Crowd-sourcing papyri fragments: The ANCIENT LIVES Project

Ancient Lives is a citizen science project inviting any volunteer to help in the transcription of papyri fragments written in Greek and coming from the city of Oxyrhynchus in Middle Egypt. This collection, owned by the Egypt Exploration Society and kept in Oxford, was gathered during several excavations in the turn of the 20th century and contains a wide variety of texts, pieces of literature as well as day-to-day documents, covering a whole millennium from the Ptolemaic period to after the Arab conquest of Egypt.

The user interface: ancientlives.org

Digital images of papyrus fragments are uploaded to the Web for anyone to transcribe. Volunteer users identify letters through shape recognition: they click letter shapes on a fragment and use the online keyboard found lower on the screen to identify the character. They can leave comments or questions on the papyrus they are transcribing by clicking on “Talk.” They can also sign in and find back their transcribed papyri in the “Lightbox.”

The consensus output

User clicks passed through character-positioning algorithm

Character consensus output

A section of the papyrus and its consensus transcription

User input is compared against expert input (including publications) as test run for algorithms.

The consensus transcription is curated by experts and University students trained in the field of Papyrology for statistics on transcription errors.

Next steps

A relaunch is in preparation for the end of 2015, including a Coptic side which will welcome pictures of texts from several collections (parchments from the White Monastery, ostraca from the Kelsey Museum, Michigan...)

The recorded shapes of individual letters invite to work on scribe handwriting recognition, in order to piece together fragments from the same text, codes, scribe, school of scribes. A starting point would be to try to put together the numerous fragments from Homer.

Pattern searches for unknown literature and documentary texts.

Some figures

Launched in July 2011.

More than 200 texts identified in the first year based on user data, among which Homer, Plutarch, Plato, Demosthenes, Thucydides, Polybius, Epictetus, Simonides, Theognis, Gospels (some unknown).

By the end of 2014, 9,288,620 characters clicked, 151,087 fragments examined.

The consensus output

User clicks passed through character-positioning algorithm

Character consensus output

A section of the papyrus and its consensus transcription

User input is compared against expert input (including publications) as test run for algorithms.

The consensus transcription is curated by experts and University students trained in the field of Papyrology for statistics on transcription errors.

Next steps

A relaunch is in preparation for the end of 2015, including a Coptic side which will welcome pictures of texts from several collections (parchments from the White Monastery, ostraca from the Kelsey Museum, Michigan...)

The recorded shapes of individual letters invite to work on scribe handwriting recognition, in order to piece together fragments from the same text, codes, scribe, school of scribes. A starting point would be to try to put together the numerous fragments from Homer.

Pattern searches for unknown literature and documentary texts.

Some figures

Launched in July 2011.

More than 200 texts identified in the first year based on user data, among which Homer, Plutarch, Plato, Demosthenes, Thucydides, Polybius, Epictetus, Simonides, Theognis, Gospels (some unknown).

By the end of 2014, 9,288,620 characters clicked, 151,087 fragments examined.

The Ancient Lives Team

Collaboration across campuses

Papyrologists and experts in Classical and early Christian literature and history

University of Minnesota: Lucy Fortson (Physics, Zooniverse), Philip Sello and Nita Krevans (Classical & NE Studies), Anne-Francoise Lamblin, HaoYu Yu, and Trevor Wennblom (Bioinformatics)

University of Oxford: Dirk Obbink, James Bruselas (Classics)

U of M postdoctoral fellows: Marco Pereale, James Bruselas (Minneapolis Futures); Geoffrey Smith, Isabelle Marthot (NEH)

Students CLA: Theresa Charesan (UROP), Rachael Cullick, Ryan Seaberg, Jessica Shao, Karl Moenschbacher, Alyssa Ramden (UROP), Kalyssa Tatroe (UROP), Alyssa Cady (GRPP)

CSE: Jared Matzke.

CEHD: Miranda Pihlaja

CUNY Graduate Center: Maura Williams.

Middle Tennessee State: Alex Williams

To go further
