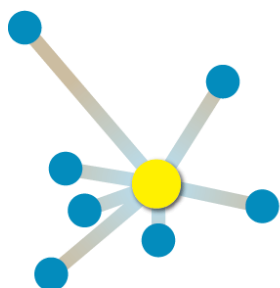European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.

ELRC Workshop Report for Bulgaria

Author(s):   **Svetla Koeva**, Institute for Bulgarian Language, Bulgarian Academy of Sciences

Hristina Dobreva, Ministry of Transport, Information Technologies and Communications of the Republic of Bulgaria

Svetlana Yurukova, DG Translation – Sofia Office

Dissemination Level: Public

Version No.: <V1.1>

Date: 14.04.2016

1. Executive Summary

The Bulgarian ELRC workshop was held on March 18, 2016, at the Representation of the European Commission in Bulgaria (House of Europe, 124 G.S. Rakovski Str., Sofia), from 08:30 a.m. to 05:00 p.m.

It was attended by 130 participants, including representatives of: the European Commission; the Bulgarian Academy of Sciences; the Ministry of Transport, Information Technologies and Communications; the Ministry of Justice; the Ministry of Finance; the Ministry of Environment and Water (and representatives from nine other ministries of the Republic of Bulgaria); the Presidency of the Republic of Bulgaria; the Council of Ministers; the National Institute of Justice; the National Revenue Agency; the Bulgarian Food Safety Agency; the Supreme Judicial Council; the National Statistical Institute; the National Customs Agency; the Executive Agency of Maritime Administration; the Executive Agency "Automobile Administration"; the Sofia University; and other participants from the public administration and the private sector, including representatives of translation and interpretation agencies, NGOs, IT companies, translators and interpreters.

The Workshop's objectives were: (a) to spread the information about the initiative for collection of the necessary language resources to be used in the creation and improvement of automated translation systems of the CEF.AT platform intended to cater to the needs of the public administrations in all EU member states, Iceland and Norway; (b) to bridge the gap between the actual needs and requirements of the public administration in Bulgaria and the potential of the machine translation systems currently offered by the European Commission; (c) to support Bulgarian language as one of the EU languages in its efforts to gain recognition and equal status.

The agenda is available on the website of the workshop: http://lr-coordination.eu/bg/bulgaria_agenda.

2. Workshop Agenda

08:30–09:00 **Registration**

09:00–09:15 **Welcome Addresses**
Svetla Koeva, Institute for Bulgarian Language – Bulgarian Academy of Sciences
Ognian Zlatev, Head of the Representation of the European Commission in Bulgaria
Hristo Hristov, Head of the Information Technology Directorate to the Ministry of Transport, Information Technology and Communication (MTITC), Republic of Bulgaria

09:15–09:30 **Opening, Aims and Objectives of ELRC**
Stelios Piperidis, ILSP/ELRC

09:30–10:00 **European Union and Multilingualism**
Svetlana Yurukova, DG Translation – Sofia Office

10:00–10:30 **Bulgarian Language and Language Technologies in Bulgaria**
Svetla Koeva, Institute for Bulgarian Language – Bulgarian Academy of Sciences

10:30–11:00 **Panel: Multilingual Public Services in Bulgaria**
Moderator: **Svetlana Yurukova,** DG Translation – Sofia Office
Panelists: **Anna Kardzheva,** MTITC of the Republic of Bulgaria; **Bogdan Mlachkov**, Ministry of Justice of the Republic of Bulgaria; **Svetla Marinova,** SOFITA Translation Agency; **Slavi Paskalev**, National Revenue Agency of the Republic of Bulgaria

11:00–11:30 Coffee break

11:30–12:00 **How Machine Translation (AT) Works**
Svetla Koeva, Institute for Bulgarian Language – Bulgarian Academy of Sciences

12:00–12:30 **How Can Public Institutions Benefit from CEF.AT?**
Spyridon Pilos, DG Translation – European Commission (videolink)

12:30–13:30 Lunch

13:30–14:00 **What Data Is Needed?**
Stelios Piperidis, ILSP/ELRC

14:00–14:30 **Legal Framework for Contributing Data**
Zheneta Rogova, Legal Directorate to the MTITC of the Republic of Bulgaria
Open Data Initiative ; Nusha Ivanova, Modernisation of the Administration Directorate to the Council of Ministers

14:30–15:00 **Panel: Language Resources for Translation to and from Bulgarian**
Moderator: **Stoyan Mihov,** Institute for Information and Communication Technologies - Bulgarian Academy of Sciences
Panelists: **Ani Rusinova,** National Revenue Agency of the Republic of Bulgaria; **Veselin Petrov,** Siela Norma Ltd.; **Kalina Tzakova,** National Institute of Justice; **Ramona Stoyanova,** SOFITA Translation Agency; **Sia Kolkovska,** Institute for Bulgarian Language – Bulgarian Academy of Sciences

