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)
Transcribing Foucault’s handwriting with Transkribus: a report

How to explore the corpus with digital tools

The Ffl platform provides heterogeneous annotations categories and description fields.
How to explore the corpus with digital tools

A network cartography of reading notes in box n°38 (V. Ventresque).
Transcribing Foucault’s handwriting with Transkribus: a report

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

Manual transcriptions to train Transkribus to decipher Foucault’s handwriting.
Transcribing Foucault’s handwriting with Transkribus: a report

Creating learning data for Foucault’s handwriting

Document segmentation (text regions, lines and base-lines)
and line by line transcription of the manuscript
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

11/6/18

Automatic transcription of Michel Foucault's reading notes by the HTR engine after training.
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.

11/6/18
Transcribing Foucault’s handwriting with Transkribus: a report

Results of the Transkribus software tests

Average success rate of the 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!