Using NLP to Create Corpus-based Vocabulary Exercises in Latin Classes

Machina Callida

Andrea Beyer Konstantin Schulz

Humboldt Universität zu Berlin

March 3rd, 2020

Funded by

DFG Deutsche Forschungsgemeinschaft

German Research Foundation

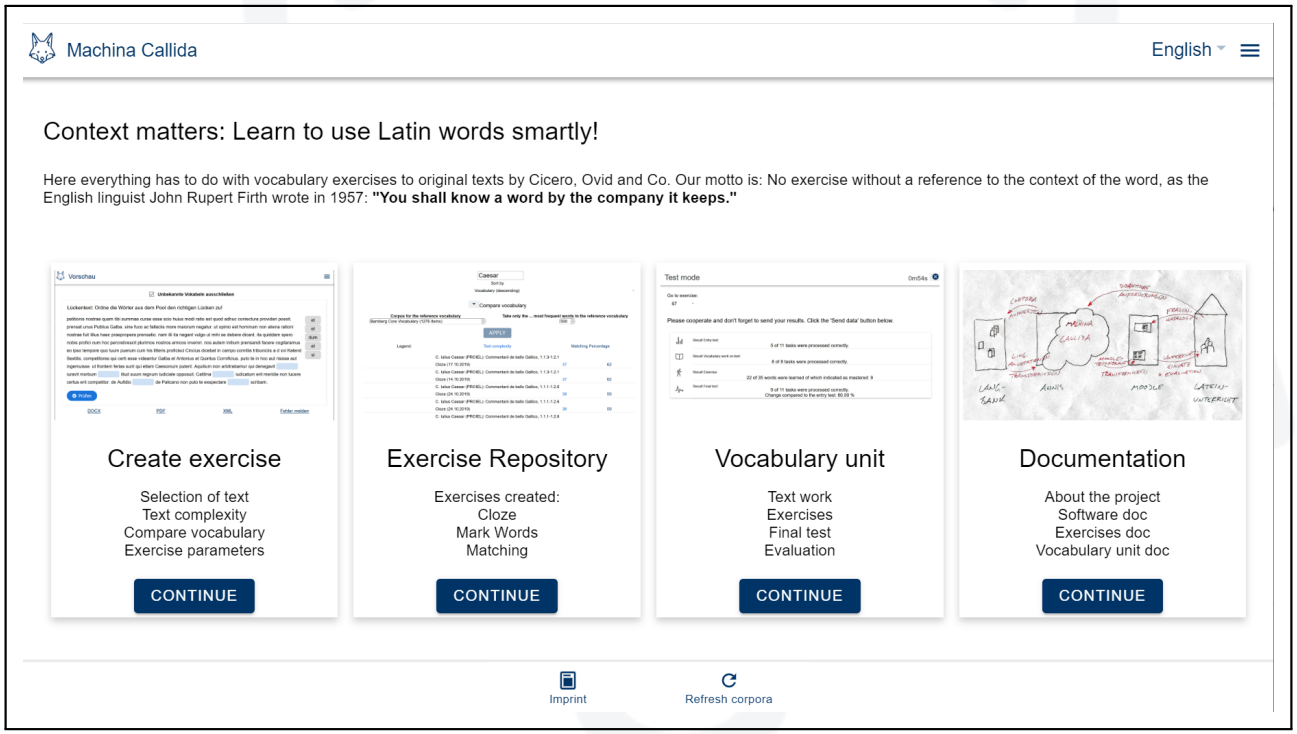
Outline

Scope

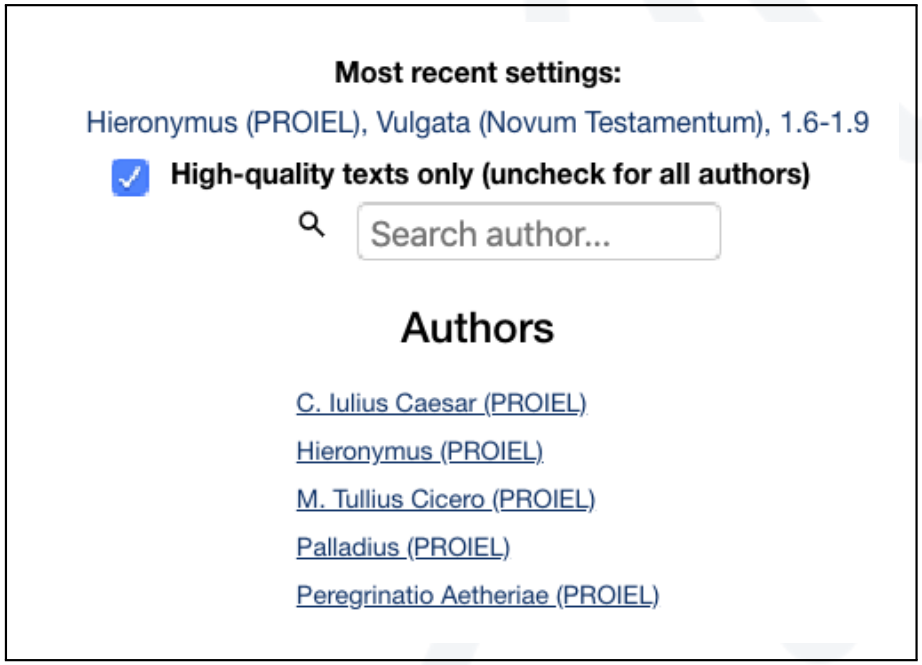
- Language: Latin
- Institution: high school
- Location: Berlin
- Target audience: teachers and intermediate learners
- Context: vocabulary training
