

Introduction à la philologie numérique

I-Introduction : Une brève histoire du livre

- Avant le 1^{er} siècle : **volumen**
- 2^e siècle apparition du **Codex**
- 3^e-5^e siècle les volumena disparaissent au profit des **codex**
- Au Moyen Age, la **production de livres** se fait essentiellement dans les scriptoria des monastères.
- 1453, première **Bible 42** imprimée par Gutenberg.
- 19^e siècle industrialisation de la production
- 21^e siècle : l'édition électronique

II— Qu'est-ce qu'une édition électronique ?

Définition :

« L'édition électronique est une forme d'édition dans laquelle le numérique joue un rôle plus ou moins important, allant de la simple mise en page d'un livre en utilisant un logiciel de PAO (Publication assistée par ordinateur) en vue de l'impression papier d'un livre jusqu'à la création et la diffusion en ligne d'œuvres. »

Wikipedia

II— Qu'est-ce qu'une édition électronique ?

Quelques projets...

1-Les chroniques du Mont-Saint-Michel.

Projet qui propose un **choix de parcours** par manuscrits.

2-Madame Bovary

Édition « génétique » qui permet de reconstituer l'histoire du texte à travers **ses brouillons**.

3-La Queste Del saint Graal

Édition multifacette d'un manuscrit médiéval avec des fonctionnalités de recherche.

4-Le projet Hyperdonat

Qui propose un appareil critique très complexe grâce aux possibilités de visualisation qu'offre une édition numérique.

II— Qu'est-ce qu'une édition électronique ?

Les différents types d'« activités » de l'édition électronique

« L'édition électronique relève de trois types d'activité bien spécifiques, correspondant à des positions différentes sur la chaîne de production du texte numérique : **certaines activités relèvent de la numérisation, c'est-à-dire le portage du texte imprimé sur support numérique**, avec différents degrés de sophistication. **D'autres relèvent de ce que nous pourrions appeler l'édition numérique** ; il s'agit de pratiques d'édition qui ne passent pas immédiatement par un support analogique — imprimé — mais **exploitent nativement les possibilités qu'offre le texte numérique**. **D'autres initiatives, enfin, que nous estimons relever de l'édition en réseau, s'appuient pour leur part sur les réseaux électroniques** pour développer de nouveaux modes de production des textes. »

Dacos, Marin et Mounier, Pierre, *l'édition électronique*, Paris, la Découverte, 2010.

II— Qu'est-ce qu'une édition électronique ?

« Le texte électronique idéal sera lisible, manipulable et citable »

Marin Dacos

Une édition électronique doit être :

