

A TEXT ANNOTATION BUILDING BLOCK

Introduction

Text annotation is crucial in many digital humanities research projects, particularly for ancient text analysis and language studies. It can facilitate analysis of ambiguous language, cultural contexts, evolving grammar in historical documents, and link texts to historical entities (events, places, persons).

Easy-to-use environments for text annotation allows scholars to share and scale-up their work. This facilitates large-scale computational analysis. For example, a text annotation environment can help scholars uncover linguistic patterns and document the evolution of ancient languages.

At GhentCDH we developed a reusable text annotation building block. It is developed with several projects in mind but can be put to use in many cases.

Needs

A general text annotation environment has a couple of requirements. Below these are grouped around visualisation, creation, adaptation and sanitation.

Visualisation

- Visualising short annotations in the text
- Visualising longer, structural annotations in the margin
- Allow overlapping annotations
- Quickly filtering annotations of interest

Creation

- Creating new annotations
- Adding meta-data to annotations
- Linking annotations to other texts or annotations

Adaptation

- Changing existing annotations via an easy-to-use UX
- Moving annotations around
- Removing annotations

Sanitation

- Flexible application dependent rules
- Quickly fix or remove invalid annotations

A text annotation building block

To address the common needs above, GhentCDH developed a reusable, flexible text annotation building block for use in web-applications.

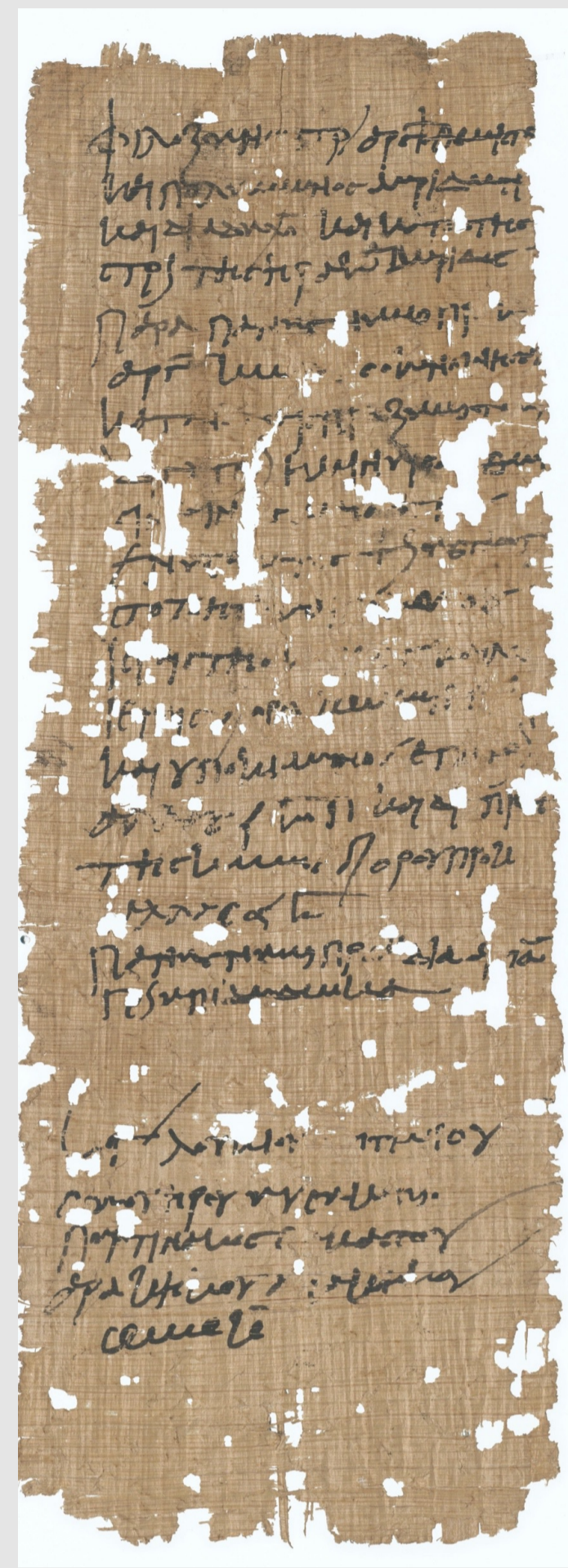
The building block is a piece of software, a software component. The component needs to be built into an application. The application – not the component – determines how the annotations look, which actions on annotations are allowed and which rules are applied to the annotations: e.g. if each annotation must snap to word-boundaries or not. It is designed with transliteration workflows in mind: there is support to show line numbers and structural annotations.

