

Transcribing Foucault's handwriting with Transkribus a report

Marie-Laure Massot

Academic engineer

CAPHÉS, UMS 3610 – CNRS/ENS Paris



Transcribing Foucault's handwriting with Transkribus a report

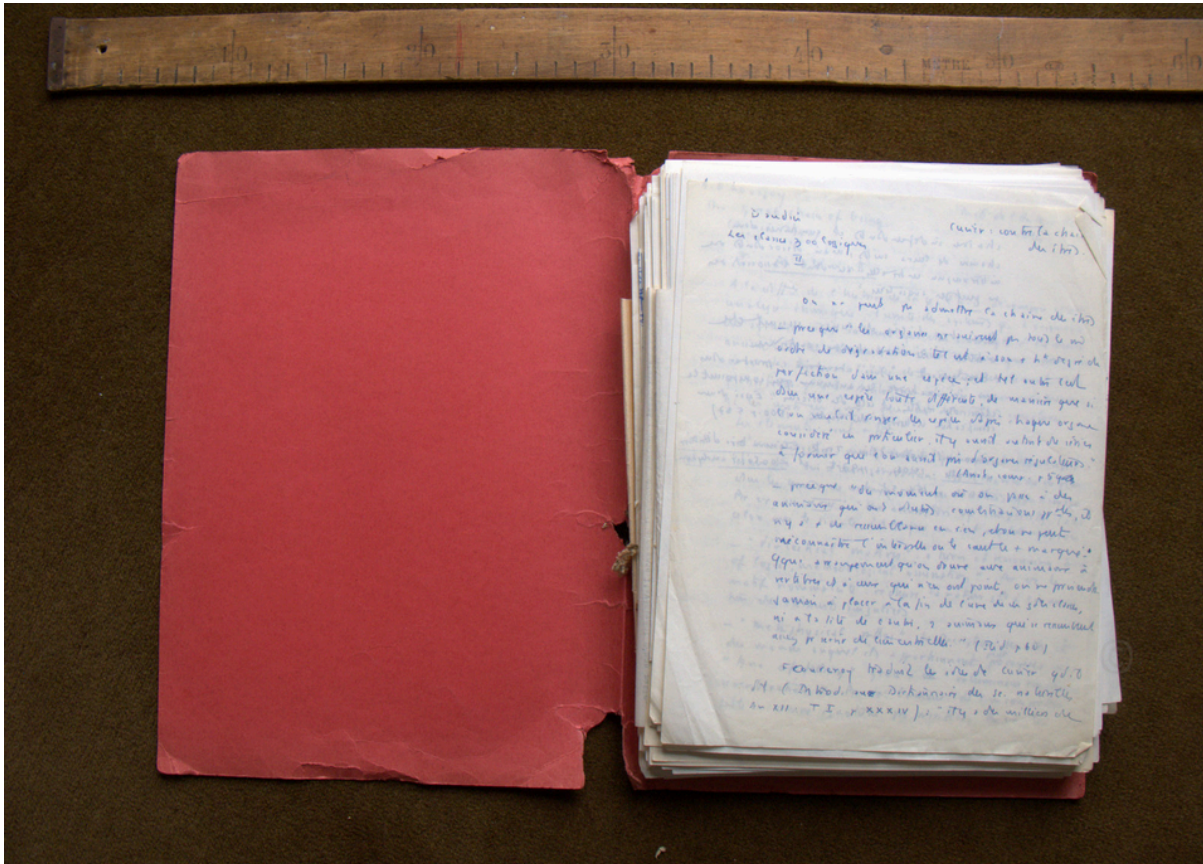
OUTLINE

- ❑ The ANR project Foucault's Reading Notes (Foucault fiches de lecture - Ffl project)
- ❑ How to explore the corpus with digital tools
- ❑ Automatic transcription of Foucault's manuscripts with Transkribus:
 1. Creating learning data for Foucault's handwriting
 2. Main difficulties encountered
 3. Assessment of the experiment and prospects



