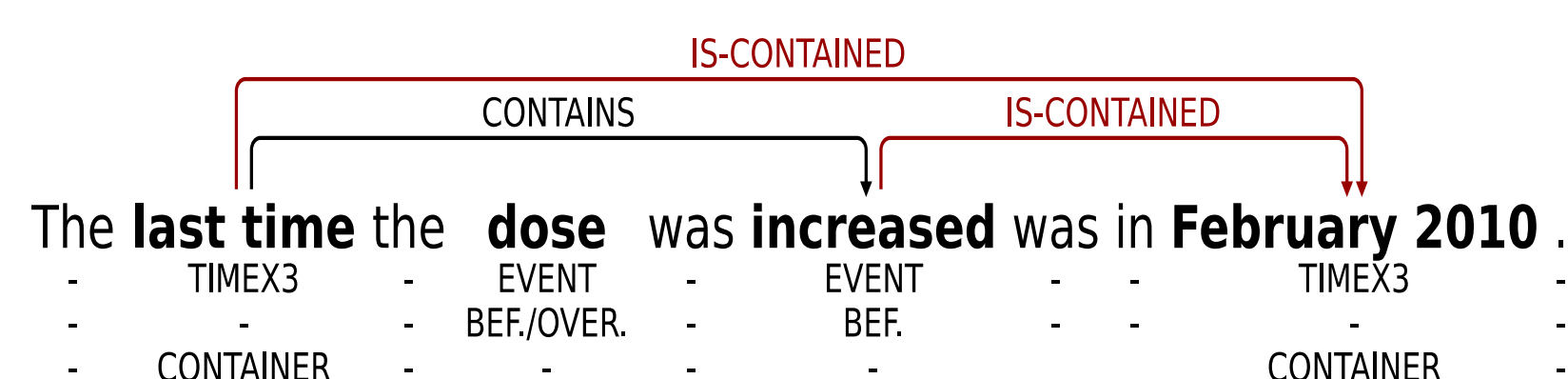
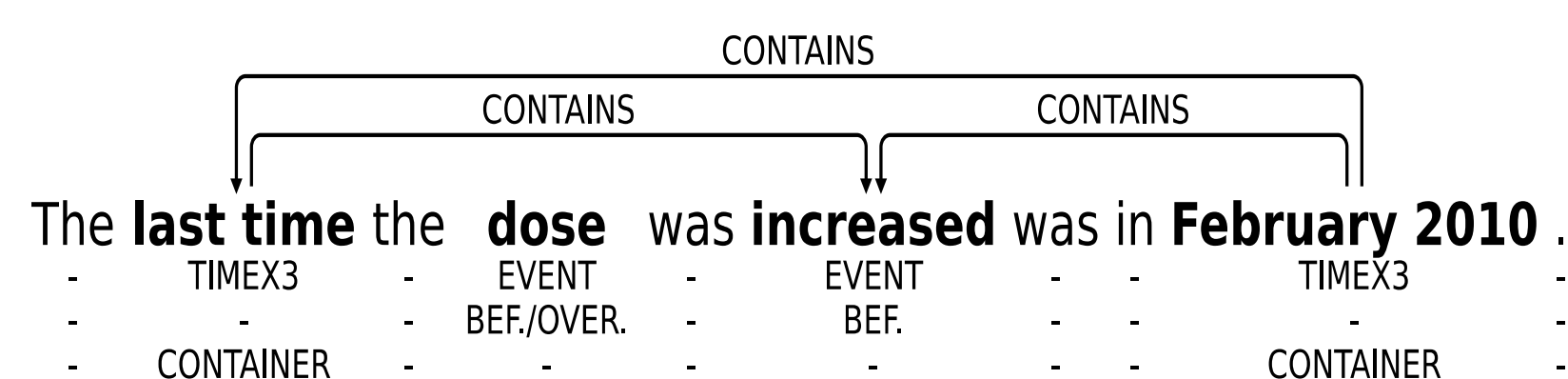
**Objective:** detect entities that are more likely to be the anchor of narrative containers



### 2 – Contains relation extraction

**Objective:** entity pair classification

**Method:** we cast a 2-category problem (contains, no-relation) as a 3-category problem (contains, is-contained, no-relation)



## Features

Feature	DocTime	Container	Contains
Entity type	✓	✓	✓
Entity form	✓	✓	✓
Entity attributes	✓	✓	✓
Entity position (within the document)	✓	✓	✓
Container model output			✓
Document type	✓	✓	✓
Contextual entity forms	✓	✓	✓
Contextual entity types	✓	✓	✓
Contextual entity attributes	✓	✓	✓
Container model output for contextual entities			✓
PoS tag of the sentence verb	✓	✓	
Contextual token forms (unigrams)	✓	✓	
Contextual token PoS tags (unigrams)	✓	✓	
Contextual token forms (bigrams)	✓	✓	
Contextual token PoS tags (bigrams)	✓	✓	

## Results – DCT Relations

	French (MERLOT)			English (THYME)		
	P	R	F1	P	R	F1
baseline	0.67	0.67	0.67	0.47	0.47	0.47
bef./over.	0.68	0.69	0.69	0.73	0.60	0.66
before	0.81	0.60	0.69	0.88	0.88	0.88
after	0.79	0.69	0.73	0.84	0.84	0.84
overlap	0.88	0.92	0.90	0.88	0.90	0.89
micro-average	0.83	0.84	0.83	0.87	0.87	0.87

## Results – Contains Relations

	French (MERLOT)			English (THYME)		
	P	R	F1	P	R	F1
baseline	0.43	0.15	0.22	0.55	0.06	0.11
no-relation	0.99	1.00	0.99	0.96	0.98	0.97
contains	0.75	0.57	0.65	0.61	0.47	0.53
micro-average	0.98	0.98	0.98	0.93	0.94	0.93

## Acknowledgements

This work was supported in part by the French National Agency for Research in Informatics Department at the under grant CABeRneT ANR-13-JS02-0009-01, by Labex DigiCosme, operated by the Foundation for Scientific Cooperation (FSC) Paris-Saclay, under grant CÔT and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676207.