

Transcribing Handwritten Documents Using Artificial Intelligence



Presentation to the Canadian Research Knowledge
Network (CRKN) Virtual Conference

Maxime Gohier, UQAR
Rénald Lessard, BAnQ
Jean-François Palomino, BAnQ

October 13, 2020



Nouvelle France
NUMÉRIQUE

IB
BQ

Presentation Plan

- Digitalization and dissemination within BAnQ
- The Digital New-France Project
- Technology for re-transcription and digital editions of handwritten documents
- BAnQ corpus as a trial project

BAnQ Mission

The purpose of the BAnQ is to collect, conserve and disseminate the documentary heritage of Quebec.

This heritage consists of millions of documents which are a testimony to the history and culture of the Quebecois.



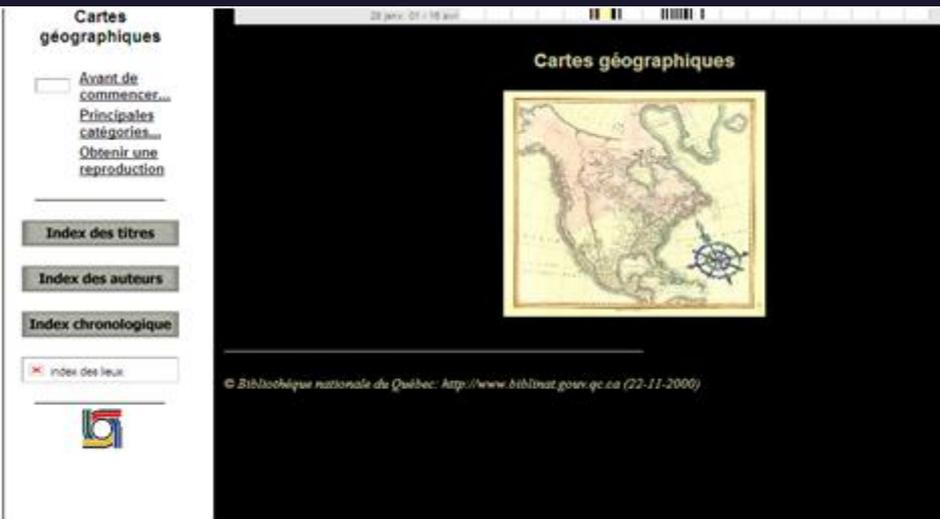
The First Revolution: Digital Collection and Dissemination

- An effective way that helps BAnQ promote research of its collections and holdings



Digitization and Dissemination within BAnQ

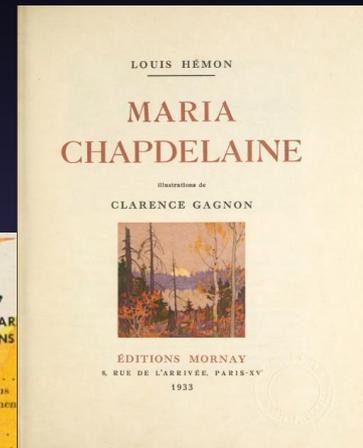
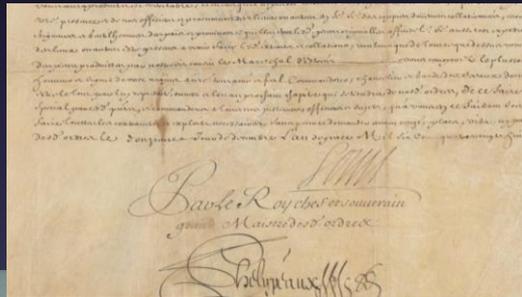
- Digital services have continued to grow over the last 20 years



Insert here a graphic on the annual volume of digitized documents or page

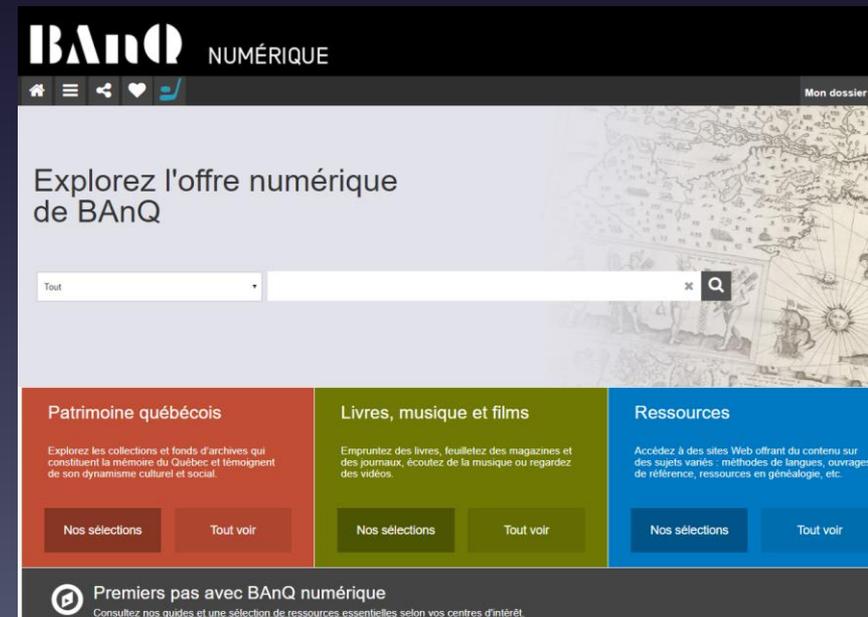
Digitization and Dissemination within BAnQ

- A variety of digitized, published and handwritten documents



Digitization and Dissemination within BAnQ

- Digitization of several million files, disseminated through the **BAnQ numérique** and the **Advitam** catalogue



The Second Digital Revolution: Optimal Character Recognition (OCR)

Launch search using keyword “Papineau” in the **BAnQ numérique**: 264,750 results !
Huge research potential.

The screenshot shows the search results page for the keyword "Papineau" on the BAnQ numérique website. The page displays 264,750 results, sorted by relevance. The search filters are: Mot clé : papineau, Patrimoine québécois, Type de document : Revues et journaux, and Texte. The results are listed in a grid view. The first three results are:

- Montebello, Papineau - Manoir Papineau**: Inventaire des Biens Culturels, Maurice Gagnon, Gérard Morisset, Jean-Paul Morisset. Vers 1950-1976. 61 éléments. Fonds ministère de la Culture et des Communications. Archives textuelles.
- François Papineau**: Papineau, François. 1797-1810. 6 éléments. Collection des petits fonds et collections d'archives manuscrites d'origine privée. Procès-verbal faisant état du bornage d'un terrain entre Charles Provost et Thomas Parris, Côteau-Saint-Antoine, 7 août [...]. Archives textuelles.
- Les deux Papineau**: David, L.O. 1840-1926, (Laurent-Olivier). Montréal : Eusébe Senécal & fils, imprimeurs, 1896. LES DEUX PAPINEAU LES DEUX PAPIN EAU PAR L.O.DAVID MONTREAL EUSÈBE SENÉCAL &* FILS, IMPRIMEURS 20, RUE SAINT-VINCENT 1896 Enregistré cou fonnémeut à l'aote du Parlement du Canada on l'année mil huit cent quatre-vingt-sei ze, par h,> O- DAVID, au Bureau du Ministre d

The right sidebar shows the search URL, the number of results for each filter, and the availability of the results. The filter "Type de document" is set to "Revues et journaux (254196)" and "Texte (10554)".