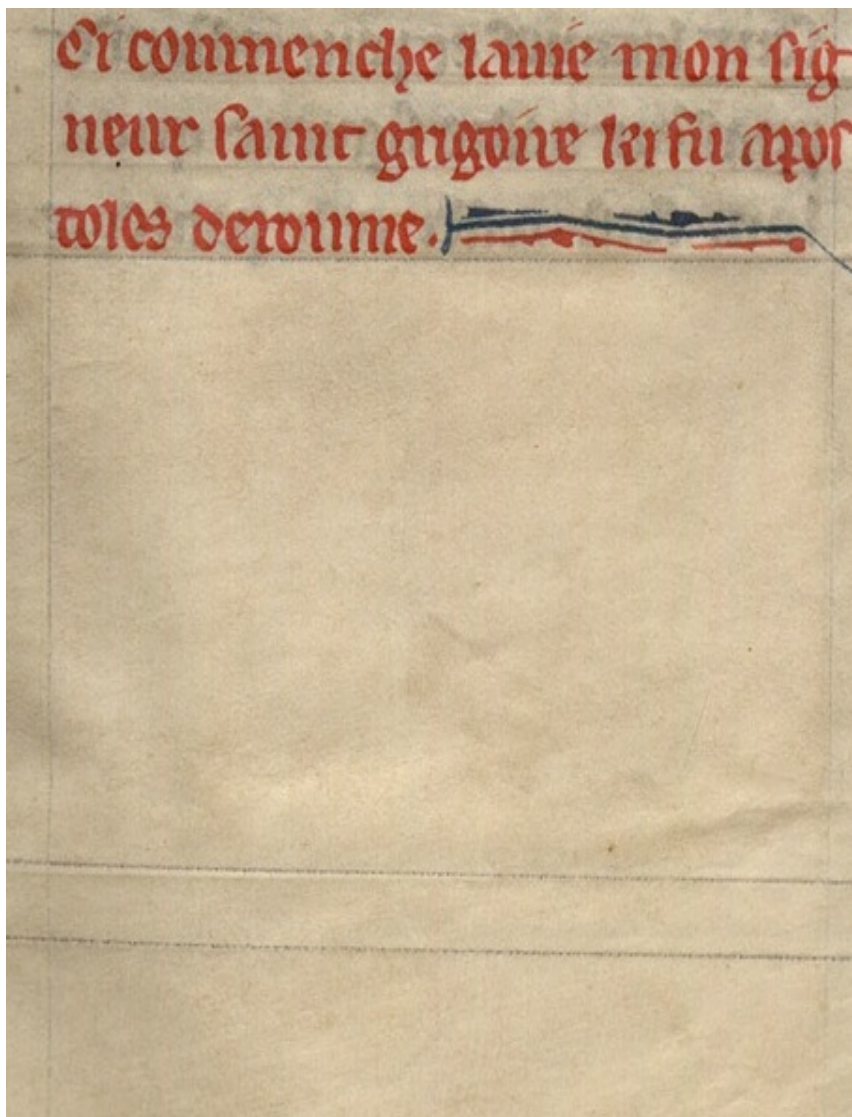
Lisible

- format ouvert et accessible sur tous les types d'appareils
- Conservable : le texte doit pouvoir être récupéré en cas de casse de l'appareil ou d'évolution de ce dernier.

Manipulable

- réutilisable
- interopérable

III— Le langage XML et la grammaire TEI



```

<hi rend="red rubric">Ci conmenche la
<choice>
  <orig>u</orig>
  <reg>v</reg>
</choice>
ie mon si<lb/>gneur
<choice>
  <orig>
    <name>
      <persName>saint grigoire </persName>
    </name>
  </orig>
  <reg>
    <name>
      <persName>saint Grigoire</persName>
    </name>
  </reg>
</choice>

```

IV— Enrichissement de l'information et multiplication des parcours de lecture

L'édition de la Vie de saint Martin :

À partir d'un document [XML](#), on peut :

- — multiplier les vues, en passant d'une [version dynamique](#) pour écran à la production papier d'une [édition traditionnelle](#) ;
- générer des [études statistiques](#) précises des phénomènes présents dans le texte et des interventions éditoriales.

Multiplication des parcours de lecture, impact sur la conception du texte dans la Recherche

BUREAU, Bruno, « Quelques réflexions sur la notion de littérarité à partir de l'édition numérique de commentateurs anciens », *Interférences. Ars scribendi*, janvier 2012.

« L'outil numérique questionne non pas tant le livre, ou même le texte, que nos modes habituels de perception de la littérarité. L'outil numérique, parce qu'il questionne notre lecture et le mode même de la lecture, ne ferme pas la porte à la littérarité, il ouvre en réalité des fenêtres sur des formes jusqu'ici restées peut-être plus dans l'ombre de l'analyse littéraire. »

DUVAL, Frédéric, « Pour des éditions numériques critiques. L'exemple des textes français », in *Le texte à l'épreuve du numérique*, vol. 73, Médiévales, Saint-Denis, France, Presses universitaires de Vincennes, Université Paris VIII, 2017, (« Médiévales »), p. 13–30.

« La force du numérique est sa modularité : une édition numérique peut proposer des parcours plus ou moins ouverts, de la consultation libre d'un dépôt d'archives à des parcours commandés par des approches et intérêts divers. »

ROBINSON, Peter, « *Toward a Theory of Digital Edition* », p.123 :

« On ne peut pas connaître l'œuvre sans les documents -- de même, on ne peut comprendre les documents sans une compréhension de l'œuvre qu'ils génèrent. On peut en tirer le principe suivant : une édition scientifique doit éclairer autant que possible chaque aspect du texte, aussi bien le texte-comme-œuvre que le texte-comme-document. Les éditions imprimées traditionnelles se sont davantage intéressées au premier. Un avantage évident des éditions numériques est qu'elles peuvent rééquilibrer le balancier en incluant un matériau bien plus riche pour l'étude du texte-comme-document qu'on ne peut y parvenir par le moyen de l'impression. »

