Université de Lorraine — IDMC
Masters SC & TAL — M1
2022 – 2023

Project Form

Variant analysis of the Ben Sira manuscripts

Supervisor:

- Laboratoire *Ecriture*
- Main supervisor: J.S Rey jean-sebastien.rey@univ-lorraine.fr
- Other supervisor: S. Robert sophie.robert2@etu.univ-lorraine.fr

Description:

1. **Global Description**: Written around 180 BC, the book of Ben Sira has been copied thousand of times over the years, and variants have been made by scribes over the copying process: some words were changed by introducing synonyms, some words were mispelled, some were added, some forgotten . . . Jean-Sébastien Rey, supervisor of this project, has recently finished working on a critical edition (an edition that compares every variants found on the manuscripts) of these manuscripts.
   The goal of this project is to perform a stastical analysis of the variants attested in the nine manuscripts found at Qumrân, Massada and in the Cairo Genizah and dating back from the first century BC to the twelve century CE. The students are expected to provide, through computational tools:
   - An alignement table (the phylology word is **collation**) for variant detection (we recommend to use the Python port of Collatex[a])
   - A report on the analysis of the variants: how many letter substitutions were there ? What letters were most frequently substituted ? How many words were switched for their synonyms ? How many words were deletted or added. . . ? We expect the students to be as creative as possible when it comes to the discovery and analysis of variants (new questions will probably arise as the project advances)

   The additional challenge of this project is that all the manuscripts are written in **ancient Hebrew**, a right to left langage, with a specific alphabet. The students will be required to adapt to this challenge, either by using Hebrew specific tools (such as Dicta) or performing parallels with the English language and use more standard tools (Spacy, . . .) (or a combination of both).
   We also encourage students to be as flexible as possible when it comes to their implementation of their solution for manuscripts analysis: rather than simply using scripted language, developing a standardized library enabling other users to use it with other manuscripts would be a huge plus for the project's success.

2. **Bibliography** (UE 705, semester 7) To our knowledge, automatic variant analysis using computational tools has not yet been done. For example of variant analysis, we refer students to the work of Parry on the analysis of the variants of the Isaiah book (Parry 2020): the computational tools should allow to perform a similar results (or at least, considerably speed-up the process) of the works performed by the philologist. A thorough classification of possible variants that students should look for can be found in the work of E. Tov (Tov 2012, students should particularly focus on pages 219 to 262 of the third edition).

3. **Implementation** (UE 805, semester 8) We imagine the implementation of the project as follows:

   (a) Data parsing of the Qumran and Cairo Genizah manuscripts from the XML TEI files, given by J.-S. Rey

   (b) Collation of data using Collatex

   (c) Identification of variants using the collation methods

   (d) Classification of variants: is it a synonym exchange ? Is it a mispell ? Has a word has been added, removed, transposed,. . . ?

   (e) Redaction of a report summarizing the students findings regarding scribal behavior and the statistical repartition of variants (expressed as rates, so for example, 10% of variants are mispells, 5% are synonym exchanges...)

   We **strongly** encourage student to develop code as generic as possible for easy re-usability (Python would be best), and if bundled as a library, the code will be uploaded on GitHub as OpenSource and advertised to other scholars.

[a] The Collatex website will provide students with a lot of information regarding alignment tables and collations

Information: Dataset will be provided by J.-S. Rey; Data is in ancient Hebrew;

Deliverables and Schedule: The expected deliverables are:

- The alignement tables of the different Ben Sira manuscripts
- The classification of the found variants
- A report regarding the findings and characterization of the variants and their statistical repartition

References

Parry, Donald W. (October 2020). “Exploring the Isaiah Scrolls and Their Textual Variants”. In: *The Journal of Theological Studies* 71.2, pp. 845–847. eprint: https://academic.oup.com/jts/article-pdf/71/2/845/38682393/flaa126.pdf.

Tov, Emanuel (January 2012). *Textual Criticism of the Hebrew Bible*. Third edition, revised and expanded. Minneapolis, Minnesota: Fortress Press.