Tech Talk

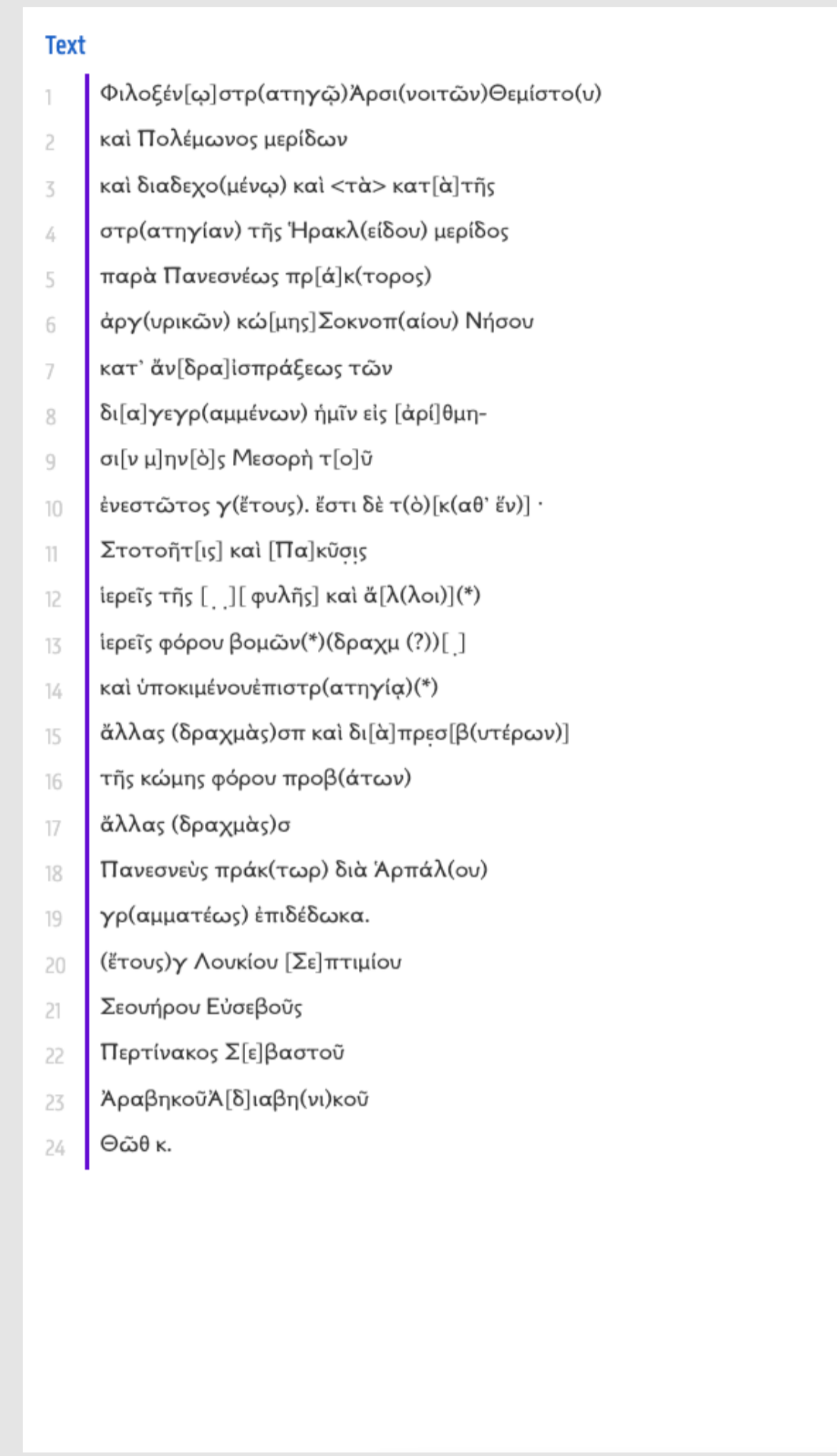
The source code is available on Github and is published under the name `vue_component_annotated_text`. Technical info:

- Programmed in **Typescript**: typed Javascript prevents some programming errors.
- It is a **Vue 3** component: a reactive framework enables a snappy UX.
- The component has flexible styling via CSS variables or styles but offers sensible defaults.
- Rules and actions are made possible **via slots and templates**. This allows flexibility in both UI as behaviour.

