

# From Annotated Text to Vocabulary Exercises

Sunoikisis Digital Classics, Summer 2019

---

Thursday June 13, 17:00 - 18:15 CEST

Neven Jovanović, Petar Soldo

University of Zagreb

Faculty of Humanities and Social Sciences

Department of Classical Philology

[github.com/SunoikisisDC/SunoikisisDC-2018-2019/wiki/Summer2019-Session11](https://github.com/SunoikisisDC/SunoikisisDC-2018-2019/wiki/Summer2019-Session11)

# The plan

---

# The plan

The task

Methods, materials, tools

Tools

Results: creating the database

Results: creating the exercises

Results: running the scripts

Discussion

## The task

---

## The background

A Greek reader for the first-year students: 60 short texts, commented and cued to the standard school grammar.

The students should master the vocabulary from these texts.

## The task

Annotate our selection of Greek texts morphologically,  
lemmatize all words.

Use that set of texts to (semi-automatically) produce vocabulary  
exercises.

## Methods, materials, tools

---

## Methods: linguistic annotation

Lemmatize words in a text

Add morphological descriptions of lemmatized words (parts of speech)



<word id="3" form="παῖς" lemma="παῖς" postag="n-s---mn-"/>

<word id="4" form="ὦν" lemma="εἰμί" postag="v-sppamn-"/>

<word id="5" form="ὁμολογεῖται" lemma="ὁμολογέω" postag="v3spie---"/>

## Methods: XML manipulation

**Reorder** words by form, lemma, part of speech

**Regroup** words by forms, lemmata, parts of speech (a group is a set of identical forms, identical lemmata etc; from it we get information about frequency of an item in the collection)

Analyze frequencies, retrieve reoccurring forms, lemmata, parts of speech

Select certain forms, lemmata, parts of speech

## Methods: XML manipulation, common tasks

Retrieve all forms from a text (but not punctuation)

For a form, retrieve lemma:  $\pi\rho\acute{\alpha}\xi\epsilon\iota\varsigma \rightarrow \pi\rho\hat{\alpha}\xi\iota\varsigma$

For a form, retrieve part of speech information:

$\pi\rho\acute{\alpha}\xi\epsilon\iota\varsigma \rightarrow \text{n} - \text{p} - - - \text{fa} -$

## Methods: XML manipulation, common tasks

For a lemma, find whether it occurs in other texts, or in another database (DC Greek Core)

For a lemma, retrieve form in a text

For a lemma, retrieve (Croatian) meaning from a lexical database (DC Greek Core)

## Methods: XML manipulation, common tasks

Format results for import into Anki, our vocabulary learning program (a simple three-fields CSV is needed)

**Spaced repetition:** we tend to remember things more effectively if we spread reviews out over time, instead of studying multiple times in one session.

## Methods: Vocabulary learning

### Make your own cards!

The act of making the card helps you remember the stuff on it.

The way you learn and the things you care about are unique to you.

## Methods: Sharing research

To distribute our work, we used a popular version control system and a popular software development platform.



## Methods: Sharing research

A **version control system** keeps record of changes to code and makes the history of these changes visible to everybody. This enables control and attribution (who did what, when?), corrections and changes of course (get back to an earlier version!).

A version control system implemented **as a software development platform** accessible over the internet enables people to share code, to work together on it from different regions and time zones, to fix bugs, propose improvements etc. etc.

This mode of operation makes programmers similar to scholars and scientists (although programmers work faster!). And scholars and scientists are aware of that.

This mode of operation also enables **replicability**, often left implicit in the humanities (“everybody who reads the book will, if clever enough, reach the same conclusions as I”).

# Tools

---

*Arethusa* / *Perseids*

# Tools: XML manipulation

BaseX / XQuery

## Tools: spaced repetition

Anki

## Tools: frequent words in (classical) Greek and Latin

### Dickinson College Core Vocabulary

“The DCC Core Vocabulary lists represent the thousand most common words in Latin and the 500 most common words in ancient Greek. They were originally composed in 2012–13 by a team at Dickinson College led by **Christopher Francese**. The lists are downloadable in various formats and licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.”



## Tools: version control, publication, preservation

Github / Zenodo (DOI [10.5281/zenodo.3244011](https://doi.org/10.5281/zenodo.3244011))

Results: creating the database

---



Results: creating the exercises

---

1. For a word form, give the lemma (standard vocabulary entry):

Q πράξεις – A πράξις

2. For a word form, give the grammatical description:

Q *πράξεις* – A noun, feminine, plural, nominative / accusative /  
vocative

3. For a lemma (of a word from the DC Greek core list), give the translation:

Q ποταμός – A rijeka

Results: running the scripts

---





## Discussion

---



# From Annotated Text to Vocabulary Exercises

Sunoikisis Digital Classics, Summer 2019

---

Thursday June 13, 17:00 - 18:15 CEST

Neven Jovanović, Petar Soldo

University of Zagreb

Faculty of Humanities and Social Sciences

Department of Classical Philology

[github.com/SunoikisisDC/SunoikisisDC-2018-2019/wiki/Summer2019-Session11](https://github.com/SunoikisisDC/SunoikisisDC-2018-2019/wiki/Summer2019-Session11)