

HTR for Codex *Genavensis* *Græcus* 44

Ariane Jambé

University of Lausanne

Ὅσοι ἄνθρωποι, τοσαῦται γινῶμαι – Harmonizing Guidelines for Handwritten Text Recognition of Ancient Greek



UNIL | Université de Lausanne



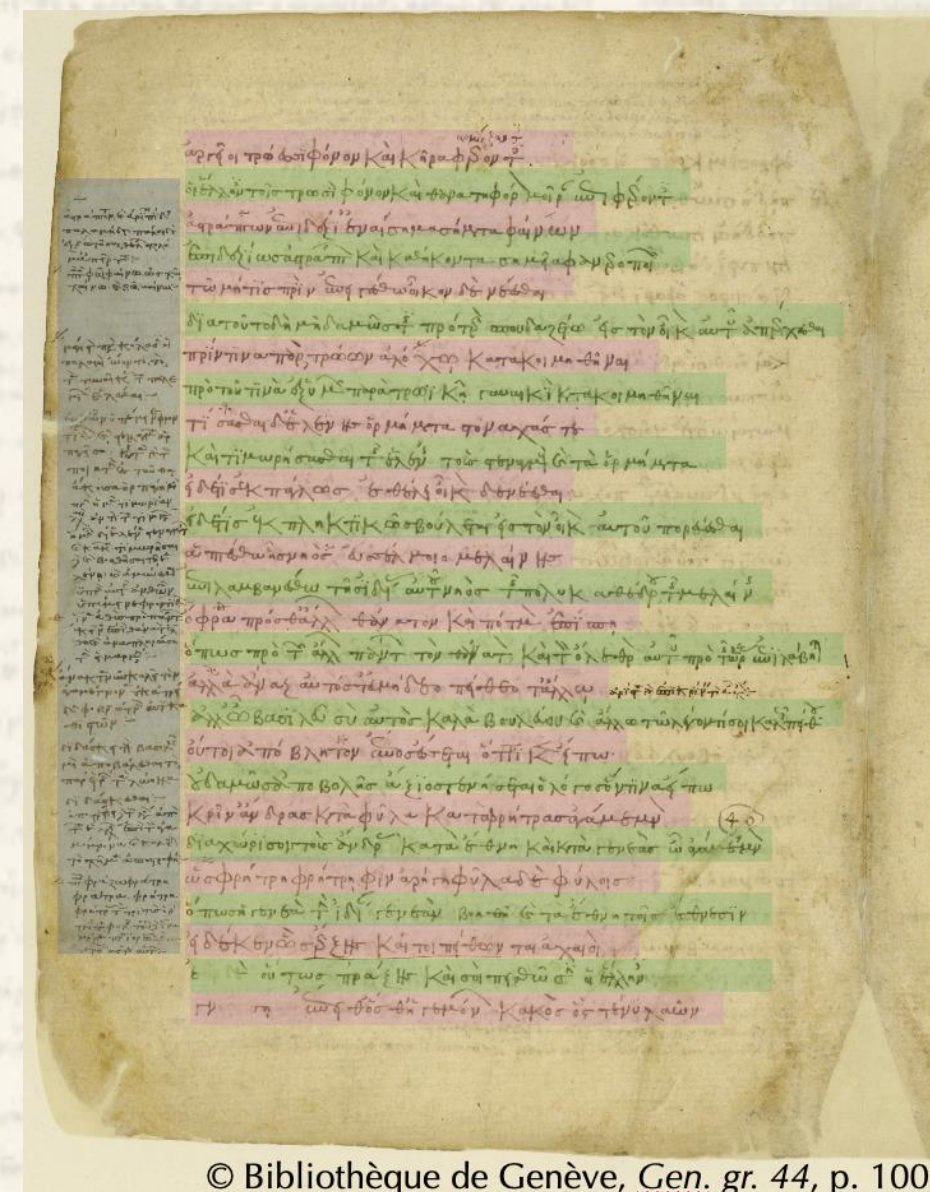
DH2025
ACCESSIBILITY
& CITIZENSHIP



ALLIANCE OF
DIGITAL
HUMANITIES
ORGANIZATIONS

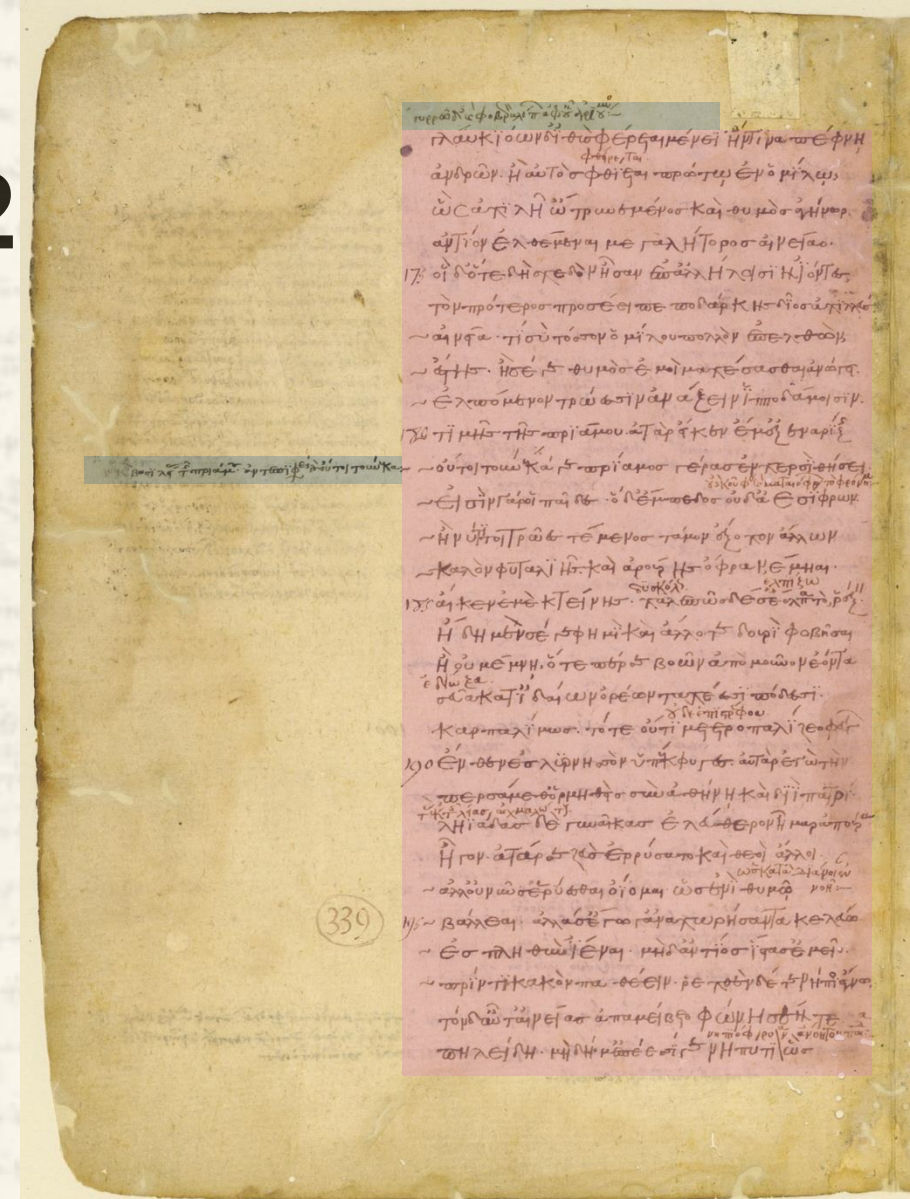
Core content /p. 1 to p. 526

- **Pink:** the *Iliad* (I, 1 to XXIV, 762)
- **Green:** up to XII, 454, each Homeric verse is followed by a **paraphrase** in Greek prose
- **Blue:** the manuscript also includes numerous **marginal scholia** and **interlinear glosses**



© Bibliothèque de Genève, Gen. gr. 44, p. 100

Core content /p. 527 to p. 802



HTR: research context

- **2020 – 2021:** Initial HTR tests with Calfa
- **2023-2024:** Opening of an eScriptorium instance at the University of Lausanne
- **2024-2029:** Post-doctoral project → develop a HTR model for the *Genavensis* (what about the paratext?)