- Problem so far:
  - Infrequent training (homework)
  - Inefficient methods (rote and cross-lingual learning, single words without context)
  - Decoupled from syntax and morphology

Solution: Vocabulary Software

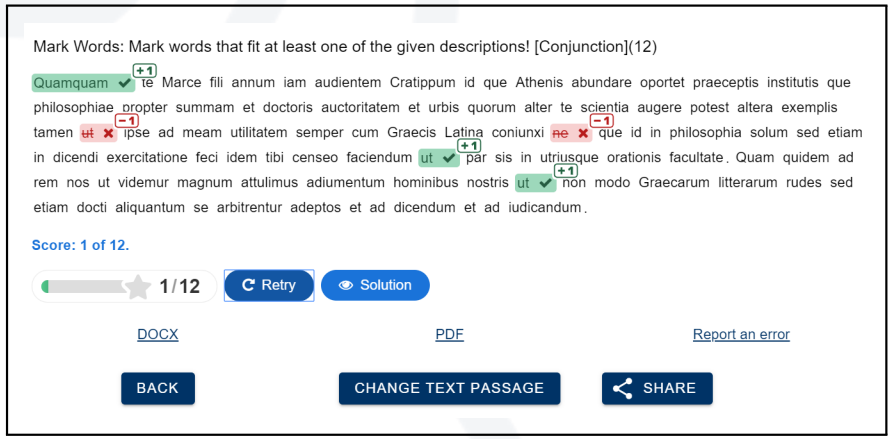
- The software is accessible at all times, on all platforms.
- It offers authentic contexts, e.g. phrases or small texts.
- It also offers ready-made exercises for instant usage.
- It supports the creating of personalized exercises.