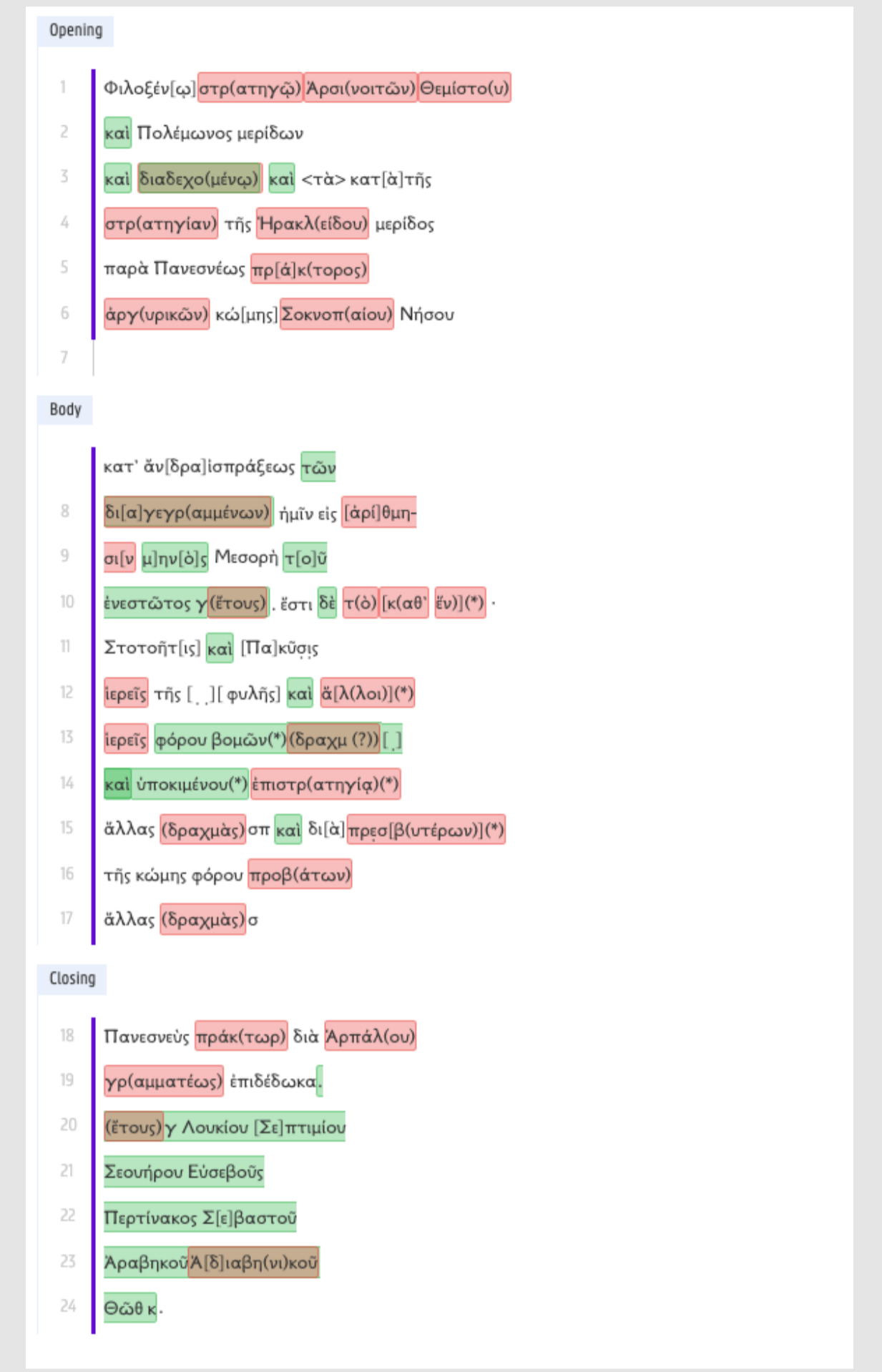
From source to understanding



a



b



c

Fig: A Greek text originating from a papyrus source (a), the same text transliterated and digitized (b), and (c) the text with several types of annotations. It shows overlapping annotations, structural annotations and different types of annotations (identified by color). The example shows text 34 from the EVWRIT data set.

Demonstration

The demo application shows a text annotation correction process. The aim is to quickly fix many annotations either automatically or flag annotations to manually check and fix.

The demo application features:

- Annotation type dependent rules to automatically suggest fixes to annotations.
- Application specific annotation visualisation
- A way to list automatically fixed annotations
- A way to quickly verify applied changes

The demo application was developed in context of the EVWRIT project.