15:00–15:30 **Coffee Break**

15:30–16:00 **Data and Language Resources: Technical Aspects and Best Practices**
Stelios Piperidis, ILSP/ELRC

16:00–16:30   **Panel: How Can We Engage?**
Moderator: **Hristina Dobreva**, MTITC of the Republic of Bulgaria
Panelists: **Svetla Koeva,** Iinstitute for Bulgarian Language – Bulgarian Academy of Sciences; **Iglika Karanikolova,** Lozanova 48 Ltd.; **Gabriela Mitova,** MTITC of the Republic of Bulgaria; **Lyubomira Toshkova**, Executive Agency of the Maritime Administration to the MTICT of the Republic of Bulgaria; **Svetlana Yurukova**, DG Translation – Sofia Office

16:30–17:00   **Wrap-up, On site Conclusions and Commitments**
Moderator: **Svetla Koeva,** Institute for Bulgarian Language – Bulgarian Academy of Sciences
Panelists: **Stelios Piperidis,** ILSP/ELRC; **Svetlana Yurukova,** DG Translation – Sofia Office; **Hristina Dobreva,** MTITC of the Republic of Bulgaria

3. Summary of Content of Sessions

3.1. Session 1: "Opening and Welcome addresses"

Svetla Koeva from the Institute of Bulgarian Language at the Bulgarian Academy of Sciences opened the event by welcoming the audience and introducing the guests and the topics to be discussed.

Ognian Zlatev, the head of the EC Representation in Bulgaria, welcomed the guests by highlighting the importance of the multilingualism as a prerequisite for access to the EU law and market. Mr. Zlatev also noted that linguistic diversity may create barriers for the public administration; thus, the creation and functionalities of the envisaged platform would be of great use for both EU institutions and the administrations of the member states.

Hristo Hristov, Director of the Information Technology Directorate to the Ministry of Transport, Information Technology and Communications, highlighted that the Ministry supports, as a partner, the European Commission's efforts into the realms of multilingualism and language technologies. Mr. Hristov shared his belief that the automated translation platform would meet the needs of the EU citizens and public administrations in various sectors, and was positive that the workshop would succeed in persuading more people to believe in the necessity of coordination efforts for collecting language resources.

3.2. Session 2: "Aims and Objectives of the European Language Resources Coordination" (ELCR)

Stelios Piperidis, on behalf of the ELRC Consortium, expressed his gratitude to the organisers of the workshop for their efforts. Mr. Piperidis noted that the workshop in Sofia was the nineteenth such event for the ELRC initiative and it had the biggest number of registered participants. He reiterated the importance of the translators' work for the European Commission as a prerequisite for the smooth work of the European institutions. He made some clarification on the aims, objectives and the use of the automated translation platform, and explained the necessity for its expansion in order to purposefully meet the needs of the EU citizens, organisations and public administration bodies.

3.3. Session 3: "European Union and Multilingualism"

Svetlana Yurukova, the Field Officer of DGT at the EC, shared some interesting facts for the official languages of the EU and the translators' work in the European institutions, together with some data about the language combinations in the European Commission. Ms Yurukova highlighted the importance of the languages for the Digital Single Market development and the need of the translations. She also noted the role of machine translation in supporting the daily work of translators working for the EU institutions, and accentuated on the opportunities in fully fledged development of the online public services and the possible benefits to the EU member states.

3.4. Session 4: "Bulgarian Language and the Language Technologies in Bulgaria"

Svetla Koeva from the Institute for Bulgarian Language, gave an overview of a number of issues related to: native speakers of Bulgarian in Bulgaria and around the world; the areas of cross-border cooperation in different sectors of economy, trade, culture and tourism, and in the programs for cross-border cooperation and the migration flow in the context of multilingualism. She supported the hypothesis that the language barriers were bigger hindrance for the European Union than the national borders. Language barriers still interfere with the various opportunities given by the Internet and block the consumer access to lots of

e-sources of information. She claimed that the language technologies and their application in various areas of life would help to overcome the language barriers. She continued with an overview of the language technologies including those that are already available and are being developed for Bulgarian, while reviewing the specific resources, products and technologies.

3.5. Session 5: "How Does Machine Translation Work?"

Svetla Koeva explained the principles behind the systems for statistical machine translation. She reviewed the properties of the natural language that might hinder the computational processing of the language data. Statistical machine translation systems make use of statistical models developed on the basis of training language data from different corpora. Statistical MT systems are trained using translated (parallel) texts and texts that have been written/spoken originally in the target language. She repeatedly highlighted the importance of the amount of language data and the dependency between the subject areas of the training data and those of the texts to be automatically translated. Other resources to help improve the quality of the MT are: dictionaries, terminology databases, ontologies, collections of person, organization, and location named entities.

