

TeiCollator : une chaîne de traitement ecdotique semi-automatisée

XXX^e Congrès International de Linguistique et Philologie Romane

Matthias GILLE LEVENSON

École Normale Supérieure de Lyon - Casa de Velázquez

prénom [point] gille [tiret] levenson [at] casadevelazquez [point] org



- 1 Introduction
- 2 Méthode
- 3 Transfert d'information et collation matérielle
- 4 Résultats
- 5 Conclusions

État de la recherche : la collation

- SPADINI 2016, TORSTEN ROEDER 2020 et NURY 2018 pour un état de la recherche et un historique complet.
- HAENTJENS DEKKER 2014 pour CollateX, l'outil d'alignement/collation probablement le plus utilisé
- BLEEKER, BUITENDIJK, HAENTJENS DEKKER et KULSDOM 2018 et BLEEKER, BUITENDIJK et HAENTJENS DEKKER 2020 pour la réflexion sur la structure et son intégration à la collation
- CAMPS, ING et SPADINI 2019 et CAMPS, ING et SPADINI 2021 pour le traitement et la catégorisation de la variation textuelle et l'intégration de cette catégorisation dans une chaîne de traitement.

État de la recherche : le modèle de Gothenburg

Modèle théorique de la collation en plusieurs étapes ¹ :

- 1 Tokénisation
- 2 Normalisation
- 3 Alignement
- 4 Analyse/Feedback
- 5 Visualisation

1. Tiré de la documentation de CollateX : <https://collatex.net/doc/>, cité dans SPADINI 2016.

État de la recherche

- Encore relativement peu de travail sur la structure (déjà cité : BLEEKER, BUITENDIJK, HAENTJENS DEKKER et KULSDOM 2018 ; BLEEKER, BUITENDIJK et HAENTJENS DEKKER 2020, mais s'éloigne de la TEI)
- Nécessité de prendre en compte l'arrivée de la transcription automatisée, OCR et HTR [KIESSLING, TISSOT, STOKES et EZRA 2019 ; KAHLE, COLUTTO, HACKL et MÜHLBERGER 2017 ; VIDAL-GORÈNE et al. 2021]

Problèmes

- Comment intégrer efficacement les résultats de la transcription automatisée au travail d'établissement du texte ?
- Comment aller au delà de la pure collation textuelle, et intégrer des document structurés au processus de collation ?
- Plus globalement, à quel point peut-on déléguer la tâche de la collation à la machine ?

- 1 Introduction
- 2 Méthode**
- 3 Transfert d'information et collation matérielle
- 4 Résultats
- 5 Conclusions

Méthode résumée ¹

- Transcription et encodage de chaque témoin en XML-TEI
- **Tokénisation et lemmatisation en place du XML-TEI [1] [2]**
- **Alignement avec CollateX [3]**
- **Production d'apparats groupés et typés [4]**
- Réinjection des apparats dans chaque transcription individuelle
- **Détection de divers phénomènes textuels : lacunes, sauts du même-au-même, transpositions [4]**
- Transferts d'information entre fichiers TEI
- **Transformation en édition consultable [5]**

1. Entre crochets, les étapes du modèle de Gothenburg

Traduire en algorithme la méthode ecdotique

On convertit ou on simule le travail ecdotique. Exemples :

- Deux mots sont identiques s'ils partagent les mêmes annotations [CAMPS, ING et SPADINI 2019]
- La *collatio* est la combinaison de deux phases distinctes [SPADINI 2016] :
 - Alignement [HAENTJENS DEKKER 2014]
 - Comparaison des chaînes et des annotations pour créer et typer les variantes [CAMPS, ING et SPADINI 2019]
- Une transposition correspond à deux suites de mots identiques dans un ordre distinct.
- Une lacune est une suite d'apparats contenant le ou les mêmes témoins sans texte.
- Un saut du même-au-même (ou omission par homéotéleute) est une lacune dont le premier mot est graphiquement proche ou identique au premier mot qui la suit.

