



**HAL**  
open science

## eScripta: A New Digital Platform for the Study of Historical Texts and Writing

Benjamin Kiessling, Robin Tissot, Daniel Stökl Ben Ezra, Peter Anthony  
Stokes

► **To cite this version:**

Benjamin Kiessling, Robin Tissot, Daniel Stökl Ben Ezra, Peter Anthony Stokes. eScripta: A New Digital Platform for the Study of Historical Texts and Writing. Digital Humanities 2019, Jul 2019, Utrecht, Netherlands. hal-02310781

**HAL Id: hal-02310781**

**<https://ephe.hal.science/hal-02310781>**

Submitted on 10 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# eScripta: A New Digital Platform for Historical Texts and Writing

B. Kiessling<sup>1,2</sup>, R. Tissot<sup>2</sup>, D. Stökl Ben Ezra<sup>1,2</sup>, P.A. Stokes<sup>1,2</sup>  
École Pratique des Hautes Études<sup>1</sup>, Université Paris Sciences et Lettres (PSL)<sup>2</sup>

## What is Scripta-PSL?

*Scripta-PSL: The History and Practices of Writing* aims at integrating the fundamental sciences that deal with written artefacts (**palaeography, codicology, epigraphy, history of the book, etc.**), with other disciplines in the humanities and social sciences (**linguistics, history, anthropology, etc.**), together with **digital and computational humanities**, around the **study of writing** throughout the **whole of human history**. Languages and writing-systems include:

- Ancient Aramaic
- Medieval Armenian
- Elamite
- Middle Iranian
- Old Javanese
- Meroitic
- Pali
- Sogdian
- Classical Tamil
- Tocharian
- Ugaritic