BAnQ is looking for potential partnerships to:

- Provide better visibility for institutions and their documentary corpus;
- Promote and facilitate research of collections and holdings;
- Promote better knowledge of the history and culture of Quebec, based on documents in BAnQ.



UQÀM

LHPM

Laboratoire d'histoire et
de patrimoine de Montréal

THÉÂTRES ASSOCIÉS

{ BnF



(...)

BAnQ is looking for potential partnerships



A partner to work with BAnQ in the next
revolution of documentary research...

The Digital Revolution and *Big Data*

A worldwide movement

Gutenberg Project

- Since 1971 more than 53,000 books have been digitally transcribed

ICMH and *Early Canadiana Online*; *Canadiana.org*

- Since 1999: 19 million Canadian print documents are available online

BAnQ numérique

- Since 2003: 5,873 journals; 5.3 million journal pages and Quebecois journals

GoogleBooks

- Since 2004: 25 millions books have been scanned in every language
- Ngram Viewer

The screenshot shows the Canadiana website interface. At the top, there is a search bar with the text "Recherche dans ce document : journaux assemblée législat" and a "Rechercher" button. To the right, there are social media icons for Facebook, Twitter, and LinkedIn, along with the text "Conditions de service" and "À propos Héritage RCDR Nous joindre". Below the search bar, there is a link "Retour à la page des résultats". The main content area displays the search results for "Journaux de l'Assemblée législative de la Province du Canada ... : [Journaux de l'Assemblée législative de la Province du Canada depuis le 28me jour de novembre 1844, jusqu'au 29me jour de mars, 1845, les deux jours inclus, et dans la huitième année du règne de notre souveraine dame la Reine Victoria : étant la première session du second Parlement provincial du Canada Montréal : L. Perrault, 1845; 504 pages avec recherche en texte intégral ?". Below the search results, there is a navigation bar with "Afficher" and "À propos" buttons. The main content area shows a preview of a document page from 1845, with the date "15 Janvier" and the page number "148". The document text is in French and discusses legislative matters.

The Digital Revolution

Google Books Ngram Viewer

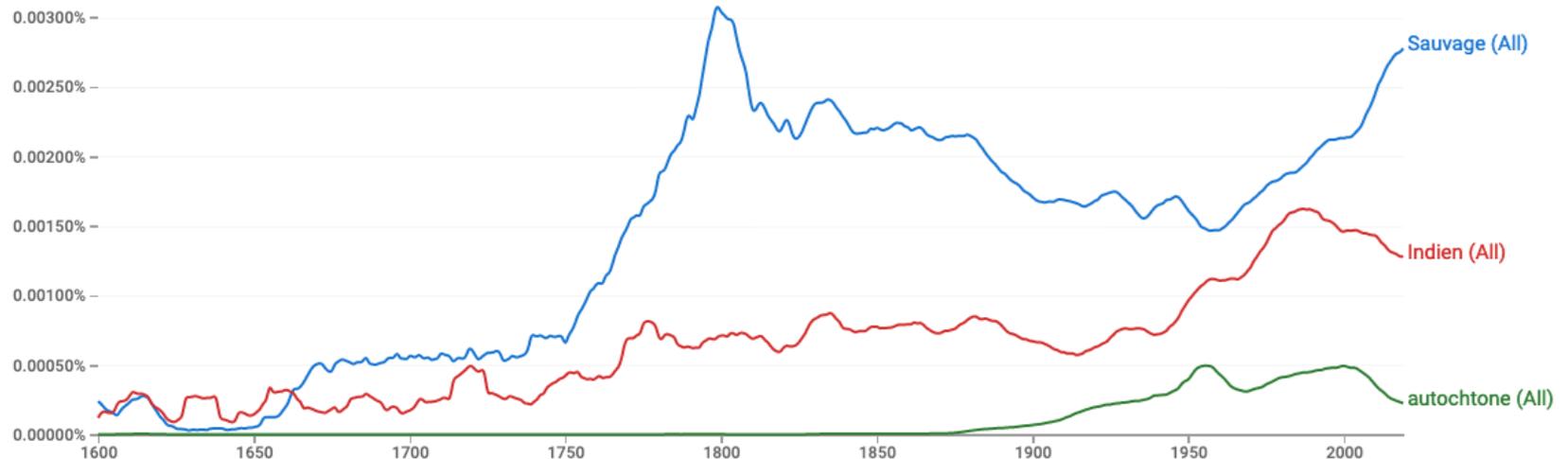
§ Sauvage,Indien,autochtone

1600 - 2019

French (2019)

Case-Insensitive

Smoothing of 5



(click on line/label for focus, right click to expand/contract wildcards)

Challenges of OCR

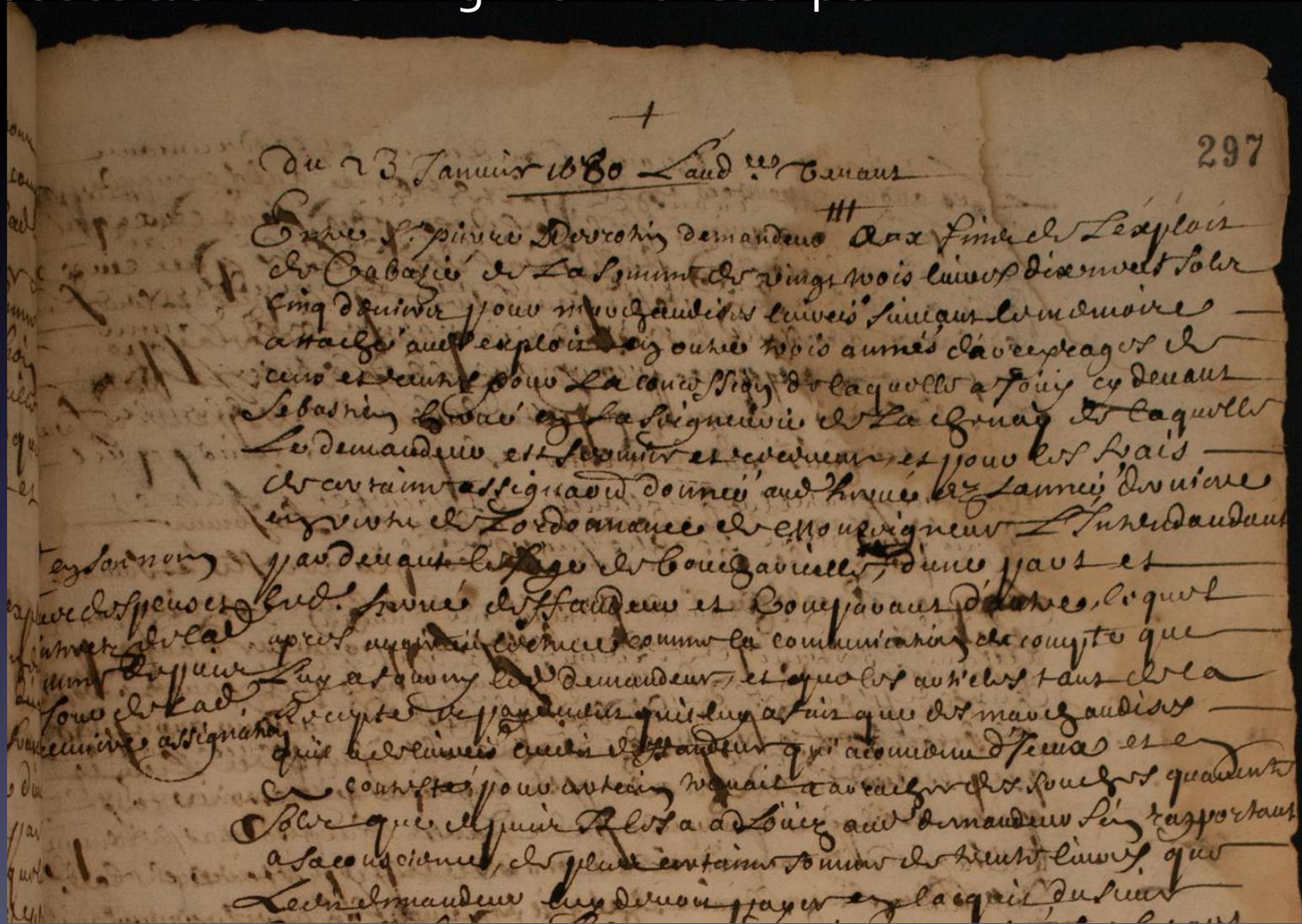
- Relative quality of OCR: Creation of an "illusionary order"

