



Plate-formes logicielles pour le TAL 3 : Unitex - CasSys

Karën Fort

karen.fort@sorbonne-universite.fr / <https://members.loria.fr/KFort/>



Quelques sources d'inspiration

- ▶ Manuel d'Unitex : <http://unitexgramlab.org/releases/3.1/man/Unitex-GramLab-3.1-usermanual-fr.pdf>
- ▶ Page de CasEN : <https://tln.lifat.univ-tours.fr/version-francaise/ressources/casen>
- ▶ Denis Maurel : son tutoriel et ses conseils

Sources

Introduction

Présentation

Historique

Démonstration

Attention

Utilisation de CasSys

Création d'une cascade

Pour finir

Bibliographie

Cascades de transducteurs :

- ▶ applique **plusieurs** graphes,
- ▶ l'un **après** l'autre,
- ▶ qui modifient le texte

Conséquences ?

Pourquoi des cascades de transducteurs

- ▶ pouvoir utiliser dans les descriptions suivantes les motifs déjà détectés
 - ▶ éviter un étiquetage non souhaité pour un motif déjà reconnu.
- ⇒ ordre de passage de ces transducteurs est donc important

Origines de CasSys

- ▶ créé en 2002, à Tours (LI), pour reconnaître les entités nommées
- ▶ puis généralisé
- ▶ puis intégré à Unitex

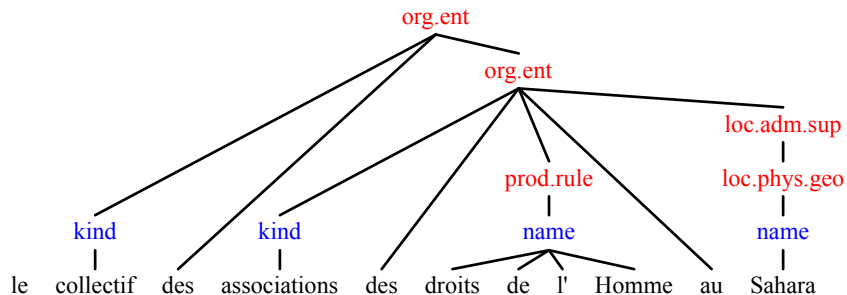
Entités nommées, késako ?

`<ENAMEX TYPE="PERSON">Henri</ENAMEX> a acheté <NUMEX TYPE="QUANTITY">300</NUMEX> actions de la société <ENAMEX TYPE="ORGANIZATION">AMD</ENAMEX> en <TIMEX TYPE="DATE">2006</TIMEX>.`

http://fr.wikipedia.org/wiki/Reconnaissance_d%27entit%C3%A9s_nomm%C3%A9es

- ▶ première création *langagière* liée à une [application](#)
- ▶ apparues pour la première fois en **1995**, lors de la campagne d'évaluation MUC 6, organisée par le département de la défense américain (DARPA)

Entités nommées : évolution



[Grouin et al., 2011]

Démo de CasSys

CasEN sur le TDM

Cascades vs composition



une cascade n'est pas bi-directionnelle

Cascades vs sous-graphes



dans une cascade, un graphe n'est pas à l'intérieur de l'autre
(plus générique)

Cascades vs sous-graphes



dans une cascade, il est impossible d'ajouter un sous-graphe

Application itérative

CasSys permet d'appliquer un graphe de manière **itérative**, tant que le texte est transformé (*Until fix point*)

Application itérative

CasSys permet d'appliquer un graphe de manière **itérative**, tant que le texte est transformé (*Until fix point*)

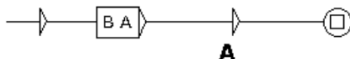


FIGURE 12.6 – Transducteur qui modifie BA en A

initial text	B	B	B	A	A	A	
itération 1		B	B	A	A	A	1 match
itération 2			B	A	A	A	1 match
itération 3				A	A	A	1 match
itération 4				A	A	A	0 match

Conséquences ?

CasSys utilise des graphes compilés (en .fst2)

Ilôts de certitude

Comment éviter que des motifs précédemment reconnus soient ambigus avec ceux reconnus par les graphes suivants ?

