



**European Language
Resource Coordination**
Connecting Europe Facility

ELRC Workshop Report for Romania



Author(s): Dan Tufiş (ICIA)
Verginica Barbu Mititelu (ICIA)
Elena Irimia (ICIA)

Dissemination Level: Public

Version No.: <V1.1>

Date: 2016-03-30



Contents

<u>1</u>	<u>Executive Summary</u>	<u>3</u>
<u>2</u>	<u>Workshop Agenda</u>	<u>4</u>
<u>3</u>	<u>Summary of Content of Sessions</u>	<u>5</u>
3.1	Opening	5
3.2	Welcome	5
3.3	Workshop Objectives	5
3.4	Europe and multilingualism	5
3.5	Languages and Language Technologies in Romania	6
3.6	Machine translation: how does it work?	6
3.7	How can Public Institutions benefit from the CEAF.AT Platform?	6
3.8	What data are necessary? Why?	6
3.9	Legal framework for data contribution”	6
3.10	Data and language resources	7
3.11	Open Discussion: How can we engage?	7
3.12	Wrap-up, on site conclusions and commitments	7
<u>4</u>	<u>Synthesis of Workshop Discussions</u>	<u>8</u>
4.1	Panel “Multilingual Public services in Romania”	8
4.2	Panel “Data and language resources in Romania”	8
<u>5</u>	<u>Workshop Presentation Materials</u>	<u>9</u>

ELRC Workshop Report for Romania

1 Executive Summary

This document reports on the ELRC Workshop in Romania, which took place in Bucharest, on the 23rd of March 2016 at the European Commission Representation headquarters in Vasile Lascăr Street, no. 31, Bucharest. The Workshop was co-organized by the Research Institute for Artificial Intelligence „Mihai Drăgănescu” of the Romanian Academy and the Representation of the European Commission in Romania. The workshop’s agenda is presented in section 2. The Workshop ELRC-Romania attracted much interest and was attended by 65 participants covering a wide range of decision makers and professionals from major ministries and public organisations. The report briefly informs about the content of each individual, interactive and panel workshop session (sections 3 and 4 respectively).

The presentations of the Romanian ELRC Workshop are available on the Workshop web-site http://lr-coordination.eu/ro/romania_agenda, together with the recordings of the respective sessions.

2 Workshop Agenda

08:00 – 09:00 *Registration*

- 09:00 – 09:10 Workshop Opening (Stelios Piperidis, ELRC/ILSP, Prof. Dan Tufiş, Director of the Institute for Artificial Intelligence, Romanian Academy)
- 09:10 – 09:20 Welcome by local representative (Liliana Comănescu, DGT Local Field Officer)
- 09:20 – 09:30 Workshop Aims & Objectives (Dr. Stelios Piperidis, ELRC/ILSP)
- 09:30 – 09:45 Europe and Multilingualism (Liliana Comănescu, DGT EC)
- 09:45 – 10:05 Languages and Language Technologies in Romania (Dr. Verginica Barbu Mititelu. Institute for Artificial Intelligence Romanian Academy)
- 10:05 – 11:00 Panel: Multilingual Public Services in România – Moderator: Alexandru IOAN (Radio Romania Actualităţi), Panelists: Marius Bostan (Minister for Communications and Information Society), Varujan Pambuccian (Member of the Romanian Parliament), Carmen Elian (Sub-Secretary of State, Ministry for Communications and Information Society), Ilie Dumitrescu (Director National Institute of Statistics), Diana Voicu (Former Secretary of State in Ministry of Justice)

11:00 – 11:30 *Coffee Break and Networking*

- 11:30 – 11:50 Machine Translation vs. Automated Translation: How does it Work? (Prof. Dan Tufiş, Institute for Artificial Intelligence, Romanian Academy)
- 11:50 – 12:20 How Can Public Services Benefit from the CEF.AT Platform? (Spyros Pilos -DGT, EC video link)