Il nous dit ensuite que nous avons mal traduit le mot "*the Indians*" qui signifie, selon lui, *les Indiens*, et non les sauvages. Si l'écrivain du *Transcript* était plus familier avec la langue française, il saurait que nous désignons plus souvent sous le nom "d'Indiens" les habitants des Indes, que les habitants des forêts du nouveau monde où la civilisation n'a pas encore pénétré ; il saurait que le mot "sauvage" a deux acceptions en français. Il ne signifie pas toujours "féroces et barbares," puis qu'on dit en français, "les sauvages civilisés, les sauvages instruits, les sauvages du Sault St. Louis, les sauvages de St. Régis, les sauvages de St. François, &c." Mais on se sert presque invariablement de ce mot pour désigner les habitants primitifs de nos forêts, les naturels du pays, toutes les peuplades sauvages, en un mot. Il nous sert

11 uous dit ensuite ue nous avnus mal ¶
Crudit le mut.* * the Indians? qui signitie, ¶
selon lni, les Indiens, ex non les sauvages. ¶
Ni l'écrivain du Transcript était plus fummi- ¶
per aveo a langue française, «1 sausait que ¶
ous d ignous plus avuveut xoux le nom ¶
« d'ludiens" les hubitante des ludes, que ¶
les habitants des tucets du nouveau moude ¶
où la civilisation n'a jus encore pénétré ¶
aurai que le mot a get deux ues ¶
ceptions en françai, l ue signilie pas ¶
toujours " féroeus et barbarue, * " pois qu' ¶
Jiten francais, * les sauvages ci vilises, jen ¶
auvages Fruits les ages du Sanit ¶
o Lonis, les sauvages de SL [Reis Jon ¶
gouvages de St, François, Le." ¶

Challenges of OCR

- More democratic access and document marginalization: the arduous task of working with manuscripts



Research Challenges Working with Handwritten Documents in the Digital Era

- Access to archives: democratization vs marginalization
- The need for paleographic expertise
- Data produced by researchers and archives is far from being uniform
- Variability of document types
- Sustainability of research data

Digital New-France:

A Large-Scale Digital Publication Project

Goal:

- Develop technology tools which will allow researchers to pool their research data, thus facilitating consultation and digital processing, as well as the continuous enrichment of research.

Historical-Literary Approach: Document as Text and Object

The materiality of the document (structure, cursive, deletions, illustrations, etc.) provide information about the writing process and the historical nature of the document.

A Collaborative *and* Research Partnership Project

Goals:

- Ensure compatibility and inter-operability of digital data generated by researchers;
- Ensure sustainability of data: allow the possibility of ongoing correction and gradual enrichment;
- Ensure that access to open-source data is based on the respect for the mandates of various archival centres.

Collaborative *and* Partnership Approach

- 1- Develop a digital repository to share documentary research by scholars;
- 2- Develop working protocols to ensure the uniformity and compatibility of data produced;
- 3- Support archival institutions in their mission to facilitate accessibility and improvement of their collections using digital tools.

The Digital New-France Project: Digital Repository and Research Tool

- Scanned images of archive documents
- Automated transcription
- Collaborative working protocols
- Structural and textual tags (encryption)
- Expertise in paleography (permanent workshop on documentary analysis by Dominique Deslandres)
- Research and consultation interface (next step)

Partners

- Library and Archives Canada
- Bibliothèque et Archives nationales du Québec
- Rare Books and Special Collections Library, University of Montreal
- Document and Archives Management Division, University of Montreal
- Museum of Civilization of Quebec
- READ Cooperative (Transkribus)

Others

- Archives nationales d'Outre-Mer (Aix-en-Provence)
- Abenaki Museum in Odanak

An Extensive Team... Destined to Grow

Management Team:

- Catherine Broué (UQAR)
- Maxime Gohier (UQAR)
- Léon Robichaud (Université de Sherbrooke)
- Jean-François Palomino (BAnQ)

Other Researchers:

- Mehdi Adda (UQAR)
- Alain Beaulieu (UQAM)
- Arnaud Bessières (Université Laurienne)
- Philippe Charland (Collège Kiuna)
- Colin Coates (Glendon College)
- Dominique Deslandres (Université de Montréal)
- Helen Dewar (Université de Montréal)
- Alexandre Dubé (Washington University)
- Gilles Havard (CNRS-EHESS)
- Daniella Kostroun (Indiana University)
- Jean-François Lozier (Université d'Ottawa)
- Marie-Christine Pioffet (York University)
- And many more to come...

Multiple Corpora

- Collections and minutes of French-Indigenous diplomatic encounters (Maxime Gohier)
- Reports on the exploration of Louisiana (Catherine Broué)
- Indigenous language dictionaries (Philippe Charland and Marie-Christine Pioffet)
- Geographical and toponymy maps (Jean-François Palomino)
- Land grids in Montreal (Dominique Deslandres)

Technology for the Digital Publication of Handwritten Documents

- The READ Cooperative (Recognition and Enrichment of Archival Documents) and Transkribus software
- XML-TEI mark-up language

Transkribus Software

- Text spatialization
- Handwritten Text Recognition (HTR)
- Mark-up language
- Research