Détection des lacunes et homéotéleutes

#Esc Q										
aderecho a+derecho VMF90SM	omisit	catyvo cautivo AG0MS0	a a SP500	omisit	omisit	omisit	omisit	omisit	Cajena cajena NP000P0	embiaronlo enviar+lo VMIS3P0+FF3MSA0
#Mad A										
aducho aducir VMF90SM	omisit	catiuo cautivo AG0MS0	a a SP500	cartalena cartalena NP000G0	omisit	los el DA0MP0	de de SP500	cartalena cartalena NP000G0	embiaronlo enviar+lo VMIS3P0+FF3CN400	
#Mad C										
aducho aducir VMF90SM	omisit	catvivo cautivo AG0MS0	a a SP500	Cartajena cartajena NP000G0	omisit	los el DA0MP0	de de SP500	Cartajena cartajena NP000G0	embiaronlo enviar+lo VMIS3P0+FF3CN400	
#Phil U										
aducho aducir VMF90SM	omisit	cativo cautivo AG0MS0	a a SP500	Cartajena cartajena NP000G0	omisit	los el DA0MP0	de de SP500	Cartajena cartajena NP000G0	embiaron enviar VMIS3P0	
#Sev R										
aducho aducir VMF90SM	omisit	captivo cautivo AG0MS0	a a SP500	omisit	omisit	omisit	omisit	Cartajena cartajena NP000G0	embiaronlo enviar+lo VMIS3P0+FF3CN400	
#Mad B										
aderecho a+derecho VMF90SM	e e cc	captivo cautivo AG0MS0	a a SP500	Cartajena cartajena NP000G0	e e cc	los el DA0MP0	de de SP500	Cartajena cartajena NP000G0	embiaronlo enviar+le VMIS3P0+FF3CS000	
#Sal J										
leuado llevar VMF90SM	omisit	omisit	a a SP500	Cartajena cartajena NP000G0	omisit	los el DA0MP0	de de SP500	Cartajena cartajena NP000G0	embiaronlo enviar+lo VMIS3P0+FF3CN400	
#Sev Z										
aducho aducir VMF90SM	omisit	catiuo cautivo AG0MS0	a a SP500	cartajena cartajena NP000G0	omisit	los el DA0MP0	de de SP500	cartajena cartajena NP000G0	embiaronlo enviar+lo VMIS3P0+FF3MSA0	
Variante	Lemme différent	Variante morphosyntactique (pos différent)	93.68%							