{blabla,.tag1+tag2+tagn}

→ crée une entrée de dico (devient un mot polylexical) pour ce qui est reconnu

Pour l'utiliser : < *NomDuGraphe* >

Corpus de test (fourni)

« Le maire de la ville a organisé une grande cérémonie le mardi 11 novembre 2014 pour modifier le nom de la rue de Paris en rue du 11 novembre 1918. Ce mardi 11 novembre prouve que nous n'avons pas oublié les anciens combattants de notre ville. D'ailleurs il n'est pas impossible qu'une Maison communale du 11 novembre voit le jour pour célébrer le 11 novembre 2018. »

Extraire les dates absolues

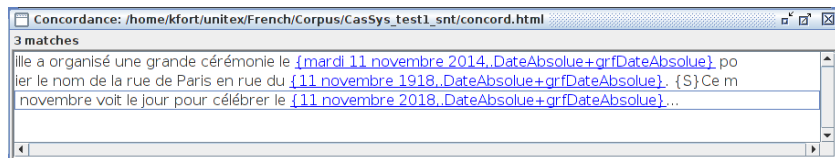
Exercice

Créer un graphe reconnaissant les dates absolues, de type :

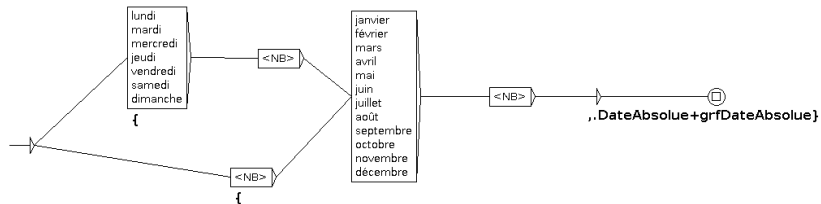
« (lundi) 30 novembre **2015** »

- ▶ annoter DateAbsolue
- ▶ ajouter la provenance de l'extraction (le nom du graphe)

(voir ci-dessous)



Graphe : dates absolues



Extraire les dates relatives

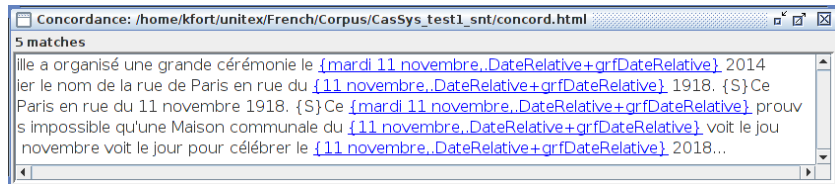
Exercice

Créer un graphe reconnaissant les dates relatives, de type :

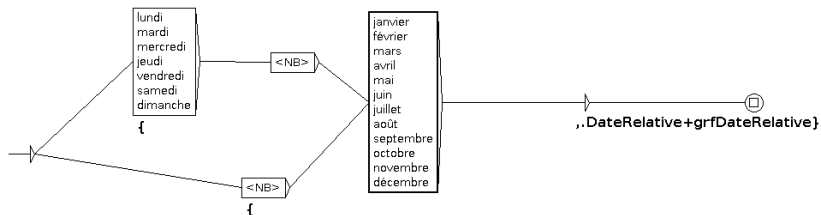
« (lundi) 30 novembre »

- ▶ annoter DateRelative
- ▶ ajouter la provenance de l'extraction (le nom du graphe)

(voir ci-dessous)



Graphe : dates relatives



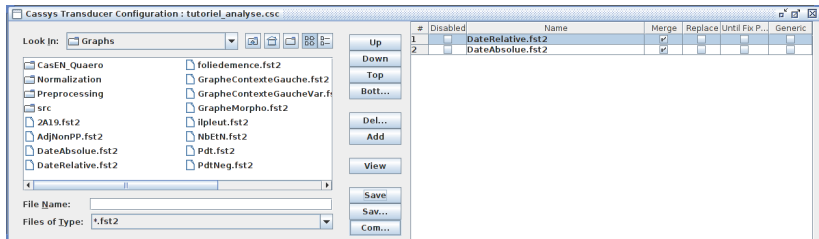


Comment faire pour que les dates soient correctement annotées ?

