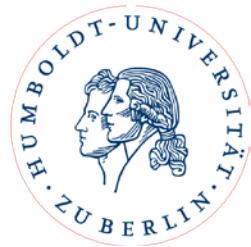


H U M B O L D T - U N I V E R S I T Ä T Z U B E R L I N



BOOK OF ABSTRACTS

International Workshop

Digital Approaches to Teaching Historical Languages (DAtTeL)

28. – 29.03.2019

Humboldt-Universität zu Berlin



Workshop 'Digital Approaches to Teaching Historical Languages' (DAtTeL) 28. – 29.03.2019 at the Humboldt-Universität zu Berlin

Conference venue

Humboldt-Universität zu Berlin
Jacob-und-Wilhelm-Grimm-Zentrum - Auditorium
Geschwister-Scholl-Straße 1/3, 10117 Berlin

PROGRAMME

Day 1: Thursday, 28.03.2019

10:00 - 11:00: **Registration & welcome coffee**

11:00 - 11:30: **Introduction** (A. Lüdeling)

11:30 - 12:30: **Keynote presentation**

Detmar Meurers & Kordula De Kuthy, Universität Tübingen

Supporting individual practice with immediate feedback: Motivation, design, and evaluation of an interactive workbook for English language learners

12:30 - 14:00: **Lunch break**

14:00 - 15:30: **Session 1** (3 talks)

- 14:00 – 14:30

Neel Smith, Sowerby, Z., Garry, J., Raheb, M. & R. Ciolek, College of the Holy Cross, USA

Corpus-specific morphological analysis for research and teaching

- 14:30 – 15:00

Monica Berti, Universität Leipzig

Computer-assisted language learning with Ancient Greek historiographical corpora

- 15:00 – 15:30 Uhr

Carolin Odebrecht, G. Schnelle & A. Lüdeling, Humboldt-Universität zu Berlin

Teaching variation and change: Using a multi-layer diachronic corpus of German to develop didactic materials

15:30 – 16:00: **Coffee break**

16:00 – 17:30: **Session 2** (4 talks)

- 16:00 – 16:30

Giulia Re, G. Mugelli, F. Boschetti & A. Taddei, Università di Pisa, Italia

Ancient Languages and DH: an annotation tool for teaching and learning

- 16:30 – 17:00

Gohar Schnelle, Humboldt-Universität zu Berlin

Corpus based teaching of Old High German. An experience report



- 17:00 – 17:30
Michael Prinz & Kyoko Sugisaki, Universität Zürich
HiVoKo – Building a corpus from handwritten lecture notes for teaching historical linguistics in practice
- 17:30 – 17:45
Saverio Dalpedri, K. Goetz & S. Skopeteas, Georg-August-Universität, Göttingen
AIEL – Ancient Indo-European languages for the 21st century

Day 2: Friday, 29.03.2019

9:00 - 10:00: Keynote

Peter Anderson, Grand Valley State University, USA

Developing the "Fifth Skill" through the reworking of historical texts

10:00 -10:30: Coffee break

10:30 – 12:15: Session 3

- 10:30 – 11:00
Andrea Beyer & Paul Reichertanz, Humboldt-Universität zu Berlin & Universität Rostock
Teaching digital literacy –Interpretieren in einer computergestützten Lehr-/Lernumgebung
- 11:00 – 11:30
Marietta Sionti, Universität Bielefeld
Architecture of an intelligent tutoring system for Ancient Greek vocabulary
- 11:30 – 12:00
Herbert Lange, Göteborgs universitet, Sweden
MULLE for Latin: Computer-generated translation exercises for Latin

12:00 – 12:30: Discussion & concluding remarks

Workshop CALLIDUS – Friday, 29.03.2019, 15:30 – 17:30

- 15:30 – 16:00: Welcome coffee
- 16:00 – 16:30 Theoretical overview
- 16:30 – 17:30 Practical issues – use cases, implications & restrictions, experience reports

Supporting individual practice with immediate feedback: Motivation, Design, and Evaluation of an interactive workbook for English language learners

Detmar Meurers & Kordula De Kuthy

Universität Tübingen

While immediate feedback is among the most powerful influences on learning and achievement, there are few opportunities for obtaining individual feedback in typical formal instruction settings. The situation is further complicated by the fact that there are substantial individual differences between learners in a typical class, making it attractive or even necessary for (groups of) learners to work on different activities. In this talk, we discuss the development of an intelligent language tutoring systems designed to address the challenge of providing immediate individual feedback to students working on a range of activities, from form-focused exercises to meaning-focused tasks. The *FeedBook* system is designed to accompany regular English instruction in German high school classes and is currently being used in 15 classes for the whole school year, for which we discuss first empirical insights. On the conceptual side, we argue that a successful approach must integrate models of language, tasks, and learners to be able to successfully characterize the expected range of well-formed and ill-formed variability in learner answers and to effectively provide the required diagnosis and feedback.

Corpus-specific morphological analysis for research and teaching

Neel Smith, Zachary Sowerby, James Garry, Michael Raheb, Richard Ciolek

College of the Holy Cross, USA

This paper presents a new approach to computational analysis of morphology. By applying corpus-specific parsers to digital Greek and Latin texts, we can systematically profile a corpus in terms of its lexicon, its surface morphology, and the generative rules mapping lexical items to tokens. We can then use these data either to prioritize pedagogical goals in order to prepare students to read a target corpus, or to extract sub-corpora tailored for specific pedagogical goals.

Several excellent Latin morphological parsers and a few options for Greek already exist.ⁱ They have been designed, however, as comprehensive systems applicable to “Latin” or “Greek” as a notional whole. Some support easily adding new vocabulary items, but the teachers and students of texts with “non-standard” morphology or orthography have few options. Normalizing a corpus to the parser’s standards is unacceptable if we want to study it from a corpus linguistic perspective, but modifying a parser’s underlying code can be as challenging as writing a new parser from scratch, and in any case is not a practical solution for most scholars of Greek or Latin.

We introduce an approach that enables maintainers of digital corpora, such as classical scholars, to construct corpus-specific parsers. The central technological component to such an approach is not a particular parser for a particular historical language, but an open-source system for building corpus-specific parsers. Our presentation illustrates two parallel systems: Kanónes, a system for constructing Greek parsers, and Tabulae, a system for constructing Latin parsers.ⁱⁱ The names are inspired by the basic task required to build a parser: compiling straightforward tables of data. Three data sets are managed in simple text files. One defines the orthography by enumerating all Unicode codepoints allowed in parseable tokens. Second, a set of delimited-text tables defines a lexicon of “stems,” recorded in the defined orthography. Third, a further set of delimited-text tables defines the inflectional rules that apply to the corpus. From these data sets, the build system composes code in the notation of the Stuttgart Finite State Tools (SFST), and compiles a binary parser (an executable finite state transducer).ⁱⁱⁱ The user can run simple scripts that in turn use the compiled parser to summarize a corpus morphologically, or find passages containing particular morphological features.

All components of a parser’s output are identified with canonically citable URN values: stems, lexical entities (“lemmata”), inflectional class (roughly corresponding to traditional categories such as “2nd declension noun”), and inflectional rule (“data singular ending is -o”). This makes it possible to identify the same features even in different corpora using different orthographic standards, and to reduce the ambiguity in morphological profiling of a corpus to instances of true morphological ambiguity.

In a corpus of Plautus, for example, the surface form *anime* can only be the vocative singular of *animus* (urn:cite2:hmt:ls:n2636 in the citable Lewis-Short dictionary from Furman University^{iv}). In a diplomatic edition of manuscripts of the Latin Psalms, “e” might represent the orthographic equivalent of classical “ae” so that *anime* could be the genitive singular, dative singular, or nominative or vocative plural of *anima* (urn:cite2:hmt:ls:n2612). A comprehensive “Latin” morphological parser would have to accept all these possibilities for analyses of *anime*. A classical Latin parser, on the other hand, could accept only “ae” as valid first-declension endings; the lexicon for a Latin parser of the psalms does not need an entry for *animus*, since that word does not appear in the Psalms. The only ambiguity this parser would identify is the identical form of four case-number combinations of the first declension noun *anima*.

This opens the door to a variety of pedagogical applications, as we will illustrate with examples from beginning and intermediate language courses. If we want to prepare students to read a target corpus, we can automate a task teachers have done by hand for centuries, and measure the frequency of vocabulary in the corpus. More generally, we can compare frequencies and distributions of *any* category of citable object. If present and aorist participles overwhelmingly outnumber perfect

participles in a target Greek corpus, we might emphasize them differently in our preparatory teaching. If a target Greek corpus includes no verb forms in the future perfect, we omit it altogether: our corpus-oriented pedagogy includes only content that appears in the corpus. We can also analyze the frequency of specific inflectional rules. Any target Latin corpus beyond a certain size is likely to include all case-number combinations of nouns (perhaps other than the vocative). Inflectional rules for various categories of third-declension i-stem nouns and adjectives on the other hand might be more or less frequent. We can vary our emphasis on these specific inflectional categories accordingly.

Finite state transducers, the engine behind our parsers, are always reversible: an analysis of a token can always be inverted to generate a token for a given analysis. We can therefore use the same transducer to determine where particular features appear in a corpus. Any analyses can therefore be used to establish priorities in our teaching, or to identify occurrences of the feature, and automate the selection of illustrative passages.

Finally, abstracting the identification of entities independently of their orthographic system creates new pedagogical possibilities. As was demonstrated at last year's International Epigraphic Congress, we can build distinct parsers for Attic Greek using (1) familiar literary orthography and (2) the epichoric alphabet of Athens.^v If we take the analysis in one system and use it to generate the tokens in the alternate system, we can effectively rewrite the (linguistically identical) text in two different orthographies. Students who have studied Attic Greek in literary orthography can work immediately with inscriptions automatically transcribed from the epichoric alphabet.

Similarly, Latinists who wish to mark long vowels with explicit macra can define a parser for this orthography. In a corpus encoded in this orthography, *animā* could be unambiguously parsed as the ablative singular of *anima* (`urn:cite2:hmt:1s:n2612`). In a corpus without editorial addition of vowel length, the ablative singular of `urn:cite2:hmt:1s:n2612` would of course be the multivalent form *anima*. Since the two tokens are recognized as equivalent in the two corpora, we can automatically construct parallel versions of the text in each orthography, and gradually wean students from the version annotated with vowel quantity.

The approach we present allows us to profile the morphology of as small or large a corpus as we choose, and to coordinate references to analyses of different corpora. It is well suited to working with digital corpora of historical texts either for corpus-linguistic research, or for rethinking our pedagogy in light of corpus-linguistic analysis of our texts.

- i In addition to the list at
https://wiki.digitalclassicist.org/Morphological_parsing_or_lemmatising_Greek_and_Latin, note also LatMor <http://www.cis.uni-muenchen.de/~schmid/tools/LatMor/> and Parsley <https://github.com/goldibex/parsley-core>.
- ii On Kanónes and its approach: Neel Smith, “Morphological analysis of historical languages”, *Bulletin of the Institute for Classical Studies* 59-2, 2016, 89-109.
- iii <http://www.cis.uni-muenchen.de/~schmid/tools/SFST/>
- iv <http://folio2.furman.edu/lewis-short/index.html>
- v Neel Smith, “Morphological parsing of Greek inscriptions in multiple dialects and alphabets” :
<https://epicongr2017.univie.ac.at/en/programme/thematic-panels/databases/>

Computer-assisted language learning with ancient Greek historiographical corpora

Monica Berti (Universität Leipzig)

The goal of this contribution is to present digital approaches to analyzing, teaching, and learning ancient Greek by using computer-assisted methods for historical languages. The discussion will be conducted with references to digital projects devoted to the creation and dissemination of text corpora about ancient Greek historiography. The presentation will also include results produced by students at the Center For Hellenic Studies (Harvard University), Furman University, and the University of Leipzig.

The first project is the Digital Fragmenta Historicorum Graecorum (DFHG), which is a corpus of 636 ancient Greek historians preserved through quotations and text reuses (*fragmenta*) in later texts. The corpus covers a period of time from the 6th century BC through the 7th century CE and is based on the digitization of the five volumes of Karl Müller's *Fragmenta Historicorum Graecorum* (FHG). It includes not only ancient Greek historiographical texts (684,977 tokens), but also Latin translations, introductions, commentaries, critical notes and indices for a total of 2,315,698 tokens.

