

Formation EnExDi 2020

Introduction: chaîne(s) de traitement(s)

Simon Gabay

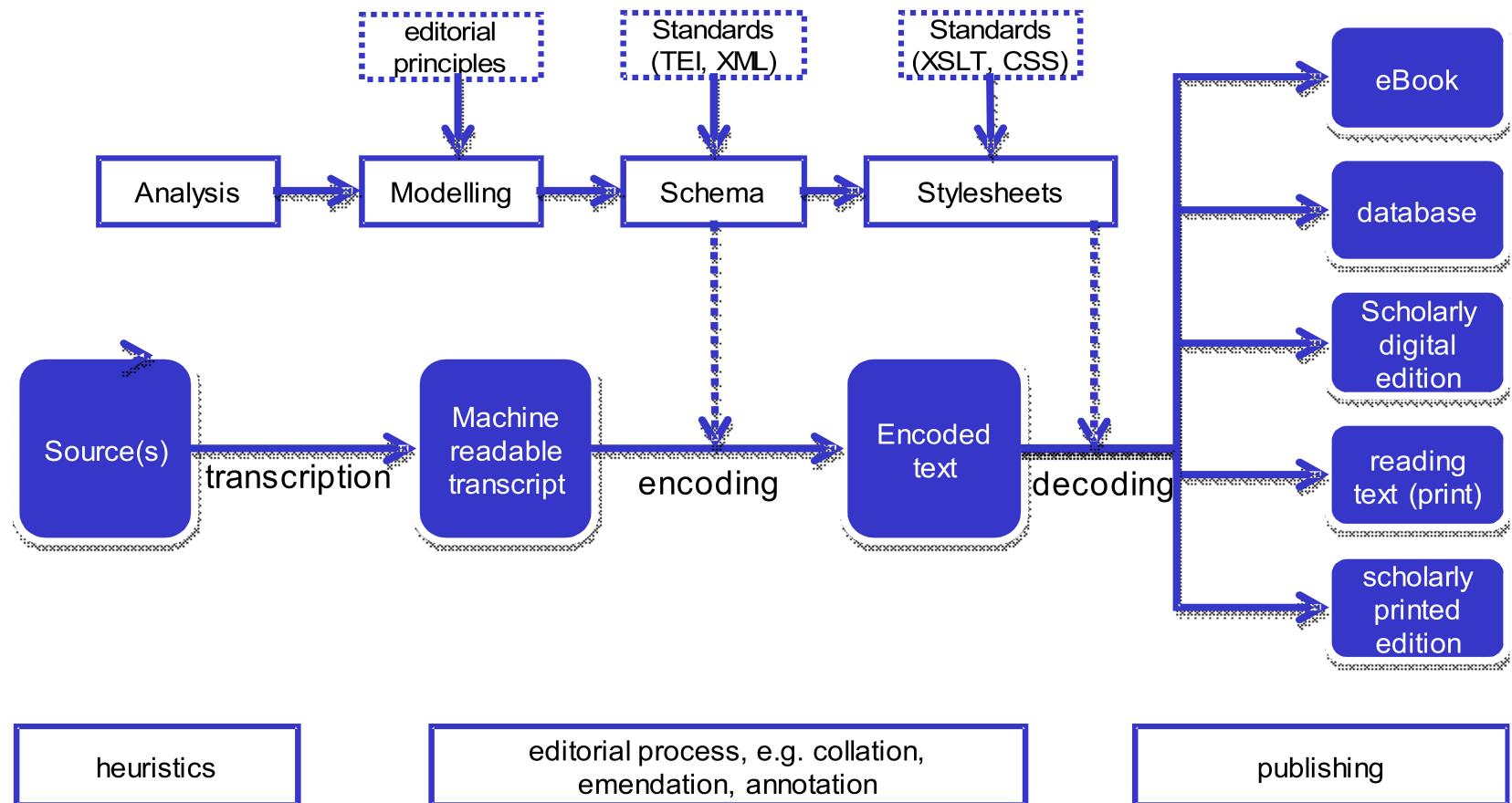
Poitiers, mardi 11 février 2020

Chaîne de traitement

On parle de chaîne de traitement (ou "flux de travail" selon la Commission générale de terminologie et de néologie) ou de *Workflow*. Comme aucune solution informatique ne permet de tout faire (à l'inverse de logiciel comme *Word* en bureautique), il faut trouver