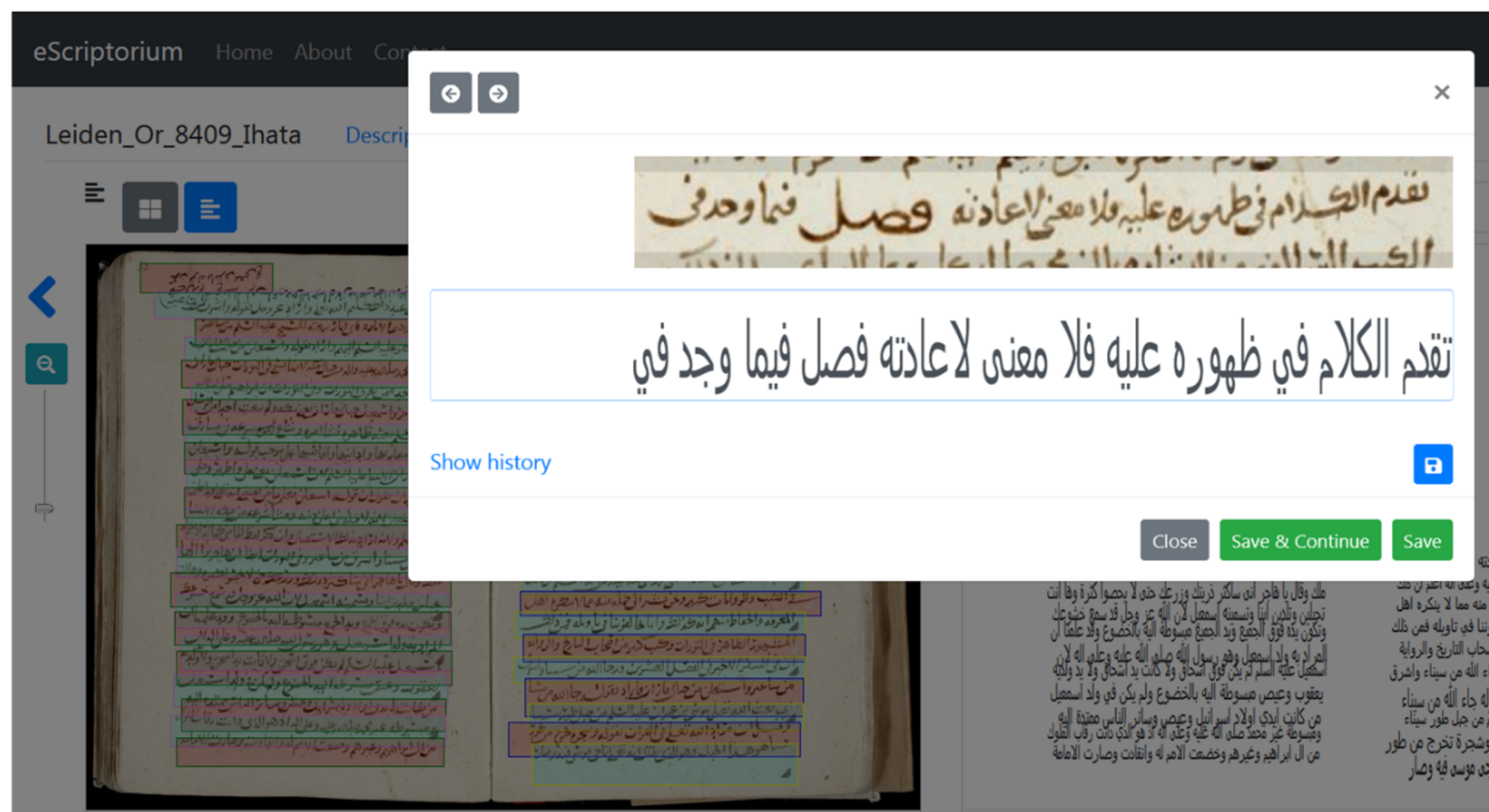
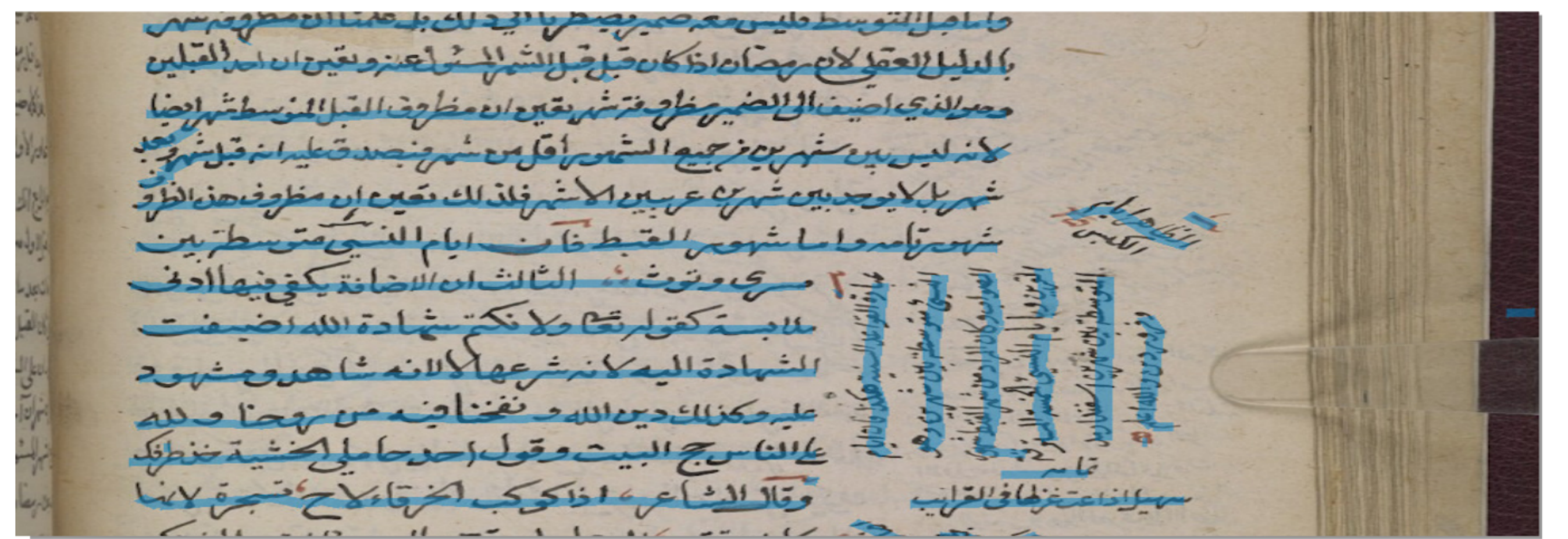
## What is eScripta?

eScripta is the **Digital Humanities component of the Scripta project**, which is developing a workflow for scholars of texts and manuscripts across **all the world's writing-systems**. Among its goals is to **reuse** components already developed in or associated with PSL and elsewhere, including **Kraken, Archetype, Pyrrha** and **TEI Publisher**. It must allow for LTR, RTL, top to bottom LTR, top to bottom RTL and beyond.