Tableau d'alignement basique (« *Regimiento de los príncipes* », chapitre III, 3, 2)

```

--rf-
<witStart xml:id="vrtPU4X" corresp="#Sal_J" previous="#pupdk60"/>
--capp ana="#not_apparat">
--<rdgGrp>
--<rdg wit="#Esc_Q #Mad_A #Mad_G #Phil_U #Sev_R #Mad_B #Sal_J #Sev_Z">
--<w ana="#same_word">a</w>
--</rdg>
--</rdgGrp>
--</app>
<witEnd xml:id="YMXQGI_j0Ueu6X" corresp="#Esc_Q #Sev_R" next="#dLPLJ1C" ana="#homeoteleuton" cert="medium"/>
</...
<idEnd xmlns="http://www.tei-c.org/ns/1.0" xmlns:tei="http://www.tei-c.org/ns/1.0" xml:id="j0Ueu6X" corresp="#Esc_Q" next="#J56Cdbq" ana="#homeoteleuton" cert="medium"/>
--<app ana="#entite_nommee">
--<rdgGrp>
--<rdg wit="#Mad_A">
--<w pos="NP000G0" ana="#annotation_manuelle"></w>
--</rdg>
--</rdgGrp>
--<rdgGrp>
--<rdg wit="#Mad_G #Mad_B">
--<w ana="#same_word">Cartagena</w>
--</rdg>
--</rdgGrp>
--<rdgGrp>
--<rdg wit="#Phil_U #Sal_J"></rdg>
--<rdg wit="#Sev_Z">
--<w>cartajena</w>
--</rdg>
--</rdgGrp>
--</app>
--<app ana="#omission">
--<rdgGrp>
--<rdg wit="#Mad_A #Mad_G #Phil_U #Sal_J #Sev_Z">
--</rdgGrp>
--<rdgGrp>
--<rdgGrp></rdgGrp>
--</app>
--<app ana="#not_apparat">
--<rdgGrp>
--<rdg wit="#Mad_A #Mad_G #Phil_U #Mad_B #Sal_J #Sev_Z">
--<w>jos</w>
--</rdg>
--</rdgGrp>
--</app>
--<app ana="#not_apparat">
--<rdgGrp>
--<rdg wit="#Mad_A #Mad_G #Phil_U #Mad_B #Sal_J #Sev_Z">
--<w>da</w>
--</rdg>
--</rdgGrp>
--</app>
<witStart xml:id="J56Cdbq_dLPLJ1C" corresp="#Esc_Q #Sev_R" previous="#j0Ueu6X"/>
</...
<idStart xmlns="http://www.tei-c.org/ns/1.0" xmlns:tei="http://www.tei-c.org/ns/1.0" xml:id="dLPLJ1C" corresp="#Sev_R" previous="#YMXQGI"/>
--<app ana="#entite_nommee">
--<rdgGrp>
--<rdg wit="#Mad_A"></rdg>
--</rdgGrp>
--<rdgGrp>
--<rdg wit="#Esc_Q">
--<w pos="NP000P0" synch="ORjEFMTjH" ana="#annotation_manuelle">
--<unclear>Sjani</unclear>
--</w>
--</rdg>
--</rdgGrp>

```

- 1 Introduction
- 2 Méthode
- 3 Transfert d'information et collation matérielle**
- 4 Résultats
- 5 Conclusions

Transfert d'information

- L'idée est de permettre de rassembler toutes les informations utiles dans le témoin final collationné.

```
<w lemma="1" pos="A00MS0">primero</w> de las Éthicas pone los oficios de las
manos que son tañer e traer e arredrar e <w lemma="empujar">empuxar</w><note
xml:id="NQwcqjKUTL" type="sources">Voir le <title>Communiloquium</title>, I,
9, 1: <quote type="primaire" xml:lang="lat">Manus enim rei publice sunt
homines militaris: qui conuenient per manus significantur. Manus enim sunt
parate ad iuuandum ex nature inperio. E quantum utilitatis habent nulli
scire licet, ait Senecat epistula xcvi: opus enim manus est tangere,
attrahere e inpellere, ait comentator i. ethi.<ref type="biblio"
target="#juandegales_SummaCollectionumSeu_1401" n="I, 9, 1, fol. 50v"
/></quote>Le <title>Communiloquium</title> ajoute une citation d'Al-Ghazâlî
qui n'est pas reprise par le glosateur.</note>, e son ellas de tal forma por
```

Témoin B, Chapitre 6.