1. Une série de solutions...
2. ... qui s'articulent correctement les unes avec les autres...
3. ... et qui correspondent à des standards.

Un exemple de chaîne de traitement pour l'édition numérique



(Quelle: Rehbein/Fritze 2012)

Source: Christof Schöch, *Digitale Textedition mit TEI, en ligne.*

Philologie numérique

L'édition numérique reprend les étapes de la philologie traditionnelle. Elle ouvre de nouvelles potentialités, malheureusement au prix d'une complexification du travail.

Retour à la renaissance, ou, comme Alde Manuce, l'humaniste maîtrise l'intégralité de la chaîne de production, de la transcription à la publication, en passant par la fabrication des outils (presse, fontes...).

L'édition numérique est avant tout une édition, et nécessite des compétences en écdotique traditionnelle.

Quelques grandes étapes

1. Transcription -> Kraken, Ocropy, Tesseract...
2. Collation -> Collatex, Juxta...
3. Analyse paléographique -> Archetype...
4. Annotation linguistique -> TreeTagger, Marmot, Pie...
5. Exploitation linguistique -> TXM, Unitex/GramLab...
6. Exploitation littéraire -> Pour les emprunts: Tracer ou Philologic
7. Indexation -> HER, GROBID entity fishing...
8. Publication -> TEIPublisher, Synoptix, LaTeX
9. Archivage -> HAL, Huma-num

Pourquoi dois-je (presque) tout faire?

L'objectif d'avoir un équivalent de *Word* n'est pas nécessairement souhaitable. Toute simplification se paye:

- Au sens propre avec l'apparition de solutions privées, donc payantes.
- Au sens figuré, avec l'enfermement dans une solution générale qui gère mal les cas particuliers.

Cependant, il existe déjà des chaînes de traitement fonctionnelles et de très grande qualité, comme [METOPES](#) à l'université de Caen.

Quelques grands principes

1. Ouvert
2. Pérenne
3. Interopérable

Un exemple d'enchaînement



Récupération des données

Le site [Dramacode](#) publie en ligne les transcriptions en XML-TEI

Dramacode, le catalogue, textes de théâtre en libre accès

Ce site hébergé sur Github permet d'exposer les textes libres de la communauté “Dramacode” (des personnes intéressées par l'étude informatisée du Théâtre classique). Les fichiers sources sont en XML/TEI, et maintenus comme des projets de code sur [Github](#). Cette page est générée automatiquement à destination des chercheurs qui ont besoin d'un point d'accès rapide pour télécharger un texte dans un format ou un autre, mais nous vous invitons surtout à retrouver l'édition sur le site du partenaire, en cliquant le lien dans la colonne Éditeur.

