

# Philologie numérique progressive dans l'environnement TXM - établissement du texte/corpus pour les chercheurs et le logiciel

Serge Heiden, ENS de Lyon

*Atelier Cahier Édition analytique*  
ENS Lyon, 30 juin – 4 juillet



# Philologie et activité d'édition traditionnelle

- Établissement du texte
- Apparat critique
- Outils éditoriaux : introduction, glossaire, index, sommaire, table des matières, etc.
- Publication et diffusion sur support papier

# Philologie et édition numérique aujourd'hui

- Reproduction du modèle traditionnel
- Circuits numériques :
  - Transcription ou OCR - traitement de texte (Word)
  - Encodage - XML-TEI (Oxygen)
  - Mise en forme (InDesign)
  - Diffusion papier (PDF) + web (Omeka ou Epub)
- **Lecture** hypertextuelle
- **Recherche par catalogue** (**métadonnées**) et en **texte intégral** (**mots**)
- Intégration dans grandes collections ou **corpus** (**données**)

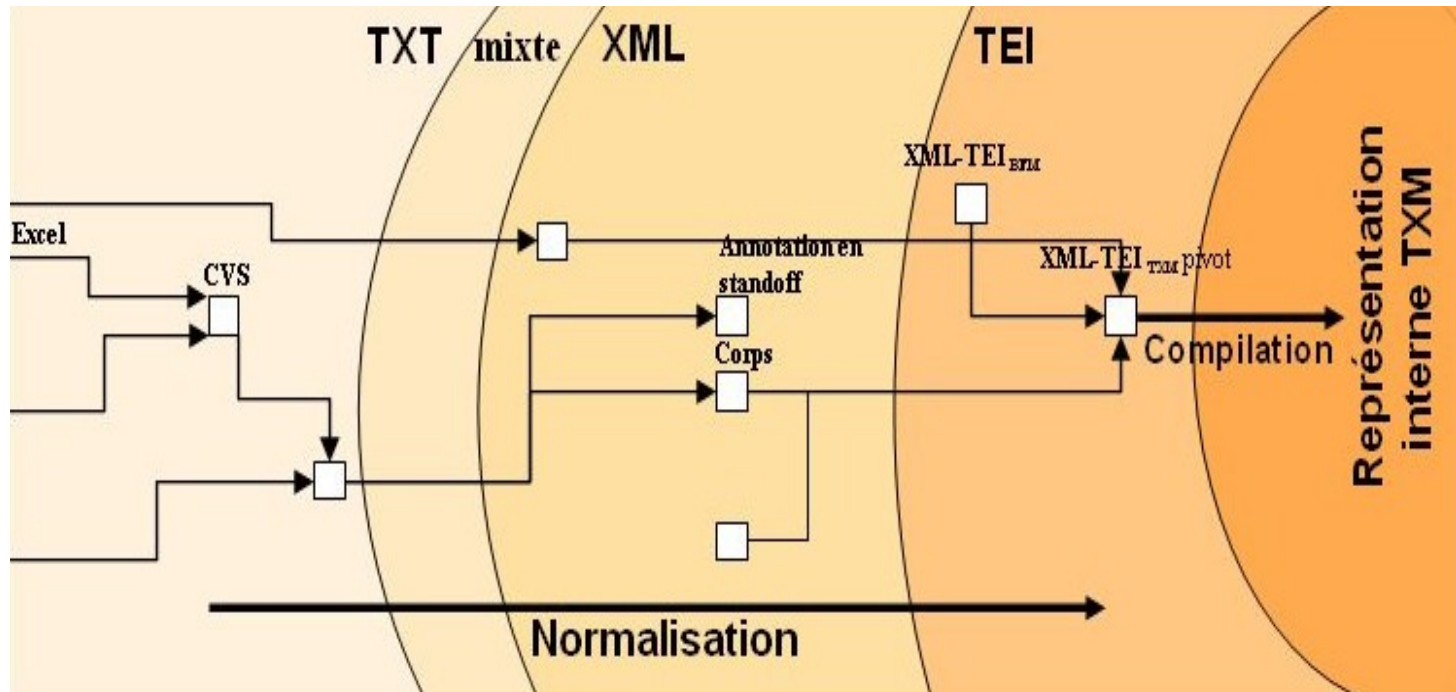
# TXM en tant qu'outil d'analyse (pour poste ou web)

- Documentaires ou qualitatifs
  - **Lecture** de l'édition des textes
  - Listes de **fréquences** : formes, lemmes... (**mots**)
  - Concordances : de **motifs** de mots combinant leurs propriétés
  - Graphique de progression
- Contrastifs ou quantitatifs
  - Analyse factorielle des correspondances, Spécificités, Classification
  - Cooccurrences
- Configurations de **corpus**
  - Sous-corpus : groupes de textes (**métadonnées**), structures **internes**, mots
  - Partitions : groupes de textes (**métadonnées**), structures internes, mots