Transcribing Foucault's handwriting with Transkribus: a report

The ANR project Foucault's Reading Notes (Ffl project)



ANR-17-CE38-0001

Transcribing Foucault's handwriting with Transkribus: a report

How to explore the corpus with digital tools

The screenshot displays the Transkribus web interface. At the top, a handwritten document is shown with the title "Exploitation des paysans au XI^e-XII^e s." and the author "Duby, Eco rural". The document text includes: "Au cours du XI^e s., « l'agriculture des champs, de bleds, etc. Dieu en richit », etc. rignent. Avec 6 ans de technique, les rignent ont moins besoin de h. de leur cour, y domptiques. Dieu au paysan en h., il préfère 1 mètre sur ce bled...".

Below the document, the interface is divided into several sections:

- Bloc-notes / mémo**: A text editor with a menu (File, Edit, View, Format) and a toolbar. It contains the text "Aucun contenu enregistré dans le bloc-note." and a "POWERED BY TRIVICE" logo.
- Décrire / annoter la fiche [b001_f0002]**: A section for describing and annotating the document. It includes a "Type de l'annotation" dropdown set to "1.1 Descripteur (no-cv-ajout)" and a "Valeur type choisie" field.
- Description et annotations de la pièce n°b001_f0002 [Boîte_001]**: A table of annotations with columns for the annotation type and the value.
- Type de la pièce**: A section for selecting the document type, currently set to "Photozoope".
- Contrôle des requêtes et données de la page**: A section for managing queries and page data, with an "Afficher" button.

| DescriptionArchivistique | | |
|--------------------------|---|------------|
| MentionDePersonne | Duby, Georges | [?] éditer |
| TermeExtrait | exploitation | [?] éditer |
| TermeExtrait | fac | [?] éditer |
| TermeExtrait | justice | [?] éditer |
| TermeExtrait | rural | [?] éditer |
| TermeExtrait | économie | [?] éditer |
| TitrePiece | Duby, Eco rurale II. Exploitation des paysans au XI ^e -XII ^e s. La justice comme fac. | [?] éditer |

| AnalyseScientifique | | |
|---------------------|------------------------------------|------------|
| Descripteur | préparation de Surveiller et punir | [?] éditer |
| Descripteur | paysannerie | [?] éditer |

| NoteEditoriale | | |
|----------------|---|------------|
| Avertissement | Reprendre les références de cette fiche | [?] éditer |

Transcribing Foucault's handwriting with Transkribus: a report

Automatic transcription of Foucault's manuscripts with Transkribus

The screenshot displays the Transkribus v1.3.7 interface. The main window shows a document page with handwritten text in French. The text is organized into sections, with some parts highlighted in blue. The transcription list at the bottom shows the following entries:

| ID | Title | Pages |
|-------|---------------------------------|-------|
| 43640 | boite22 | 1140 |
| 42205 | boite20 | 1225 |
| 35689 | boite51_diff | 628 |
| 23825 | Fonds_Foucault_BnF_boite-1-jpeg | 698 |
| 8619 | icp1 | 2 |

The transcription list also includes the following text:

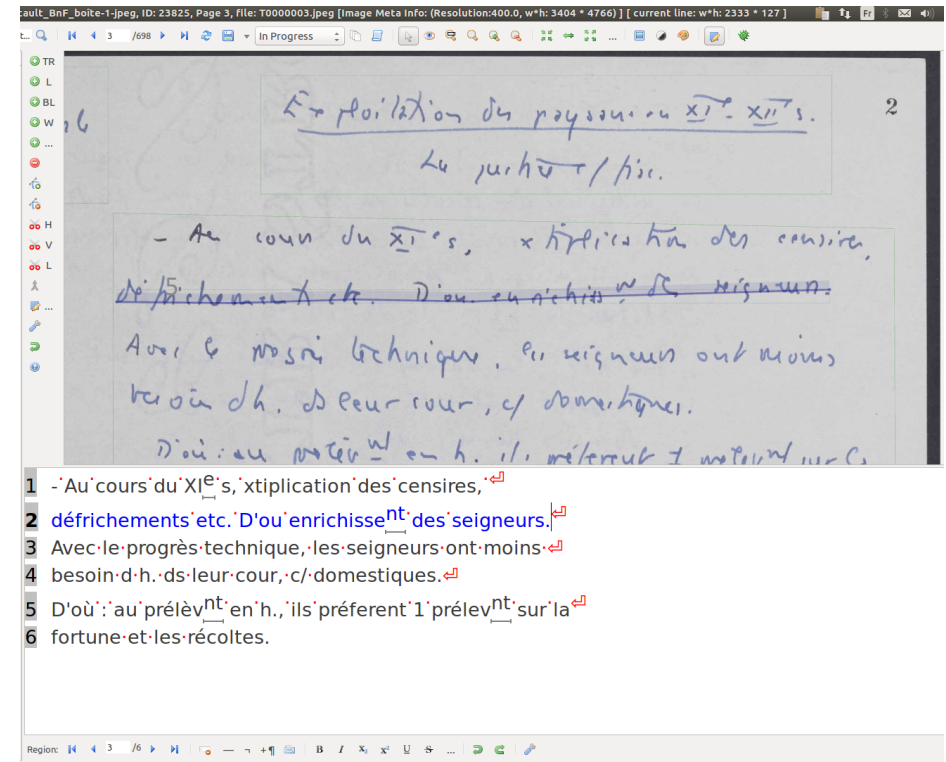
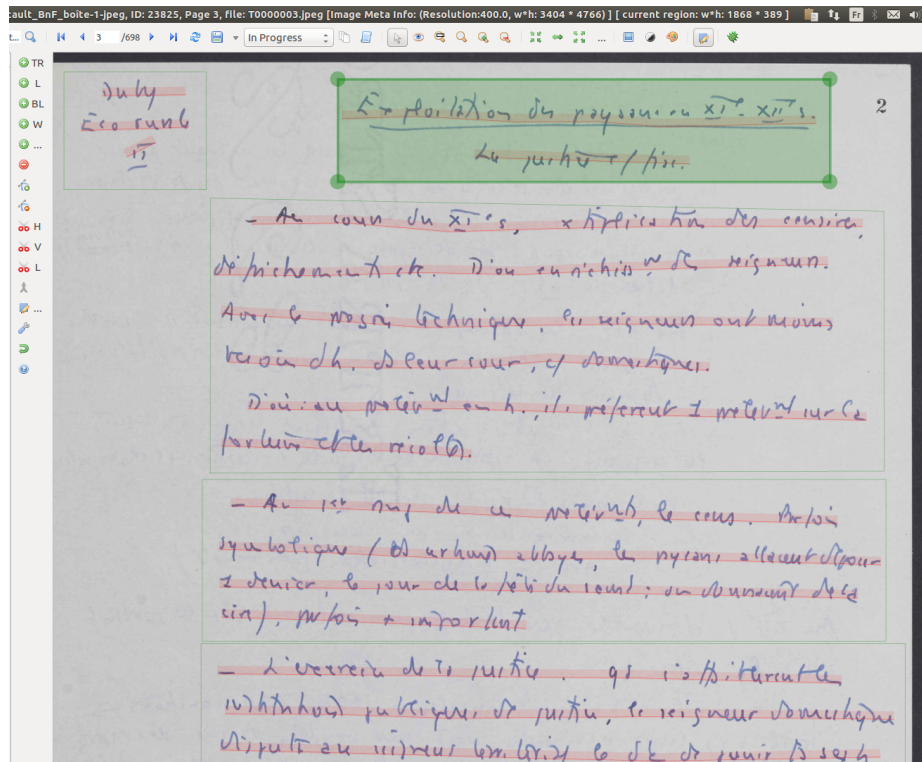
- 1 - Au cours du X^es, xtiplication des censires,
- 2 défrichements etc. D'ou enrichisse^{nt} des seigneurs.
- 3 Avec le progrès technique, les seigneurs ont moins
- 4 besoin d-h. ds leur cour, c/ domestiques.