The second project is the Digital Athenaeus and is focused on the ancient Greek work entitled *Deipnosophists* (or *Sophists at Banquet*) written by Athenaeus of Naucratis between the 2nd and the 3rd century CE. This work is arranged in 15 books, counts 264,750 tokens and is a rich collection of quotations and text reuses (*fragmenta*) of Greek authors belonging to many different genres of Classical literature (historians, comic poets, philosophers, grammarians, tragic poets, etc.). This is a pilot experiment for providing an inventory of authors and works cited and preserved by ancient Greek and Latin authors, and for implementing a data model for identifying, analyzing, and citing uniquely instances of text reuse in ancient Greek sources (Berti et al. 2016b and 2016c). The aim of this project is to expand it to other Greek texts in order to produce a corpus of annotated text reuse data of ancient Greek with a focus on historiography.

The goal of these projects is to deal with new needs in research, teaching, and learning of Classical languages by making use of digital data. This kind of work is addressed to the three groups of users identified by the organizers of the DAtTeL workshop:

- *Scholars*, who need to access and evaluate the growing amount of digital historical corpora of ancient Greek. This group includes humanists interested in the historical and historiographical content of the data, linguists interested in historical linguistic phenomena, and computer scientists interested in applying text mining techniques to historical corpora.
- *Students*, who learn ancient Greek in high schools and academic programs and who need to access and analyze digital data of historical corpora. The goal is to give them the possibility to learn Classical languages while also learning elements of linguistics and computer science in order to experiment with blended learning methods.
- *Teachers*, who need to ingest digital data in their teaching programs. Historical corpora offer them a wide variety of examples and exercises for establishing and implementing new teaching methods for ancient Greek and Latin.

This contribution will provide examples from the above mentioned projects in order to show how to use computer-assisted language learning methods for ancient Greek (Martin-Berti 2017). The contribution will focus on the following topics:

- Learning ancient Greek by correcting OCR outputs. The DFHG is the result of optical character recognition (OCR) of the FHG printed volumes, which contain texts in Greek and Latin. In order to improve high-quality OCR of ancient Greek texts in a high-performance computing environment, the

DFHG offers a reach collection that students can proofread with two results: 1) learn the correct spelling of ancient Greek in 19th century critical editions by comparing the original printed text with the OCR output, and 2) provide scholars with cleaned data for producing high quality OCR outputs that are now essential to expand the number of historical text corpora (see the project Lace. Polylingual OCR editing). The DFHG project offers an interface for OCR proofreading (fig. 1).

The screenshot shows a web-based application titled "Edit OCR Errors". At the top, it displays the title "Volumen primum HELLANICUS HELLANICI FRAGMENTA ΑΡΓΟΛΙΚΑ fr. 39". Below this is a table with four columns: "witness", "fragment text", "latin translation", and "commentary".

witness	fragment text	latin translation	commentary
Schol. Apollon. I, 131, et Phavorin.:	Ὑλας, Ἡρακλέους ἐρώμενος, σιδός θειοδάμαντος τοῦ Δρύοπος (Schol. Theocr. In argumento Idyllii 18, corrupte exhibet δρυτόμων). Οὗτος Μνασέας. Ἑλλάνικος δὲ θειομένητα θειοδάμαντος καλεῖ.	Hylas, Herculis amasius, filius erat Theodamantis, Dryopis. Ita Mnaseas. Hellanicus vero Theomenetem theodamantis filium nominat (vel potius: Theomenetem Theodamantis loco nominat).	«Hellanicī sententia haud dubie fuit alia, et locus emendandus est sic: Ἑλλάνικος δὲ θειομένητα (vel θειούντην) ἀντὶ θειοδάμαντος καλεῖ. Nam in iisdem scholis ad v. 1207 exstat θειομένους. Hylae pater igitur Hellanicō auctore fuit θειούντης.» Sturz.

Below the table, there is a field labeled "submitter: _____". Underneath this is a reCAPTCHA verification box with the text "I'm not a robot" and a checkbox. To the right of the checkbox is the reCAPTCHA logo and the text "reCAPTCHA Privacy - Terms". At the bottom right of the form is a "Submit" button.

Fig. 1. DFHG OCR editing environment

- Learning ancient Greek with translation alignment. The DFHG is a collection of more than 7,000 Greek *fragmenta* with Latin translations by Karl Müller. Given the chronological range and the variety of genres of the source texts of the fragments, students can learn the language by comparing DFHG Greek texts with their Latin translations and by actively trying to align them. This work can be expanded outside the DFHG collection by looking for translations into contemporary spoken languages. Translation alignment editors like the one developed by the Alpheios Project provide students with interfaces for manual alignment and data export. As for OCR, the goal is twofold: 1) learn ancient Greek by a careful examination of translations of the Greek *fragmenta* of the DFHG, and 2) provide the community of scholars with manual translations pairs that can be used as training data for experimenting with automatic translation alignment of historical texts (Yousef-Berti 2015 and Berti et al. 2016a).
- Morpho-syntactic annotations of ancient Greek. The DFHG and the Digital Athenaeus offer an integration with Morpheus, which is the morphological parsing and lemmatising tool of the Perseus Project, and with the OpenNLP POSTagger for Ancient Greek, which is tested on the data of the Ancient Greek Dependency Treebank. These tools provide scholars, students, and teachers with semi-automatic annotation of words with tags corresponding to the part of speech they represent (Celano et al. 2016). This integration offers students many possible exercises for testing their knowledge of the morphology and the syntax of ancient Greek. With the aid of the instructor, students can also actively produce their own morpho-syntactic annotations of Greek texts of the DFHG and the Digital Athenaeus providing the community with more data for a further disambiguation of semi-automatic generated results (see the Arethusa Treebank Editor developed by the Alpheios Project). Given that there are Greek lost works preserved by Latin authors, the DFHG collection includes also Latin *fragmenta* that can be used for exercises to learn Latin.
- Named entities recognition and disambiguation. Texts of the DFHG and the Digital Athenaeus are interesting for the huge amount of named entities they include (e.g., names of people, gods, authors, titles of works, places, etc.). Both projects provide lists of automatically disambiguated named entities that still need a further work of manual disambiguation. This work can be performed by students with the help of their teachers in order to generate different outputs: lists of ancient Greek proper names

and their relationships for learning ancient onomastics and prosopography, and for experimenting with social network analysis; lists of ancient Greek author names and work titles for producing a library catalog of Classical literature and learning its complexities (see the Perseus Catalog); lists of place names to be disambiguated and geo-located in order to create maps for visualizing the geography of the *fragmenta* of the FHG collection and of the *Deipnosophists* of Athenaeus of Naucratis. External web based platforms for semantic annotations like Recogito, WebAnno, and INCEpTION can be used by students for named entities and geographic annotations. Fig. 2 shows the example of the map of the DFHG corpus.

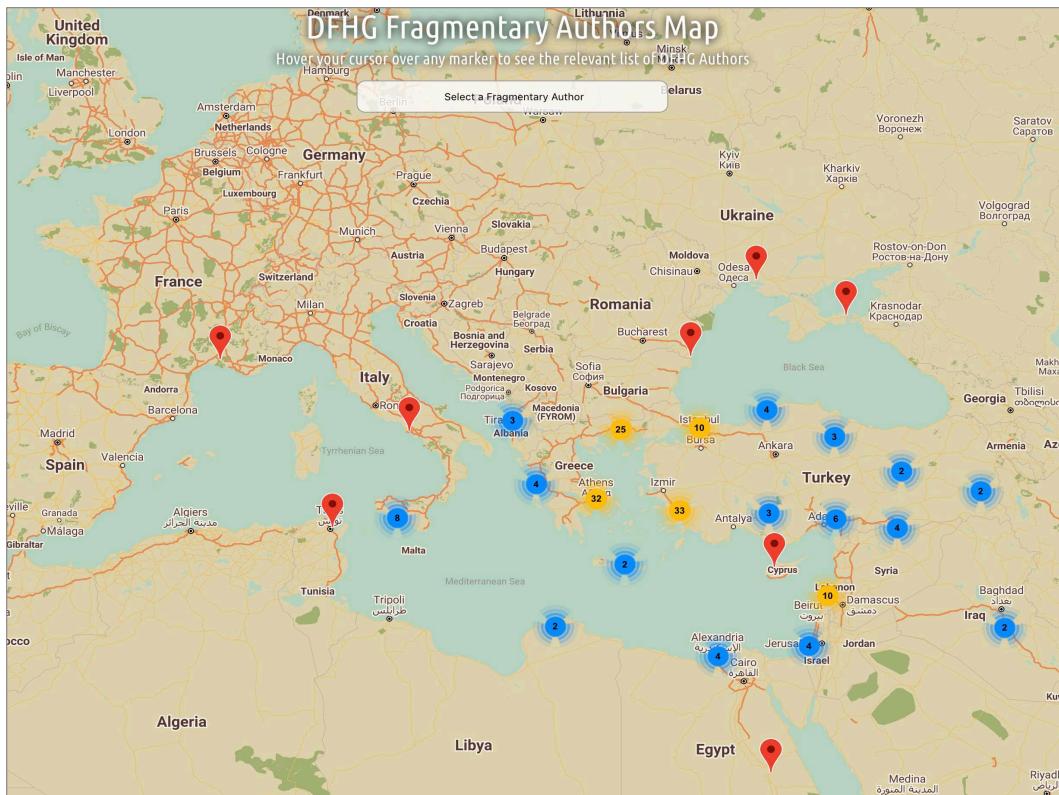


Fig. 2. DFHG Fragmentary Authors Map

These topics show possibilities for locating exercises and examples from ancient Greek and Latin texts that instructors can use for applying digital approaches to teaching historical languages. The idea behind these topics – and behind other projects in the digital humanities and digital philology – is to actively involve students in their learning process by offering them and their teachers the possibility to produce data for communities of other students and scholars. Challenges and complexities of this process are represented by the implementation of environments that are suitable for these tasks, and by the creation of evaluating and assessing methods that guarantee high quality standards of learning and data production. The DFHG and the Digital Athenaeus projects don't offer immediate solutions to these challenges and complexities, but aim at being part of a discussion concerning computer-assisted learning and teaching methods for historical languages that involve scholars, students, and teachers in order to improve the quality of digital data of historical corpora.

References

- Alpheios Project: <http://alpheios.net>
- Ancient Greek Dependency Treebank: https://perseusdl.github.io/treebank_data
- Arethusa Treebank Editor: <https://alpheios.net/pages/tools>
- Digital Athenaeus: <http://www.digitalathenaeus.org>
- Digital Fragmenta Historicorum Graecorum (DFHG): <http://www.dfhg-project.org>
- Berti M.; Crane, G.R.; Yousef, T.; Bizzoni, Y.; Boschetti, F.; Del Gratta, R. (2016a): “Ancient Greek WordNet Meets the Dynamic Lexicon: the Example of the Fragments of the Greek Historians”. In *Proceedings of the Eighth Global WordNet Conference, Bucharest, Romania, January 27-30, 2016*. Ed. by V.B. Mititelu; C. Forăscu; C. Fellbaum; P. Vossen. Bucharest, 34-38.
- Berti M.; Blackwell, C.W.; Daniels, M.; Strickland, S.; Vincent-Dobbins, K. (2016b): “Documenting Homeric Text-Reuse in the *Deipnosophistae* of Athenaeus of Naucratis”. In *Digital Approaches and the Ancient World*. Ed. by G. Bodard; Y. Broux; S. Tarte. BICS Themed Issue 59:2, 121-139.
- Berti M.; Almas, B.; Crane, G.R. (2016c): “The Leipzig Open Fragmentary Texts Series (LOFTS)”. In *Digital Methods and Classical Studies*. Ed. by N.W. Bernstein; N. Coffee. DHQ Themed Issue 10:2.
- Celano G.G.A.; Crane, G.R. ; Majidi, S. (2016): “Part of Speech Tagging for Ancient Greek”. In *Open Linguistics* 2, 393-399.
- INCEpTION: <https://inception-project.github.io/>
- Lace. Polylingual OCR editing: <http://heml.mta.ca/lace>
- Martin T.R.; Berti, M. (2017): “Open Greek and Latin Data for the Challenges of the Fragmentary State of the Primary Sources for the Pentekontaetia”. In *Mouseion: Journal of the Classical Association of Canada* (special issue on Open Greek and Latin) 14:3, 409-436.
- Müller, K. (1841-1873): *Fragmenta Historicorum Graecorum (FHG)*, I-V, Paris.
- Perseus Catalog: <http://catalog.perseus.org>
- Perseus Project: <http://www.perseus.tufts.edu>
- Recogito: <https://recogito.pelagios.org>
- Yousef T.; Berti, M. (2015): “The Digital Fragmenta Historicorum Graecorum and the Ancient Greek-Latin Dynamic Lexicon.” In *Proceedings of the Workshop on Corpus-Based Research in the Humanities (CRH), 10 December 2015 Warsaw, Poland*. Ed. by F. Mambrini; M. Passarotti; C. Sporleder. Warszawa, 117-123.
- WebAnno: <https://webanno.github.io/webanno>

Teaching Variation and Change: Using a Multi-Layer Diachronic Corpus of German to Develop Didactic Materials

Carolin Odebrecht, Gohar Schnelle, Anke Lüdeling
Humboldt-Universität zu Berlin

Our study is concerned with the (semi-automatic) development of corpus-based teaching materials for university courses of Early New High German (ENHG) and the early stages of New High German (NHG; all in all we look at the time period between 1550 and 1900). Due to external factors such as the invention of letter press printing, Luther's translation of the Bible, the founding and development of universities and, later, the evolution of the sciences, as well as other political and social factors, this period is characterized by a large increase of writers, texts and text types (Pörksen 2004, Habermann 2001, Ágel & Hennig 2006, among many others). Because of this, the period in question is a particularly interesting one in the linguistic history of German. We find a large amount of variation on all linguistic levels which slowly, over several centuries, converge to a common standard for some systemic aspects of the language (such as orthography or syntax, see e.g. Besch 2003) as well as the evolution of many formal and informal registers (Ágel & Hennig 2006).