↓N°↑	↓Éditeur↑	↓Auteur↑	↓Date↑	↓Titre↑	↓Genre↑	↓Actes↑	↓Vers↑	↓Téléchargements↑
1	Théâtre Classique	(anonyme)	1722	L'actrice nouvelle	Comédie	1	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
2	Théâtre Classique	(anonyme)	1739	Thélamire	Tragédie	5	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
3	Théâtre Classique	(anonyme)	1772	Le prix d'un moment	Comédie	1	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
4	Théâtre Classique	(anonyme)	1772	Les bonnes gens	Comédie	1	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
5	Théâtre Classique	(anonyme)	1772	Oh ! voilÀ bien le diable ! comédie	Comédie	1	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
6	Théâtre Classique	(anonyme)	1784	Le club des dames ou le retour de descartes	Comédie	1	prose	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
7	CELLF, Bibliothèque dramatique	Abeille, Gaspard	1681	Lyncée, tragédie	Tragédie	5	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
8	Théâtre Classique	Aigueberre Jean Dumas d'	1729	Pan et doris	Pastorale	1	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
9	Théâtre Classique	Aigueberre Jean Dumas d'	1729	Polixène	Comédie	1	vers	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html
10	Théâtre Classique	Aigueberre Jean Dumas d'	1729	Trois spectacles	Prologue	1	prose	TEI, epub, kindle, markdown, iramuteq, paroles, TXM, html, fragment html

La TEI

```
<div1 type="act" n='1'><head>ACTE I</head>
<div2 type="scene" n='1'><head>SCÈNE PREMIÈRE. Oreste, Pylade.</head>
<sp who="ORESTE"><speaker>ORESTE.</speaker>
  <l id="1">Oui, Pylade, il est vrai, la valeur et l'adresse</l>
  <l id="2">Ont de l'Asie enfin fait triompher la Grèce. </l>
  <l id="3">De tous côtés en foule on vient dans cette Cour, </l>
  <l id="4">Croyant d'Agamemnon célébrer le retour,</l>
  <l id="5">Et toi-même suivant le zèle qui te guide, </l>
  <l id="6">Pour voler dans Mycènes as quitté la Phocide.</l>
  <l id="7">Cependant cher Ami, tes soins sont superflus. </l>
  <l id="8">Les Troyens sont vengés : mon Père ne vit plus.</l>
</sp>
<sp who="PYLADE"><speaker>PYLADE.</speaker>
  <l id="9">Il n'est plus ! Ô disgrâce à jamais déplorable !</l>
  <l id="10">Des caprices du fort exemple mémorable !</l>
  <l id="11">Monarque infortuné ! Mais de grâce, Seigneur, </l>
  <l id="12">Quand, comment a péri cet illustre Vainqueur ?</l>
</sp>
```

Traitement avec XSLT

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:xs="http://www.w3.org/2001/XMLSchema"
    exclude-result-prefixes="xs"
    version="2.0"
    xpath-default-namespace="http://www.tei-c.org/ns/1.0">