1. Creating learning data for Foucault's handwriting
2. Main difficulties encountered
3. Assessment of the experiment and prospects

Transcribing Foucault's handwriting with Transkribus: a report

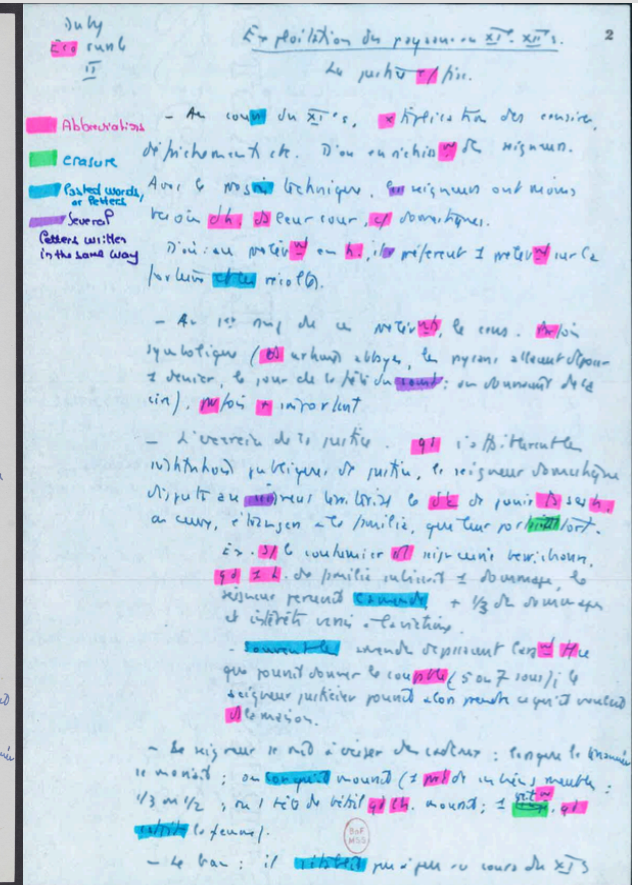
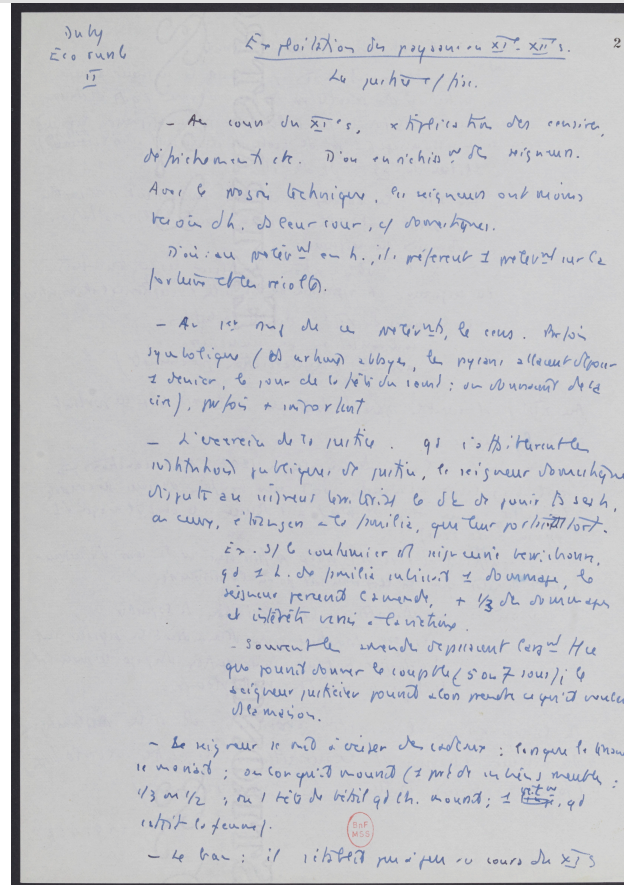
Creating learning data for Foucault's handwriting



