Plateforme TXM - Feature # 3522

Statut:	New	Priorité:	Normal
Auteur:	Serge Heiden	Catégorie:	Import
Créé :	30/11/2023	Assigné à:	
Mis-à-jour :	18/12/2023	Echéance:	
Sujet:	VOEUX corpus, enhance		
Description			
FR:			
<ul style="list-style-type: none">- encoder les paragraphes dans les textes de 1975 et 1987- homogénéiser les guillemets « » "- améliorer tokenisation des clitiques et t euphoniques			
Tokenizer			
<ul style="list-style-type: none">- tirets : https://forge.cbp.ens-lyon.fr/redmine/issues/3005- clitiques : https://forge.cbp.ens-lyon.fr/redmine/issues/2940, https://forge.cbp.ens-lyon.fr/redmine/issues/217- t euphonique : https://forge.cbp.ens-lyon.fr/redmine/issues/3090- common cases : https://forge.cbp.ens-lyon.fr/redmine/issues/1779- abréviations et mots multi-tokens : https://forge.cbp.ens-lyon.fr/redmine/issues/3520			

Historique

#1 - 30/11/2023 13:38 - Serge Heiden

- Description mis à jour

#2 - 30/11/2023 13:59 - Benedicte Pincemin

- Description mis à jour

#3 - 18/12/2023 09:56 - Matthieu Decorde

- Version cible changé de TXM 0.8.4 à TXM 0.8.3

#4 - 18/12/2023 10:03 - Matthieu Decorde

- Version cible changé de TXM 0.8.3 à TXM 0.8.4