Teaching materials for ENHG (such as e.g. Hartweg & Wegera 2015) illustrate the internal variation and the developments with examples from authentic texts. However, often there are not enough texts and examples to show the possible variants and their influencing factors in sufficient detail. We therefore want to discuss how deeply annotated corpora can be exploited by teachers and learners of ENHG/NHG to find examples for, analyze and teach different linguistic phenomena as the need arises. Using the RIDGES corpus¹, we illustrate our approach in several case studies. The corpus is diachronic, deeply annotated, and each text is marked with rich metadata. All of the information in the corpus can be accessed via the search tool ANNIS². In our presentation, we focus on graphemic and lexical variation; our approach can, however, be easily extended to other levels of grammar.

Our first case study deals with graphemic variation. The data in the RIDGES corpus is transcribed diplomatically, and the corpus is annotated with a normalized layer (*undi*, *unti*, *unnti*, *vnti* etc. are all mapped to the normalized form *und*, and'; see Odebrecht et al. 2016 and Belz et al. 2018 for more information in the construction of this layer). Figure 1 shows the ratio of different spellings per normalized word form. The decline of the ratio clearly shows the convergence towards an accepted orthography. Students of historical German can use these layers to find examples of developments that interest them, e.g. examples for the evolution of accepted standards in the spelling of vowel quality and quantity or the change from prosodically motivated virgulas to syntactically motivated commas.

The second case study is concerned with different types of multi-word elements. Each word can be studied in its syntagmatic and paradigmatic context. The graphemic variation (especially in ENHG) makes it difficult to find 'the same' word. Again, we can use the normalized layer to calculate different types of association measures and see how associations and patterns change over time. We can also use part-of-speech tags to calculate frequencies and associations of more abstract patterns. Figure 2 shows different syntactic and lexical continuations of *X ist gut*, 'X is good'.

- with a preposition: *X ist gut wider das Haarausfallen*, 'X is good against hair loss'
- with a dative object: *X ist gut dem der gebissen ist*, 'X is good for he who has been bitten'
- with a subordinate clause: *X ist gut wenn man das haupt damit salbt*, 'X is good if one anoints his head with it'

¹ korpling.org/ridges, Odebrecht et al. (2016).

² <http://corpus-tools.org/annis/>, Krause & Zeldes (2016).

Each of these continuation patterns can be easily identified in the corpus and analyzed further. The second option – which was frequent in ENHG – is not (easily) available in NHG. Figure 3 shows the different continuation patterns that can follow the expression *ist gut*. Starting from this lexical example, the gradual loss of dative objects for similar structures can then be analyzed using part-of-speech tags.

In our presentation, we will illustrate in detail how students and teachers of ENHG/NHG are able to use a deeply annotated diachronic corpus to find patterns of variation and change. We believe that this is especially interesting for a historical language stage that is close enough to the current variety of a language because students may not be able to appreciate the subtle differences in meaning and pragmatics of a historical variety as long as they are able to understand (or think they can understand) the text.

References

Belz, Malte; Odebrecht, Carolin; Perlitz, Laura; Schnelle, Gohar & Voigt, Vivian (2018): Annotationsrichtlinien zu Ridges Herbology Version 8.0, Humboldt-Universität zu Berlin. URL: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/download-files/pubs/ridgesv8_2018-04_06.pdf

Ágel, Vilmos / Hennig, Mathilde (Hrsg.) (2006): *Grammatik aus Nähe und Distanz. Theorie und Praxis am Beispiel von Nähetexten 1650-2000*. Tübingen: Niemeyer.

Besch, Werner (2003) Die Entstehung und Ausformung der neuhighdeutschen Schriftsprache/Standardsprache. In: Werner Besch and Anne Betten and Oskar Reichmann and Stefan Sonderegger (Hrsg.) Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Volume 3, p. 2252–2296.

Habermann, Mechthild (2001) Deutsche Fachtexte der frühen Neuzeit: naturkundlich-medizinische Wissensvermittlung im Spannungsfeld von Latein und Volkssprache. Berlin: Walter de Gruyter.

Hartweg, Frédéric and Klaus-Peter Wegera (2005): *Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit*. Tübingen: Narr.

Krause, Thomas & Zeldes, Amir (2016): ANNIS3: A new architecture for generic corpus query and visualization. in: *Digital Scholarship in the Humanities 2016* (31).

<http://dsh.oxfordjournals.org/content/31/1/118>

Odebrecht, Carolin, Malte Belz, Amir Zeldes, Anke Lüdeling und Thomas Krause (2016). RIDGES Herbology - Designing a Diachronic Multi-Layer Corpus. In: *Language Resources and Evaluation*. DOI: 7/s10579-016-9374-3.

Pörksen, Uwe (1998). Deutsche Sprachgeschichte und die Entwicklung der Naturwissenschaften. — Aspekte einer Geschichte der Naturwissenschaftssprache und ihrer Wechselbeziehung zur Gemeinsprache. In: Werner Besch and Anne Betten and Oskar Reichmann and Stefan Sonderegger (Hrsg.) Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Volume 2,1, p. 193–210.

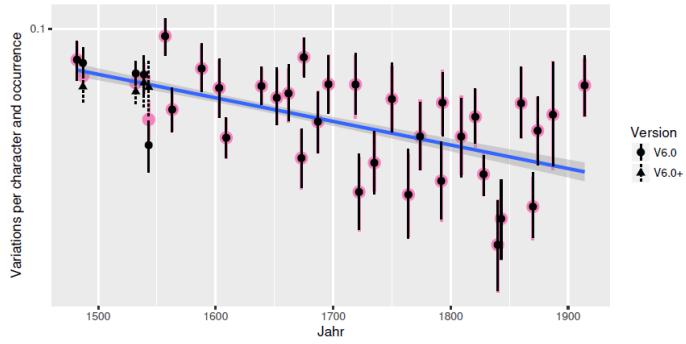


Figure 1: The ratio of different spellings for a word in the RIDGES corpus.



Figure 2: Possible syntactic continuations of *ist gut* ‘is good/benefits’ in the RIDGES corpus.

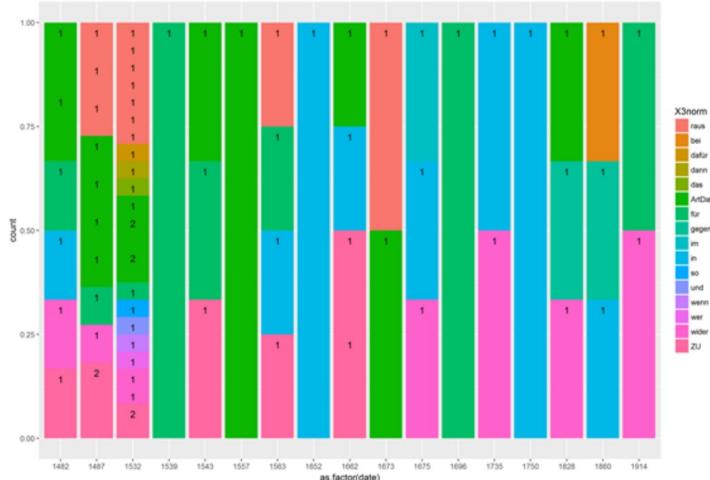


Figure 3: Possible continuations of the expression *ist gut* ‘is good/benefits’ in the RIDGES corpus over time.

Ancient Languages and DH: an annotation tool for teaching and learning

G. Re, G. Mugelli, F. Boschetti, A. Taddei

Who we are

This abstract presents a system of annotation of ancient Greek texts, which can be used in schools and university both for research and didactic purposes. The annotation system Euporia has been worked out in a collaboration between the CoPhiLab, CNR of Pisa and the *Laboratorio di Antropologia del Mondo Antico* (LAMA), University of Pisa.

The aim of this cooperation is to improve the use of digital tools as a support for the study of ancient languages. Learning Greek and Latin needs not only a strong linguistic and historical background, but also the enhancement of semantic, morphological and morphosyntactic aspects. Such an accomplished knowledge can be provided to students by working on digital annotation of texts.

Euporia: features and applications

Euporia is a *user-centred* annotation system for ecclotic and hermeneutics. The *user-centred* approach is based on the adoption of Domain Specific Languages, allowing the user to have a flexible syntax and an open tagset in order to annotate linguistic, morphological, syntactic, semantic, stylistic data, and textual variants. Every DSL is structured in accordance to the needs of the annotator.

In a second time, the tag system set by the user is converted in TEI-XML by a *parser*, and the references to the passages of the original text are converted in a CTS-URI.

The annotation system can be used in exercises of contrastive translation, and to annotate different layers of interpretation of the text proposed by the annotator or deduced by secondary literature.

Today, there are many customized versions of Euporia, used in different levels of academic education and more recently in two high schools. We started using the annotation system as a support for PhD researches, then we tested it in teaching in the *Laboratorio di lingua Greca*, with first-year students of ancient Greek language in the University of Pisa.

EuporiaRAGT (Rituals in Ancient Greek Tragedy)

EuporiaRAGT project has been worked out in the wider frame of a doctoral research on forms and functions of rituals in Greek tragedy of V b.C. Tragic texts allow the reader to go back to ritual performances in their historical shape, and to put in a relation the dramatized form of the rites with the religious reality of V century Athens. Therefore, EuporiaRAGT's aim is to compare rites as they are performed in tragedy with rituals as they are known by ancient sources.

The thematic annotation shows crossed references to some dynamics of tragic ritual actions and elements of the staging, i.e. characters' mimic, onstage/ offstage performances, etc.

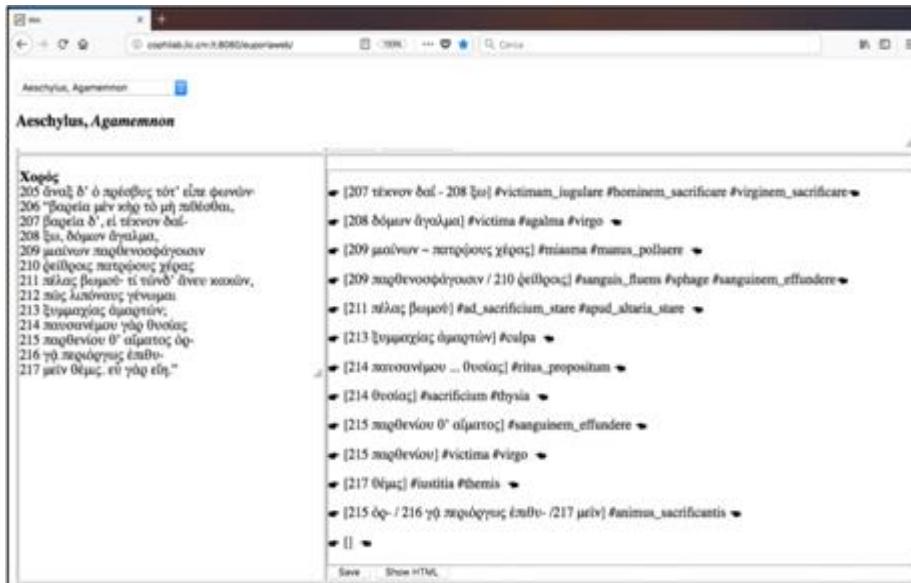


Fig. 1. Annotation in EuporiaRAGT

The annotation has a *bottom-up* approach, which *does not* entail the existence of a predetermined tag-set, but allows the user to build it up during its research. The *bottom-up* approach provides a database of annotations which allows the user to model the information following the methodological assumptions of Anthropology of the Ancient world. The originally unstructured tags are structured *a posteriori* in a domain ontology, modelled on the features of ritual practices.