Créer la cascade

Exercice

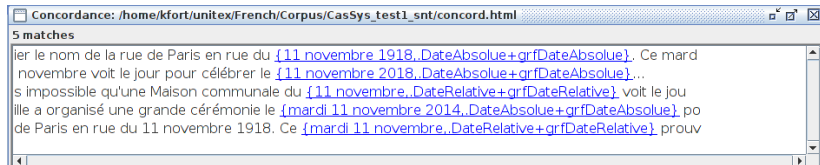
Créer la cascade reconnaissant les entités nommées du corpus de test (qui va contenir les graphes créés ici)



Appliquer la cascade

Exercice

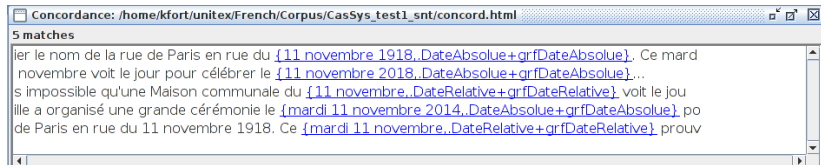
Appliquer la cascade reconnaissant les dates du corpus de test



Appliquer la cascade

Exercice

Appliquer la cascade reconnaissant les dates du corpus de test



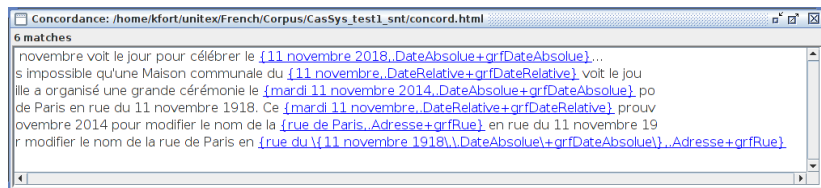
? Pourquoi ça ne marche pas ?

Extraire les noms de rues

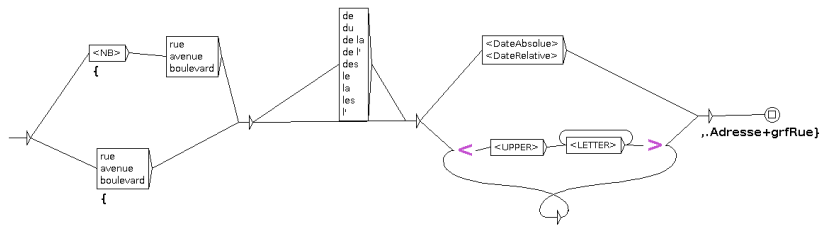
Exercice

Créer un graphe reconnaissant les noms de rues, de type : « (28) rue Serpente »

- ▶ annoter NomRue
- ▶ ajouter la provenance de l'extraction (le nom du graphe)



Graphe : rues

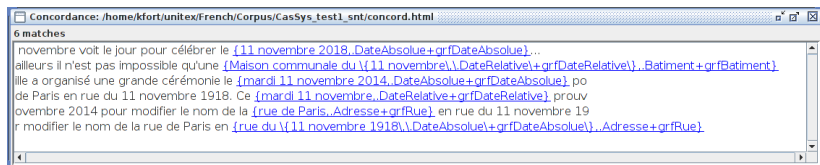


Extraire les noms de bâtiments

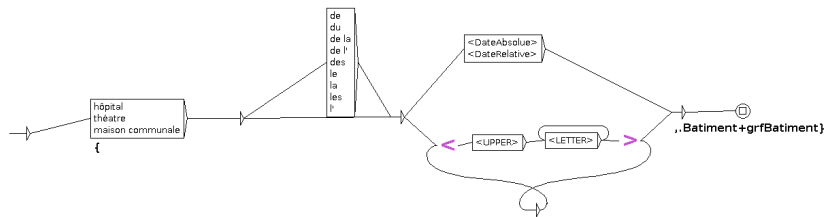
Exercice

Créer un graphe reconnaissant les noms de bâtiments, de type :
« hôpital de la Pitié Salpêtrière »

- ▶ annoter NomBat
- ▶ ajouter la provenance de l'extraction (le nom du graphe)



Graphe : bâtiments



Sources

Introduction

Utilisation de CasSys

Création d'une cascade

Pour finir

CQFR : Ce Qu'il Faut Retenir

Bibliographie



- ▶ principes de CasSys
- ▶ utilisation de CasSys



Grouin, C., Rosset, S., Zweigenbaum, P., Fort, K., Galibert, O., and Quintard, L. (2011).

Proposal for an extension of traditional named entities : From guidelines to evaluation, an overview.

In Proceedings of the 5th Linguistic Annotation Workshop, pages 92–100, Portland, Oregon, USA.

Poster.