3.6. Session 5: "How Can Public Institutions Benefit from the Automated Translation Platform CEF.AT?"

Spyridon Pilos from the DG Translation, participated in the workshop through videoconferencing. Mr. Pilos explained the role of the MT, the profile of its users and how the public administration may already benefit from using the existing MT@EC service, offered by the European Commission. He listed various online services that are already connected to the service and gave guidance on how officials in public administrations can have access to it.

3.7. Session 7: "What Data is Needed?"

Svetla Koeva adverted that for improvement of the quality of services provided by the automated translation platform CEF.AT, the machine translation systems should be trained on high-quality data in all official languages of the countries that participated in the Connecting Europe Facility. The needed data can be found in: corpora: monolingual and multilingual corpora with comparable, parallel and aligned documents; dictionaries: monolingual and multilingual dictionaries with lexical units being aligned and provided with various information; translation memories: language databases that contain human translations made by professionals; language and translation models: statistical data about the word sequences observed in a language and the translation correspondences of words and phrases in languages in a language pair.

3.8. Session 8: "Legal Framework for Contributing Data"

Zheneta Rogova from the Legal Directorate to the MTITC of the Republic of Bulgaria, provided the details of the requirements stipulated in the Access to Public Information Act, security of public data and work with them.

Nusha Ivanova from the Modernisation of the Administration Directorate to the Council of Ministers discussed in detail the requirements of the legal framework. She paid special attention to the file formats for sharing the information. The Bulgarian open data portal of the Council of Ministers was given as an example as it was shared on an international platform and the code could be used by any user that had access to this platform

3.9. Session 9: "Data and Language Resources: Technical Aspects and Best Practices"

Stelios Piperidis summarized the types of data needed and he addressed a number of important legal and technical issues in the process of sharing data for machine translation

purposes. He gave an explanation on the legal issues of use and reuse of the language resources by other users of the platform. Some legal issues about the use of personal data were also discussed. He went on to present some high-level technical specifications on data formats and potential further processing. In conclusion, he gave an overview of the ELRC portal and the various functionalities it offers in terms of depositing language data and posting technical and/or legal questions.

4. Synthesis of Workshop Discussions

4.1. Panel 1: "Public Administration in Bulgaria and Multilingualism in Europe"

Svetlana Yurukuva made some introductory remarks into the first discussion panel by putting the focus on a number of issues concerning working languages, the type of the translated documents, the difficulties encountered by the translators, the organization of the translators' work, storage of the data and the documents. She also mentioned the priorities with respect to working with and introducing new languages over the past five years, some ideas for possible upcoming changes, the documents to be translated, and the potential of the automated translation.

Anna Kardzheva from the Ministry of Transport, Information Technologies and Communications (MTITC) of the Republic of Bulgaria, was the first panelist. She confirmed that the employees at the MTITC had already used the machine translation platform - for example, for the translation of EU legal documents. She also mentioned the languages which are most often translated from - English, French and German. Translations are made by an external contractor - a translation company that was selected through public procurement. Ms Kardzheva said that, in her observation, documents that were most often translated were incoming documents but given the future Bulgarian presidency of the EU Council in the second half of 2018, she expected an increasing need for translation of outgoing documents. She also pointed out that EU's legislative proposals had to be translated by the Bulgarian public administration to prepare country's comments and opinion, as before the official translations made by the translators of the EU institutions are made available.

Slavi Paskalev from the National Revenue Agency of the Republic of Bulgaria, claimed that the documents to be translated were most often in English, German, French, more rarely in Polish and Greek. The NRA works with both incoming and outgoing documents but most often these were business documents. Several people are responsible for coordination of both oral and written translations. When lacking internal human resources and capacity, the Agency uses the services of external translation agencies. NRA maintains two databases – for internal and for external translations, respectively. The database contains documents from the past ten years. The NRA also maintains glossaries.

Svetla Marinova from Sofita Translation Agency gave some examples of specific documents requested by different public administration bodies (covering different sectors). She pointed out that the exclusive copyright owner of the documents (and the translations) was the contracting entity and thus the company could not share and provide them to third parties.

Bogdan Mlachkov from Ministry of Justice of the Republic of Bulgaria noted that the Ministry mainly deals with documents in English. As the legal terminology used in Bulgaria reflected a different legal system, if a term with the same meaning was found in both languages, it was rather a coincidence. Therefore, the translation of legal texts is closer to adaptation and is not a translation *per se*. Currently, the Ministry works on specification and clarification of these issues in partnership with an external company. Mr Mlachkov noted that the machine translation platform would be very useful for the Ministry's work, and confirmed that the Ministry did not maintain records for the translated documents, yet.