# Sources textuelles analysables par TXM

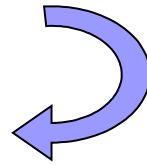
- **Corpus de textes écrits** :
  - Presse-papier
    - >> TXT (Unicode)
    - >> XML/w
    - >> XML-TEI (BFM, BVH, Perseus, TextGrid, etc.)
  - éditions multiples : niveaux diplomatiques
  - multimédia : images folio ou colonnes
- Corpus de transcriptions de l'**oral** :
  - synchronisation, tours de parole
  - multimédia : audio/vidéo
- Corpus **alignés** : traduction, version

Importer des sources dans TXM  
= préparer+équiper+intégrer  
→ pour l'analyse



# Chaîne d'import type

- Source = répertoire de fichiers
- Textes brut
- Textes XML
- Textes étiquetés
- Contrastes : sous-corpus & partition
- Contextes
- Propriétés lexicales



TreeTagger



Import



Analyse

# Exemple d'import – annotation inline

```
207 <lb n="52"/>sa hanste brise par asteles.  
208 <lb n="53"/>E Gorm<supplied rend="crochets">un</supplied>d ad l'espee traite,  
209 <lb n="54"/>si l'ad feru sur le heaume :  
210 <lb n="55"/>la teste en fist voler a destre,<pb n="6"/>
```

```
2103 <lb n="53"/>  
2104 <s n="17" id="s_17">  
2105 <w id="w_360">  
2106 <txm:form>E</txm:form>  
2107 <interp resp="#txm" type="#ttpos">CONcoo</interp>  
2108 <interp resp="#txm" type="#ttlemma">--</interp>  
2109 </w>  
2110 <w id="w_361">  
2111 <txm:form>Gorm<supplied rend="crochets">un</supplied>d</txm:form>  
2112 <interp resp="#txm" type="#ttpos">NOMpro</interp>  
2113 <interp resp="#txm" type="#ttlemma">Gormund</interp>  
2114 </w>  
2115 <w id="w_362">  
2116 <txm:form>ad</txm:form>  
2117 <interp resp="#txm" type="#ttpos">VERcjpg</interp>  
2118 <interp resp="#txm" type="#ttlemma">--</interp>  
2119 </w>  
2120 <w id="w_363">  
2121 <txm:form>l&apos;</txm:form>  
2122 <interp resp="#txm" type="#ttpos">DETdef</interp>  
2123 <interp resp="#txm" type="#ttlemma">--</interp>  
2124 </w>
```




# Modèle de données de TXM

- A. Unités textuelles : roman, article, entretien...
  - métadonnées : titre, auteur, date, domaine, genre... (dimensions sociolinguistiques)
- A. Structures internes : sections, paragraphes...
  - *plans textuels* : notes, titres, discours direct, multi-locuteur, multilingue partiel, hors texte
- A. Unités lexicales : forme graphique, lemme, catégorie grammaticale...
- B. Édition (retour au texte) : propre, *tierce*
- C. Synchronisation transcription / audio-vidéo
- D. Alignement (entre langues, entre versions, texte/images de manuscrits)

# Encodage progressif et analyses

	Presse- papier	TXT	XML/w	XML-TEI
<i>Unités textuelles</i>	-	fichiers	fichiers	fichiers
<i>Métadonnées</i>	-	CSV	CSV	teiHeader
<i>Structures</i>	-	-	toutes	spécifique
<i>Plans</i>	-	-	pré- XSL	spécifique
<i>Unités lexicales</i>	brut	brut	<w>?	<w>?

*Partners TEI sources* |  = XSLT

# Encodage progressif et Analyses

	Presse- papier	<i>métadonnées</i> TXT	XML/w <i>structures, mots</i>	XML-TEI
<i>Unités textuelles</i>	-	fichiers	fichiers	fichiers
<i>Métadonnées</i>	-	CSV	CSV	teiHeader
<i>Structures</i>	-	-	toutes	spécifique
<i>Plans</i>	-	-	pré- XSL <i>plans</i>	spécifique
<i>Unités lexicales</i>	brut	brut	<w>?	<w>?