By *query* SPARQL, the *user* can do complex queries on the ontology (fig. 2). It is also expected to create a *user interface*, in order to make queries into the annotations database.

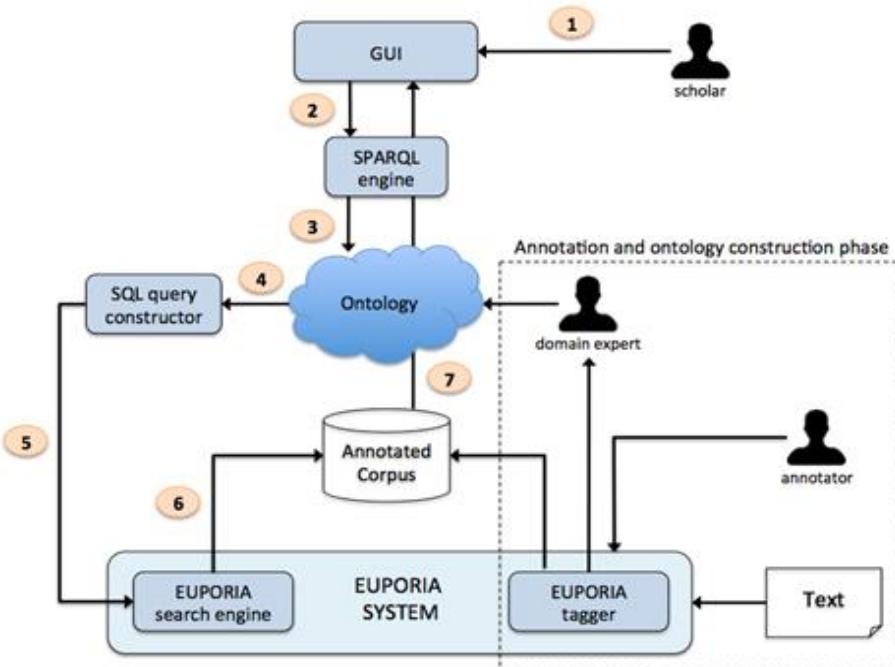


Fig. 1. EuporiaRAGT system

EuporiaLanGT (*Landscape in Greek Tragedy*)

Beside a *user-centred* annotation mode, Euporia is thought as a tool for collaborative annotations on the same *corpus*. If the work on EuporiaRAGT represents a *personomy*, a cooperative annotation produces a *folksonomy*, which provides wider possibilities in modelling the ontology.

This collaborative aspect is carried out by EuporiaLanGT, created by CophiLab and LAMA as a support for a PhD project on the ritual functions of landscapes in Greek tragedy.

The research question of this thematic and semantic annotation of tragedies is to find *patterns* which combine one or more divinities with one or more ritual actions, within the frame of a specific natural environment.

The annotation of the tragic *corpus* is going to mark: *a)* elements of the tragic landscapes that are linked to ritual actions; *b)* divinities of the Greek *pantheon* who are recipient of cults in a particular space; *c)* the forms of religious thought that associate a given god to a specific ritual landscape.

The annotation language of Euporia *LanGT* is, unlike RAGT's, semi-structured and based on 3 main tags:

1. **#evidence**: marks a natural element explicitly mentioned in the tragic text, valid in specific context
2. **#ontology**: states a general assumption, as the relation between a divinity and a specific landscape
3. **#association**: throws back to the plan of mental associations operated by the spectators of Athenian drama.

According to the needs of the user, the annotation can point out *variae lectiones*, by the tag **@vl**, i.e. different philological interpretations with precise references to manuscripts and to secondary literature (**#bibl**):

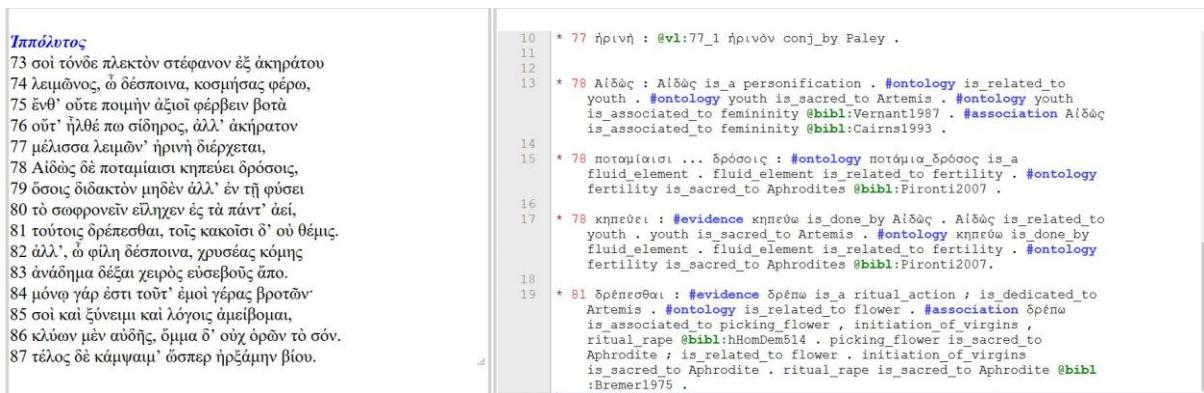


Fig. 3. Annotation and tagset in EuporiaLanGT

Experiments in teaching

Euporia has been used as a support for ancient Greek teaching in the *Laboratorio di Lingua greca*, attended by first-year students of the University of Pisa.

In 2016/2017, the work has been conducted on *Odyssey* XXIV. Students have been asked to use keywords in order to mark metrics, linguistic or stylistic elements, and formularity in the text.

Homerus, *Odyssea* 24

35 τὸν δ' αὐτὲς ψυχὴ προσεφώνεεν Ἀτρεῖδαιο-
36 "δίληι Πήλεος τέλε θεοῖς ἐπείνελα." Ἀγγέλοι
37 ὃς θάνεις ἐν Τροΐῃ ἐκάς Ἀργεῶς· ἀμφὶ δέ σ' ἄλλοι
38 κτείνοντα Τρώων καὶ Ἀχαιῶν τις ἀριστος,
39 μαρνάντων περὶ σείο σὺ δὲ ἐν στροφάλιγγι κονίς
40 κείο μέγας μεγαλώστι, λελασμένος ἵππουσάνων.
41 ἡμεῖς δὲ πρόπταν ἦμας ἐμφανώμεθ', οὐ δέ κε πάλμαν
42 πανσάμεθε ποτολέμου, εἰ μὴ Ζεὺς λαβάται πανοεν.
43 αὐτῷ δέ τοι σ' ἐπὶ νῆσος ἔνεικαντεν ἐκ πολέμου,
44 κάτθεμεν ἐν λεγέσσοις, καθήσαντες χρόα καλῶν
45 ὑδατί τε λαρῷ καὶ ἀλλεψατο· πολλὰ δέ σ' ἀμφὶ
46 δάκρυα θερμά χέον Δαναοι κείροντο τα κάταις.
47 μῆτρο δέ ἐξ ἀλός ἥλιος σύν ἀθανάτης ἀλίσουν
48 ἀγγέλις ἀλισσα· βοή δέ ἐπι πόντον δρόμῳ
49 θεοποιη, ὑπὸ δέ τρόμος Ἑλλαβε πάντας Ἀχαιούς·
50 καὶ νῦν ἀναβάντες ἔβαν κολας ἐπὶ νῆσος,
51 εἰ μὲν ἀνήρ κατέρυπε παλαί τε πολλὰ τε εἰδύν,
52 Νέστορο, οὐ καὶ πρόσθεν ἀρίστη φαίνετο βουλή·
53 δι σφιν ἐνθρόνεων ἀγόρησατο καὶ μετέπειν·
54 "Τοχεοῦθ, Ἀργεῖοι, μὴ φεύγετε, κοιροὶ Ἀχαιῶν·
55 μῆτρο δέ ἀλός ἥλιος σύν ἀθανάτης ἀλίσουν
56 ἔρχεται, οὐ παύδος τενθῆτος ἀντώνουσα."

- [35 τὸν ... Ἀτρείδαο] #formula ➔
- [35 Ἀτρείδαο] #spondeo_Vsede #forma_ionica ➔
- [35 προσφέντων] #distrazione ➔
- [36 δῆλει ... Ἀχιλλέν] #olodattilico #formula ➔
- [37 ἐν Τροῇ ἔκατον Ἀργεος] #formula ➔
- [37 δῆλοι ... ~38 μίες] #perbato #enjambement ➔
- [38 Τρόπον καὶ Ἀχαιῶν] #formula ➔
- [39 ἐν στροφάλιγγι κονίς] #formula ➔
- [39 στο] #forma_ionica ➔
- [42 πτολέμου] #forma_cipriota ➔
- [44 λεχέσσον] #forma_elocica ➔
- [45 ὑδατί ... ἀμφὶ] #olodattilico #dieresi_bucolica ➔
- [45 καὶ ... ἀλειφάτη] #correptio_epica ➔
- [49 ὑπὸ ~ Ἐλλαζή] #tmesi ➔
- [53 μετέειπεν] #distrazione ➔

Fig.4. teaching with Euporia

This experiment returned positive feedbacks. First of all, students have experienced a process of *close reading* of the homeric text. Besides, the completely autonomous selection of keywords and textual passages allowed them to focus on some peculiarities of the language and on specific textual phenomena. This increased their awareness of some problems in the translation and visibly strengthened their ability of memorizing single words.

Since the annotation system is not a systematic analysis but a collection of a bunch of cases for specific learning needs, it matches with the morphological information available online in the *Perseus* and the *Alpheios* projects.

In 2017/2018, a second experiment started. The annotation dealt with some staging aspects in Sophocle's *Ajax*. Students were asked to mark all the occurrences in which Ajax's dead body was represented as an onstage object with scenic and ritual features. Furthermore, they had to point out the syntactic function of words relating to the hero's corpse as an *agent*, or a *patient* of the actions performed by other characters.

It has been noticed that, in some cases, the *bottom-up* tagset created by the students matched with the EuporiaRAGT one, while in other cases it was divergent. This is a desirable result for us, because it shows the great variety of solutions provided by a collaborative annotation, and represents a concrete example of Euporia's interoperability.

Towards Collaborative Annotation: ongoing projects

In 2018 a new ongoing project involves first-year students in Greek Language and Literature. The annotation plan deals with Aeschylus' *Agamemnon*.

In this case, students have been asked to choose the annotation topics by themselves. In this way, the teacher evaluated previous knowledges of the students and adjusted the complexity of the work.

The topics deal both with morphology (doric elements, *hapax legomena* in choral sections, linguistic peculiarities) and semantic (semantic field of guilt, war, blood, gender-inverted epithets).

Students work alone or in pairs on the same topic, but they are not aware of the annotation choices of their colleagues. Working in couples provides occasions of *peer learning* and allows us to observe the gradual formation of a *folksonomy*.

Working on Euporia, students are learning Aeschylean lexicon and style faster, and at same time the teacher is modelling his didactic activity according to the progress made.

Aside from experiences in University teaching, from 2018/2019 Euporia is part of some ongoing projects in secondary schools. Three Italian *Licei Classici*, located in different areas of the country have been chosen to be part of the experiment.

Teachers will model research questions, select texts and build the DSL according to the classes' needs. The school class will become an actual *collaborative environment*, where students and teachers can *interact* and *cooperate* together in organizing the work and interpreting the text.

Eventually, the project will underline statistic data relating to the local impact of DH tools for didactic purposes on the whole territory, and highlight *best practices* adopted in specific areas.