Transfert d'information

```

<app ana="#graphique">
  <rdgGrp>
    <rdg wit="#Esc_Q #Mad_G #Sev_R #Mad_B #Sal_J #Sev_Z">
      <w>arredrar</w>
    </rdg>
    <rdg wit="#Mad_A">
      <w>aredrar</w>
    </rdg>
  </rdgGrp>
</app>
<app ana="#lexicale">
  <rdgGrp>
    <rdg wit="#Esc_Q #Mad_B">
      <w>e</w>
    </rdg>
  </rdgGrp>
  <rdgGrp>
    <rdg wit="#Mad_A #Mad_G #Sev_R #Sal_J #Sev_Z">
      <w>o</w>
    </rdg>
  </rdgGrp>
</app>
<app ana="#not_apparat">
  <rdgGrp>
    <rdg wit="#Esc_Q #Mad_A #Mad_G #Sev_R #Mad_B #Sal_J #Sev_Z">
      <w>enpuxar</w>
    </rdg>
  </rdgGrp>
</app>
<note xml:id="NQwcqJKUtl" type="sources" ana="#injected" corresp="#Mad_B">Voir le
<title>Communiloquium</title>, I, 9, 1: <quote type="primaire" xml:lang="lat">Manus enim rei
publice sunt homines militaris: qui conuenient per manus significantur. Manus enim sunt parate
ad iuuandum ex nature imperio. E quantum utilitatis habent nulli scire licet, ait Senecat
epistula xcvi: opus enim manus est tangere, attrahere e inbellere, ait comentator i. ethi.<ref
type="biblio" target="#juandegales_SummaCollectionumSeu_1401" n="1, 9, 1, fol. 50v"
xml:id="d3h8bHK"/></quote>Le <title>Communiloquium</title> ajoute une citation
d'Al-Ghaz6#226;16#238; qui n'est pas reprise par le glosateur.</note>
<pc xml:id="hMwbRsa">;</pc>

```

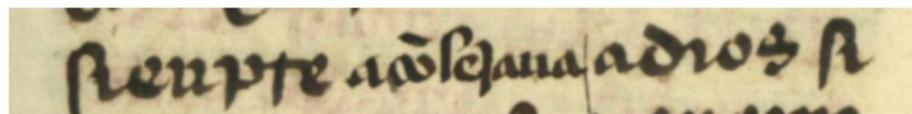
Résultat de la collation (témoin base : G)

Collation matérielle

- Ce transfert d'informations peut permettre la collation matérielle du texte (marques de lecture, gloses, corrections, changements de folios, de colonne, etc). Exemple :



Ms. 15304, Fundación Lázaro Galdiano, Madrid, fol. 245v



Ms. II/215, Real Biblioteca, Madrid, fol. 418r

Collation matérielle



Ms 15304, Fundación Lázaro Galdiano, Madrid, fol. 245v

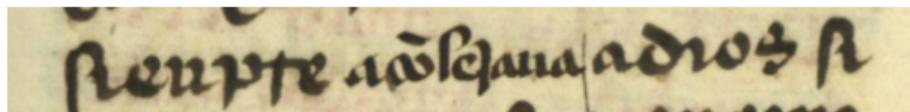
```

/>tremor a enxemplo de dauid que<lb break="yes" facs="#facs_y1X37h" xml:id="elem_y1X37h"
copy0f="file:/home/mgl/Bureau/These/Edition/hyperregimiento-de-los-principes/Dedans/XML/analyse_linguistique/Mad_A.xml#facs_y1X37h"
/>sienpre<space xml:id="0arJbBUnCZ" type="in_witness"/><note xml:id="NZrE0mzJZs"
type="variante" anchored="true">Ici, un espace apparaît et le mot est clairement
manquant, comme ce qu'il semble s'être passé dans le témoin G! Les deux témoins sont
donc liés d'une façon ou d'une autre.</note> a dios <w lemma="si" pos="CS" part="N"
>si</w><w lemma="acometer" part="N">acome<lb break="no" facs="#facs_bGNyEp"
xml:id="elem_bGNyEp"
copy0f="file:/home/mgl/Bureau/These/Edition/hyperregimiento-de-los-principes/Dedans/XML/analyse_linguistique/Mad_A.xml#facs_bGNyEp"
/>teria</w> a sus enemigos segunt que<lb break="yes" facs="#facs_pnWcc5"

