



The goal of the project is to develop language technology resources, tools and services for Slovene.

The results are aimed at companies, research institutions, and a wider public.

## Upgrade of corpora used for LT

Renewed and larger **language resources** for machine processing, available under open access. New processing pipelines and tools will enable quicker and simpler updating Slovene corpora in the future.

## Speech technology

A **speech database** and **speech recognition software** will be developed, together with supporting tools. A free speech recognition service web site will be launched at the end of the project.

## Semantic technology

A central **Digital Dictionary Database** combining different types of Slovene language data will be compiled, and offered under open access. Also, a knowledge base will be developed.

+

**Resources and tools for:**

- word-sense disambiguation and semantic shift recognition
- automatic summarization and question answering
- relation extraction and semantic analysis

## Terminology portal

A **web portal** combining terminological resources and offering a server infrastructure for the creation of new terminological resources as well as tools for terminology extraction from specialized corpora.

+

**A terminology helpdesk for:**

- terminological queries
- guidelines with concise instructions for creating terminological resources and sample collections

## Machine translation

**English-Slovene parallel corpora** will be collected and offered under open access.

**Neural machine translation software** will be developed and included in the free web portal.

## Long-term language technology infrastructure

The Slovene CLARIN.SI infrastructure will oversee compliance with international standards of language resources created in the project and for their long-term archiving in its certified repository. It will also integrate the developed corpora into its online concordancers.



The project Development of Slovene in a Digital Environment (Slovene: *Razvoj slovenščine v digitalnem okolju, RSDO*) is co-financed by the Republic of Slovenia and the European Union under the European Regional Development Fund. The operation is carried out under the Operational Programme for the Implementation of the EU Cohesion Policy 2014–2020.