## Line Detection with CNNs

Automatic **line detection** of complex layouts is provided using **Convolutional Neural Networks**, and Ground Truth training data can be generated using the interface or separately. Straightforward layouts can be managed by a simpler system which does not require training.

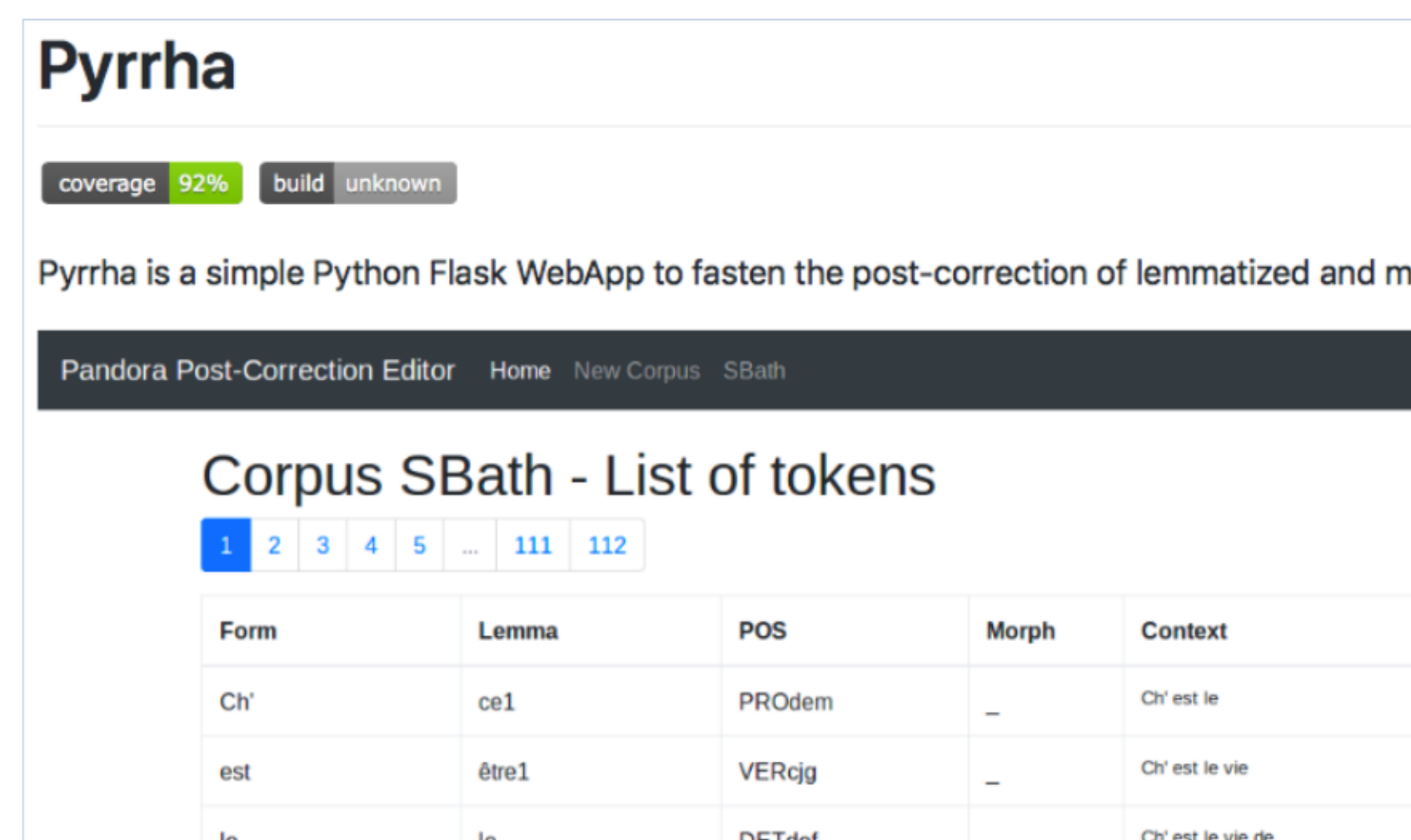
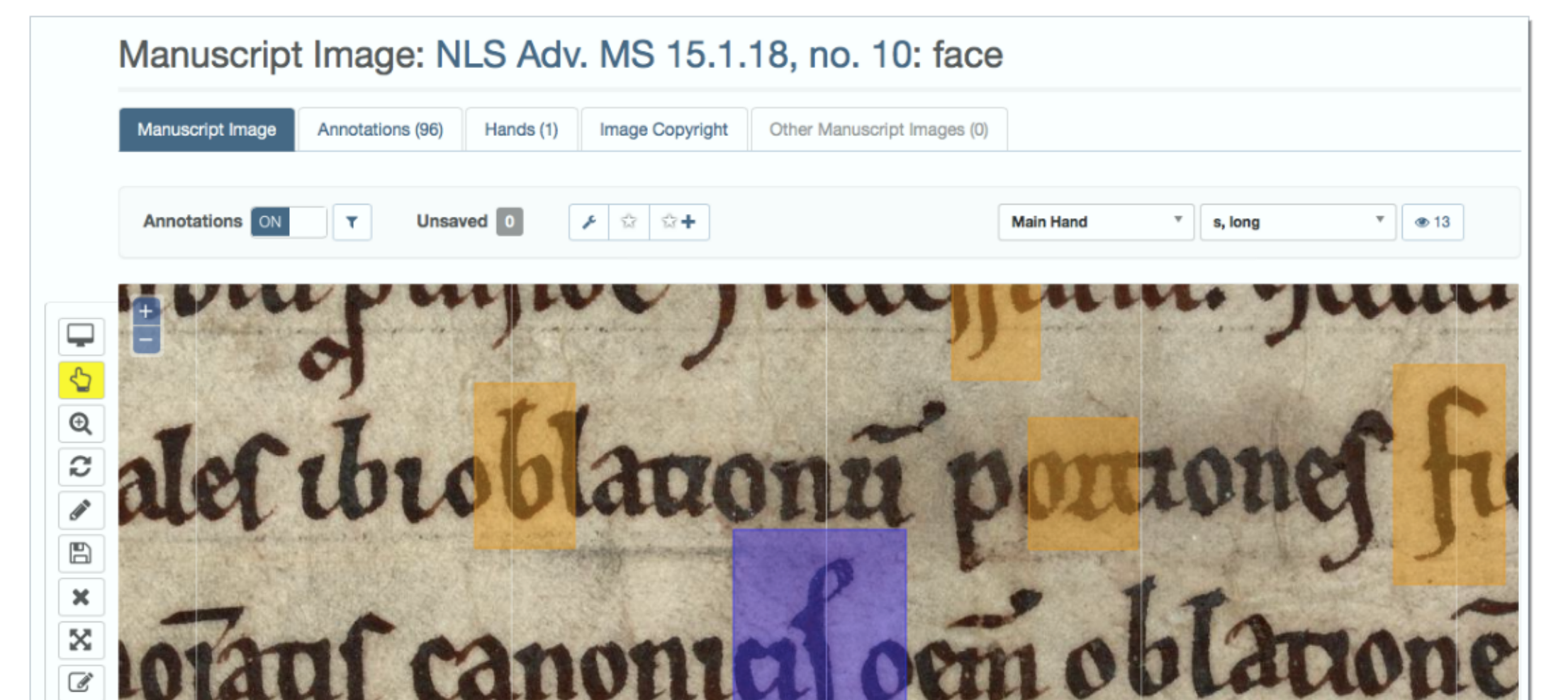


