



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable D3.2.23

Task 3

ELRC Workshop Report for Slovenia



Author(s):	Simon Krek, Teja Goli
Dissemination Level:	Public
Version No.:	V1 Final
Date:	2022-11-29



Contents

1	Executive Summary	3
2	Workshop Agenda	4
3	Summary of Content of Sessions	6
3.1	Welcome and introduction	6
3.2	eTranslation	6
3.3	Update from the ELRC	6
3.4	Machine Translation for Slovenian	6
3.5	Speech Technology for Slovenian	7
3.6	Semantic Technology for Slovenian	7
3.7	Slovene Terminology Portal	7
3.8	CEF FedTerm Project	7
3.9	National program on encouraging the development and use of artificial intelligence by 2025	7
3.10	Q/A session	7
4	Synthesis of Workshop Discussions	8
5	Country Profile: Language data creation, management and sharing	8

1 Executive Summary

The 3rd ELRC workshop in Slovenia, was collocated with the **European Language Grid workshop**, and it focused on recent developments in Slovenian LT, mainly supported by an extensive national language technology project – **Development of Slovene in a Digital Environment** (DSDE - <https://www.slovenscina.eu/en>, 2020-2023) funded by the Slovene Ministry of Culture and the European Regional Development Fund. The topics covered by the DSDE project and presented at the workshop included Machine Translation, Speech Technology, and the Slovenian Terminology Portal. Additionally, the FedTerm project, funded by the Connecting Europe Facility, was presented covering international aspects of terminology activities in Slovenia.

The current National program on encouraging the development and use of artificial intelligence was presented by a member of the Government Office for Digital Transformation. In addition to the members of ELRC and ELG projects presenting activities from the two initiatives, other presenters were mainly the coordinators of different activities in the DSDE project, and the government officials covering Artificial Intelligence programs. The whole program was recorded and is available on the Videlectures.net web site: http://videlectures.net/elrc2022_ljubljana/.

2 Workshop Agenda

The final workshop program in English

	ELRC WORKSHOP
08:00 – 09:00	Registration
09:00 – 09:15	Welcome and introduction <i>Simon Krek, Jozef Stefan Institute and Centre for Language Resources and Technologies, University of Ljubljana</i>
09:15 – 09:40	eTranslation <i>Miha Žličar, European Commission</i>
09:40 – 10:05	Update from the ELRC and new tools on the CEF platform <i>Thierry Declerck, DFKI, European Language Resource Coordination</i>
10:05 – 10:30	Panel on ELRC and eTranslation seen in the Slovenian Context
10:30 – 10:45	Coffee Break
10:45 – 11:00	Presentation of Machine Translation for Slovenian <i>Iztok Lebar Bajec, Faculty of Computer and Information Science</i>
11:00 – 11:15	Presentation of Speech Technology for Slovenian <i>Marko Bajec, Faculty of Computer and Information Science</i>
11:15 – 11:30	Presentation of Semantic Technology for Slovenian <i>Slavko Žitnik, Faculty of Computer and Information Science</i>
11:30 – 11:45	Slovenian Terminology Portal <i>Miro Romih, Amebis</i>
11:45 – 12:00	FedTerm CEF project <i>Andraž Repar, Jozef Stefan Institute</i>
12:00 – 12:15	National programme on encouraging the development and use of artificial intelligence by 2025 <i>Samo Zorc, Government Office for Digital Transformation</i>
12:15 – 12:45	Q/A
12:45 – 13:30	Lunch Break
	ELG WORKSHOP
13:30 – 14:00	The European Language Grid and European Language Equality Projects <i>Penny Labropoulou, Institute for Language and Speech Processing/R.C. "Athena", Greece</i>

14:00 – 14:20	Panel on ELG and Synergies with the Slovenian LT community as presented in the morning sessions
14:20 – 14:40	Conclusions

3 Summary of Content of Sessions

3.1 Welcome and introduction

The first presentation by **Simon Krek**, the organizer of the workshop, coordinator of the Development of Slovene in a Digital Environment project, and ELRC Technology NAP in Slovenia, included the overview of LT activities in Slovenia, in particular: National program on encouraging the development and use of artificial intelligence by 2025 (NpUI) which was adopted in May 2021, Resolution on the national language policy program 2021-2025 (ReNPJP21-25) adopted in the parliament in June 2021, and the Research Infrastructure Development Plan by 2030 (NRR1 2030) adopted in May 2022.

In addition to this, the Development of Slovene in a Digital Environment project (2020-2023) was introduced. The presentation finished with the conclusions of the Report on Slovenian¹ of the European Language Equality project.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

3.2 eTranslation

The presentation of eTranslation service was delivered by **Miha Žličar**. MT@EC was explained in the context of eTranslation, the CEF.AT program, and the future Digital Europe program. NLP tools other than MT@EC system were mentioned and the two typical processes were described: through an API and by a human translator. The use of different translation models was explained, together with some examples. At the end, Named Entity Recognition tool was mentioned and relevant contacts were given.

(We did not receive the consent form from the presenter, and therefore the presentation is not publicly available.)