```

Encodage en TEI

Collation matérielle



Ms. II/215, Real Biblioteca, Madrid, fol. 418r

```

>fazienda</w> con temor e con tremor, a enxemplo de <w lemma="david"
pos="NP000P0">David</w>, que sienpre <w><subst>
  <space/>
  <add hand="identique" place="inline" type="correction">aconsejava</add>
</subst></w><note type="particulier" xml:id="HgAJWEnofé"
fac="/home/mgl/Bureau/These/Edition/hyperregimiento-de-los-principes/Dedans/Facsimiles/Mad_G/aconsejava_add.png"
>Ici, le terme <quote>aconsejava</quote> semble avoir été ajouté après, par
une main similaire: il est écrit dans la ligne, mais avec une taille de
caractères plus petite, comblant un emplacement qui semble avoir été laissé
vide pour une raison inconnue.</note> a Dios <w lemma="si" pos="CS">si</w>
<w lemma="acometer" pos="VHIC350">acometería</w> a sus enemigos, segund que de

```

Encodage en TEI

Mise en regard d'éléments matériels

```

-<app ana="#graphique">
-<rdgGrp>
-<rdg wit="#Esc_Q #Mad_A #Mad_G #Sev_R #Sal_J">
  <w>siempre</w>
  <space xml:id="OarJbBUnCZ" type="in_witness" ana="#injected" corresp="#Mad_A"/>
  +<note xml:id="NZrEOmzJZs" type="variante" anchored="true" ana="#injected" corresp="#Mad_A"></note>
</rdg>
+<rdg wit="#Mad_B"></rdg>
+<rdg wit="#Sev_Z"></rdg>
</rdgGrp>
</app>
-<app ana="#omission #lexicale">
-<rdgGrp>
-<rdg wit="#Mad_G">
  -<w ana="#annotation_manuelle">
    -<subst>
      <space xml:id="gW4OVGI"/>
      <add hand="identique" place="inline" type="correction">aconsejava</add>
    </subst>
  </w>
  +<note type="codico" xml:id="HgAJWEnofe" facs="/home/mgl/Bureau/These/Edition/hyperregimiento-de-los-principes/Dedans/Facsimiles/Mad_G/aconsejava_add.png" ana="#injected" corresp="#Mad_G">
</rdg>
</rdgGrp>
-<rdgGrp>
-<rdg wit="#Esc_Q #Sev_R #Mad_B">
  <w>demandava</w>
</rdg>
+<rdg wit="#Sal_J #Sev_Z"></rdg>
</rdgGrp>
-<rdgGrp>
  <rdg wit="#Mad_A"/>
</rdgGrp>
</app>

```

Résultat du processus : une collation matérielle ?

- 1 Introduction
- 2 Méthode
- 3 Transfert d'information et collation matérielle
- 4 Résultats**
- 5 Conclusions

nin de verdaderos omnes, ca los Romanos en su fecho syenpre mostraron el contrario, según que cuenta Tullio en el primero libro de los Ofiçios, do dize que quando el rrey Marcos fue aderecho¹ e captivo² a ^{QR} Cartagena e los de³ ²³ Cartagena embiaronle a Rroma para que fuese pleytes en rredemir los captivos que ellos tenían de Rroma, e tomásenle jura que tornase a ellos. Et el dicho Marco vino a ^{QR} Rroma, e entró en el concejo de⁴ ²⁸ Rroma, e dio por consejo a los Romanos que non rredimiesen los captivos nin embiasen a rrehenes por ellos por rrazón que los que están ^{QR} captivos eran viejos, como quier que ellos fuesen buenos cavalleros, e los que⁵ avian de embiar en arrehenes⁶ eran mançebos⁷ e buenos⁸ para lidiar : et todos los Romanos tomaron este consejo, e quando quisieron rretener sus parientes e sus amigos, non lo quiso el consentir, e más quiso tornar a muy cruels enemigos e a grandes penas que quebrantar la fe e el omenaje que avia puesto : en que dio enxemplo a todos que avian de guardar fe e verdat a sus enemigos⁹ .

²³Saut du même au même ici pour QR.