V – Travailler à partir de données numériques

Analyse de réseau

Franco Moretti, *Distant reading*, London, Verso, 2013

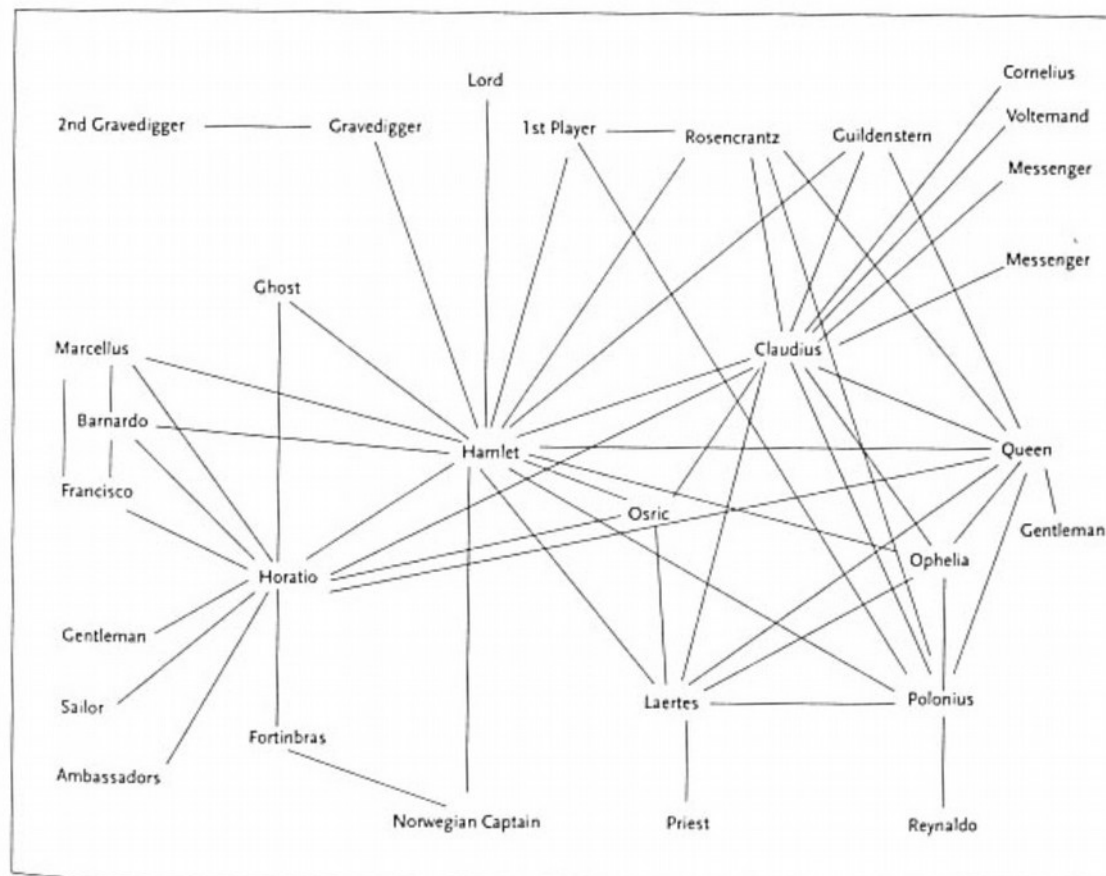
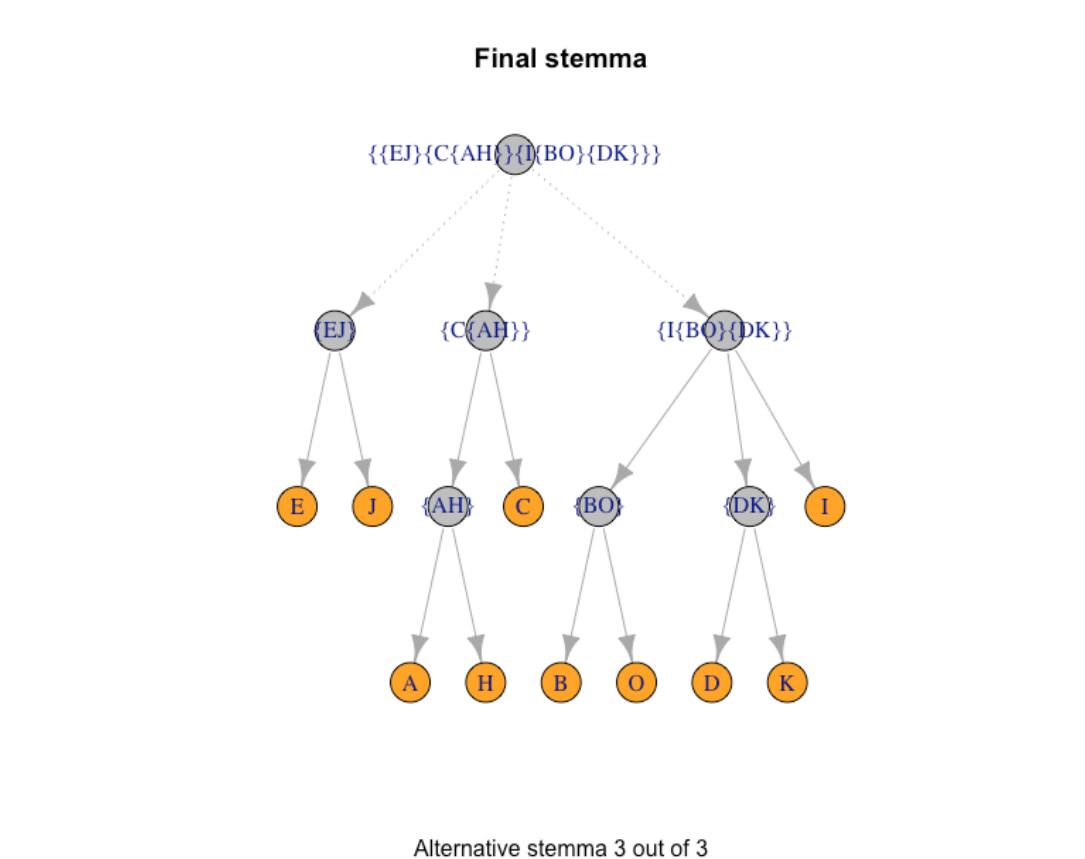


