Hugh McElroy
xyxlosphongium@gmail.com
www.palaeographer.org

Ut silicis venis abstrusum excuderet ignem

A workflow of free tools to create student editions of Latin texts from scanned PDFs
(SCS Annual Meeting 2023 - Ancient Makerspaces)

There is a rich vein of Latin that remains to be mined in scanned PDFs of early modern Latin editions. The pedagogical potential of these texts is huge, catering as they do to interests (in the fantastical, in a plot-driven romance, in science) our students already have. Who wouldn't want to read a natural history of dragons or a 16th century verse drama based on Helidorus' *Aithiopika*? If those don't appeal, there are myriad works of history, science, geography, poetry, and drama waiting online.

One fairly simple, easily and freely accessible workflow to transform these texts, or selections from them, into editable text with vocabulary and reading help for students follows.

a. Identify a text you want to read with students (promising sources I've browsed include the Library of Congress and other national libraries, the Internet Archive, Google Books, the Anarchist Nubiology Squad Library)
b. Optical character recognition from scanned PDF (Rescribe)
    i. Human correction of OCR output (your favorite text editing software; a second screen -or tablet on a stand- is very useful)
        1. [find and replace] for
            a. historical characters (long s, digraphs)
            b. resolve abbreviations (and all the ways the OCR has interpreted characters indicating *-que*)
            c. unnecessary or obsolete punctuation and diacritical marks
            d. miscellaneous patterns of repeated error
        2. manual correction
c. Lemmatization (Bridge Lemmatizer)
    i. Generate lemmatization sheet
        1. edit in your favorite spreasdsheet software
        2. manual completion and correction (Google Sheets can be slow with large data sets like the Bridge Dictionary sheet)

3. Logeion is a great resource for filling in lemmata missing from the dictionary sheet

d. Macrons, if you're into that kind of thing (Johan Winge's Macronizer)
    i. Logeion is a great resource, especially the LaNe entries, for checking ambiguities
e. Putting it all together
    i. Decide what vocabulary to include
        1. DCC Core/Other Vocab Lists (a number are available at the Bridge)
    ii. Choose a format
    iii. Format text, vocabulary, and other reading help
    iv. Print or put online for students

URLS for print handout

OCR
www.rescribe.xyz

Lemmatization
https://bridge.haverford.edu/lemmatizer/
https://logeion.uchicago.edu/

Macrons
https://alatius.com/macronizer/
https://logeion.uchicago.edu/

Commentary Sandbox (one option for online formatting)
https://iris.haverford.edu/sandbox/

Anarchist Nubiology Squad Library
http://suntuwekane.memoryoftheworld.org/