Findings: ELE Report on the Bulgarian Language

Svetla Koeva (Institute for Bulgarian Language, Bulgarian Academy of Sciences)
svetla@dcl.bas.bg

08-11-2022 STOA Workshop Towards full digital language equality in a multilingual European Union
http://european-language-equality.eu
Bulgarian

- Census (2021): population 6,519,789
- Eurostat (2021): 1,071,590 Bulgarians abroad
- South Slavic language
- Many regional dialects
- National Language
- The official alphabet is Cyrillic
- Official EU Language
Bulgarian Language Technology

- **Tools and services**
  - Fragmentary support: Text Processing, Speech Processing, Image / Video Processing, Information extraction and Information Retrieval, Translation Technologies, Natural Language Generation

- **Language resources**
  - Fragmentary support: Text Corpora, Parallel Corpora, Lexical resources
  - Weak support: Multimodal Corpora, Language Models

- Many technologies are still **not available** (human-computer interaction, multimodal processing, language generation, etc.)
- There are **no ready-to-use** applications (summarization, question answering, speech recognition, etc.)
- The machine translation technologies for Bulgarian are still **underdeveloped**
- There are **no accessible and reliable** speech-to-text systems for Bulgarian, especially ones that work in real time.
- There is **no dedicated funding** for Bulgarian Language Technologies
There is a need for open real-time services for machine translation from and to Bulgarian combining text and speech, **taking into account context, communicative purposes, and different environments**.

Speech and text technologies for Bulgarian have to be combined with technologies for other modalities: real time image and video processing **working simultaneously in a multilingual environment**.

Natural language understanding and generation of Bulgarian have to become part of multilingual and multimodal processing.

A **convenient and well-regulated access to data** is needed for the development of new products, applications, and services.

Digital **Bulgarian** needs **large-scale, long-term support**, harmonised and coordinated with the support for all European languages.

The **vision for the future** is **high-quality Language Technology for all European languages** that supports political and economic unity through cultural diversity.
The European Language Equality project has received funding from the European Union under the grant agreements № LC-01641480 – 101018166 (ELE) and № LC-01884166 – 101075356 (ELE 2).

Thank you very much

Svetla Koeva
(Institute for Bulgarian Language, Bulgarian Academy of Sciences)
svetla@dcl.bas.bg

08-11-2022 STOA Workshop Towards full digital language equality in a multilingual European Union
http://european-language-equality.eu