Contacts and affiliations:

Giulia Re, PhD student, member of LAMA, University of Pisa (giulia.rex90mail.com)

Gloria Mugelli, PhD, member of LAMA, University of Pisa (gloria.mugelli@gmail.com)

Federico Boschetti, ILC (Istituto di Linguistica Computazionale) CNR Pisa, (federico.boschetti@yahoo.com)

Andrea Taddei, Professor of Greek Language and Literature, Anthropology of Ancient Greece, senior member of the LAMA, University of Pisa (andrea.taddei@unipi.it)

Bibliography

Alvoni G., *Scienze dell'antichità per via informatica. Banche dati, Internet e risorse elettroniche nello studio dell'antichità classica*, Bologna 2002.

Boschetti, F., Del Gratta, R. e M. Lamé, *Computer Assisted Annotation of Themes and Motifs in Ancient Greek Epigrams: First Steps*, in *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014*, Pisa, 9-10 December 2014, (a cura di) R. Basili, A. Lenci, B. Magnini, Pisa 2014, pp. 83-86.

Condello F, *La scuola giusta*. In difesa del liceo classico, Milano 2018.

Fabrizi P., *La rete degli antichisti*, Pisa 1999.

Favini L., *Un modello, un metodo*, in *Nuova Secondaria* n. 1 (settembre 2006), pp. 79-84.

- Gibbs, F., Owens, T., *Building better digital humanities tools: toward broader audiences and user-centered designs*, DHQ 6, 2 (2012).
- Khan, A. F. et al., *Restructuring a Taxonomy of Literary Themes and Motifs for More Efficient Querying*, MATLIT 4 (2), Coimbra 2016a, pp. 11–27.
- Khan, A. F. et al., *Leveraging a Narrative Ontology to Query a Literary Text*, in *OASIcs-OpenAccess Series in Informatics*, 2016b, pp. 10:1-10:10.
- Mastandrea, P., *Gli archivi elettronici di Musisque deoque. Ricerca intertestuale e cernita fra varianti antiche (con qualche ripensamento sulla tradizione indiretta dei poeti latini)*, in *Poesia latina. Nuova e-filologia. Opportunità per l'editore e per l'interprete*, (a cura di) P. Mastandrea e L. Zurli. Roma 2009, pp. 41–72.
- Mugelli, G., Boschetti, F. et al. 2016, *A user-centred design to annotate ritual facts in ancient greek tragedies*, in *BICS* 59 (2), pp. 103–20.
- Mugelli, G., Khan A. F., et al. 2017, *Designing an Ontology for the Study of Ritual in Ancient Greek Tragedy*, in *Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017)*, Montpellier, pp. 96-105.
- Parr, T., *Language implementation patterns: create your own domain-specific and general programming language*, Raleigh NC 2010.
- Taddei, A. *Testi e contesti. Per una didattica della letteratura greca nei nuovi licei*, Pisa 2011.

Corpus Based Teaching of Old High German.

An Experience Report

Gohar Schnelle, Humboldt-Universität zu Berlin,
Institut für deutsche Sprache und Linguistik, Lehrstuhl für Geschichte der deutschen Sprache

In my talk I want to show how the Referenzkorpus Altdeutsch (Donhauser et al.) can be used in teaching Old High German (OHG) in conjunction with conventional textbooks to promote the ability to abstract and critical thinking.

I will present an experience report about the corpus based teaching of a seminar "Althochdeutsch" (Old High German) held in the winter semester 2018/2019 at the Humboldt-Universität of Berlin using the example of the Old High German nominal morphology (cf. Donhauser 2015).

The Referenzkorpus Altdeutsch (Version 1.1) is a deeply annotated corpus based on the complete OHG and old saxon editions and searchable via the search and visualization platform ANNIS (Krause & Zeldes 2016).

The diachronic and diatopic unbalanced, incomplete and thematically one-sided transmission of OHG (Fleischer 2006) makes an explicit thematization of variation in teaching necessary. The descriptively oriented grammars, dictionaries and introductory works (Braune & Reiffenstein 2004; Schützeichel 2006, Splett 1993; Sonderegger 2003), however, do not always adequately reflect this variation and provide only selective and inconsistent information about quantitative distributions of the variants.

In a variation-linguistic OHG seminar these specifics of OHG need to be addressed and made tangible by an iterative comparison of the information contained in the grammar with quantitative and qualitative corpus data.

The concept of the seminar will be presented on the basis of the topic "Nominal Flexion".

The semester is divided into three thematic sections:

1. Context and variation of OHG
2. OHG Grammar
3. Corpus linguistic annotation of an OHG text

In the first section, students are familiarised with the extra-linguistic framework conditions and varietal forms of OHG. Each student is given a specific OHG text. Students are encouraged to apply their knowledge by analyzing the extra-linguistic conditions of their given text (time and place of origin, length, content, Latin dependency).

In the second section students are taught grammatical topics through the interplay of grammars and corpus data.

In the field of nominal flexion, the OHG paradigms are the first topic. As a first step, the students examine how the 15 paradigms which have been laid out and described in the OHG grammar (Braune & Reiffenstein 2004) are quantitatively distributed in their respective texts. Through the exchange with their fellow students, they learn to identify and describe similarities and differences in the texts. After an intralingual evaluation of the differences, various factors (time of origin, place of origin, context) for these differences can be discussed on the basis of the exact metadata knowledge that each student possesses about his text. As a result, the students work out which information is valid for OHG as a whole and which information varies in accordance with various extra-linguistic factors.

As a second step, students familiarize themselves with the realizations of individual morphemes of a paradigm and create an inflection inventory for a word stem on the basis of the evidence. Subsequently, they examine how and in what form the flexion morphemes collected in the individual text are mentioned in the grammar of Braune & Reiffenstein 2004 and whether any variation is mentioned.

Individual grammatical topics are rounded off by a summary for the entire OHG language period in the form of chronographic maps (Donhauser & Zeige, in Ersch.), which visualize the diachronic and diatopic distribution of variants of a particular phenomenon.

The third section pursues two objectives: on the one hand, students practice grammatical analysis and interpretation as well as the use of relevant reference works through confrontation with an authentic OHG text, thus applying their acquired knowledge. On the other hand, they also become familiar with the working steps of corpus creation by applying guidelines and making criteria-driven decisions for annotation.

References:

- Braune, Wilhelm & Ingo Reiffenstein (2004):** Althochdeutsche Grammatik. 1. Laut- und Formenlehre. Tübingen: Max Niemeyer.
- Donhauser, Karin (2015):** Das Referenzkorpus Altdeutsch: Das Konzept, die Realisierung und die neuen Möglichkeiten. In: Gippert, Jost/Gehrke, Ralf (Hg.): Historical Corpora. Challenges and Perspectives (=CLIP 5). Tübingen, 35-49.
- Donhauser, Karin; Gippert, Jost; Lühr, Rosemarie;** ddd-ad (Version 1.1), Humboldt-Universität zu Berlin. <http://www.deutschdiachrondigital.de/>. <http://hdl.handle.net/11022/0000-0007-C9C7-6>
- Donhauser, Karin & Lars Erik Zeige:** Bildgebende Verfahren zur Analyse des Referenzkorpus Altdeutsch. Das Konzept der chronographischen Karte. (i. Ersch.).
- Fleischer, Jürg (2006):** Zur Methodologie althochdeutscher Syntaxforschung. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur* 128: 25–69.
- Krause, Thomas & Zeldes, Amir (2016):** ANNIS3: A new architecture for generic corpus query and visualization. In: Digital Scholarship in the Humanities 2016 (31).
- Rössler, Paul (2012):** Die Grenzen der Grenzen. Sprachgeschichtsperiodisierung zwischen Forschung und Lehre. In: Peter Maitz (Hg.): Historische Sprachwissenschaft. Erkenntnisinteressen, Grundlagenprobleme, Desiderate. Berlin: De Gruyter (Studia Linguistica Germanica, 110), S. 153–167.
- Schrodt, Richard (2004):** Althochdeutsche Grammatik. 2. Syntax. Tübingen: Niemeyer.
- Schützeichel, Rudolf (2006):** Althochdeutsches Wörterbuch. 6. Auflage. Berlin [u.a.]: de Gruyter.
- Sonderegger, Stefan (2003):** Althochdeutsche Sprache und Literatur. 3. Aufl. Berlin, New York: Walter de Gruyter.
- Splett, Jochen (1993):** Althochdeutsches Wörterbuch: Analyse der Wortfamilienstrukturen des Althochdeutschen, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes.. Berlin [u.a.]: de Gruyter.

HiVoKo – Building a corpus from handwritten lecture notes for teaching historical linguistics in practice

Michael Prinz, Kyoko Sugisaki
University of Zurich

University lecture notes have hardly been studied by German linguists, although they provide insight into the communicative conditions of university tuition and the academic languages in general (Prinz 2017). In this abstract, we (a) report the development of a corpus of 18th and 19th century lecture notes carried out during three semesters at the University of Zurich and (b) discuss the didactic implications for teaching historical linguistics.

Although there are many research projects (e.g. Verbmobil project in Reithinger & Kipp 1998) in which linguistics students are engaged in transcribing and annotating texts for the creation of a corpus, there are few projects (e.g. RIDGES: Register in Diachronic German Science, cf. Springmann et al. 2018) whose focus is on analyzing and annotating older texts with an annotation guideline within a Bachelor's or Master's study of linguistics. In our seminar on German historical linguistics, we have developed HiVoKo, a lecture corpus from various disciplines with the students. HiVoKo (diachronic lecture corpus) consists of samples (6000-12.000 tokens per lecture) of handwritten university lecture notes from the Age of Enlightenment to the 19th century (lectures delivered by Thomasius, Kant, Lichtenberg, v. Humboldt and others): <https://www.ds.uzh.ch/de/forschung/projekte/hivoko.html>. The corpus offers an empirical basis for linguistic investigation such as:

- communicative needs of different academic faculties and disciplines,
- multilingualism and language choice in academic communication,
- written and spoken language in the history of academic communication,
- scientific terminology.

Although most historical lectures exist only in the form of manuscripts, some of the notes chosen for the corpus are available as printed or digitized editions. However, these editions are often not the exact copy of the handwritten text, i.e. some of the linguistic properties such as writing variations are lost. Therefore, the students were taught to decipher and transcribe the original handwritten texts, comparing them with the printed and digitized texts. The transcription was conducted in form of a simplified XML format so that some textual properties such as headers and character types could be annotated during the transcription. The transcribed texts with XML mark-ups were then automatically segmented into tokens (Sugisaki 2017) and converted to Microsoft Excel documents for further annotations. In the seminar, the following annotation was performed by the students:

- normalized word form,
- types of compound forms,
- text-linguistic and pragmatic properties such as intertextuality and person deixis,
- sociolinguistic aspects of multilingualism and code-switching by means of the “biscriptality” (Kurrent vs. Antiqua) of the text (Bunčić et al. 2016).

For the students, we have provided a guideline for the annotation and assistance during the procedure. At the end of the seminar, the Excel documents with the annotations were compiled to NoSketch Engine (Rychlý 2007), so that the students were able to use the corpus for their term papers.

In the presentation, we will further discuss didactic challenges of creating a corpus with students, such as learning to read historical handwritten texts or the proper use of corpus-linguistic tools, and some prospects for learning linguistics in practice.

References

- Bunčić, D., S. L. Lippert, & A. Rabus (2016). *Biscriptality: A sociolinguistic typology*. Heidelberg: Winter (Akademiekonferenzen 24).
- Klein, T., & Dipper, S. (2016). Handbuch zum Referenzkorpus Mittelhochdeutsch. *Bochumer Linguistische Arbeitsberichte*, 19.
- Prinz, M. (2017). Anmerkungen zur Vorlesungspraxis und Unterrichtssprache im 18. Jh. Paper from the conference *Geschichte der Fach- und Wissenschaftssprachen. Identität, Differenz, Transfer*. University of Würzburg, 13.10.2017.
- Reithinger, N., Kipp, M. (1998). Large Scale Dialogue Annotation in Verbmobil. In *Workshop Proceedings of ESSLI 98*, pages 1–6, Saarbrücken.
- Rychlý, P. (2007). Manatee/Bonito - A Modular Corpus Manager. In *1st Workshop on Recent Advances in Slavonic Natural Language Processing*. p. 65–70.
- Springmann, U., A. Lüdeling, C. Odebrecht, and Th. Krause (2018). Das RIDGES-Korpus. Ein diachrones, tief annotiertes Mehrebenenkorpus aus Kräuterkundetexten. In M. Prinz & J. Schiewe, eds., *Vernakuläre Wissenschaftskommunikation. Beiträge zur Entstehung und Frühgeschichte der modernen deutschen Wissenschaftssprachen* (Lingua Academica 1), Berlin/Boston: de Gruyter, 473–478.
- Sugisaki, K. (2017). Word and sentence segmentation in German: Overcoming idiosyncrasies in the use of punctuation in private communication. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL)*.