## OCR/HTR with Kraken

Kraken is used for handwritten **text recognition (HTR)** to allow the automatic analysis and transcription of manuscripts as well as printed books. It is designed to accommodate **multi-lingual documents and non-Latin scripts** including non-alphabetic writing-systems and RTL, top to bottom, bottom to top and bidirectional writing. Manual transcription can be entered directly or imported, as a final text or for Ground Truth, and trained models can be imported or exported for sharing and reuse.

## Deep Annotations

The immediate next step is '**deep**' **structured annotations** of images and text. Based on the **Archetype model**, these are not unstructured labels but are instead linked directly **structured project-specific models** of the objects being annotated. This enables scholars to perform powerful **searching**, such as queries for all graphs with ascenders.

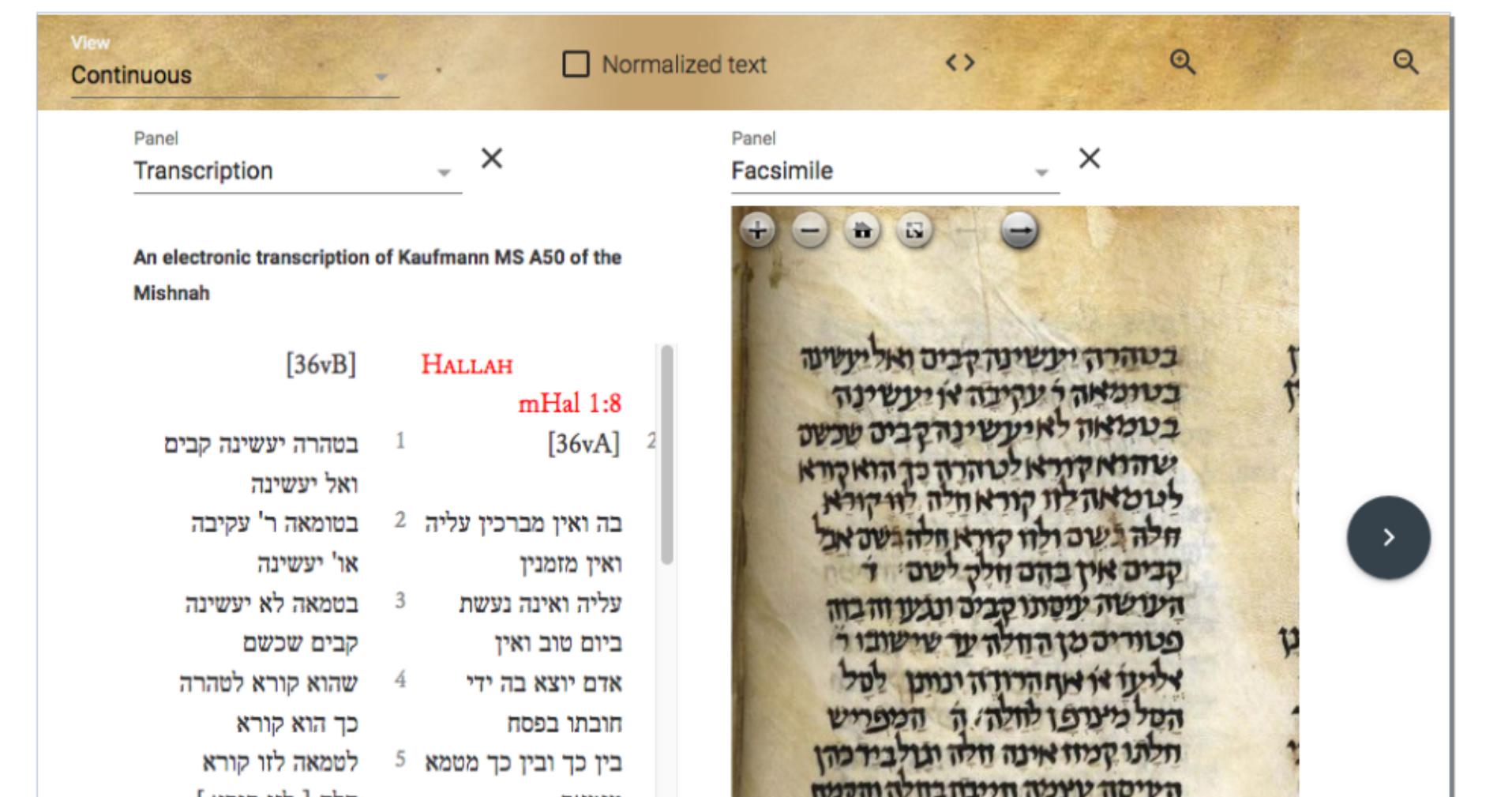


## Post-Correction

Correcting Linguistic Markup with Pyrrha

## Publication

Presenting Texts with TEI Publisher



For updates, visit our blog at [escripta.hypotheses.org](https://escripta.hypotheses.org), or see the code at [gitlab.inria.fr/scripta](https://gitlab.inria.fr/scripta) (MIT Licence)

Scripta-PSL is a Strategic Interdisciplinary Research Initiative (IRIS) of PSL, supported by the École pratique des hautes études (EPHE) and the École française d'extrême-orient (EFE), in association with the École normale supérieure (ENS), the École nationale des chartes (ENC), the École des hautes études en sciences sociales (EHESS), the Collège de France and the Institut de recherche et d'histoire de textes (IRHT, CNRS).