Transcribing Foucault's handwriting with Transkribus: a report

Main difficulties encountered

- Foucault's handwriting: reading notes with a lot of ambiguous **abbreviations**, several letters written in the same way, **erasure**, misspelled, **pasted words**, etc.
- Illegible words were identified with TEI encoding
- Getting to grips with the software and average time per card
- Software Ergonomics
- Multi-handed transcripts



Transcribing Foucault's handwriting with Transkribus: a report

Automatic transcription

Loaded doc: bolte51_diff, ID: 35689, Page 51, file: T0000424.jpeg [Image Meta Info: (Resolution:400.0, w'h: 3496 * 4918)] [current line: w'h: 2394 * 252]

Search current document...

TR 162
L 166
BL
W
...
H
V
L
...
167

Faute de parents qui punissent les enfants
Pon qu'il se sont blessés ds des jeux, si bien que
les enfants cochent leurs blessures, qui denentent
+ grave.
"L'on ne doit presque jamais battre les
enfants, car, sans compter que c'est les avilir et
le ravater au rang des malheureux, et qu'on
leur inspire des sentiments bas, ainpans,
suvent le mensonge et pa. des vices encore +
gdts, c est qu'il est très ivident quecela est

5 desanfants
6 1762.
7 - Faute des parents qui punissent les enfants
8 166
9 lorsqu'ils se sont blessés ds des jeux, si bien que
10 les enfants cochent leurs blessures, qui denentent
11 + grave.
12 167
13 - "L'on ne doit presque jamais battre les
14 enfants, car, sans compter que c est les avilir et
15 le ravater au rang des malheureux, et qu'on
16 leur inspire des sentiments bas, ainpans,
17 suvent le mensonge et pa. des vices encore +
18 gdts, c est qu'il est très ivident quecela est

Loaded doc: bolte51_diff, ID: 35689, Page 45, file: T0000418.jpeg [Image Meta Info: (Resolution:400.0, w'h: 3524 * 4936)] [current line: w'h: 2299 * 224]

Search current document...

TR
L
BL
W
...
H
V
L
...
1776
122

si sur la sant et
léducatim medat des filles
le lait de la mère
Le lait de la mère
L'aliment qui convient le mieux aux enfants
en ga jusqu a ce que leurs forces soient accrues
eu x de pouvoir digérer des aliments solides est
icontestibint le lait et surtt celui de la mère
si elle est saine Le lait des mammrelles de la
mère, qui n est qu'une gradation du lait utérain
123
est la nourriture la + analogue la constitution
parhieulture de l'infant. La nature, qui proportionne
tjrs les ressources au besom ... donne à chaque

1 anit
2 Essai sur la sant et
3 Le meilleur alement pr lanfant
4 léducatim medat des filles
5 le lait de la mère
6 17 76
7 122
8 1L'aliment qui convient le mieux aux enfants
9 en ga jusqu a ce que leurs forces soient accrues
10 eu x de pouvoir digérer des aliments solides est
11 icontestibint le lait et surtt celui de la mère
12 si elle est saine Le lait des mammrelles de la
13 mère, qui n est qu'une gradation du lait utérain
14 123
15 est la nourriture la + analogue la constitution
16 parhieulture de l'infant. La nature, qui proportionne
17 tjrs les ressources au besom ... donne à chaque