AIEL – Ancient Indo-European Languages for the 21st century

Saverio Dalpedri, Götz Keydana, Stavros Skopeteas
Georg-August-Universität Göttingen

OVERVIEW/AIMS. The study of ancient Indo-European languages has a long tradition within the history of linguistics. The findings of historical-comparative linguistics have been cumulated with the developments of general linguistics in recent years, leading to several new resources for the available data, to corpus studies with the methods of current empirical research, and to new hypotheses about the related phenomena inspired by current linguistic knowledge. The aim of AIEL is to strengthen these developments and to create a freely accessible online platform for the teaching of ancient Indo-European languages and the transmission of their findings. Such platform includes teaching units for twelve historical languages: (Old-)Albanian, Ancient Greek, Old Irish, Old Church Slavonic, Old Latin (i.e. pre-classical Latin), Avestan, (Old-)Lithuanian, Classical Armenian, Gothic, Hittite, Vedic Sanskrit and Tocharian.

CONTENTS. The AIEL teaching platform offers a didactically informed introduction to the resources and grammars of ancient Indo-European languages. It is made up by three main modules:

(a) Introductory module:

This module includes introductions to the individual languages in video and text form. The introductions are conceived by international specialists who are recorded in a professional video studio (at least two experts per language). Each teaching unit includes an introduction to the language situation (language community, history, varieties) and to the documentation of the language (available sources, writing system, milestones of linguistic research, lexica) as well as a grammatical outline (phonology, morphology, syntax).

In contrast to the grammars already available for specialists in Indo-European studies, the grammatical description is purely synchronic and is built up according to an onomasiological, or top-down, structure. Up-to-date grammatical information of the respective languages is sketched out and questions which are relevant for the current linguistic research are discussed, e.g. the layers of the prosodic hierarchy, valency changing strategies, or the reflexes of the information structure. It is also to be expected that new insights into the structure of the individual languages will be gained, which can live up to contemporary general linguistics. All descriptions are conceived and written according to a uniform structure; this makes it possible to compare the outlines down to their individual parts, so that the relevant features can be found quickly and easily through the individual chapters.

(b) Illustrative module:

The illustrative module contains different types of illustrations that give the learner an insight into the linguistic or cultural-historical material. This includes one illustrative text per language, discussed along the common levels of linguistic analysis (text in the original orthography, in phonological transcription, in morphemic transcription, provided with glosses and free translation). Furthermore, this module contains language maps, which – depending on the language – show the dispersal of the community of speakers in space, the findspots of documents and/or the distribution of varieties. These are prepared by the GIS Department of Göttingen University's Geographical Institute to ensure a professional and harmonious layout. Finally, culturally relevant artefacts are offered as visual material (e.g. photos of inscriptions, manuscripts, papyri, etc.) and integrated as hyperlinks into the language maps.

(c) Exercise module:

This module contains exercises for the learners and is prepared mainly by students of our study program under the supervision of instructors of Indo-European linguistics. In cooperation with experts in university didactics, this material is made accessible through the e-Learning Management System OpenILIAS.

USER GROUPS. The target audience of the teaching platform includes *researchers*, *teachers* and *learners* of linguistics. Especially the grammatical descriptions take linguistics concepts and methods for granted. They are designed to make the relevant findings of the individual research paradigms accessible to the general public of linguistics, i.e. they do not require the technical jargon or methods proper to a community of experts (e.g. of Indo-European studies or phonological theory). The group of *researchers* will then include not only Indo-Europeanists, but also specialists in general linguistics and especially typologists, who will find data from ancient Indo-Indo-European languages systematized for the first time. The systematic, typologically oriented interpretation of the linguistic material is a core characteristic that sets AIEL apart from a traditional teaching and learning platform. Further user groups can certainly answer questions of a general and encyclopaedic nature in those parts of the teaching platform that do not require linguistic knowledge, e.g. language maps, illustrations, descriptions of the language situation, etc. In addition, the AIEL platform is aimed at students at our university. They will approach the teaching units in the modality of blended learning, so that they will be able to ask questions, deepen additional topics and give short presentations in classroom-based courses. However, they will not only act as *learners*, but will also play an active role in shaping the platform, e.g. by designing interactive exercises, by compiling sections on individual topics about the history or the grammar of the presented ancient Indo-European languages, or by preparing collections of online resources and important bibliographical references to the individual languages.

Developing the "Fifth Skill" through the reworking of historical texts

Peter Anderson

Grand Valley State University, Allendale, USA

Translation from the target language into L1 has sometimes been described as the "fifth skill" in language learning. While it is clear that translation as an assessment practice in language learning and language acquisition presents significant challenges in terms of reliability and validity, translation can be a key methodological tool for promoting vocabulary acquisition and competency with a variety of syntax, both simple and complex. In a similar way, recasting as a pedagogical strategy has been shown in almost every context to be at best problematic and at worst so stressful for the learner as to obstruct learning and acquisition. Nevertheless, some recent studies show that certain kinds of form-focused recasts (oral or written), when supported by learner attention strategies, can lead to significant gains in acquisition among students. In this paper Anderson proposes a potential model for using translation into Latin as a kind of "fifth skill" that avoids the traditional models of so-called prose composition and instead incorporates some best practices around recast, learner attention, and writing-to-learn strategies; these strategies are deployed using Agostino Dati's renaissance style manual *Elegantiolae* and early modern Latin translations of Marcus Aurelius. This model, aligned with the potentials of annotated text corpora that CALL offers, can promote student activity in translation at every level, from vocabulary acquisition, to building competencies in syntax, to the study of both rhetorical structures and style.



Teaching Digital Literacy – Interpretieren in einer computergestützten Lehr-/Lernumgebung

Andrea Beyer (Humboldt-Universität zu Berlin), Paul Reichenanz (Universität Rostock)

Der Vortrag präsentiert einen in den klassischen Sprachen bisher vernachlässigten Ansatz, um die Interpretationskompetenz der Lernenden durch eigenständige (editorische) Aktivitäten auszubauen.¹ Exemplarisch wird gezeigt, wie Studierende der klassischen Philologie mithilfe von Grundkenntnissen in modernen, webbasierten Sprachen (XML, HTML, CSS) lateinische Texte mehrdimensional zu interpretieren lernen. Hierbei steht zunächst die Visualisierung von literarischen Mustern im Vordergrund, wonach in einem zweiten Schritt weiterführende Optionen (linguistische, kulturelle, fachwissenschaftliche Kommentierung) im Rahmen von Editions- oder Forschungsentscheidungen möglich sind.

Den didaktischen Hintergrund für diesen Ansatz bildet die Diskussion um *digital literacy*. *Digital literacy* wird im Anschluss an Gilster (1997), Martin (2006) und Bawden (2008) als ein sehr weit gefasstes Konzept verstanden, das keineswegs auf eine spezielle Technologie oder eine besondere Darstellungsform der Information begrenzt ist. „Digitale Literalität“ bezieht sich daher nicht nur auf einzelne Kompetenzen wie z.B. *information literacy* oder *media literacy*, sondern ist in Fortführung von Humboldts Idee der allgemeinen Menschenbildung als digitale Bildung des Einzelnen zu verstehen, die neben zentralen Kompetenzen auch Grundfertigkeiten (Lesen, Schreiben, kritisches Denken), Orientierungswissen (Hintergrundwissen zu Informationen) und Haltungen (Lebenslanges Lernen, Werte) umfasst. Diese (digitale) Erweiterung des bisher vor allem im Kontext der klassischen Sprachen unterstützten Bildungskonzepts gilt es nun in der Lehre umzusetzen, um Lernende auch im Digitalzeitalter in ihrer zeitgemäßen Selbstbildung zu unterstützen, ohne dabei den fachlichen Schwerpunkt außer Acht zu lassen.

Der zu gleichen Teilen fachdidaktisch und fachwissenschaftlich motivierte Ansatz kann als neuartig beschrieben werden, weil das Studium der klassischen Philologie in Deutschland, das vornehmlich Lehrer ausbildet, traditionell beim Erwerb der sprachlichen und literarischen Fähigkeiten auf analoge Materialien und Methoden setzt. Diese fachspezifischen Fähigkeiten zu entwickeln, ist meist Aufgabe kleinschrittiger Lektüre und (vor allem zu Beginn) mühsamer

¹ Dieser Ansatz folgt der Idee, dass eine Edition „a theory of the text, whether consciously or unconsciously“ hervorbringt (Birnbaum 2018, Folie 3). Da Interpretationen zu einem (subjektiven) Textverständnis führen sollen, wird auch bei ihnen eine „Theorie des Textes“ konstruiert, so dass editorische Tätigkeiten geeignet erscheinen, ein Verstehen des Textes zu fördern.

Interpretationen, da größere Zusammenhänge in den Texten oft unerkannt bleiben. Dass aber der gesamte Prozess des *close readings* durch das *marking up* eines Textes gemäß der TEI-Richtlinien erleichtert wird, ist bereits mehrfach festgestellt worden.²

Der hier vorgestellte Ansatz macht diese Tatsache für die Klassische Philologie fruchtbar: Studierende übersetzen und interpretieren in gewohnter Weise, um anschließend ihre Erkenntnisse mittels HTML und CSS zu visualisieren. Durch das neu gewonnene Medium werden sodann Reflexionsebenen erschlossen, die über das gewöhnliche Maß hinausgehen, wie z.B. eine Kritik der Digitalisierung selbst.³

Der Vortrag zeigt anhand eines Beispiels – ein kleines Korpus aus Ovids Amores – Umsetzungsmöglichkeiten (auch hinsichtlich einer Verankerung als Modul in der universitären Lehre) und diskutiert zentrale Fragen: Welche Voraussetzungen benötigen die Lehrenden und Lernenden? Wie und in welchem Maß trägt dieses Vorgehen dazu bei, die Literaturkompetenz von Studierenden nachhaltig zu verbessern? Welche fachübergreifenden Fähigkeiten werden im Gegensatz zu einer traditionellen Übersetzungsübung erworben?

Schließlich erläutert ein Ausblick, wie der Ansatz auf den schulischen Kontext übertragen und somit ebenso für Schüler und Schülerinnen attraktiv gestaltet werden kann.

Bibliographie:

- Bawden, David (2008): Origins and concepts of digital literacy. In: *Digital literacies: Concepts, policies and practices* 30, S. 17–32.
- Birnbaum, David. *Editions*, Vortrag gehalten am 19.7.2018, Boston, Folie 3, https://moodle.hu-berlin.de/pluginfile.php/2353776/mod_resource/content/1/day04.pdf.
- Fyfe, Paul (2011). “Digital Pedagogy Unplugged.” *Digital Humanities Quarterly* 5 (3). - <http://www.digitalhumanities.org/dhq/vol/5/3/000106/000106.html> (zuletzt aufgerufen: 08.11.2018).
- Gailey, A. (2014). “Teaching Attentive Reading and Motivated Writing through Digital Editing.” *The CEA Critic* 76 (2): 191-199, bes. 194.
- Gilster, Paul (1997): *Digital literacy*. New York, NY u. a.: Wiley.
- Iantorno, L. A. (2014). “Introducing Digital Humanities Pedagogy.” *The CEA Critic* 76 (2): 140-146, bes. 143.
- Kennedy, Kara (2017). *A Long-Belated Welcome: Accepting Digital Humanities Methods into Non-DH Classrooms*. DHQ 11,3 – <http://www.digitalhumanities.org/dhq/vol/11/3/000315/000315.html> (zuletzt aufgerufen am 08.11.2018)
- Martin, A. (2006): Literacies for the Digital Age. In: Martin, A.; Madigan, D. (Hrsg.): *Digital Literacies for Learning*, S. 3–25. London: Facet Publishing.
- Singer, K. (2013). “Digital Close Reading: TEI for Teaching Poetic Vocabularies.” *The Journal of Interactive Technology & Pedagogy*. Issue 3. – <https://jtip.commons.gc.cuny.edu/digital-close-reading-tei-for-teaching-poetic-vocabularies/> (zuletzt aufgerufen: 08.11.2018).