Authentic and context-based exercises

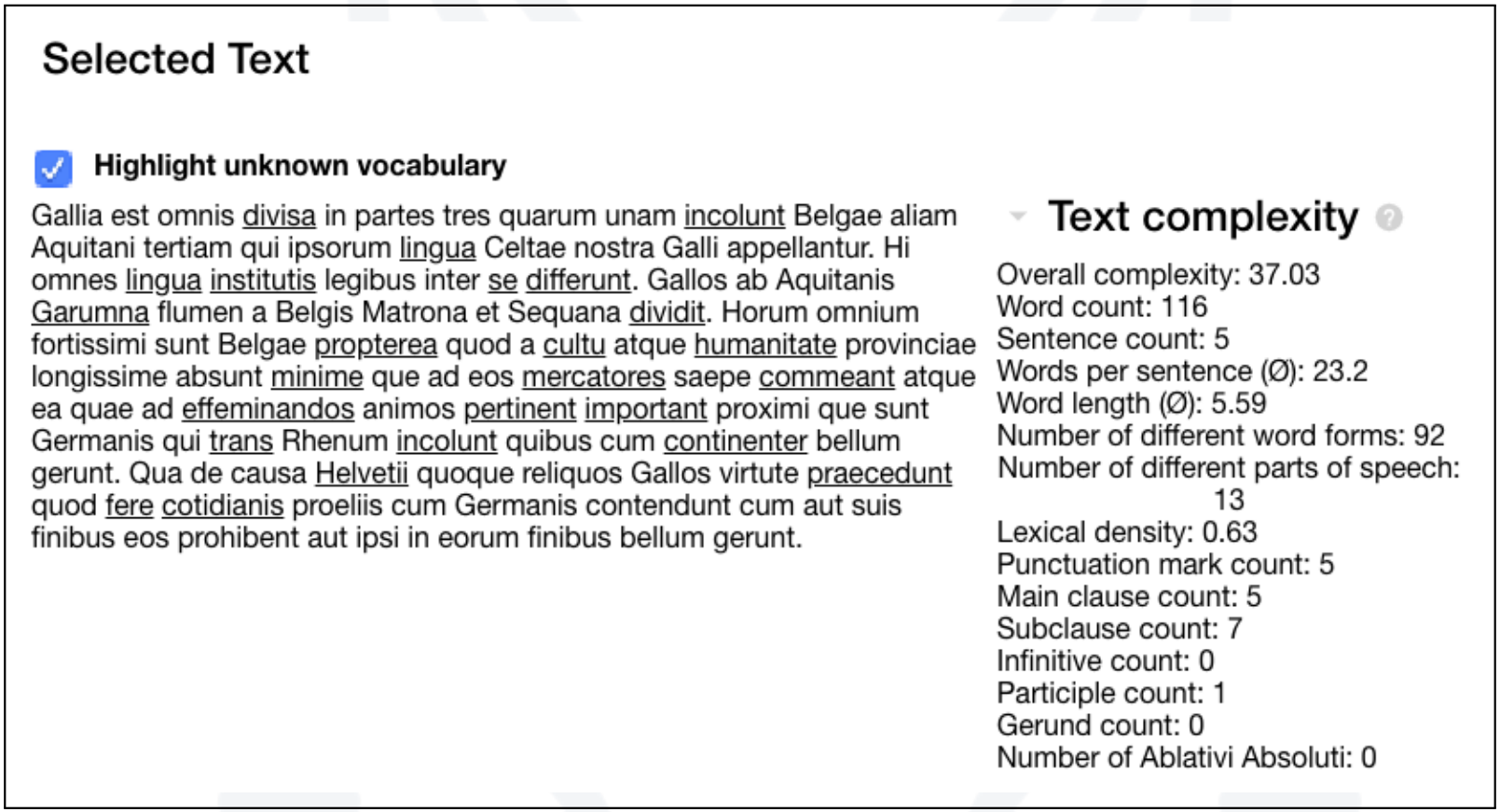


- Exercises to be selected from more than 100 Latin authors.
- The corpus is of varying quality depending on the annotations (manual vs. automatic).



