Lithuania

Report on preparing texts for ELTeC

how was the project organized ?

Arūnas Gudinavičius

Saulius Keturakis

Vincas Grigas

Selected all possible works from epaveldas.lt
Split all the list in to three equal parts
Discussed via emails on how to do it following the requirements

what is the current state of the collection ?

- **Texts** - 25 (red). Could be uploaded up 30 texts. Nothing more to include :(We had a ban on Lithuanian language until 1905.

- **Words** - 636132. Average words in a text - 25445. Shortest of all counties.

- **Gender** - male 18 and female 7. No possibility to select. Just the fact.

- Most of texts are short (19), medium (3) and long (2, red). No source to select more longer texts.

- **Period** - most of works from 1900-20 (14), after allowing texts in Lithuanian.

- **E5C** - 55.38. Not red, but not green also.

how were texts selected?

- Had no possibility to select.

- Just grabed everything what met the criteria.

It is true that the selection of the LT corpus was not very selective, as the texts were lacking in quantitative terms.

The novel as a genre in Lithuanian was very late compared to the larger European literature, so reaching the prescribed volume of texts is impossible.

- The source is **Lithuanian cultural herita**

https://www.epaveldas.lt/en/home

how did you learn how to use TEI ?

From scratch on TEI, no previous knowledge

Had previuos knowledge on XML, HTML

- Looked for examples at ELTEC (already were submissions)
- Participated in seminars organised by the COST
- Looked for information on Internet - **https://tei-c.org/guidelines/**

how were texts processed?

- Use TXT as a source
- Convert into TEI P5 XML scheme with https://oxgarage.tei-c.org

For editing and finalising:

S. Keturakis used – Oxygen XML Editor (has lots of scenarios)

A. Gudinavičius – Atom (has an error checker)

V. Grigas – Notepad++ and Atom

what problems did you encounter?

- Too **few texts in Lithuanian** met all the criteria.

It was a big pity and strong blow to our pride

- The problem is the **genre unequal nature** of the texts contained in ELTEC.

Small language issue – not only novels, but also another prose or even drama. The set becomes unsympathetic from the genre's point of view. Not a slice of novels, simply a slice of literature in general.

- **Wrong interpretation** what has to be disclosed using certain TEI tags

Got valuable support from **Lou Burnard. Many thanks!**

- Have some works in **Bounced** folder. Do not know how to proceed.

https://github.com/COST-ELTeC/ELTeC-lit/tree/master/bounced

Man who saved Lithuanian language

Was one of the main organizers of the illegal book-smuggling at the time of the Lithuanian press ban (1864–1904).

Bielinis is informally referred to as the King of Book Smugglers.

In Eastern Prussia, more than 3 500 000 copies of publications in Lithuanian were published between 1864 and 1896.

Almost everything had to be smuggled into Lithuania using very dangerous routes.