Transcribing Foucault's handwriting with Transkribus: a report

Assessment of the experiment and prospects

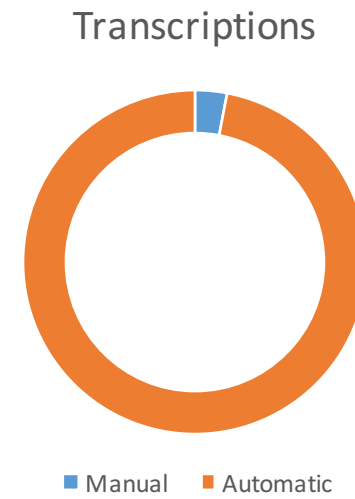
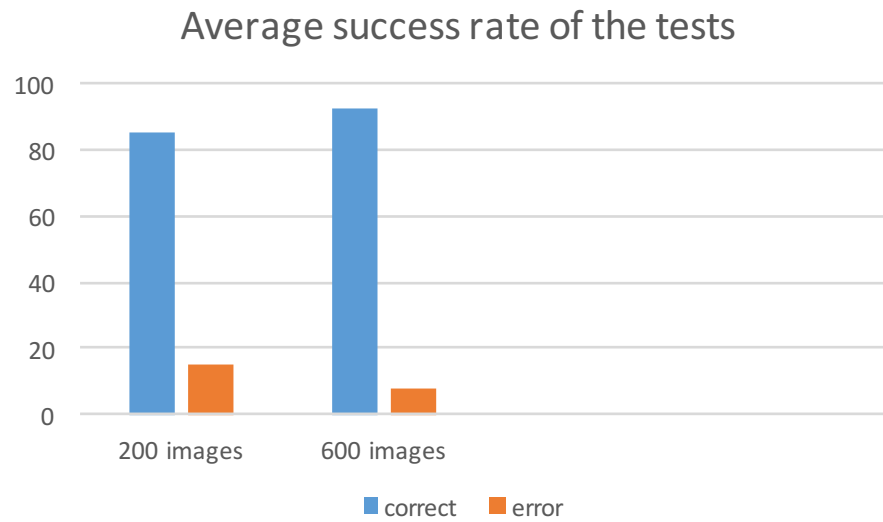
- Training of **600 images** transcribed **manually**
- Average success rate of **92% on the characters**
- Future corrected transcripts can be fed back into Transkribus to complete the training and **increase the accuracy rate**
- Results of the Transkribus software tests very encouraging
- Producing **automatic transcripts for all images**
- Using the Omeka platform to set up a **collaborative correction system**
- Despite their imperfection Automatic transcription usable to perform « **full text** » search in images or texts
- « **Fuzzy** » search : search for keywords (similarity between words and terms the user wants to retrieve)



A textual database that can be verified, documented, enriched and explored.

Transcribing Foucault's handwriting with Transkribus: a report

Results of the Transkribus software tests



Very encouraging! The HTR model can be improved!

Transcribing Foucault's handwriting with Transkribus a report

Marie-Laure Massot, Arianna Sforzini, Vincent Ventresque. *Transcrire les fiches de lecture de Michel Foucault avec le logiciel Transkribus : compte rendu des tests*. June 2018: available on HAL

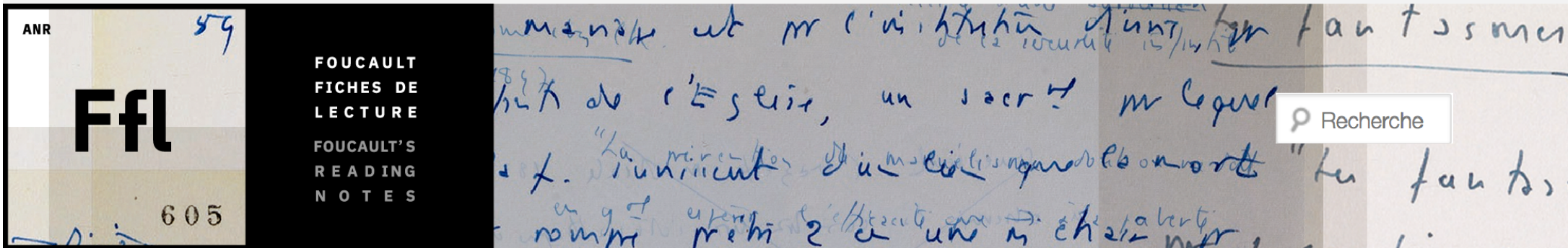
<https://hal.archives-ouvertes.fr/hal-01794139v2>

Coming soon

Transcribing Foucault's handwriting with Transkribus. November 2018.



<https://ffl.hypotheses.org/>



THANKS FOR YOUR ATTENTION!