3.3 Update from the ELRC

ELRC updates were presented by **Thierry Declerck** (DFKI). It was a standard presentation in English, available also on [Videolectures.net](https://videolectures.net).

3.4 Machine Translation for Slovenian

The presentation was delivered by **Iztok Lebar Bajec**, professor at the Faculty of Computer and Information Science, University of Ljubljana, and the coordinator of the Machine Translation Work Package in the DSDE project. He presented the technological background of the work: the parallel corpus used, pre-processing steps, training procedures, models developed using the chosen framework (NeMo), and evaluation results.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

¹ https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_31__Language_Report_Slovenian_.pdf

3.5 Speech Technology for Slovenian

The presentation was delivered by **Marko Bajec**, professor at the Faculty of Computer and Information Science, University of Ljubljana, and the coordinator of the Speech Technology Work Package in the DSDE project. He presented the general status of the development of speech technology for Slovene, and the goals of the Work Package in the project. He demonstrated current speech recognition solutions for automatic subtitling (national TV), and for the translation of lectures in real time (at the University of Ljubljana), as well as speech synthesis.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

3.6 Semantic Technology for Slovenian

The presentation was by **Slavko Žitnik**, professor at the Faculty of Computer and Information Science, University of Ljubljana, and coordinator of the Semantic Technology Work Package in the DSDE project. He presented the current status of the following tasks: named entity recognition, coreference resolution, word sense disambiguation, semantic shift detection, automatic summarization and question answering.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

3.7 Slovene Terminology Portal

The presentation was delivered by **Miro Romih**, CEO of the Amebis Language Technology company and the coordinator of the Terminology Work Package in the DSDE project. He presented the status of the development of the new terminology portal for online editing and searching terminology, and the accompanying tools: terminology extraction module and an online concordancer for specialized corpora.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

3.8 CEF FedTerm Project

The presentation was delivered by **Andraž Repar** from Jožef Stefan Institute. He presented the FedTerm project, funded by the Connecting Europe Facility project and coordinated by Tilde S.A. (Latvia). The project has developed a federation of terminology portals in Europe. The Slovenian partner (Jožef Stefan Institute) set up a Slovenian node at <https://fedterm.ijs.si/>.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

3.9 National program on encouraging the development and use of artificial intelligence by 2025

Samo Zorc, representative of the Government Office for Digital Transformation, presented the solutions in the National programme on encouraging the development and use of artificial intelligence by 2025, in particular those related to language technology.

The presentation is publicly available on [Videolectures.net](https://videolectures.net).

3.10 Q/A session

The panelists were the presenters from the previous session: Iztok Lebar Bajec, Marko Bajec, Slavko Žitnik, Miro Romih, Samo Zorc and Andraž Repar, as well as **Špela Vintar**, professor at the Department of Translation Studies

at the Faculty of Arts (University of Ljubljana) who represented the newly established study program “Digital Linguistics”. The discussion was moderated by Simon Krek.

4 Synthesis of Workshop Discussions

The discussion was about: (1) the use of (copyrighted) subtitles from the OPUS page for the purposes of training machine translation models, (2) the question of tonemic accentuation in speech synthesis, (3) accessibility of technologies for people with different disabilities (e.g. dyslexy) and (4) the question of funding language technology study programs.

5 Country Profile: Language data creation, management and sharing

In the overall state of technology support for selected European languages chart Slovene is positioned rather low, in 23rd place, close to the better supported non-official EU languages such as Basque or Galician, well behind the comparable official EU languages such as Estonian, Latvian and Lithuanian, and still further from languages like Swedish, Greek or Hungarian which have approx. five times the number of speakers as Slovene.

Slovenia as an EU state will need to encourage the creation of resources, tools and services, and provide a stable environment for LT development. Currently, most of the long-term activity is financed through the CLARIN.SI infrastructure which is a major achievement compared to the situation ten years ago, but still not enough if the current trends in AI are taken into account, with Natural Language Understanding as the goal.

There are several state planning documents that could lead to better support for LT for Slovene in the future. The first is the Resolution on the National Programme for Language Policy 2021-2025 which includes an extensive section on language infrastructure covering topics such as language description (dictionaries, lexical resources), standardization, terminology, multilingualism, language technologies, digitization, special needs. The total sum of funds assigned to these topics is approximately 15M€.

Another important document covering language technology is the National programme on encouraging the development and use of artificial intelligence by 2025. Language technology is mentioned in several sections and there are funds assigned to topics covering LT. However, it is not clear what portion of funding will be assigned to LT in a narrower sense, or which mechanisms will be used.

In general, one can conclude that (a) the support for Slovene is comparable with other languages with a similar status, (b) there is a general awareness in the governmental bodies that LT for Slovene should be supported in the future, (c) the LT community is growing, also through new educational initiatives such as the MA study of Digital Linguistics (Faculty of Arts, University of Ljubljana), (d) there is an infrastructural support, mainly through CLARIN.SI infrastructure at the Jožef Stefan Institute, which also covers all other stakeholders through CLARIN.SI consortium. However, more efforts are needed in the future to bring the existing support closer to the one available for other (official EU) languages.