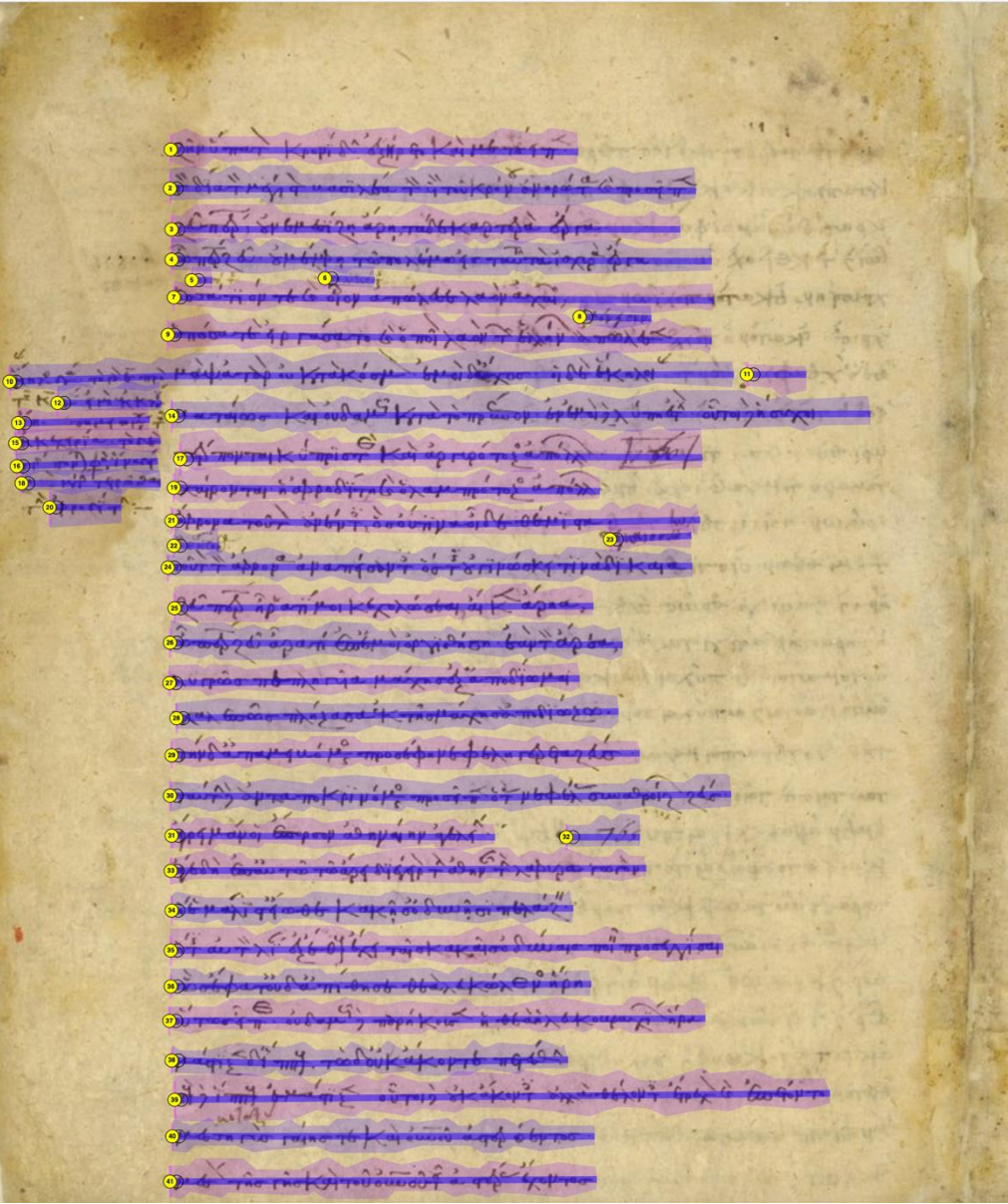
Early phase (*Calfa*)

- Approximately 70 pages processed
- Excellent results (97% average accuracy)
- Methodological bias:
 - resolved abbreviations;
 - *Iliad* and paraphrase = 2 distinct corpora
 - Scholia / paratext not taken into account

Postdoc phase

Methodology:

- eScriptorium model ✗
- Fine-tuning of other(s) model(s) ✗
- Integration of the Calfa-trained model into eScriptorium ✓



Layout: a (non-)issue? HTR point of view

‘Layout Analysis is mandatory to get to the text level, but also to the page level and document level. It allows us to get information about the document itself, namely, for downstream tasks’

Alix Chagué, Hugo Scheithauer, “ATR Step 4 – Layout Analysis”. In: Anne Baillot, Mareike König (ed.): *Automatic Text Recognition. Harmonising ATR Workflows*, 10.05.2024, <https://harmoniseatr.hypotheses.org/1215>.

Layout: a (non-)issue? Codicological POV

‘La trascrizione dei testi commentati [è] un’operazione di elevata complessità tecnica nella quale, data **l’impossibilità di predefinire un modello di «pagina standard»** valido per la totalità del codice, pianificazione e controllo esecutivo si rivelano indissolubilmente associati.’

Marilena Maniaci, *Archeologia del manoscritto: metodi, problemi, bibliografia recente*, Rome, 2002, p. 102-103.

‘Lehrbücher, die über Generationen hinweg von Hand zu Hand gingen und ihrem Zweck entsprechend unbefangen gebraucht (und verbraucht) wurden, sind am Ende nicht mehr aus einem Guß. (...) Die individuellen Aufschwellungen bilden die Formen eher um, zerstören hier und da das Gleichmaß eines bewußt regulierten Textbildes.’

Gehrardt Powitz, “Textus cum commento”, *Codices manuscripti*, no 5, 1979, p. 87.

To keep in mind for the Guidelines...

Standardization vs. actual use of a manuscript

- How do we define "layout"?
- What qualifies as "marginalia"?
- Can we apply a controlled vocabulary?

p. 184 (with paraphrase)

p. 297 (with paraphrase)

- A. the numbering of the quire (Gen I, 13th century)
- B. the alphanumeric numbering of the Book of the *Iliad* (Henri II Estienne, 16th century)
- C. the numbering referring to the corresponding page of the 1541 edition of the *Iliad* printed in Basel by Johannes Herwagen (Henri II Estienne, 16th century)
- D. the numbering of the verses of the *Iliad* (Henri II Estienne, 16th century)
- E. page numbering of the manuscript (19th century)
- F. the numbering of the Book in Arabic numerals (date unknown)

