Adjusting LERA For The Comparison
Of Arabic Manuscripts Of *Kalila wa-Dimna*

Beatrice Gründler

[beatrice.gruendler@fu-berlin.de](mailto:beatrice.gruendler@fu-berlin.de)
Freie Universität Berlin

Marcus Pöckelmann

[marcus.poeckelmann@informatik.uni-halle.de](mailto:marcus.poeckelmann@informatik.uni-halle.de)
Martin Luther University Halle-Wittenberg

**Introduction**

In this paper, we present the collaboration between the pilot project *Kalila wa-Dimna – Wisdom encoded*¹ and LERA.² In the project's first phase, devoted to experimenting with existing tools and identifying necessary adjustments, we adopted and generalized LERA for the classical Arabic language. This modification worked well and thus will become a cornerstone for future research within the ERC-Advanced Grant Project “AnonymClassic” (Gruendler 2017).³ The benefit is already apparent, yielding first observations of the text's development, and the tool was successfully applied in an undergraduate academic course at the Seminar for Semitic and Arabic Studies of FU Berlin.

*Kalila wa-Dimna – Wisdom encoded*

Using Sanscrit sources, *Kalila wa-Dimna* was composed in Middle Persian (version lost) in the 6th century CE and ultimately translated into in forty languages worldwide. Its key phase is the Arabic translation from the 8th century, from which all later translations derive. But this version has proliferated in many variants, which prevents a conventional critical edition by stemma (Gruendler 2013). This project seeks to assess via a (partial) synoptic critical edition the range of variation between selected Arabic manuscripts of this work. These derive from the 13th to the 19th century CE.

---

1 E-Learning/E-Research project, located at and funded by Freie Universität Berlin, homepage [http://www.geschkult.fu-berlin.de/e/kalila-wa-dimna](http://www.geschkult.fu-berlin.de/e/kalila-wa-dimna)
2 Information and a demonstrator can be found at [https://lera.uzi.uni-halle.de](https://lera.uzi.uni-halle.de)
LERA – Locate, Explore, Retrace and Apprehend complex text variants

LERA is an interactive, digital tool for analyzing variations between multiple versions of a text in a synoptic manner with several differences to other well known collation tools (Schütz and Pöckelmann 2016). It was first developed for printed texts of the French Enlightenment (Bremer et al. 2015) within the SaDA-project⁴ at Martin Luther University Halle-Wittenberg and since then adopted to other texts and languages.

The tool follows three major steps for generating the synopsis. The first is a segmentation of the given manuscripts. In the case of *Kalīla wa-Dimna*, the text-units are narrative steps. Second, an automatic alignment of these segments is done by the built-in algorithm. The researcher can intervene afterwards by moving, cutting or merging the segments if necessary. The final step is the detailed comparison of the aligned segments’ words by the system. The identified variants are highlighted with colors depending on the kind of difference. Besides, a variety of filters are available, e.g., hiding all changes that purely concern punctuation. On this basis, a comprehensive and easily readable apparatus is generated. The proto-edition thus produced can be downloaded in several formats.

Moreover, LERA provides further assistance for the analysis of the variants. Most prominent is CATview (Pöckelmann et al. 2015), a graphical representation of the alignment that facilitates overviewing and navigating within the synopsis.⁵ It is also associated with the word clouds and search functions of LERA.

Modification of LERA for *Kalīla wa-Dimna*

In this project, LERA made its debut in classical Arabic, which has required some language-specific adoptions. Processing the Arabic alphabet was rather uncomplicated, because the system already uses Unicode. Regarding the backend, the processes for tokenizing, indexing (for search) and language recognition were extended. On the other hand, modifications for the frontend included adding a font for the alphabet, displaying the correct writing direction, and revising the download function.

More important, some specific needs for the *Kalīla wa-Dimna* project have already been implemented. LERA now allows the manual alignment by experts using unique segment identifiers, which are encoded within the manuscripts’ XML/TEI files. On this basis, we also added identifiers for the segments that can be edited and displayed in the synoptic view. On major improvement is the visualization of transposed segments. They occur if the order of the segments within one manuscript was changed or when similar segments appear somewhat distant to each other, but aligning them is blocked due to other aligned segments. We included an option to display copies of

---

⁴ See the project’s homepage: [https://sada.uzi.uni-halle.de](https://sada.uzi.uni-halle.de)
⁵ Further information on CATview is also available at [https://catview.uzi.uni-halle.de](https://catview.uzi.uni-halle.de)
them in the synopsis that will be shown grayed and are linked to their actual position. They will also appear in CATview.

In respect to the project's goal to investigate the interrelation of the manuscripts of *Kalīla wa-Dimna*, we developed two new modes for coloring the variants. The first one only highlights passages that are unique to one manuscript, which points to some independence of the copyist-redactor regarding the other manuscripts. The second mode only highlights those passages that appear in exactly two manuscripts. Finding such pairs is important evidence that suggests that both manuscripts are related to each other.

Another helpful extension is the so called 80%-filter, which leads to treating words as identical if they share at least 80% of their letters. This approximating similarity measure is grounded in the property of the Arabic language that words with identical roots tend belong to one semantic field.

**Benefits**

LERA could be adjusted for the first phase of this interdisciplinary collaboration. Based on this, we already made interesting observations: against our assumption, the first analysis shows that there are no distinct groups of manuscripts. Instead, variations fluctuate, forming continua in which some manuscripts cumulatively assemble reformulations that appear scattered among others.

Furthermore, in the summer semester of 2017 the project was integrated into an undergraduate academic course on *Kalīla wa-Dimna* at Freie Universität Berlin. The students used the synopsis to explore the variants of five aligned manuscripts in class and wrote papers applying this method individually.

**Conclusion**

With the work presented here, we established a foundation for a comprehensive analysis of *Kalīla wa-Dimna*. Owed to the text’s complex history and manifold variants, this ambitious project is planned for a timespan of ten years. With the ongoing research, more features of analysis will be needed. This includes an advanced utilization of language specific information for the comparison, e.g., a root extraction for Arabic words. Moreover, the comparison and visualization of manuscripts of *Kalīla wa-Dimna* in other language is being considered. Finally, the functionality to comment on the identified variants is crucial for their scientific investigation.
References