    <xsl:output method="text" omit-xml-declaration="yes" indent="no"/>

    <xsl:template match="TEI">
        <xsl:apply-templates/>
    </xsl:template>

    <xsl:template match="teiHeader"/>
```

Texte nettoyé

Oui, Pylade, il est vrai, la valeur et l'adresse
Ont de l'Asie enfin fait triompher la Grèce.
De tous côtés en foule on vient dans cette Cour,
Croyant d'Agamemnon célébrer le retour,
Et toi-même suivant le zèle qui te guide,
Pour voler dans Mycènes as quitté la Phocide.
Cependant cher Ami, tes soins sont superflus.
Les Troyens sont vengés : mon Père ne vit plus.

Il n'est plus ! Ô disgrâce à jamais déplorable !
Des caprices du fort exemple mémorable !
Monarque infortuné ! Mais de grâce, Seigneur,
Quand, comment a péri cet illustre Vainqueur ?

Texte nettoyé

Oui, Pylade, il est vrai, la valeur et l'adresse
Ont de l'Asie enfin fait triompher la Grèce.
De tous côtés en foule on vient dans cette Cour,
Croyant d'Agamemnon célébrer le retour,
Et toi-même suivant le zèle qui te guide,
Pour voler dans Mycènes as quitté la Phocide.
Cependant cher Ami, tes soins sont superflus.
Les Troyens sont vengés : mon Père ne vit plus.

Il n'est plus ! Ô disgrâce à jamais déplorable !
Des caprices du fort exemple mémorable !
Monarque infortuné ! Mais de grâce, Seigneur,
Quand, comment a péri cet illustre Vainqueur ?

Texte annoté

0 Oui
0 ,
0 Pylade
0 ,
0 il
0 est
0 vrai
0 ,
0 la
0 valeur
0 et
0 l'
0 adresse
0 Ont
0 de
0 l'
GEO-B Asie
0 enfin
0 fait
0 triompher
0 la
GEO-B Grèce
0 .

Calcul des scores par pièce

GEO

Asie	5
Grèce	15
Mycènes	34
Phocide	1
Ilion	11
mer Egée	1
Argos	8
Troie	19
Crète	2
Sparte	9
Athènes	2
Corinthe	1
Ithaque	1
Epire	1
Europe	1

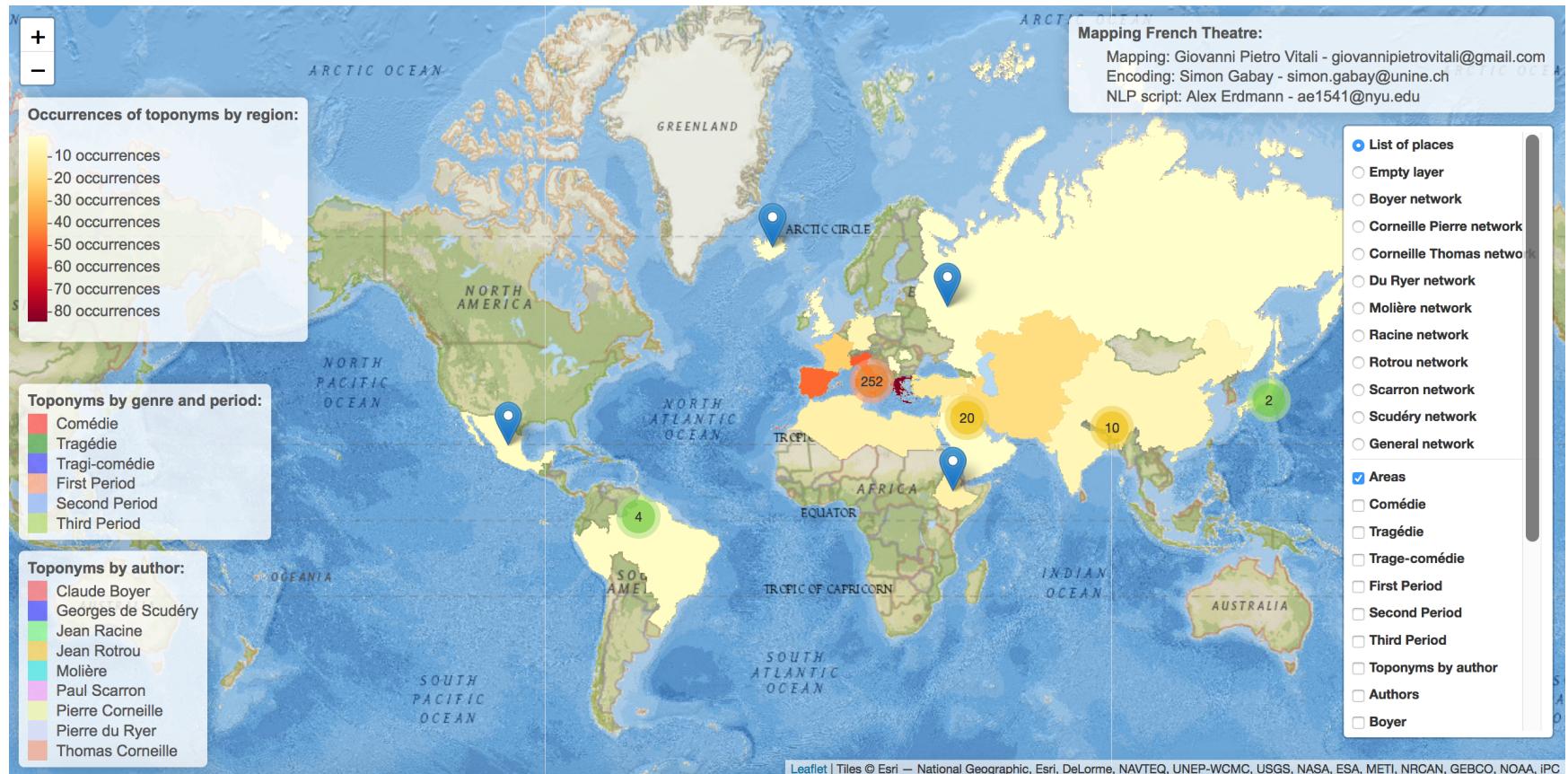
Géoréférencement

BOYER_AGAMEMNON	5 Asie	Asia	Continent		29.84064, 89.29688	http://www.geonames.org/6255147
BOYER_AGAMEMNON	15 Grèce	Greece	Pays	Grèce	Europe 39, 22	http://www.geonames.org/390903