Server Overview Layout Metadata Tools

Logout maxime_gohier@uqar.ca

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

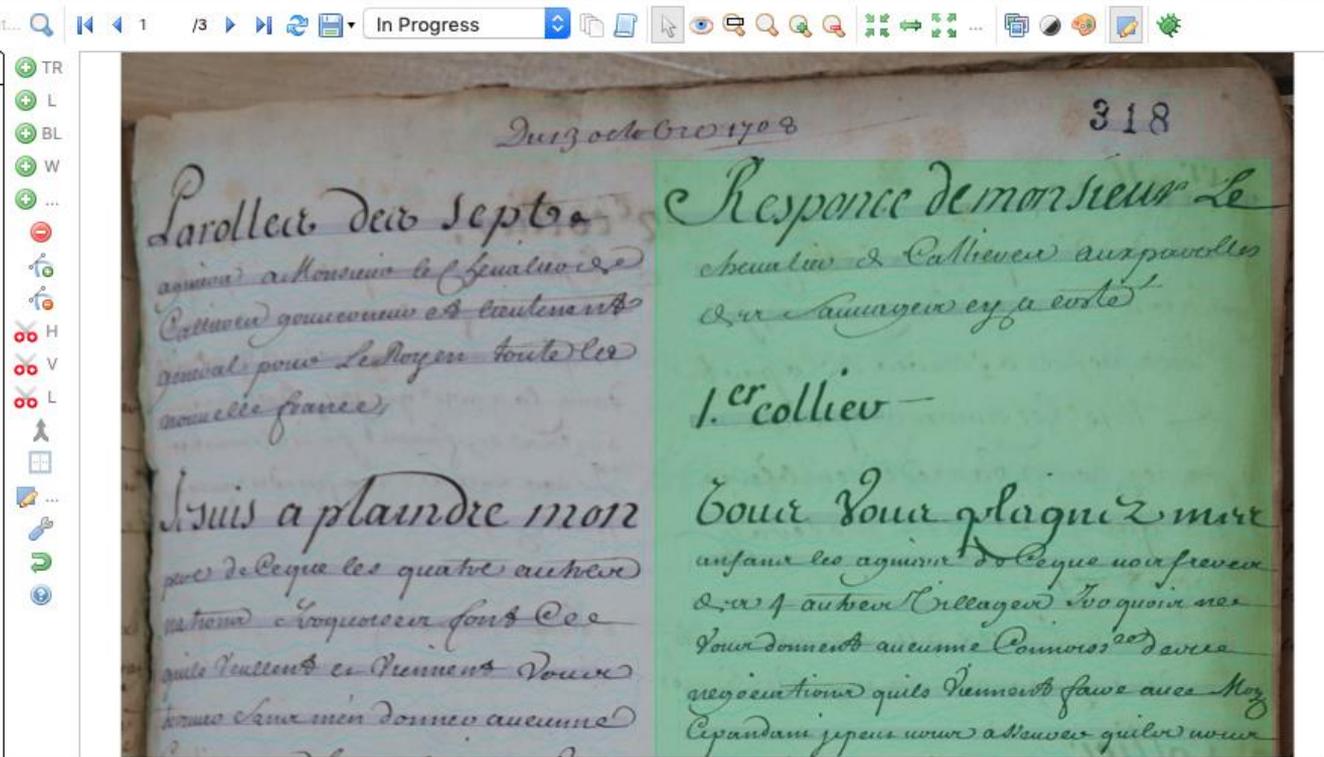
Collections: ANOM, F3 (34476, Owner) Col-ID

Documents

1-80 / 80

ID	Title	Pages	Upload
1468...	ANOM, F3, vol. 7, f. 116r-138v	46	davie
1468...	ANOM, F3, vol. 7, f. 93r-93v	2	davie
1468...	ANOM, F3, vol. 7, f. 52r-59v	16	davie
1468...	ANOM, F3, vol. 7, f. 48r-51v	8	davie
1468...	ANOM, F3, vol. 7, f. 44r-47v	8	davie
1468...	ANOM, F3, vol. 7, f. 38r-41v	8	davie
1468...	ANOM, F3, vol. 7, f. 03r-06v	8	davie
1468...	ANOM, F3, vol. 7, f. 01r-02v	4	davie
1468...	ANOM, F3, vol. 2, f. 444r-447r	7	davie
1468...	ANOM, F3, vol. 2, f. 423r-427r	9	davie
1468...	ANOM, F3, vol. 2, f. 413r-416r	7	davie
1468...	ANOM, F3, vol. 2, f. 410r-412r	5	davie
1468...	ANOM, F3, vol. 2, f. 320r-325r	11	davie
146...	ANOM, F3, vol. 2, f. 318r-319r	3	davie
1468...	ANOM, F3, vol. 2, f. 317r-317v	2	davie
1468...	ANOM, F3, vol. 2, f. 315r-316v	4	davie
1468...	ANOM, F3, vol. 2, f. 314r	1	davie
1468...	ANOM, F3, vol. 2, f. 310r-312v	6	davie