²⁸Saut du même au même ici pour GUR.

desiliare dicuntur. Spiritum effugium et in concursu ad primis impetu aduersari, inclinatū opem ferre laborant filiores. Otiosum officium [en] partu uictorio conseruare eis qui non cradela, non timorant fuerunt. Nisiue aere salua et premis laudibus exuere. Non enim his adspiciendū est qui dicunt: « Dulu an atriis, qui in hanc requirit? » [GUILLEAUME DE CONCILES 1929, p. 37-38]

³ (Q)

Indiat hoc Regulus qui captus a pennis, care de captiuitate conuulsit Romanos etiam est, lauatū et rediitiam, pennis et uenit captiuis reddendis non conuulsi, deside cum reuocaretur^{QR} ab amicis, ad appetitū redire maluit quam filios hanc dicit filiores. [ibid., p. 38]

La source de Guillaume de Conches est ici effectivement Caesiro. Voir Clément Giraud, *CÉSIRE* 1913, I, XIII, p. 42. avec le personnage de Regulus. On pourrait ici élargir l'identification de la famille de manuscrits qui contiennent cette variation. Par ailleurs, on ne pourra manquer de s'étonner sur l'argument de l'âge qui s'est pas dans la source latine et qui semble contredire l'exemple précédent sur Alexandre et l'importance de l'expérience dans la chevalerie.

126 non | Ce mot est ajouté au dessus de la ligne dans le témoin B. 127 que los que estincaitos | Ce passage est ajouté au dessus de la ligne dans le témoin B.

121 de Q|ZZ|ov.AGUR 121 su AGURB|Z|ov.Q 121 fecho AGURB|Z|fechos Q 122 libro Q|GRB|Z|ov.U 122 do QAGURB|Do &Z 122 Marcos QGRB|Martos U marcos AZ 122 aderecho Q|leuado./ adicho AGURZ 123 e B|ov.QAGURZ 123 Cartagena¹ GB|U|Z|cartagena.A 123 e B|ov.AG|U|Z 123 Cartagena² GBUR|Z|cartagena.A Cajena (?) Q 123 embiaronle B|embiaron U embiaronlo QGR|Z.A 123 fueze QAGB|fueze.Z fueze UR 124 tomásenle B|tomásenle QGUR|AZ 124 tornase QAGUR|Z|tomase R 124 dicho QAGURB|Z|dicho^{QR} R 125 Marco QGR|AZ.B|Martos U 125 en Q|B|Z|encl.A 125 el Q|B|Z|ov.A 125 concejo B|concejo Q|Z.A 126 rredimiesen F|GURB|Z.A|rredimiese Q 126 rredimiesen QAGURZ|rredimiesen a R| 126 a AGURB|ov.Q|Z 126 estan B|estaron QGUR|AZ 127 como QAGURB|como.Z 127 fuesen QAGURB|fuesen.Z 127 que QGURB|Z|ov.A 128 en QGURB|en.A 129 e QRB|AZ|Et.G.ov.U 129 quando B|quando QAGURZ 129 quando.B|quando lo QAGURZ 129 amigos QGURB|Z|amigos.A 129 amigos QAGURB|amigos.É.R 129 lo QAGURB|ov.R 129 el GURB|el.AZ.ov.QR 129 e³ AURZ|Et.GR.mas.Q 129 más GURB|AZ|ov.Q 129 quiso QUR|quiso elcl.AZGR| 130 tornar AGURB|Z|tomar Q 130 la QAGRB|Z|ov.U 130 fe QURB|Z|ov.G 130 avia QGURB|Z|havia.A 131 todos AGURB|Z|todos lo Q

- 1 Introduction
- 2 Méthode
- 3 Transfert d'information et collation matérielle
- 4 Résultats
- 5 Conclusions**

Limites

- Peu adapté aux traditions très hétérogènes (problème ecdotique avant tout?);
- Surtout utile pour les longs textes en prose (l'alignement macroscopique doit être fait à la main);
- Le tokéniseur fonctionne pour le castillan et le latin, non testé sur les autres langues;
- Très dépendant de la qualité du lemmatiseur ;
- Pour l'instant, ne produit que des apparats au mot ;

