



Plate-formes logicielles pour le TAL 3 : Unitex - CasSys 2

Karën Fort

karen.fort@sorbonne-universite.fr / <http://karenfort.org>

12 octobre 2018



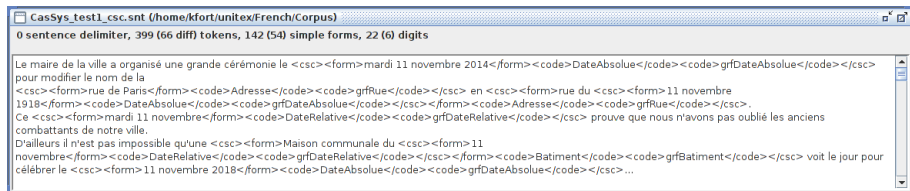
Quelques sources d'inspiration

- Manuel d'Unitex : <http://unitexgramlab.org/releases/3.1/man/Unitex-GramLab-3.1-usermanual-fr.pdf>
- Page de CasEN :
http://tln.li.univ-tours.fr/Tln_CasEN.html
- Denis Maurel : son tutoriel et ses conseils

À savoir

- le nettoyage des balises est effectué à l'aide d'une cascade dite « de synthèse »
- une cascade de synthèse s'applique sur le fichier résultat de l'analyse :
 - ▶ nomDeFichier_**csc.txt**
- le résultat d'une cascade de synthèse est stocké dans un autre fichier :
 - ▶ nomDeFichier_**csc_csc.txt**

Fichier résultat : _csc.txt



```

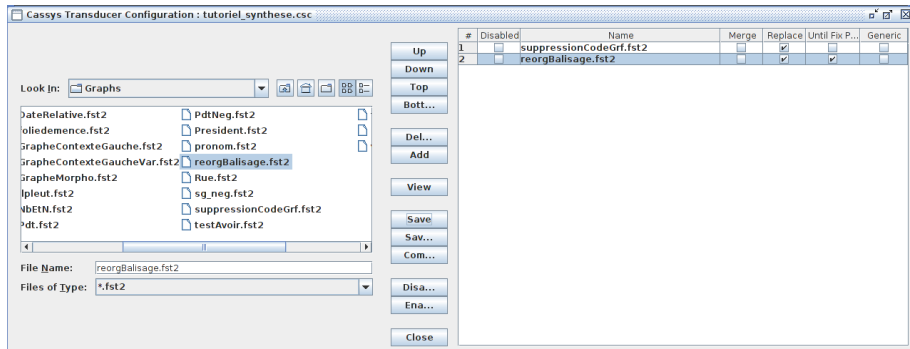
CasSys_test1_csc.snt (/home/kfort/unitex/French/Corpus)
0 sentence delimiter, 399 (66 diff) tokens, 142 (54) simple forms, 22 (6) digits

Le maire de la ville a organisé une grande cérémonie le <csc><form>mardi 11 novembre 2014</form><code>DateAbsolue</code><code>grfDateAbsolue</code></csc>
pour modifier le nom de la
<csc><form>rue de Paris</form><code>Adresse</code><code>grfRue</code></csc> en <csc><form>rue du <csc><form>11 novembre
1918</form><code>DateAbsolue</code><code>grfDateAbsolue</code></csc></form><code>Adresse</code><code>grfRue</code></csc>.
Ce <csc><form>mardi 11 novembre</form><code>DateRelative</code><code>grfDateRelative</code></csc> prouve que nous n'avons pas oublié les anciens
combattants de notre ville.
D'ailleurs il n'est pas impossible qu'une <csc><form>Maison communale du <csc><form>11
novembre</form><code>DateRelative</code><code>grfDateRelative</code></csc></form><code>Batiment</code><code>grfBatiment</code></csc> voit le jour pour
célébrer le <csc><form>11 novembre 2018</form><code>DateAbsolue</code><code>grfDateAbsolue</code></csc> ...