Figure 1: The Hamlet network

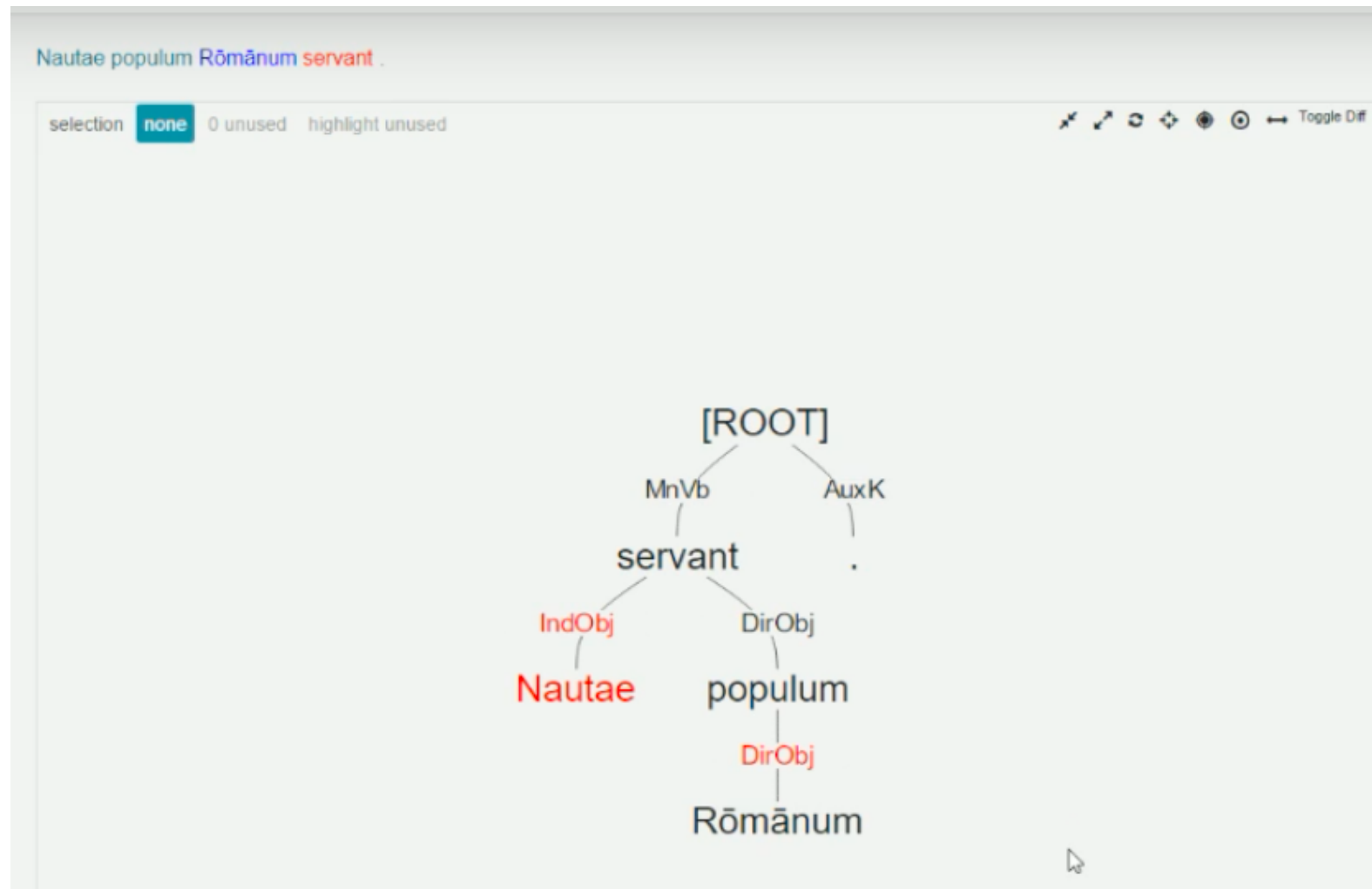
V – Travailler à partir de données numériques

Génération automatique d'un stemma à partir d'un encodage XML TEI à l'aide du package stemmatology pour R



V – Travailler à partir de données numériques

*Outils numériques d'analyse de la langue : le [treebank](http://alpheios.net/pages/tools/#alpheios-treebank-editor) –
<http://alpheios.net/pages/tools/#alpheios-treebank-editor>*



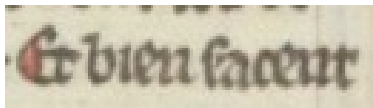
V – Travailler à partir de données numériques

Outils numériques d'analyse de la langue : les enjeux

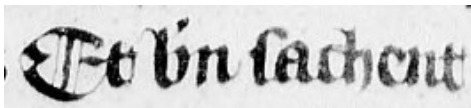
Assister l'étude linguistique : repérer les traits de langue picarde

Confusion entre le son [s] et le son [k]

- sacent (ms 412 e la BnF)



- sachent (ms 411 de la nF)



V – Travailler à partir de données numériques

Assister l'étude morphosyntaxique : repérer les traits de langue picarde

Confusion morphologique entre les pronoms personnels objets masculins et féminins

Vie saint Nicolas, §10

« **Cele feme** prist ses maris, si **le** mena au leu ou messires seinz Nicholas arrestoit adonques et qant il i vint, si li cheï as piez et si li dist : “Home Deu, soceur a **ceste chetive feme**, car li deables **le** tormente molt griefment et travaille.” Lors se coucha messires seinz Nicholas en orissons envers nostre Signor por celi et tantost com il l’ot finie, li issi li deables del cors, si fu sanee. Cil qi ce virent en gracierent nostre Signor et la dame repeira a sa meson tout loant Deu »

V – Travailler à partir de données numériques

Pandora, une proposition de solution pour la lemmatisation et d'étiquetage morphosyntaxique automatisée

_ un algorithme de lemmatisation automatique basé sur des réseaux neuronaux qui permet :

- pas de lexique ni de règles prédéfinies ;
- apprentissage en contexte sur des corpus annotés ;
- représentation sémantique via des vecteurs de mots (word2vec) ;

V – Travailler à partir de données numériques

Pyrrha, une interface de correction de l'étiquetage du lemmatiseur
Pandora, <http://dev.chartes.psl.eu/ppa>