100 Filter



1-1 318

2-1 Du 13 octobre 1708

3-1 Parolles des sept

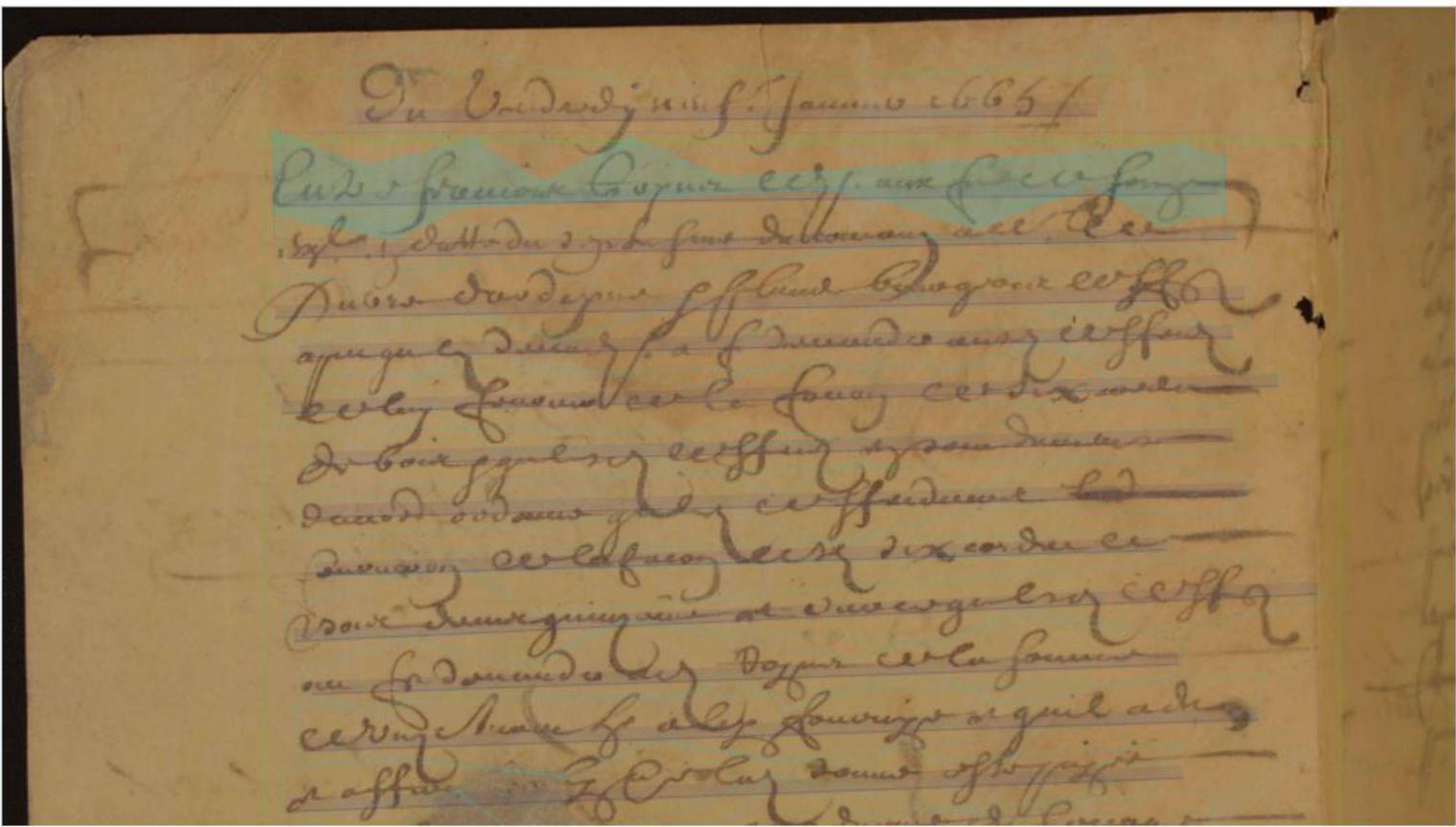
3-2 agniers a monsieur le chevalier de

3-3 callieres gouverneur et lieutenant

3-4 general pour le Roy en toute la

3-5 nouvelle france

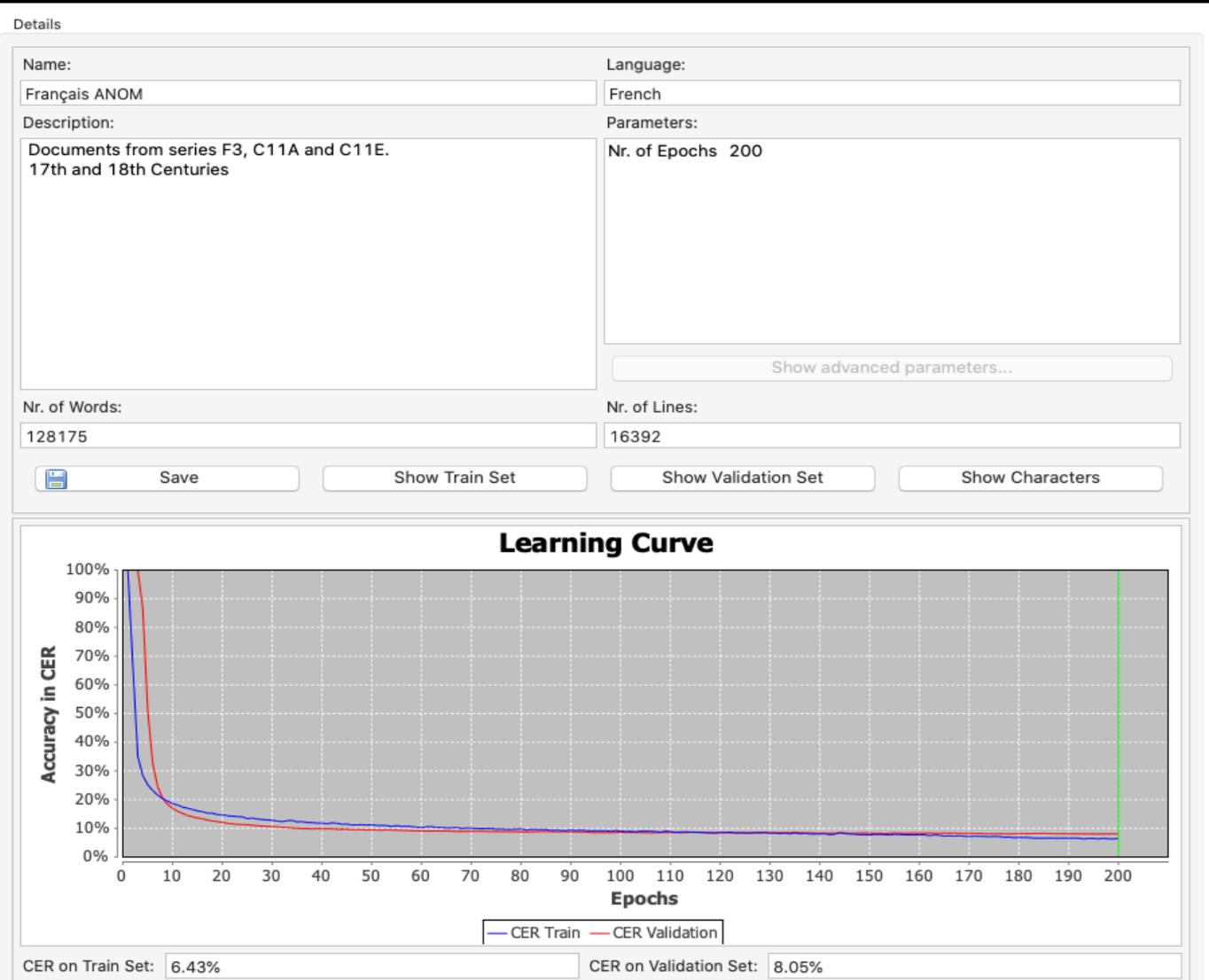
Transkribus Software and Functions



- 1-1 Du vendredy neuf^e janvier 1665
- 2-1 Entre francois Roynet dr aux fins de son
- 2-2 Expl en datte du septiesme du courant a le. de
- 2-3 Pierre dardeynes & Claude bourgeois des^r
- 2-4 apres qu led demdr a f demande ausd desfend
- 2-5 de luy fournir de La facon de dix cordes
- 2-6 de bois & qu lesd desfends en sont demeures

Spatialization of Transcriptions Within the Image

Handwritten Text Recognition (HTR)



Details

Name:

Relation de 1668

Language:

French

Description:

18th Century

Parameters:

Nr. of Epochs 50

Show advanced parameters...

Nr. of Words:

13856

Nr. of Lines:

1873



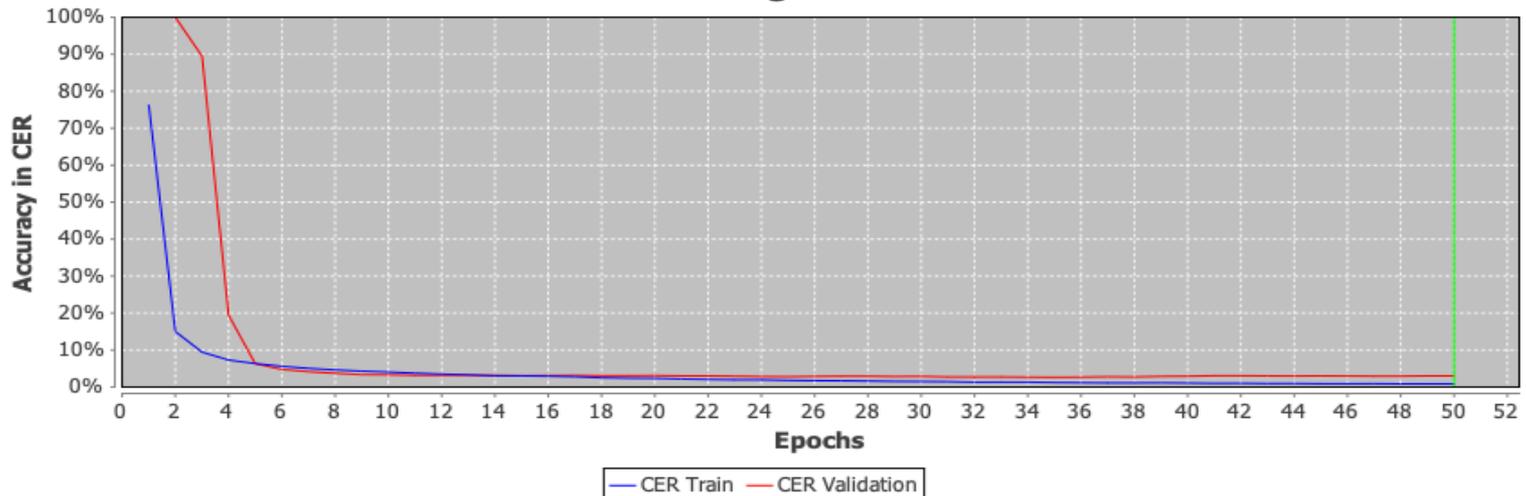
Save

Show Train Set

Show Validation Set

Show Characters

Learning Curve



CER on Train Set: 0.87%

CER on Validation Set: 3.02%

Name:

Théodore De Pincier

Language:

French

Description:

Journal d'arpentage de Théodore de Pincier, 1798-1799 (35 pages)

Parameters:

Nr. of Epochs 200

Show advanced parameters...

Nr. of Words:

6277

Nr. of Lines:

558



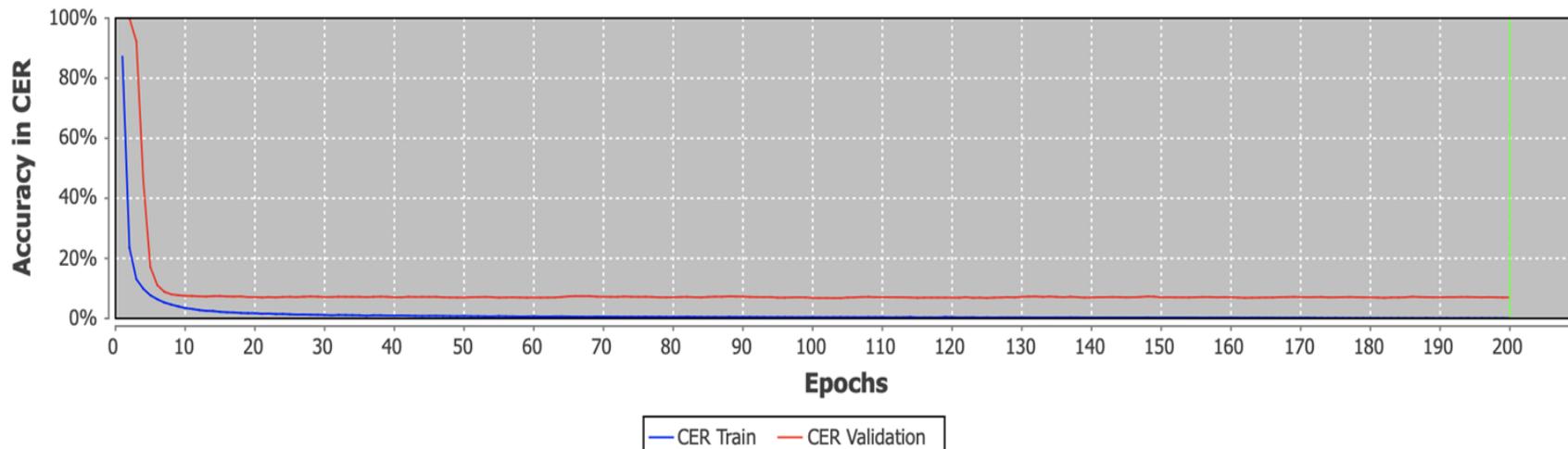
Save

Show Train Set

Show Validation Set

Show Characters

Learning Curve



CER on Train Set: 0.11%

CER on Validation Set: 6.98%

Transcription-Enriched "Spatialized" PDF Files

[Handwritten text in French, mostly illegible due to cursive script]

Defaillit

[Handwritten signature]

[Handwritten text in French, mostly illegible due to cursive script]

du xiiiie desd mois & an a le. de Charles Martin
desfr apres qu led dr a ft demande aux desfr
de la somme de vingt une livres ts a chacun
a payer de trente livres ts pour le louage de la
maison appt au nomme St. George suivant le
transport ft par led dr aud desfr de louage
d'Icelle maison Ensemble de Reparer la Closture
des pieux a la Reserve de vingt ung pieux dune part
et den a donné autre part ld desfr est dem
daccord ordonne que led desfr payera
Donc lad somme de vingt une livres ts
& ce dans quinzaine dhuy a faulte fera fr lad closture
Dans Le premier temps commode a la defense

**Lesd et aux desps
Dessailly**

**DeMonchy
Greffier**

dud Jour et an que

Entre charles LeMoyn dr aux fins
de son expl du xiiiie du pn mois & an
a le de pierre Raguedeau desfendeur
apres qu led Dr a ft damande aud desfendeur
de la somme de Cent dix livres ts
tant pour marchandises fournyes par led
dr aud desfr, quargent deub a lesglise
de ce lieu duql le dr est dispensable
& que led desfr a ft demande de Conte

What is the Text Encoding Initiative (TEI)?



“XML-TEI provides a framework to represent everything we deem important about a text and not uniquely its appearance, so that software can process different characteristics that have been identified, thus generating new visualizations and new perspectives.”

(Lou Burnard, *Qu'est-ce que la Text Encoding Initiative?*, Marseille, OpenEdition Press, (n.d.) [en ligne:](généré le 04 novembre 2015), p. 10.

XML-TEI Mark-Up Language

- Identifying and indexing of identified entities (surnames, place names, ethnic names) and dates (numbered or text format);
- Identification and resolution of abbreviations;
- Identification of ambiguous passages and alternate proposals;
- Identification of structural components of text (paragraphs, page numbers, titles, signatures, etc.);
- Identification of non-textual entities (tables, graphs, illustrations, etc.);
- Identification of successive interventions regarding a document (additions and deletion of passages, changes to cursive style);
- Clarifications concerning material support and its state of conservation;

Structural tags are used to provide a summary of the form.

- Titles and subtitles
- Paragraphs
- Page numbers
- Signatures
- Legend
- Drawings
- Greeting or salutation
- etc.

Transkribus v1.8.0 (19_09_2019_14:16), Loaded doc: ANOM, F3, vol. 2, f. 318r-319r, ID: 14

Server | Overview | Layout | **Metadata** | Tools

Document | **Structural** | Textual | Comments

Page type:

Links:

Selected element type:

Structure Type

Type of selected:

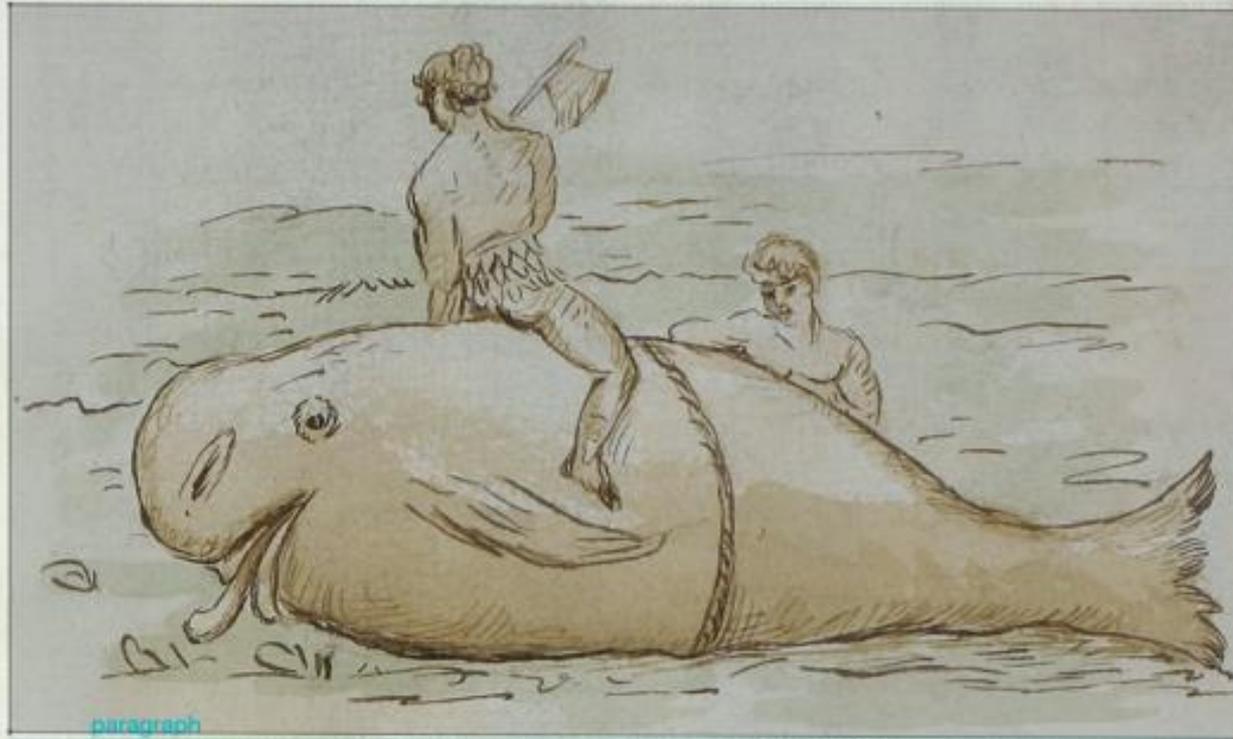
Draw struct type Draw default colors

Structure type	Color	Shortcut
paragraph		<input type="button" value="⊕"/>
heading		<input type="button" value="⊕"/>
caption		<input type="button" value="⊕"/>
header		<input type="button" value="⊕"/>
footer		<input type="button" value="⊕"/>
page-number		<input type="button" value="⊕"/>
drop-capital		<input type="button" value="⊕"/>
credit		<input type="button" value="⊕"/>
floating		<input type="button" value="⊕"/>
signature-mark		<input type="button" value="⊕"/>
catch-word		<input type="button" value="⊕"/>

Layout

Type	Structure	Text	ID
▼TextRegion	page-number		region_...
Line		318	r111
▼TextRegion	header		region_...
Line		Du 13 octobre 1708	r112
▼TextRegion	paragraph		TextReg...
Line		Parolles des sept	r114
Line		agniers a monsieur le chevalie...	r116
Line		callieres gouverneur et lieutenant	r118

Identification and Description of Non-Textual Entities



paragraph

se courent contre le froid. Qui leur monstreroit à
vivre, & leur enseigneroit le labourage des terres,
& autres choses, ils apprendroient fort bien: car
il n'y a de gens si bêtes qui ont leu sagement d'yeux

Textual mark-ups are used to take into account the contents and to facilitate tracking.

The screenshot displays a transcription software interface. The main window shows a document with various tags applied to the text. A 'Text based Tagging' dialog box is open, showing a list of tags and their specifications. The dialog box has a 'Tags' section with a table of tag specifications and a 'Props for tag: 'date' - value: '13 octobre 1708'' section with a table of properties.

Tag	Value	Text	Properties
1	date	13 octo Du 13 octobre 1708	year: 1708 day: 1
2	textStyle	Parolles Parolles des sept agniers a mons	fontSize: 0.0 kern
3	organization	agniers Parolles des sept agniers a mons	type: nation
4	person	de monsieur le chevalier de calliere	
5	person	callieres le chevalier de callieres gouvern	
6	place	nouvell Roy en toute la nouvelle france	
7	textStyle	Je suis : nouvelle france Je suis a plaindre	fontSize: 0.0 kern
8	organization	iroquois autres nations iroquoises font ce	
9	organization	Sonnont vont chez les Sonnontouans et r	
10	organization	onneyou pasent pas les onneyoust elle ne	
11	person	corlard fait sans que corlard (cest le goi	
12	person	onnontio desclave du Cots de onnontio et	
13	organization	harassic Coste des anglois, harassion ne	
14	person	de callie monsieur le chevalier de calliere	
15	textStyle	1 sauvages cy a costé 1er collier v	fontSize: 0.0 kern
16	textStyle	er sauvages cy a costé 1er collier v	fontSize: 0.0 kern
17	textStyle	collier cy a costé 1er collier vous vous	fontSize: 0.0 kern
18	textStyle	vous vo costé 1er collier vous vous plaig	fontSize: 0.0 kern
19	organization	agniers mes anfans les agniers de ce qu	
20	organization	iroquois autres villages iroquois ne vous	
21	abbrev	connois donnent aucune connoissce des	
22	textStyle	ce aucune connoiss ce des negocia	fontSize: 0.0 kern
23	unclear	harassic apporté et que sy harassion quil	
24	person	harassic apporté et que sy harassion quil	
25	place	montreé paffé cet esté a montreal y a ma	
26	abbrev	derere trouvates point lanné derere a m	
27	textStyle	ere point lanné derere a montreal pc	fontSize: 0.0 kern
28	place	montreé lanné derere a montreal pour rat	
29	place	lac de rendre par le lac de champlain c	
30	place	champl par le lac de champlain ce quy le	
31	unclear	put vous avez les put mal fait les un:	

All Tags	Color	Shortcut
abbrev	Red	
add	Cyan	
Address	Green	
blackening	Black	
date	Blue	
del	Light Blue	
div	Dark Blue	
gap	Cyan	
organization	Magenta	
person	Yellow	
place	Purple	
sic	Yellow	
speech	Pink	
supplied	Red	
unclear	Orange	
work	Green	

Props for tag: 'date' - value: '13 octobre 1708'

day	13
month	10
year	1708

Easy-to-use, including:

- Copy-paste-click
- Tags and mark-ups can be personalized at all times

Textual Mark-Ups are Used to Underscore the Writing Process and the Historical Authenticity of Documents

Transkribus v1.6.0 (28_02_2019_14:59), Loaded doc: Lettre de Condé à La Salle, ID: 50492, Page 2, file: Lettre a L-Ç La Salle de Chantilly le 28 janvier 1684 1.JPG [Image Meta Info: (Resolution:180.0, w*h: 2520 * 3776)]