```

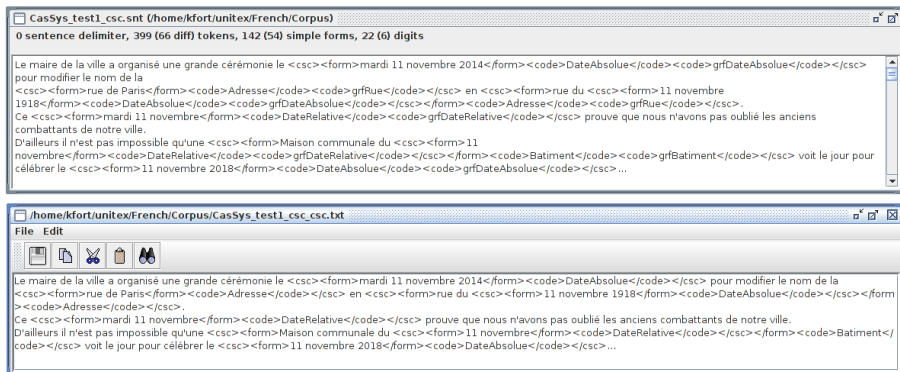
- que remarquez-vous ?
- que faut-il enlever ? pourquoi ?
- comment ?

Cascade de synthèse

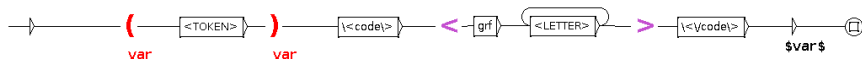


Nettoyage : enlever les marqueurs de déboguage

grfRue, grfDateAbsolue, etc.



Grappe : enlever les marqueurs de déboguage

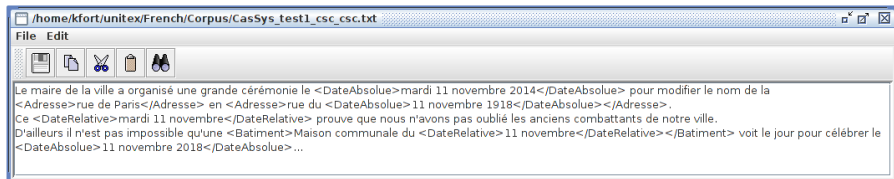
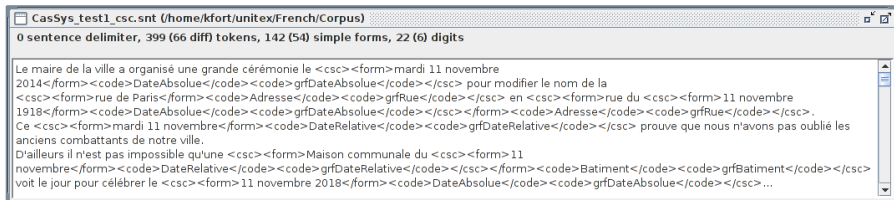


Réorganisation : créer les balises finales

`< DateAbsolue >` au lieu de :

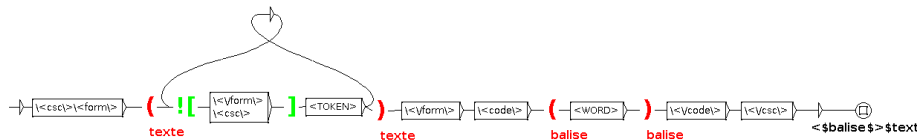
`< csc >< form >mardi 11 novembre`

`2014< /form >< code > DateAbsolue < /code >< /csc >`



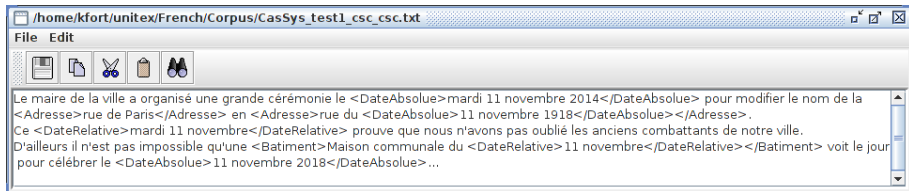
Tip : utiliser les contextes droits

Graphe : créer les balises finales

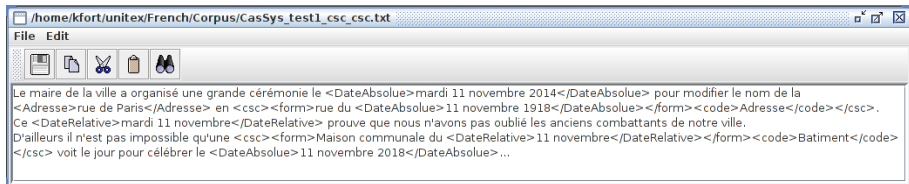


Until fix point : effet

Avec :



Sans :



1 Sources

2 Nettoyage

3 Pour finir

- CQFR : Ce Qu'il Faut Retenir



- principes de CasSys
- utilisation de CasSys