BOYER_AGAMEMNON	34 Mycènes	Mycenae	Ville	Grèce	Europe 37.7300345711, 22.7540710459	https://pleiades.stoa.org/places/570491
BOYER_AGAMEMNON	1 Phocide	Phocis	Région	Grèce	Europe 38.5, 22.25	https://pleiades.stoa.org/places/541048
BOYER_AGAMEMNON	11 Ilion	Ilion	Ville	Grèce	Europe 39.740892, 20.552904	https://pleiades.stoa.org/places/530898
BOYER_AGAMEMNON	1 mer Egée	Aegean Sea	Mer	Grèce	Europe 37.4126570185, 25.8014773544	https://pleiades.stoa.org/places/560221
BOYER_AGAMEMNON	8 Argos	Argos	Ville	Grèce	Europe 37.631561, 22.719464	https://pleiades.stoa.org/places/570106
BOYER_AGAMEMNON	19 Troie	Troy	Ville	Anatolie	Asie 39.9574327844, 26.2384586227	https://pleiades.stoa.org/places/550595
BOYER_AGAMEMNON	2 Crète	Creta	Pays	Grèce	Europe 35.0627201667, 24.9469437222	https://pleiades.stoa.org/places/981517
BOYER_AGAMEMNON	9 Sparte	Sparta	Ville	Grèce	Europe 37.077905, 22.4272985	https://pleiades.stoa.org/places/570685
BOYER_AGAMEMNON	2 Athènes	Athens	Ville	Grèce	Europe 37.9728191495, 23.7237438519	https://pleiades.stoa.org/places/579885
BOYER_AGAMEMNON	1 Corinthe	Corinth	Ville	Grèce	Europe 37.8980016198, 22.8999795422	https://pleiades.stoa.org/places/570182
BOYER_AGAMEMNON	1 Ithaque	Ithaca	Île	Grèce	Europe 38.44193, 20.638399	https://pleiades.stoa.org/places/530905
BOYER_AGAMEMNON	1 Épire	Epirus	Région	Grèce	Europe 39.5771060055, 20.589832047	https://pleiades.stoa.org/places/530871
BOYER_AGAMEMNON	1 Europe	Europa	Continent		Europe 48.69096, 9.14062	http://www.geonames.org/6255148

Traitements

```
## 3. Corneille.Thomas
addCircleMarkers(lng == corneille_T$lng, ~
  lat == corneille_T$lat, ~
  color == "Green", ~
  weight == 1, ~
  radius == 14, ~
  fillOpacity == 0.6, ~
  group == "Corneille.Thomas", ~
  label == ~paste("The toponym", ~
    corneille_T$occurrence, ~
    "is mentioned in the work", ~
    corneille_T$titre, ~
    "by", ~
    corneille_T$auteur)) %>%
```

Publication



Analyse: Racine vs Scarron

