Call for Papers

Workshop
Processing Language Variation:
Digital Armenian (DigitAm)

2022 International Conference on Language Resources and Evaluation (LREC 2022)
Marseille, France, June 20, 2022

The workshop **Processing Language Variation: Digital Armenian (DigitAm)** will be held within the framework of the 2022 International Conference on Language Resources and Evaluation (LREC 2022)**,** Marseille, France, **afternoon June 20, 2022**. DigitAm is in line with the international conference Digital Armenian first held in Paris in 2019 and is organized by the team of DALiH[1] project: Digitizing Armenian Linguistic Heritage (DALiH): Armenian Multivariational Corpus and Data Processing.

Workshop Objectives

A significant gap exists for the availability of NLP resources for different languages with a few languages having quasi-complete NLP coverage and many others being under-resourced (or no-resourced at all). Besides, the under-resourced languages can often have variation either at synchronic (dialects, oral vernacular varieties) or diachronic level (ancient variants of a target language) for which resources can be completely absent especially if no written tradition exists for a target variety. The workshop will focus on processing and reutilization of NLP resources for under-resourced languages with variation in general, with a particular attention to the Armenian language data.

Current state-of-the-art NLP approaches open up remarkable perspectives not only to exploit the available NLP resources of the well-resourced languages for the under-resourced ones, but also to recycle the existing resources of a target language for its varieties (multivariational resources) instead of processing target language/variety-based new NLP resources from scratch.

Workshop Topics

Main workshop foci will be the interoperability of NLP and linguistic resources and tools in particular (but not limited to) for multivariational under-resourced languages, multivariational corpora designing and functionality, the evaluation of language scalar variation and the degree of interoperability relevance, language variety identification and distance measuring.

- interoperability of NLP and linguistic resources and tools in particular (but not limited to) for multivariational under-resourced languages /transfer learning to under-resourced language,
    - morphological annotation resources
    - syntactic annotation resources
    - text-to-text alignment resources
    - speech-to-text alignment resources
- multivariational corpora designing and functionality,
- the evaluation of language scalar variation and the degree of interoperability relevance,
- language variety identification and distance measuring,
- various linguistic resources and tools (dictionaries, electronic libraries, corpora, manuscript collation etc.) for under-resourced languages, in particular for Armenian (but not limited to).

Submission

Proposals must:
- consist of **4 to 8 pages, references excluded**
- comply strictly to the LREC stylesheet
- be formatted as a PDF
- submitted via START Conference Manager: https://www.softconf.com/lrec2022/Armenian/

Submissions can be of two types: oral and poster papers. Oral presentations will be of 15 minutes and posters will be presented in a poster session. There is no difference in quality between oral and poster presentations. Only the appropriateness of the type of communication (more or less interactive) to the content of the paper will be considered. The type of the submission will be selected by the programme committee.

Proceedings

Workshop Proceedings will be published on the LREC 2022 website and they will include both oral and poster papers, in the same format. All the proceedings should be strictly in compliance with the to the LREC stylesheet.

Working languages: English or French

Timelines

Submission deadline (extended): **April 26, 2022**
Notification of acceptance: **May 3, 2022**
Submission deadline for camera-ready versions: **May 23, 2022**
Submission deadline for sending the Workshop Proceedings: **May 31, 2022**

Identify, Describe and Share your LRs!

- Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about "Sharing LRs" (data, tools, web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of sharing LRs, linked to the LRE Map for their description, may become a new "regular" feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.

- As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2022 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number (ISLRN, www.islrn.org), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.

Organizing Committee

Victoria Khurshudyan, Inalco, Sedyl, CNRS
Anaid Donabedian, Inalco, Sedyl, CNRS
Chahan Vidal-Gorene, École Nationale des Chartes-PSL
Nadi Tomeh, LIPN, Université Sorbonne Paris Nord
Damien Nouvel, INALCO, ERTIM

Programme Committee

Victoria Khurshudyan, Inalco, Sedyl, CNRS, IRD
Anaid Donabedian, Inalco, Sedyl, CNRS, IRD
Chahan Vidal-Gorene, École Nationale des Chartes-PSL
Nadi Tomeh, LIPN, Université Sorbonne Paris Nord
Damien Nouvel, Inalco, ERTIM
Emmanuel Cartier, LIPN, Université Sorbonne Paris Nord
Thierry Charnois, LIPN, Université Sorbonne Paris Nord
Ilaine Wang, Inalco, ERTIM
Vladimir Plungian, Vinogradov Russian Language Institute, Russian Academy of Sciences
Timofey Arkhangelskiy, University of Hamburg