



COST ACTION IS1305 European Network of e-Lexicography

Location: Austrian Academy of Sciences, Dr. Ignaz Seipel-Platz, 1010 Vienna, Austria

Present:

WG 2 members: Vladimír Benko, Jesse de Does, Katrien Depuydt, Kseniya Egorova, Jens Erlandsen, Yan Greub, Andrew Hawke, Vera Hildenbrandt, Alexandre Huber, Attila Mártonfi, Nathalie Mederake, Tobias Roth, Klaas Ruppel, Elena Tamba, Lidija Tanturovska, Toma Tasovac (8 members of WG2 excused, 13 members: no feedback)

Guests: Jack Bowers, Martin Everaert, Oddrun Grønvik, Halldora Jonsdóttir, Kristian Kankainen, Ilan Kernermann, Dirk Kinable, Rosana Lotski, Chris Mulhall, Maria Pober, Eveline Wandl-Vogt

Meeting on Friday 13-02-2015, 9.00-13.00

1. Opening & welcome

Vera Hildenbrandt opened the meeting and welcomed the participants.

2. Organisational matters

Currently, WG2 has 37 members from 18 countries.

New vice chair of the WG is Toma Tasovac who replaces Vladimír Benko. Vera thanked Vlado and welcomed Toma.

The participants agreed on the agenda which, differing from the announced agenda, included one more talk from Oddrun Grønvik.

3. Keynote

Jack Bowers (INRIA): Recommendations for encoding etymological information using TEI

Jack discussed the lack of markup models and standards to model etymological information and made concrete proposals on how the TEI Guidelines could be extended to accommodate better and more consistent handling of etymological data.

4. Presentations:

- a) Lidija Tanturovska** (Institute of Macedonian Language “Krste Misirkov”, Sts. Cyril and Methodius University): An overview of potential Macedonian retro-digitised dictionaries

As a new member of the Action, Lidija provided a useful overview of Macedonian dictionaries that could be digitized in the future and raised the difficult question of copyright issues with more recent dictionaries.

- b) Klaas Ruppel** (Institute for the Languages of Finland): Overview over retro-digitised dictionaries of the finish languages

The Institute for the Languages of Finland no longer publishes printed dictionaries; bonuses to its different online dictionaries are conjugation and declension tables, maps, timetables.

- c) Attila Mártonfi**: Overview over Hungarian dictionaries

Attila's vision is a meta-dictionary of 132 dictionaries (255 volumes) of different type (explanatory, phraseological, technical, etymological, dialectal, taxonomic, slang, authors')

- d) Elena Tamba** (Romanian Academy of Iasi): Digitization of Romanian dictionaries

Elena discussed steps in retro-digitization (scanning, image processing, OCR processing.)

- e) Oddrun Grønvik** (University of Oslo): Taking older dictionaries in the electronic age

Oddrun presented an overview of retro-digitization projects in Norway with a special focus on links between facsimiles and transcriptions, searching and sorting mechanisms as well as transcription difficulties.

- f) Nathalie Mederake** (Göttingen Academy of Sciences and Humanities): Some thoughts on enrichment of retro-digitized dictionaries

Nathalie discussed different possibilities of content-related enrichment of dictionaries (for example the building of quotations catalogues as possibility of interlinking dictionaries).

Local organizers of the meeting will make all presentations available online.

Recurring topics in the presentations showed the necessity of discussing

- the pre-digitization analysis according to the users' needs and expectations
- the classification of "digital" dictionaries (paper, paper + PDF, full text)
- copyright questions
- the interlinking of dictionaries
- questions of data enrichment
- visualization techniques and tools.

All these issues have consequences for the retro-digitization process and will be taken up by the working group for further analysis and discussion in subsequent meetings.

5. Progress Report

Unfortunately, time was too short to give insights in the progress of the task groups. Therefore, it should be referred to the annual report.

6. Division of Tasks / Subgroups (preparation of the next meeting)

Leaders of task group 1 (overview of dictionaries that have been and are to be retro-digitised) are Katrien Depuydt and Jesse de Does. Members interested in participating in this task group are Janusz S. Bień, Joanna Bilińska, Dominik Brückner, Attila Mártonfi, Nathalie Mederake, Klaas Ruppel, Mateusz Sarnecki.

Leader of task group 2 (description of the process of retro-digitisation) is Vera Hildenbrandt. Members interested in participating in this task group are Marius Clim, Kseniya Egorova, Ioana Galleron, Lidija Tanturovska, Geoffrey Williams.

Leader of task group 3 (development of standards for the encoding of information) is Toma Tasovac. In the general discussion of Task Group 3, a decision was made to collect samples of encoded dictionary data (in TEI, but also other formats) which would reflect the heterogeneity of lexical data (structural, syntactic, semantic) and to create a publicly accessible repository of multilingual examples, which will be used not only during the Summer School, but as a general scholarly resource. No such resource, focusing on retro-digitized dictionaries, currently exists. Members of the working group agreed to contribute to this collection.

Nominated leader of task group 4 (investigating relevant information categories to be added to the dictionary in order to make the dictionary content more readily accessible and interoperable) is Nathalie Mederake.

7. AOB

There was no AOB.

8. Closing

Toma and Vera closed the meeting and thanked the participants for their presence and cooperation.