² Vgl. Kennedy (2017: 25) mit Iantorno (2014: 143), Fyfe (2011: 12 und 15), Gailey (2014: 194) u.a.

³ Als Basis für die Adaption dient Singer (2013).

Kurzbiographien

Andrea Beyer studierte Latein, Sport und Informatik, absolvierte im Anschluss das Referendariat und erwarb über das BMBF die Qualifikation für Deutsch als Zweitsprache. Von 2006-2017 arbeitete sie als StR an einer Berliner Schule und promovierte von 2016-2018 bei Prof. Dr. Stefan Kipf („Das Lateinlehrbuch aus fachdidaktischer Perspektive – Theorie, Analyse, Konzeption“). Sie ist wissenschaftliche Mitarbeiterin im DFG-Projekt CALLIDUS. Ihre Forschungsschwerpunkte sind die Lehrbuchforschung, Sprachbildung und der Wortschatzerwerb im LU, die Digitalisierung, Design-Based Research und Andragogik.

Paul Reichtanz studierte von 2010-2017 Latein, Griechisch und Geschichte an den Universitäten Rostock und Groningen. Sein durch das Kurt von Fritz-Wissenschaftsprogramm THEORIA geförderte Promotionsprojekt (2017-2020) untersucht, wie römische Briefautoren der Späten Republik und der frühen Kaiserzeit Rat erteilen, insbesondere in den beiden *Epistulae ad Caesarem senem de re publica*. Seine Forschungsinteressen liegen, neben dem Umgang mit Authentizität und Autorschaft, auf der Frage, wie die Digitalisierung in Forschung und Lehre der Klassischen Philologie angewendet werden können.

Architecture of an Intelligent Tutoring System for Ancient Greek Vocabulary

Marietta Sionti

Excellence Cluster - Cognitive Interaction Technology, Bielefeld University

marietta.sionti@uni-bielefeld.de

In the current abstract, we are presenting ongoing work for designing the architectural scheme of an Intelligent Tutoring System (ITS), which would allow adaptive learning of Ancient Greek vocabulary. Due to machine learning techniques, adaptivity is the main advantage of ITSs, which would allow us to offer individualized material, procedure and hint-giving to users depending (i) on the similarity of Ancient Greek (AG) vocabulary to their mother tongue (Modern Greek, Indo-European or other languages' speakers) and (ii) the stage they are in. Different approach and hint-giving phraseology (Sionti et al, 2012; 2018) should be given to gymnasium/high school students -who are still taught how to achieve metacognition of learning procedure -compared to AG teachers and specialists, who are mainly interested in their domain-specific texts, e.g. Galinos texts for doctors. Since this is an initial step to formulate the ITS, we consider that the DATTEL workshop would provide valuable feedback and collaboration.

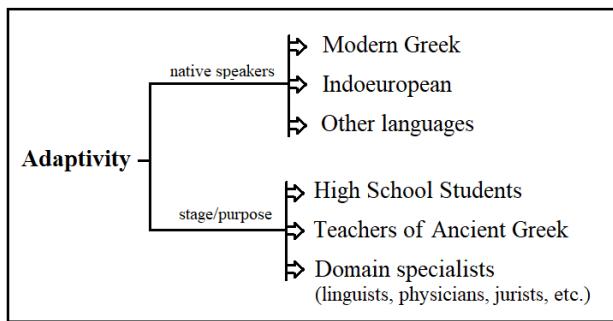


Figure 1. Adaptivity's levels

Architecture of ITSs: According to Freedman (2000), the term *intelligent tutoring system* is quite broad and includes a computer program, which can display artificial intelligence features and be deployed in the educational process. A conventional architecture model consists of four structural components:

- *Cognitive domain model* (according to Freedman *domain model*, according to Corbett *domain knowledge*). It has been used for teaching fractions in mathematics (Koedinger & Corbett, 2006) or planning (Carbonell, 1970).
- Student model
- *Teaching model* (Freedman), (*pedagogical module*, Koedinger & Corbett). Student and teaching models restrict thinking on a particular problem to be solved.
- *Learning environment or user interface* by Freedman (2000) and *problem solving environment* by Koedinger & Corbett.

In particular, learning environment determines the activities that will teach problems' solutions. It usually includes an editor, which can receive text or graphics and shows the student's responses/movements, while the teaching model is significantly more complex. It supports the process of resolving the problem primarily by checking the correctness of student responses and secondarily by controlling the process through allocation of systematic errors. In case of a wrong answer or systematic errors this pedagogic intelligent tutoring system detects the need for assistance, helps students identify their weakness and indicates hints for self-correcting.

The need for assistance depends on the educational level of the student. ITSs use advanced machine learning features to increase productive learning behaviors, in order to support both cognitive and metacognitive process. Specifically, cognitive process is subject to the correct ratio between information and assistance offer and it is useful for all stages of learners, while metacognitive process serves the fundamental principle of learning, namely self-correcting and it is of particular importance for high school students, who would benefit by different strategies to improve self-correcting, such as the procedural repetition and the socratic dialogue (Aleven & Koedinger, 2000).

Concerning the above mentioned parameters of existing ITS, the aroused question is suitable provision of helping. Is it preferable to provide assistance at a student's request or the system should decide the right time to provide advice? Currently, ITS give advice mainly after students' request. According to Baker (2004), there is a possibility that students 'game' the system, in order to avoid thinking. So, they prefer quick guess or frequent help request, without

prior exhaustion of their cognitive resources. On the other hand, a student may avoid seeking for help when he really needs it. In this point, we come to the second part of our question. How and when can a tutor (computer or human) identify the need for hint giving? The intervention is a difficult task, defined by dialogue turns between students and teachers. In the case of our software, we will use simpler automatic hint giving techniques such as segmentation of each individual problem and helping at the end of each sub-problem depending on the time of each phase's completion or failed attempts.

An additional parameter for robust hint giving is the different age of students, in relation to that of aforementioned research. Regarding age group, the majority of literature in ITS focused on high school or college students. Presumably, based on this point we could explain the discrepancy concerning personalized principle. According to Clark & Mayer (2011), instructions and advice should be written in the informal first or second person, while McLaren (2008), unlike his intuition, claimed that only formal style aligns with cognitive theories of learning. Both approaches can be true depending on the age of the students. Linguistic style of intervention is of paramount importance regardless of real or virtual classroom environment and is responsible for the success or failure of learning process, depending on several factors, e.g. age and educational background.

Game features for better engagement for high school students: In order to enrich the design of our tool for high school students only, we incorporate the methodology of (serious) educational games in the student model, which includes a game-like approach. Artificial intelligence has long studied individual educational games, which are not embedded in a general educational model. According to Connolly, Boyle, MacArthur, Hainey and Boyle's review paper (2012:662, 667), there is lack in coherence and organization in literature, which complicates our understanding of the effects of games in learning. However, concerning learning and behavioral outcome, they found that the most frequent results of games were strong motivation and knowledge acquisition followed by cognitive skills and behaviour change, both for entertainment and serious games. Opposite to the ordinary games, serious games do not exclusively aim to entertain but to achieve the learning objective, while maintaining a pleasant atmosphere (Houser & DeLoach, 1998). Due to amusement, students interact with the exercises excluding the apprehension of an error, which will likely result in penalties, such as traditional tests (Bizzocchi & Paras, 2005). Moreover, serious educational games fulfil most of the seven conditions of an educational environment: (a) motivation, (b) reinforcement of challenge and undiminished interest, through playful presentation and endogenous fantasy that excites the student, (c) specific and obvious targets, (d) high level of interaction and feedback, through hint-giving, (e) direct occupation with the subject, (f) evasion of disruptive factors that will divert attention from the subjective experience and (g) offer of suitable tools to support the above mentioned goals, while the latter two conditions are related solely to software implementation.

Moreover, Amory, Naicker, Vincent & Adams's (1999: 319) attempt to link educational with playful elements, a desirable combination for our software. Therefore, in figure 2, we modify their game and visualization space according to our design. Abstract concepts appear in white circles and promote educational goals while specific notions (black circles) implement the goals. So both spaces, game and visualization, serve educational elements and problems in a joyful learning environment.

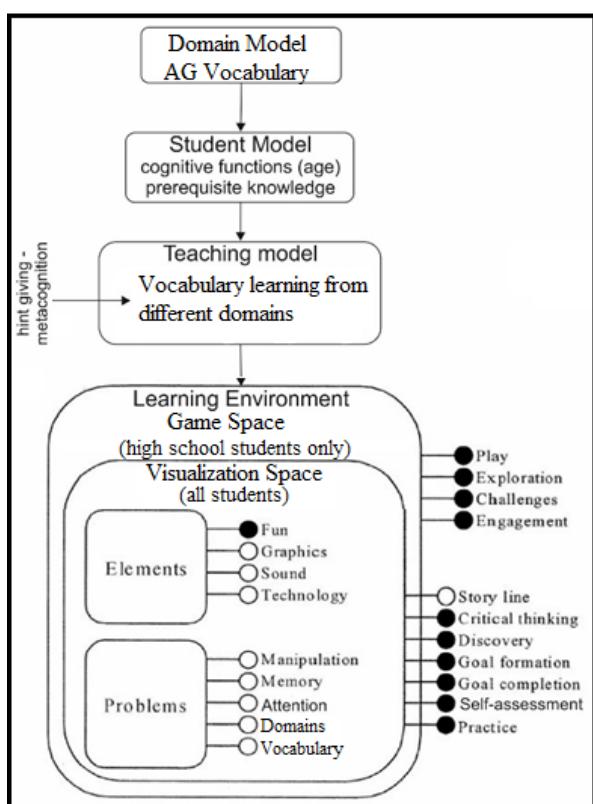


Figure 2. proposed architecture (ITS AG Vocabulary)

References

- Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help?. In *Intelligent Tutoring Systems*, 292-303.Berlin: Springer.
- Amory, A., Naicker, K., Vincent, J., & Adams, C. (2002). The use of computer games as an educational tool: Identification of appropriate game types and game elements. *British Journal of Educational Technology*, 30(4), 311-321.
- Baker, R. S., Corbett, A. T., Koedinger, K. R. & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: when students game the system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 383-390). New York: ACM Press
- Bizzocchi, J., & Paras, B. (2005). Game, motivation, and effective learning: An integrated model for educational game design. In *Proceedings of Digital Games Research Conference (2005)*. Utrecht: Utrecht University
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *Man-Machine Systems, IEEE Transactions on*, 11(4), 190-202.
- Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Wiley. com.Conference: *Changing views—Words in play*. [Retrieved 30-11-2010, from <http://ir.lib.sfu.ca/handle/1892/1623>]
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661-686.
- Freedman, R. (2000). Plan-based dialogue management in a physics tutor. In *Proceedings of the sixth conference on Applied natural language processing*, 52-59. Association for Computational Linguistics.
- Houser, R. & DeLoach, S. (1998). Learning from Games: Seven Principles of Effective Design. *Technical Communication: Journal of the Society for Technical Communication*, 45(3), 319-29.
- Koedinger, K. R. & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. *The Cambridge Handbook of the Learning Sciences*, 61-78.
- Sionti, M., Ai, H., Rosé, C. P., & Resnick, L. (2011). A framework for analyzing development of argumentation through classroom discussions. *Educational technologies for teaching argumentation skills*, 28-55.
- Sionti, M., Schack, E. & Schack, T. (2018). Hint-giving phraseology for Computer Assistive Systems. In ICL 2018-The challenges in digital transformation of education, Kos 2018. In press by Springer series "Advances in Intelligent Systems and Computing"

MULLE for Latin: Computer-Generated Translation Exercises for Latin

Herbert Lange

Computer Science and Engineering

University of Gothenburg and Chalmers University of Technology

`herbert.lange@cse.gu.se`

1 Introduction

The MUSTE Language Learning Environment (MULLE)¹ is a universal and flexible software application to automatically generate language learning exercises, like translation exercises, based on computational grammars (Lange and Ljunglöf 2018a).

One of its first applications is the generation of translation exercises in the context of teaching Latin as a foreign language at high-school or undergrad university level (Lange 2018). Here we present the ideas behind the systems and show which features make it particularly applicable to teach historic languages like Latin.