**Partners TEI sources** | **= XSLT**



Merci

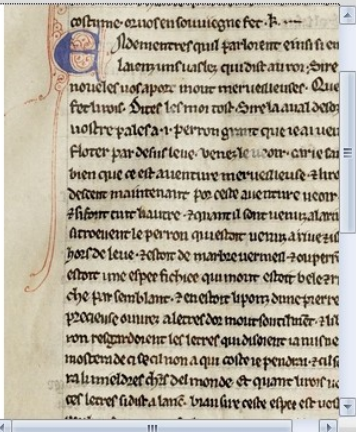
<http://textometrie.ens-lyon.fr>

[textometrie@ens-lyon.fr](mailto:textometrie@ens-lyon.fr)

# Copie d'écran de la version portail – Concordance de motif & Édition synoptique

afc Mon profil | Se déconnecter | Aide | Contact | fr

Accueil | GRAAL: 'espee' [pos="VER.\*"] x



[S 7] **E** Ndemetres quil parloient einfi si ent laienz un vaslez qui dist au roi . Sire nouleuf uof apozt mout merueilleuf . Quelel fet li roif . Dites les moi tost . Sire la aual defoz

10 uofre pales a .i. perron grant que se ai ueu floter par defus leue . Venez le ueoir . car se fau bien que ce est aenture merueilleuse . et roif descient maintenant por ceste aenture ueoir . et sifont tuit haute . et quant il font uenez alaru

15 sifroeuent le perron que estoit uenez ariueciffu : hozf de leue . et estoit de marbre uermel . et pouz estoit une **espee fichiee** qui mout estoit bele : che par semblant . et estoit lizonz dune pierre precieufe ouurez a letres d'or mout sou

20 ron resgardoient les letres qui disoient ia nulne moftera de ci se cil non a qui coste ie pendrai . et ces lettres si dist a Lancelot : « Bia u sire ces par bon droit . car ie sa i bien que vos es

[S 7] **E** Ndemetres qu'il parloient einfi si e laienz uns vaslez qui dist au roi . Sire noueles uos apozt mout merueilleuses . fet li rois . Dites les moi tost . Sire la aual

10 uostre pales a .i. perron grant que ie ai floter par desus l'eu . Venez le ueoir , ca bien que ce est aenture merueilleuse . et descient maintenant por ceste aenture et si font tuit li autre , et quant il sont ven

15 si troeuent le perron qui estoit uenez a hors de l'eu , et estoit de marbre verme estoit une **espee fichiee** qui mout esto che par semblant , et en estoit li ponz d'i precieuse ouurez a letres d'or mout sou

20 ron resgardoient les lettres qui disoient m'ostera de ci se cil non a qui costé ie p ra li mieldres **cheualiers** del monde . Et ces lettres si dist a Lancelot : « Bia u sire par bon droit . car ie sa i bien que vos es

Seigneur, en contrebas de votre palais, il y a un gros bloc de pierre que j'ai vu flotter sur l'eau. Venez le voir, car je sais bien que c'est là un prodige extraordinaire. " Et le roi descend aussitôt pour voir ce prodige, et tous les autres font de même ; et quand ils ont rejoint la rive, ils trouvent la grosse pierre qui est parvenue jusque là et s'est échouée ; et c'était une roche de marbre vermeil, et dans le bloc était fichée une épée qui paraissait splendide et de grande valeur ; et la garde en était d'une pierre précieuse gravée très finement de lettres d'or ; et les barons regardaient l'inscription qui disait : 'Nul jamais ne m'ôtera d'ici sinon celui au côté duquel je prendrai, et il sera le meilleur chevalier du

page | GRAAL - ms-col ...

Requête : 'espee' [pos="VER.\*"] | Chercher | Réglages

Référence	Contexte gauche	Pivot	Contexte droit
1 graal_cm, col. 161a, l. 17	de l'euve, et estoit de marbre vermeil, et ou perron estoit une	espee fichiee	qui mout estoit bele et riche par semblant, et en estoit li ponz d'
2 graal_cm, col. 161a, l. 23	quant li rois voit ces lettres si dist a Lancelot: «Biau sire ceste	espee est	vostre par bon droit, car je sai bien que vos estes li mieldres chevaliers
3 graal_cm, col. 161b, l. 4	, i essaieroz vos por ce que je le voil. Ne mie por l'	espee avoir	» Et il gite la main et prent l'espee par le heut et
4 graal_cm, col. 162c, l. 30	Galaad: «Sire veez ci l'aventure donc je vos parlai, a ceste	espee tere	fors de cest perron ont hui failli des plus proisiez chevaliers de mon ostel qui
5 graal_cm, col. 162c, l. 35	n'ert pas lor, et por la grant seurté que je avoie de ceste	espee avoir	n'en aportai je point a cort si come vos peustes veoir.» Et
6 graal_cm, col. 171a, l. 29	efforce tant qu'il lor fait guerpir place, et les atorne tiex a l'	espee trenchant	que armeure ne les puet garantir qu'il ne lor face le sanc saillir des
7 graal_cm, col. 182b, l. 12	asena en cel leu meïsmes ou il l'avoit asené au commencement, et l'	espee fu	legiere et bone si cola legierement parmi la teste puis que li cuirs fu entamez
8 graal_cm, col. 205d, l. 18	nomme, et quant il le conoist si le deffie et li cort sus l'	espee trete	, et li done si grant cop com il puet amener del brant. Quant

19 occurrences | Nombre de lignes : 100 | Exporter