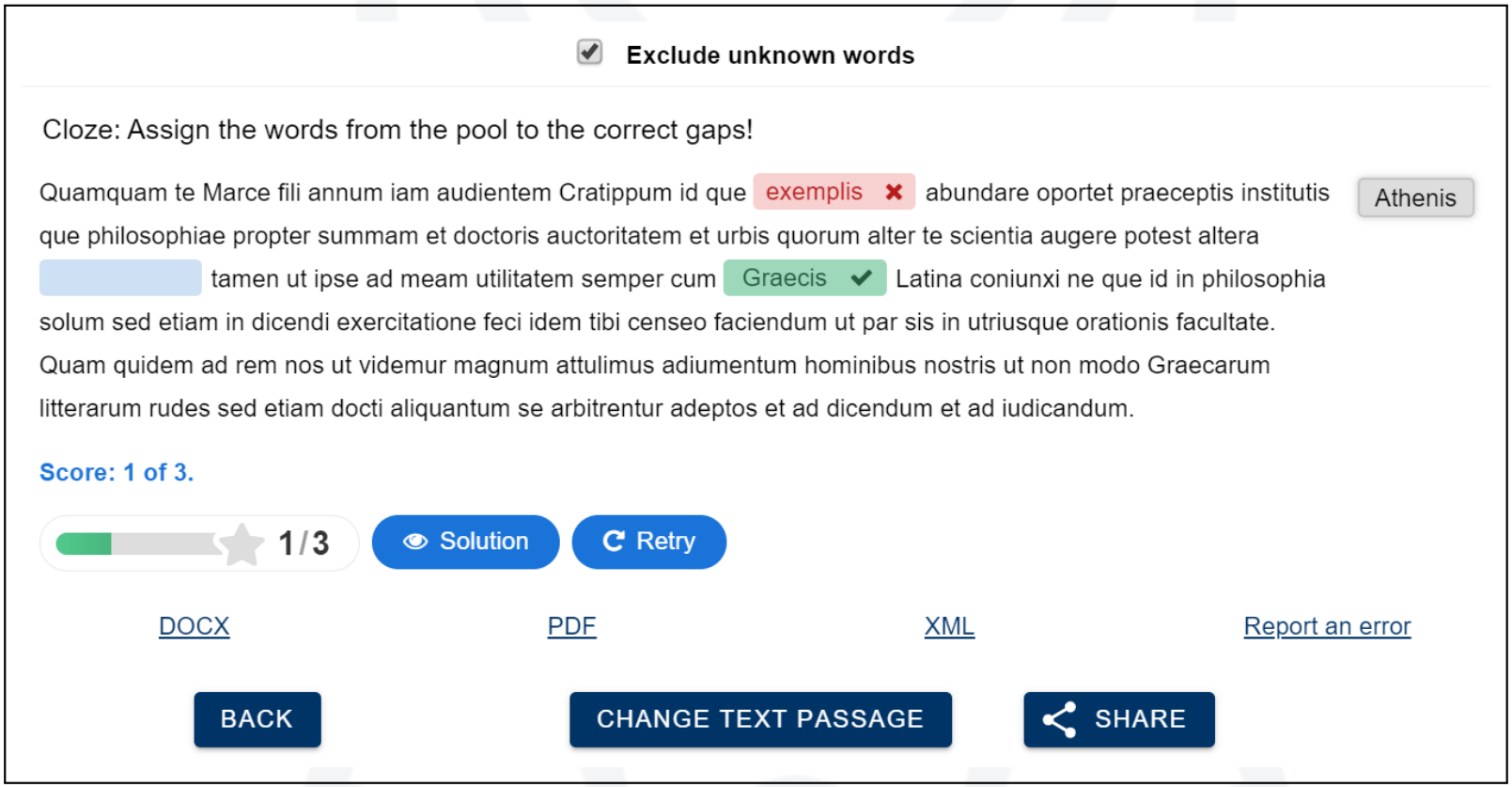
- The exercises work intra-lingually, e.g. cloze.
- The exercises focus on form, function or meaning, e.g. mark words.