Server Overview Layout Metadata Tools
Document Structural Textual Comments

Tags of current Transcript

Tag	Value	Text	Properties
1	place	Chantilly A Chantilly le 28. Janer 1684	
2	date	28. Janer A Chantilly le 28. Janer 1684 11	year: 0 day: 0 month: 0
3	textStyle	er Chantilly le 28. Janer 1684 11.	fontSize: 0.0 kerning: 0 sup
4	person	sr de St Le sr de St Auban ma escrit pour	
5	abbrev	sr Le sr de St Auban ma	expansion: sieur
6	textStyle	r Le sr de St Auban ma	fontSize: 0.0 kerning: 0 sup
7	abbrev	St Le sr de St Auban ma escrit	expansion: Saint
8	textStyle	t Le sr de St Auban ma escrit	fontSize: 0.0 kerning: 0 sup
9	textStyle	avec tenir en mon nom avec Mademo	strikethrough: true fontSize
10	del	avec tenir en mon nom avec Mademo	
11	textStyle	Mademi en mon nom avec Mademoisell	strikethrough: true fontSize
12	del	Mademi en mon nom avec Mademoisell	
13	del	et il ma enfans au baptesme et il ma fait	
14	textStyle	et il ma enfans au baptesme et il ma fait	strikethrough: true fontSize
15	add	ce que j qu'il seroit bon ce que je luy ay z	
16	textStyle	jay volontiers. Et jay je vous ay choi	strikethrough: true fontSize
17	del	jay volontiers. Et jay je vous ay choi	
18	abbrev	fr ay choisy pour fr pour moy en m expansion: faire	
19	textStyle	pour mc choisy pour fr pour moy en ma p	strikethrough: true fontSize
20	add	pour mc choisy pour fr pour moy en ma p	
21	del	pour mc choisy pour fr pour moy en ma p	
22	add	en ma p pour fr pour moy en ma place toi	
23	abbrev	ql place tout ce ql faudra pour pou expansion: qu'il	
24	textStyle	pour tout ce ql faudra pour pour cela	strikethrough: true fontSize
25	del	pour tout ce ql faudra pour pour cela	
26	abbrev	parer vostre merite parer qu a cause expansion: particulier	

Tags

All Tags Show all Customize...

Tag specification	Color	Shortcut
date	Blue	
del	Green	
div	Yellow	
gap	Red	
organization	Purple	
person	Magenta	
place	Cyan	

Props for tag: no tag selected

Previous Next Apply to selected

A Chantilly le 28. Jan^{er} 1684. 11.

Le s^r de S^t Auban ma escrit pour me prier
de vouloir faire tenir en mon nom avec
Mademoisell vn de ses enfans au baptesme
ce que je luy ay accordé volontiers. Et jay je vous ay choisy
pour moy en ma place tout ce ql
pour fr pour moy en ma place tout ce ql

1-1 A Chantilly le 28. Jan^{er} 1684
1-2 11.
2-1 Le s^r de S^t Auban ma escrit pour me prier
2-2 de vouloir faire tenir en mon nom avec
2-3 Mademoisell vn de ses enfans au baptesme
2-4 et il ma fait entendre dire qu'il seroit bon
2-5 ce que je luy ay accordé volontiers. Et jay je vous ay choisy
2-6 pour fr pour moy en ma place tout ce ql

Transkribus v1.8.0 (19_09_2019_14_16), Loaded doc: Rasles-Dictionar of the Abenmakis Language, ID: 61743, Page 26, file: 27.jpg [image meta info: (Resolution: 1.0, w.h: 1045 - 1473)]

Server Overview Layout

Logout maxime_gohier@uqar.ca

Document Manager User Manager

Versions Jobs

Recent Documents User activity

Collections: Dictionnaires (1) Col-ID

Documents

1-13 / 13

Doc-ID

ID	Title
2136...	Test Mélissa
1713...	TRAINING_TESTSET_De La Brosse
1657...	Le petit catéchisme de Québec
1465...	Akitami Aukhigan
1301...	St-Mark
1299...	Pobatami kidwogan
1292...	P8batammi Linto8ganal
1258...	Kagakimzouiasis_1832
1170...	Abenakis Indian Legends, Gram
97592	New Familiar Abenakis and Eng
61743	Rasles-Dictionary of the Aben
53262	Louis Nicolas-Grammaire algonq
50362	Dictionnaire De La Brosse

100 Filter

1-5 cela est amer

1-6 PSaqlupagat

1-7 Je te trouve amer

1-8 nissaitauvo de ei

2-1 Ameçon p. hameçon

2-2 ma glii aen

The database consists of various corpora ensuring cross reference searches (ex. toponymy or terms derived from Aboriginal languages in the glossaries).

Logout maxime_gohier@uqar.ca

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: Cartes (14125, Owner) Col-ID

1-11/11

ID	Title	Pages	Uploader
66706	Amérique septentrionale par N. Sa...	1	maxime_gohi
65331	0001_00736003	1	maxime_gohi
63907	Carte de l'Amerique septentrionale...	1	maxime_gohi
63906	Carte Lahontan Jovenzano	1	maxime_gohi
53546	Cartes	17	catherine_br
51113	Carte de la decouverte du Sr Jolliet	1	maximegohie
51112	Carte de la decouverte du Sr Jolliet	1	maximegohie
51111	Carte de la Decouverte du Sieu...	1	maximegohi
51110	dossier sans titre 2	1	maximegohie
51108	dossier sans titre	1	maximegohie
45880	Carte géographique de la nouvelle...	1	maximegohie

100 Filter



- 1-8 assez de place
- 2-1 [Monseigneur Le Comte de Frontenac, Conseiller du Roy en sès Conseils Gouverneur et Lieutenant general pour sa Maiesté en Canada, Acadie, Isle de Terrebonne et autres pays de La](#)
- 2-2 France septentrionale
- 2-3 Monseigneur
- 2-4 Cest avec bien de la ioye que iay le bonheur aujourd'huy de vous presenter cette Carte qui vous fait connoitre la scituation des rivieres et des Lacs sur lesquels on navigue au travers du Canada ou Amerique septentrionale qui a
- 2-5 plus de 1200 lieux de LEst a LOuest, Cette grande riviere qui porte le nom de Riviere Colbert pour avoir esté decouverte ces dernieres annees 1673 et 1674 par les premiers ordres que vous me donnates entrant dans votre gouvernement

Old maps contain a great deal of text in which the search function is not normally available.

Collective Construction

- A collective decoding operation
- Open data format
 - Possibility of correcting and improving transcription
 - Use of mark-ups is potentially infinite
 - The open nature of TEI-Consortium (possibility of adding tags depending on the research interests)
- Transkribus and READ-Coop
 - Open software
 - Cooperative model: funding and cost of use
 - From training to technology
- The more researchers continue to participate, the greater will be the utility of the database.

The BAnQ Corpus

Falls within scope of this Project

- Decrees issued by the Intendant of New-France (approximately 11,000 pages have already been scanned)
- Notary records (digital processing of these records is underway by Professor Dominique Deslandres)
- Geographical maps from New-France

Benefits for BAnQ

- BAnQ is active in a project that uses the most up-to-date tools of artificial intelligence
- BAnQ then receives textual retranscriptions produced by handwritten language recognition and can use these texts within its search engines (BAnQ numérique and Advitam)

Questions Arising During Project

- What are the rules to manage and share retranscriptions between researchers and partners?
- What is an acceptable level of error in light of re-dissemination in the BAnQ search engines (BAnQ numérique and Advitam)?
- What corpora should have priority?
- What form will be used for a potential central repository for these retranscriptions using the sharing and dissemination tool specific to each partner institution?

The Floor is Yours...