2 General Mechanics

We divide the language learning task into lessons following a long tradition in language learning. Each lesson is based on a separate grammar which is used to generate a set of exercises. To finish a lesson the student has to finish a certain number of exercises within this lesson.

In a translation exercise the user is presented with two sentences in different languages, one language the student is learning and one that is used to teach this language. In a tradition originating from Logic we call the first language the object language and the second one the meta-language. To both sentence one or several syntax trees are assigned which correspond to the analyses of the sentences according to the grammars. These syntax trees are language independent and can be used to represent sentences across languages.

The student can then change one of the sentences by using simple point-and-click actions on the surface string. The editing operations, like changing words or phrases, inserting new ones or deleting existing ones, are accessible from list of suggestions. They are translated to operations on the underlying trees, following a method presented in (Ljunglöf 2011). The grammar guarantees at each editing step that the new sentence is still grammatical according to it.

The translation task is finished as soon as at least one of the syntax trees of the two sentences matches with a syntax tree for the other one.

The system itself poses no restrictions on where the grammars for the lessons are coming from, as long as they are multilingual in at least the two languages, the meta- and the object language. One way to get lesson grammars is to extract them from an existing textbook used in language classes (e.g. Lindauer, Westphalen, and Kreiler 2000; Ehrling 2015). The lessons in such a book usually consist of a text fragment, an explicit list of vocabulary to be learned as well as some explanation of grammatical phenomena and exercises to be solved on paper.

¹<https://github.com/MUSTE-Project/MULLE>

We adopt a process to convert the information given in the lesson text and the vocabulary list into a grammar that can be used in the language learning application (Lange and Ljunglöf 2018b) (Figure 1):

1. The first step is to adapt a lexicon for the textbook lesson, which often is given as an explicit vocabulary list in the book. It is possible to use existing reliable lexical resources or morphological information (Détrez and Ranta 2012) in this step.
2. The next step is to create syntax trees for all sentences in the text. This can either be done manually or semi-automatically. To automate this process, each sentence is parsed using syntactic resources (Ranta 2009) augmented with the lexicon from step 1. Because of syntactic ambiguities this might result in several possible trees, so afterwards the correct syntax tree, i.e. the desired analysis of the sentence, has to be selected manually. This involves some linguistic knowledge from the person creating the grammar.
3. Finally, the grammar that describes the text fragment in the textbook can be extracted. This can be done straightforwardly by reading off the grammar rules from the internal nodes in the trees. This will usually result in an over-generating grammar, so different techniques to reduce the over-generation such as merging several rules into one can be applied.

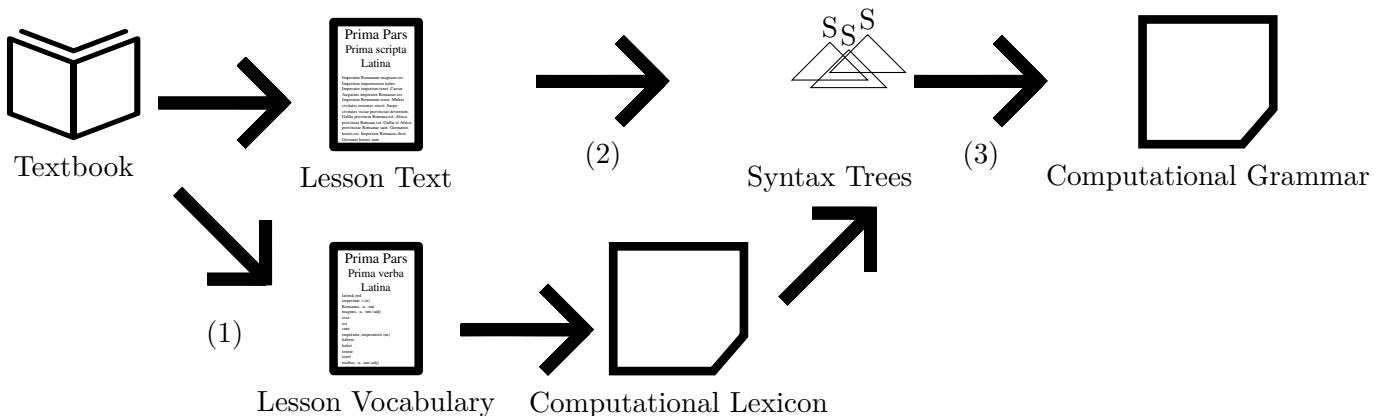


Figure 1: Process for the extraction of a grammar from a textbook lesson

This method is only one of several possible approaches which can even be used to include corpora in the generation of text exercises.

3 Usability for Historic Languages

A few aspects of our current approach speak in favor of using it in teaching Latin and other historic languages.

The first aspect that speaks in favor of our approach is that the main competence required among learners of historic languages is mostly translation competence. This competence usually is trained by translating texts and text fragments on a piece of paper. Especially in early stages of learning a language coming up with translation exercises that match the learner level and have a connection to authentic language data is challenging. The approach we present generates a large quantity of flexible and reliable language exercises that are placed in a familiar context given by the content of the language classes (Lange and Ljunglöf 2018b).

The second point is that learning historic languages can lead to a negative learner motivation even up to states of anxiety (Dimitrijevic 2017; A. Takahashi and H. Takahashi 2015). The method we intend to use to tackle this problem is the use of gamification (Deterding et al. 2011), the augmentation of potentially tedious tasks with features of games to make them more enjoyable. Several approaches to gamification have been suggested including GameFlow (Sweetser and Wyeth 2005), of which we adopted the following concepts: *Concentration*, i.e., minimizing the distraction from the task with a

goal-directed user interface design, *Challenge* by assigning a score to the exercises by using a timer, *Control* by providing an intuitive way to modify the sentence, *Clear goals* by adopting a lesson structure and show the progress within the language learning process, and *Immediate feedback* with a color schema to show the progress within an exercise(Lange and Ljunglöf 2018a). Most of them can be seen in the screenshot in Figure 2.

Finally, the source for the teaching material can be addressed. The method for extracting grammars seems to work well for existing textbooks. That means it can be used in a straightforward way at least in the cases where the historic language has a tradition of being taught in usual language classes and with existing textbooks. However, the method can in principle be used to extract a grammar from any corpus. This allows for broader learning material but lacks the implicit pedagogics that is already present in the lesson structure of textbooks.

4 Conclusion

Even though the design of our language learning environment allows, due to its flexibility, the application in various areas of Computer-Assisted Language Learning, it seems specifically suitable for historic languages.

This is among other things due to the main focus on translation exercises which is suitable for historic languages where the main competence required is translation. Also, different from other modern approaches to language learning a very limited amount of language data is required to generate a large quantity of translation exercises.

The data that can be used in the translation exercises can either be sourced from the content of an existing language class or from authentic texts.

An evaluation still remains for the future but we already collected feedback in collaboration with teachers and students of a beginner's class in Latin at the University of Gothenburg. We got positive and encouraging responses from both groups which can be used as a source for further development.

References

- Deterding, Sebastian et al. (2011). "From Game Design Elements to Gamefulness: Defining "Gamification"". In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. New York: ACM, pp. 9–15.
- Détrez, Grégoire and Aarne Ranta (2012). "Smart Paradigms and the Predictability and Complexity of Inflectional Morphology". In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. EACL '12. Avignon, France: Association for Computational Linguistics, pp. 645–653. URL: <http://dl.acm.org/citation.cfm?id=2380816.2380895>.
- Dimitrijevic, Dragana (2017). "Latin Curricula, Attitudes and Achievement: An Empirical Investigation". In: *Proceedings of the 19th International Colloquium on Latin Linguistics*.
- Ehrling, Sara (2015). *Lingua Latina novo modo – En nybörjarbok i latin för universitetsbruk*. University of Gothenburg.
- Lange, Herbert (Sept. 2018). "Computer-Assisted Language Learning with Grammars. A Case Study on Latin Learning". Licentiate thesis. Gothenburg: Department of Computer Science and Engineering. University of Gothenburg.
- Lange, Herbert and Peter Ljunglöf (2018a). "MULLE: A Grammar-based Latin Language Learning Tool to Supplement the Classroom Setting". In: *Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA '18)*. Melbourn. Australia: Association for Computational Linguistics, pp. 108–112. URL: <http://aclweb.org/anthology/W18-3715>.
- (2018b). "Putting Control into Language Learning". In: *SIGCNL 2018*. Maynooth, Ireland.
- Lindauer, Josef, Klaus Westphalen, and Bernd Kreiler (2000). *Roma, Ausgabe C für Bayern, Bd.1*. C.C. Buchner.

- Ljunglöf, Peter (2011). “Editing Syntax Trees on the Surface”. In: *Nodalida’11: 18th Nordic Conference of Computational Linguistics*. Riga, Latvia.
- Ranta, Aarne (Dec. 2009). “The GF Resource Grammar Library”. In: *Linguistic Issues in Language Technology* 2(2). URL: <https://journals.linguisticsociety.org/elanguage/lilt/article/view/214/158.html>.
- Sweetser, Penelope and Peta Wyeth (July 2005). “GameFlow: A Model for Evaluating Player Enjoyment in Games”. In: *Computers in Entertainment (CIE)* 3.3, pp. 3–3. URL: <https://dl.acm.org/citation.cfm?id=1077253>.
- Takahashi, Ayumi and Hideki Takahashi (Aug. 2015). “Anxiety and Self-Confidence in Ancient Language Studies”. In: *Niigata University Language and Culture Research Department Bulletin*.

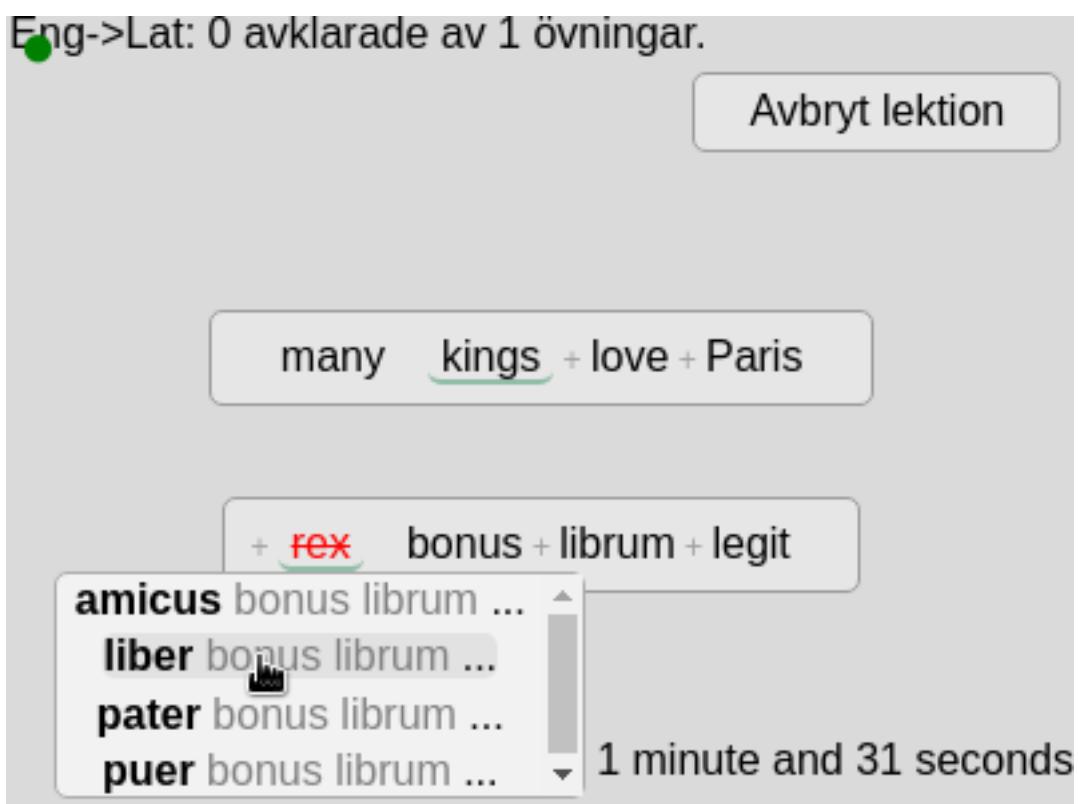


Figure 2: Screenshot of the application showing various concepts of gamification like *Concentration*, *Challenge* and *Immediate Feedback*