4.2. Panel 2: Language Resources for Automated Translation to and from Bulgarian

Stoyan Mihov remarked that the language resources might have a significant role in the development not only of machine translation systems, but of other language technologies, as well (e.g., spell checkers).

Sia Kolkovska from the Institute for Bulgarian Language reviewed the principles for developing Bulgarian National Corpus - a very large repository of language data for Bulgarian and 47 more languages (5.4 mil. words all together), supplied with extensive metadata descriptions and annotations. The Corpus is available for search on the web and a part of it can be freely downloaded.

Ani Rusinova from the National Revenue Agency talked about the organisation of NRA's work regarding the translation. Translations are made by the staff of the National Revenue Agency but if the resources are not enough, the Agency subcontracts an external translation agency.

Vesselin Petrov, a representative of Siela Norma, proposed for companies with an experience in the sector to be involved in the process of coordination of the language resources and improvement of the automated translation platform. Mr. Petrov maintained that there should be reward for the shared resources by the legal entities while noting that Siela Norma was ready to provide data in different formats. The translations were made by their internal staff – they very rarely use services of external contractors. The company uses software that was specifically designed to deal with legal framework and regulations.

Kalina Tzakova from the National Institute of Justice (NIJ) talked about the NIJ's work on magistrates' training. She maintained that the use of machine translation systems would assist the NIJ in its international activities and would significantly improve the quality of services they offer.

Ramona Stoyanova, CEO of the Sofita Translation Agency, highlighted the important role of the translation memories and maintained that their use in or together with the MT systems would be extremely useful. She made some remarks on the issues regarding the quality of the translations stating that it also depends on the translation source. Ms Stoyanova expressed her hope that the MT service would significantly improve the quality of the translators' work.

4.3. Panel 3: "How Can We Engage"? & Wrap-up, On Site Conclusions and Commitments

Participants in this last panel were: Iglika Karanikolova, Lozanova 48 Ltd.; Gabriela Mitova, MTITC of the Republic of Bulgaria; Lyubomira Toshkova, Executive Agency of the Maritime Administration to the MTICT of the Republic of Bulgaria; Svetlana Yurukova, DG Translation – Sofia, European Commission; Stelios Piperidis, ILSP/ELRC.

Hristina Dobreva asked the audience whether they considered appropriate to notify and inform the heads of their departments and the higher staff at their institutions about the issues discussed at the workshop, and all attending guests gave a positive response.

Lyubomira Toshkova made a proposal for a deadline of the provision and collection of the available language resources.

Hristina Dobreva added that the heads of the institutions and organisations had to be officially informed for the collection of the data.

Iglika Karanikolova confirmed, on behalf of the CEO of the Lozanova 48 Ltd., that the translation agency was ready to help the initiative.

Overall, it was stressed that the public institutions have a huge amount of data that they cannot provide it on their own decision but only after a sanction from a higher institution – an order from the Minister to the heads of the responsible directorates/departments, and respectively, from a higher institution or a superior body to the Minister. There is a need to determine by official channels the person and/or the department in an institution entitled to provide documents, what procedures to be employed when providing documents, and the person and/or department that can determine exactly which documents to be provided.

5.   Workshop Presentation Materials

Presentations are available on the website of the workshop in the Agenda section: http://lr-coordination.eu/bg/bulgaria_agenda

Photos are available at: http://ibl.bas.bg/ELRC-in-Sofia/

Information in media:

Покана за семинар „Координиране на езиковите ресурси в Европа" (Информационен портал за неправителствените организации (НПО) в България/The Information Portal of Nongovernmental Organizations (NGOs) in Bulgaria: ngo.bg)

Информация за посещението на семинара на представители на НАБИС (nalis.bg)

МТИТС участва в семинар за бъдещето на платформа за автоматичен превод на официалните езици на ЕС (сайт на Министерство на търговията, информационните технологии и съобщенията/website of the Ministry of Transport, Information Technologies and Communications)

Нова платформа за автоматичен машинен превод в ЕС (Българско национално радио/Bulgarian National Radio: bnr.bg)

Платформата CEF.AT ще прави автоматичен превод на официалните езици на ЕС (itjobs.bg)

Европейската комисия отваря платформата си за автоматичен превод за публичните администрации на страните в ЕС (klassa.bg)

Създава се нова платформа за автоматичен превод на официалните езици на Европейския съюз (ProjectMedia.bg)

Машини помагат на ЕС за превода (в. „Аз Буки")

ЕК: Комисията отваря платформата си за автоматичен превод за публичните администрации на страните в ЕС (Информационна агенция „Фокус"/Focus Information Agency: http://www.focus-news.net/)

Национален семинар на проекта на Европейската комисия „Координиране на езиковите ресурси в Европа" (сайт на Българска академия на науките/website of the Bulgarian Academy of Sciences: http://www.bas.bg/)