Preparing the exercises



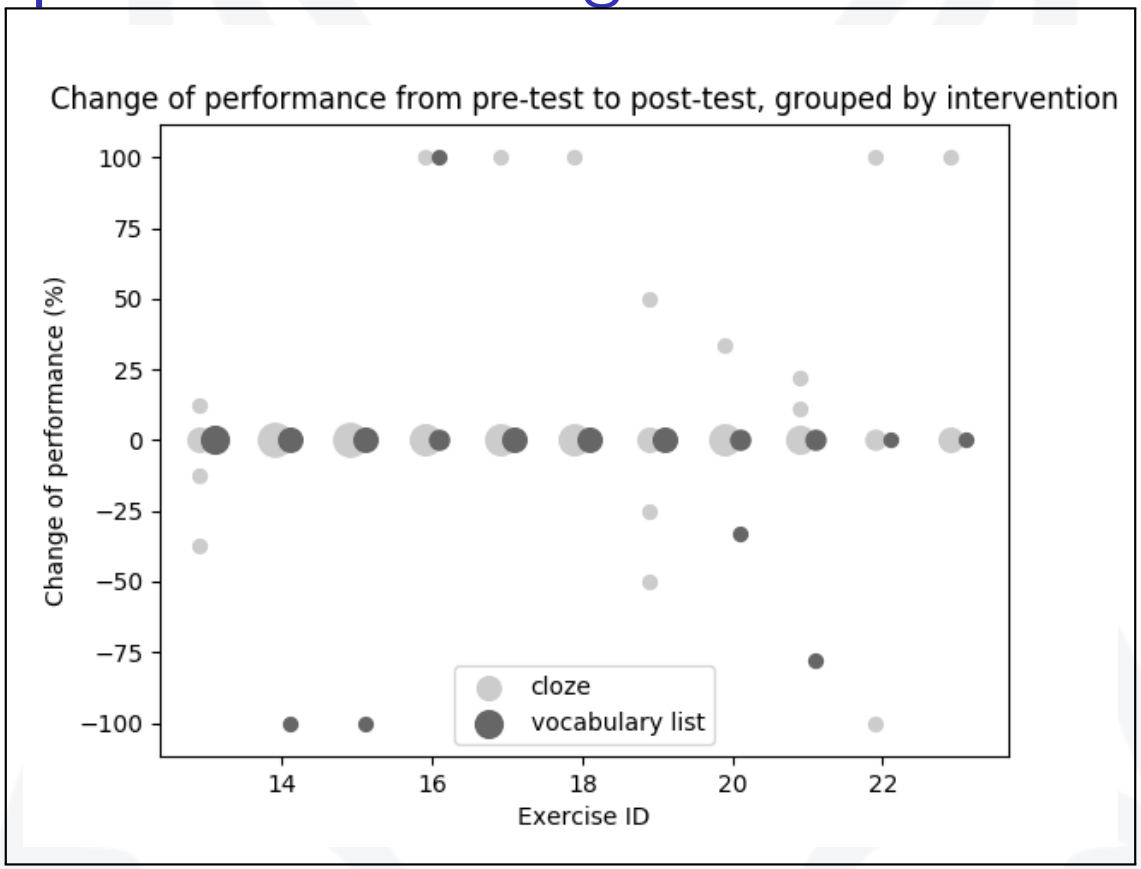
- You can compare the text to your personal vocabulary.
- You can assess the linguistic complexity of the chosen text.
- You can set different parameters: interaction type, linguistic phenomena, instructions.

Working with the exercises



- You get an automatic feedback (and scores).
- You can share a link or use exercises stored in the repository.

Study on corpus-based learning: Cloze vs. word list



Comparison of students' performance before and after interventions. Larger markers indicate multiple students with the same percental change. We may conclude carefully that solving cloze exercises is correlated with improvements in vocabulary competence (more light grey dots above the baseline), as opposed to traditional rote learning.

Conclusions

1. Performance improvements are not restricted to just matters of word choice or the translation of single words. Instead, they even relate to more complex tasks like reading comprehension or translation of multiword expressions.
2. Therefore, lexical competence is to be understood in a broad sense, i.e. including certain aspects of syntax and morphology.
3. Finally, corpus-based vocabulary training seems to be beneficial for a learner's lexical competence.

Try our software (Machina Callida)
`https://korpling.org/mc` !