Pandora Post-Correction Editor New Corpus SaintMartin ▾							
Corpus SaintMartin - List of tokens							
1	2	3	4	5	...	188	189
Form	Lemma	POS	Morph	Context	Similar	Save	
De	de	PRE	DEGRE=-	De seint Martin mout doit on doucement et volentiers	4	Save	
seint	saint1	ADJqua	NOMB.=s GENRE=m CAS=r DEGRE=p	De seint Martin mout doit on doucement et volentiers le	109	Save	
Martin	Martin	NOMpro	NOMB.=s GENRE=m CAS=r	De seint Martin mout doit on doucement et volentiers le bien	55	Save	
mout	mout	ADVgen	DEGRE=-	De seint Martin mout doit on doucement et volentiers le bien oïr	47	Save	
doit	devoir	VERcjg	MODE=ind TEMPS=pst PERS.=3 NOMB.=s	De seint Martin mout doit on doucement et volentiers le bien oïr et	16	Save	
on	on	PROimp	PERS.=3 NOMB.=s GENRE=n CAS=n	De seint Martin mout doit on doucement et volentiers le bien oïr et entendre	6	Save	
doucement	douz	ADVgen	DEGRE=-	De seint Martin mout doit on doucement et volentiers le bien oïr et entendre ,	1	Save	
et	et	CONcoo	DEGRE=-	De seint Martin mout doit on doucement et volentiers le bien oïr et entendre , car	760	Save	
volentiers	volentiers	ADVgen	None	De seint Martin mout doit on doucement et volentiers le bien oïr et entendre , car par	8	Save	
le	le	DETdef	None	seint Martin mout doit on doucement et volentiers le bien oïr et entendre , car par le	185	Save	
bien	bien	ADVgen	None	Martin mout doit on doucement et volentiers le bien oïr et entendre , car par le bien	68	Save	
oïr	oïr	VERinf	None	mout doit on doucement et volentiers le bien oïr et entendre , car par le bien savoir	9	Save	
et	et	CONcoo	DEGRE=-	doit on doucement et volentiers le bien oïr et entendre , car par le bien savoir et	760	Save	
entendre	entendre	VERinf	None	on doucement et volentiers le bien oïr et entendre , car par le bien savoir et retenir	8	Save	
,	,	PONfbl	None	doucement et volentiers le bien oïr et entendre , car par le bien savoir et retenir puet	622	Save	
car	car	CONcoo	None	et volentiers le bien oïr et entendre , car par le bien savoir et retenir puet l	61	Save	

Bibliographie sélective

BARBIER, Frédéric, Histoire du livre en Occident, Paris, France, A. Colin, 2012.

BURNARD, Lou, What is the Text Encoding Initiative? How to add intelligent markup to digital resources, OpenEdition Press, 2014.

CLÉRICE, Thibault et CAMPS, Jean-Baptiste, « Pandora A (language independent) Tagger Lemmatizer for Latin and the Vernacular », [En ligne : http://www.academia.edu/35076560/Pandora_A_language_independent_Tagger_Lemmatizer_for_Latin_and_the_Vernacular].

DACOS Marin et MOUNIER Pierre Mounier, L'édition électronique, Paris, la Découverte, 2010.

DARNTON, Robert, Apologie du livre : demain, aujourd'hui, hier, trad. Jean-François Sené, Paris, Gallimard, 2010.

MERLOT, Michel, « La Bibliothèque multimédia contemporaine », in Lieux de savoir, dir. Christan Jacob, Paris, A. Milchel, 2007, p.637-652.

MORETTI, Franco, Distant reading, London, Verso, 2013.

PIERAZZO, Elena, Digital scholarly editing: theories, models and methods, Farnham Burlington, Ashgate, 2015.

ROBINSON, Peter, « Where We Are with Electronic Scholarly Editions, and Where We Want to Be », Jahrbuch für Computerphilologie, vol. 5 / 5, 2003, p. 126– 146.

SEGRE, Cesare, « Lachmann et Bédier. La guerre est finie », Actes du XXVII^e Congrès international de linguistique et philologie romane, eds. Éva Buchi, Jean-Paul Chauveau et Jean-Marie Pierrel, Strasbourg, 2016, p. 15–28.