Conclusions

- On a proposé une méthode de collation qui permet d'utiliser la TEI en entrée ;
- Cette méthode permet de commencer à travailler sur de l'alignement de phénomènes matériels (le texte restant le pivot de l'alignement) ;
- L'outil adopte globalement les présupposés de la méthode bédieriste ;
- L'utilisation de la machine pour la collation permet une formalisation d'une partie de la méthode ecdotique ;

Travaux envisagés par la suite

- (en cours) rédaction d'un schéma/ODD « universel » ;
- Réduire la dépendance à CollateX (utilisé pour l'alignement uniquement) qui n'est plus en développement actif ;
- Intégrer le sémantisme à la collation (modèles de langues et plongements de mots) ;
- Travailler sur l'alignement macro.

Lien vers le dépôt

`https://gitlab.huma-num.fr/mgillelevenson/tei_collator`

Merci!

Merci de votre attention!

Références

- Elli BLEEKER, Bram BUITENDIJK et Ronald HAENTJENS DEKKER. « Marking up Microrevisions with Major Implications : Non-linear Text in TAG ». *Balisage : The Markup Conference*. 2020.
- Elli BLEEKER, Bram BUITENDIJK, Ronald HAENTJENS DEKKER et Astrid KULSDOM. « Including XML Markup in the Automated Collation of Literary Text. ». *XML Prague 2018*. Prague, République Tchèque, 2018.
- Jean-Baptiste CAMPS, Lucence ING et Elena SPADINI. « Collating Medieval Vernacular Texts : Aligning Witnesses, Classifying Variants ». en. *DH2019 Digital Humanities Conference*. Utrecht, 2019. URL : <https://dh-abstracts.library.cmu.edu/works/10074> (visité le 07/11/2020).
- Jean-Baptiste CAMPS, Lucence ING et Elena SPADINI. *falcon. A processing workflow for automated collation*. Version 0.0.1-alpha. 2021. URL : <https://github.com/CondorCompPhil/falcon>.
- Ronald HAENTJENS DEKKER. « Computer Automated Collation with CollateX and Python ». *Presentation given at DH Benelux in The Hague (2014)*, p. 12-13.
- Philip KAHLE, Sebastian COLUTTO, Günter HACKL et Günter MÜHLBERGER. « Transkribus. A service platform for transcription, recognition and retrieval of historical documents ». *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. T. 4. IEEE, 2017, p. 19-24.
- B. KIESSLING, R. TISSOT, P. STOKES et D. Stökl Ben EZRA. « eScriptorium : An Open Source Platform for Historical Document Analysis ». *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*. 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). T. 2. Sept. 2019, p. 19-19. DOI : 10.1109/ICDARW.2019.10032.
- Elisa NURY. « Automated Collation and Digital Editions From Theory to Practice ». Thèse de doct. Codir. Elena PIERAZZO et Victoria MOUL. Londres : King's College, 2018. URL : https://kclpure.kcl.ac.uk/portal/files/105283803/2018_Nury_Elisa_1337422_ethesis.pdf (visité le 17/06/2019).
- Elena SPADINI. « Studi Sul Lancelot En Prose ». Thèse de doct. Dir. Arianna PUNZI. Rome : Sapienza Università, 2016.
- Leopoldina TORSTEN ROEDER. « Juxta Web Service, LERA, and Variance Viewer. Web Based Collation Tools for TEI ». *RIDE 11 (2020)*. DOI : 10.18716/ride.a.11.5. URL : <https://ride.i-d-e.de/issues/issue-11/web-based-collation-tools/> (visité le 23/11/2021).
- Chahan VIDAL-GORÈNE, Noémie LUCAS, Clément SALAH, Aliénor DECOURS-PÉREZ et Boris DUPIN. « RASAM. A Dataset for the Recognition and Analysis of Scripts in Arabic Maghrebi ». *International Conference on Document Analysis and Recognition*. Springer, 2021, p. 265-281.