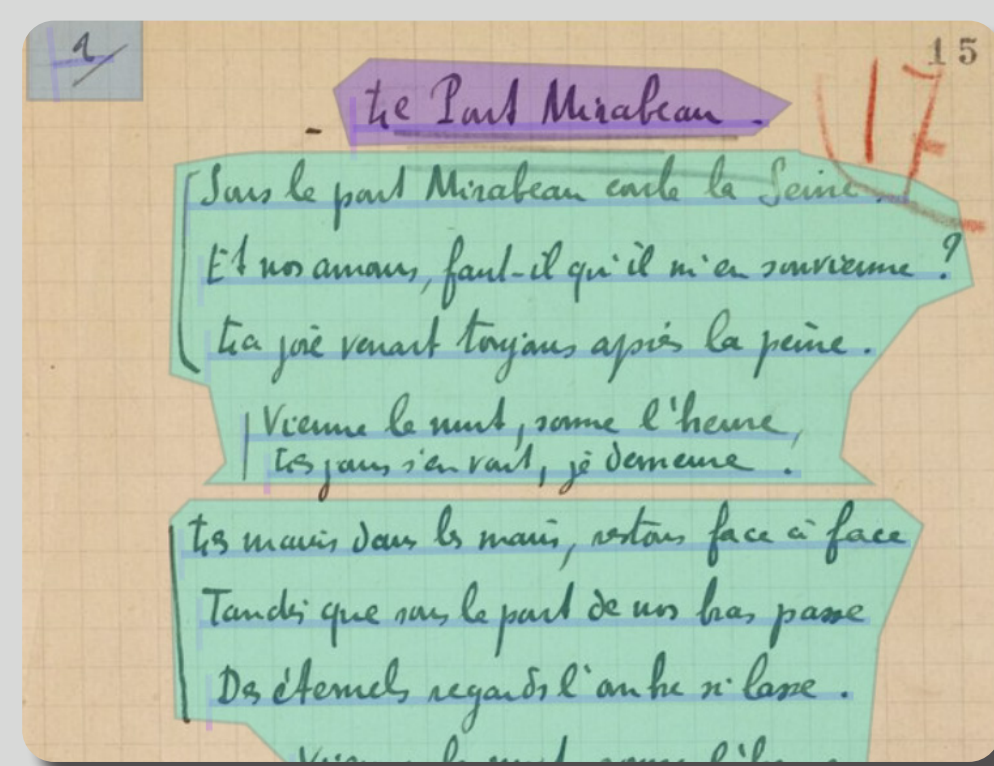
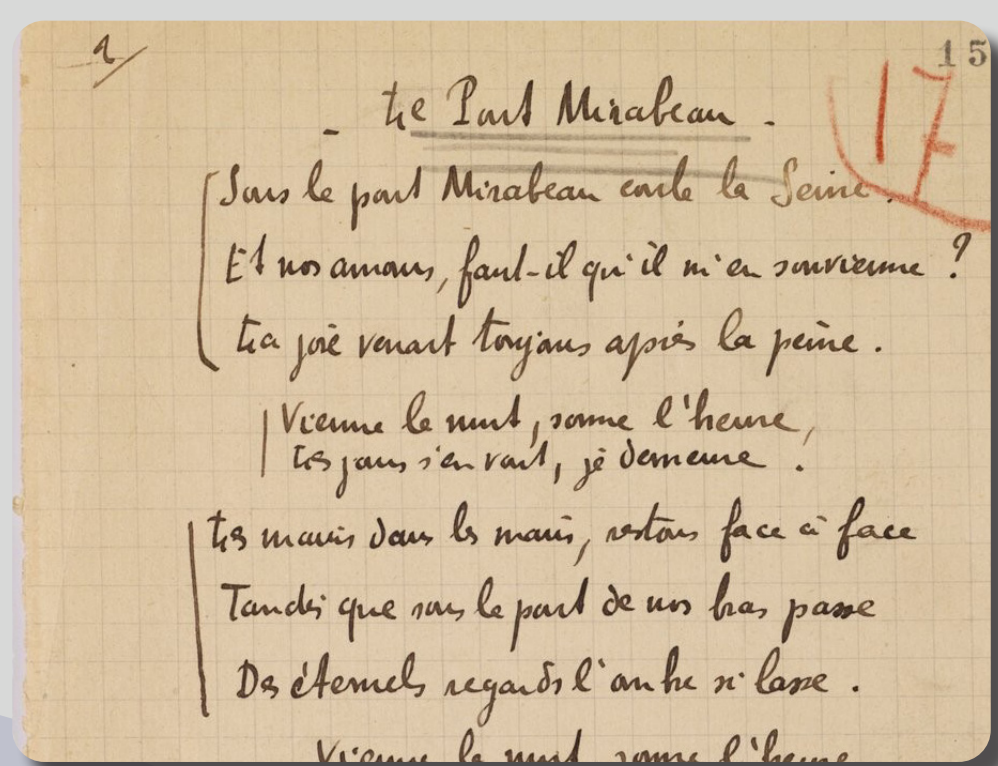




Automatic Text Recognition

Extracting Text From Scanned Images



ATR-Video Tutorials for the Humanities

This project leverages Automatic Text Recognition (ATR) technologies to transform extensive text corpora into searchable transcriptions. We have developed a series of **video and text tutorials** that detail the application of ATR technology, providing step-by-step guidance through the typical ATR workflow and explaining main concepts and terms.

Our aim is to equip researchers, librarians, and archivists with the **knowledge and tools** to efficiently convert images of textual heritage materials into editable text, enhancing their ability to engage with, and utilise these resources effectively.

The tutorials are available on YouTube, on our blog, and on DARIAH Campus in English with subtitles in **English, German, and French**.



Your ATR-Journey starts here!

Textual sources, primarily handwritten, and found in archives or libraries, are critical for humanities research. These sources, whether scanned by researchers or digitised by cultural institutions, need to be accessible in full text to facilitate (computational) analysis. With ATR, learning technologies help create a **transcription of extensive corpora**.

01

Getting Started with ATR

02

Get Images: Where and How to Get Images of Textual Sources?

03

Image Optimisation

04

Layout Analysis

05

Text Recognition and Post-ATR Correction

06

Choosing an End Format and Making Your Project Open and Reusable

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



@harmoniseatr@fedihum.org



<https://campus.dariah.eu/>