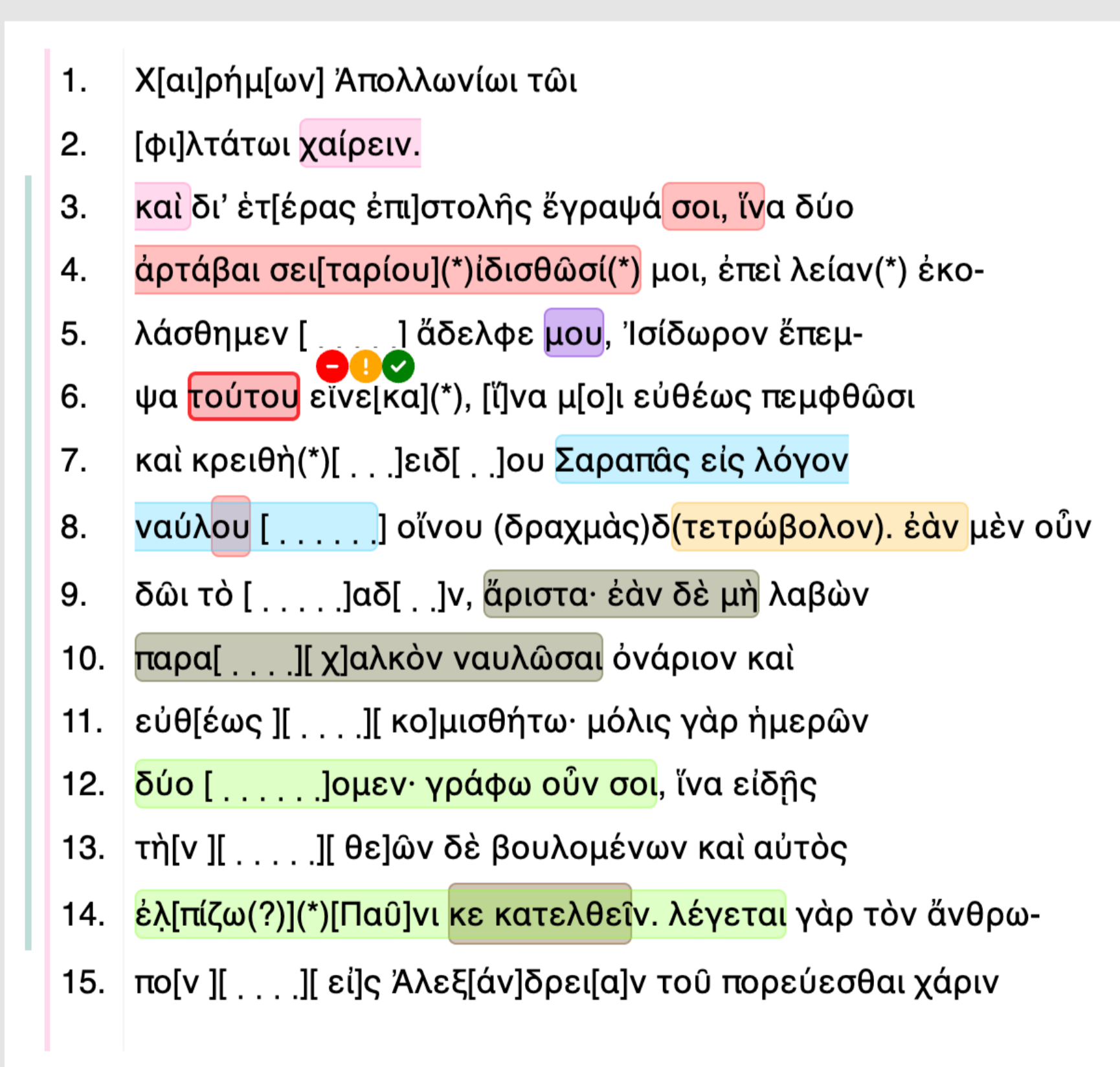


Fig: This demo application show support for visualisation of different annotation types, selection of annotations, and a modification workflow - see the icons next to the annotation.

Alternatives

Inception

- A collaborative web-based environment for text annotation
- UX is problematic
- Flexibility is limited
- Ready-to-use environment

Recogito JS Text Annotator

- Component still in development
- No documentation
- No concept of line numbers: digital text annotation vs texts linked to physical documents

Conclusions

Does your research project need a text annotation environment? Check out the GhentCDH services and contact us.

Thanks to

The text annotation work is supported and used by the following projects:

- ANNOPHIS
- EVWRIT
- MELA

And many more projects to come. Also supported by the FWO international research infrastructure grant CLARIAH-VL+.

Contact

Joren Six - joren.six@ugent.be

ghentcdh.ugent.be

GhentCDH

GhentCDH