12:20 – 13:20 *Lunch Break and Networking*

- 13:20 – 13:40 What Data is Needed? Why? (Prof. Dan Cristea, Univ. "A. I. Cuza" Iaşi)
- 13:40 – 14:00 Legal Framework for Contributing Data (Prof. Mihai Floroiu, Univ. "Dunărea de Jos", Galaţi)
- 14:00 – 14:55 Panel: Data and Language Resources in Romania – Moderator: Gabriel Vasile (Market Watch Chief Editor), Panelists: Prof. Dan Tufiş, Institute for Artificial Intelligence, Romanian, Academy, Laura Mihăilescu (Head of the Translation Department, European Institute of Romania), Dan Matei (Vice-President of the Romanian Association for Terminology), Marius Nicolaescu (Head of IT Department, Executive Agency for Higher Education, Research, Development and Innovation)

14:55 – 15:25 *Coffee Break and Networking*

- 15:25 – 15:50 Sharing Data and Language Resources: Technical Aspects and Best Practices (Stelios Piperidis – ELRC/ILSP)
- 15:50 – 16:20 Interactive Session: How Can We Engage? (Prof. Dan Tufiş, Institute for Artificial Intelligence, Romanian Academy, Stelios Piperidis, ELRC/ILSP; Liliana Comănescu, DGT Local Field Officer)
- 16:20 – 16:30 Summary and next steps (Prof. Dan Tufiş, Institute for Artificial Intelligence, Romanian Academy, Stelios Piperidis, ELRC/ILSP; Liliana Comănescu, DGT Local Field Officer)

3 Summary of Content of Sessions

3.1 Opening

Mr. Stelios Piperidis, the ELRC representative for the region and Mr. Dan Tufiş, the Romanian ELRC National Anchor Point (NAP) opened the event by welcoming the audience and introducing the key persons in conceiving and organizing the event.

3.2 Welcome

Ms Liliana Comanescu, the DGT Local Field Officer, thanked the audience for participating in the workshop and pointed out the success of the event, which showed that the topic is of great interest and may help identify the needs and share interests. She also stressed the importance of translation and translation quality, and invited everybody to participate and to share their hopes and questions about the importance of Language Resources for Language Technologies.

3.3 Workshop Objectives

Mr. Stelios Piperidis, the ELRC representative, went through the workshop objectives and its logistics. He stressed first the multilingual aspect of Europe and reiterated that there is one important barrier, namely the language barrier, which prevents the EU's market to become a really single market. He also introduced the CEF scene from the perspective of such a multilingual Digital Single Market strategy. He identified current multilingual challenges in the European public services and in the business sector in general and stressed the support of the EC to digital multilingualism. He introduced ELRC and explained its relation to CEF and CEF.AT, while he briefly presented the main stakeholders, principles and goals of this endeavour and he also stressed the main points of potential collaboration between ELRC and the public sector in view of the multilingualism support within the EU. Mr. Stelios Piperidis finished his talk with a strong statement: Supporting our languages is supporting Europe and supporting Europe is supporting our languages.

3.4 Europe and multilingualism

Ms. Liliana Comănescu, the DGT representative in Romania showed the European scene which has grown from an original 4 languages to 24 official languages and 60 major regional/minority languages pointing out that all languages are equal in Europe and that Europe is committed to multilingualism. She highlighted the advantages of member state collaboration with CEF and CEF.AT, among which are free automated translation, easier sharing of information, and more accessible public services for all European citizens. Liliana Comănescu continued with a brief overview with regard to the translation services in the EU, i.e. the volume of translated documents, the number of appointed and freelance translators, and the tools they use in their everyday translation process. She also reported on the challenges in terms of machine translation and in view of the multilingual Digital Single Market, stressing the need for multilingual support and securing digital inclusion.

ELRC Workshop Report for Romania

3.5 Languages and Language Technologies in Romania

The aim of this presentation, given by Ms Verginica Mititelu from the Romanian Academy Research Institute for Artificial Intelligence “Mihai Drăgănescu”, was to present the multilingual dimension as it manifests in the domains of activity in Romania: number of languages spoken by the population, languages taught as mother tongues in schools, languages spoken at the borders and in the activities at the borders, languages spoken in commerce and trade, languages involved in the migration affecting Romania. It was obvious that many European languages are involved in various activities. Language technology is rather a domain of research in Romania than a domain of activity.

3.6 Machine translation: how does it work?

Mr. Dan Tufiş, director of the Romanian Academy Research Institute for Artificial Intelligence “Mihai Drăgănescu”, started the presentation distinguishing between the terms “machine translation” and “automated translation”. After reminding the audience about the language complexity, and thus of the translation complexity, he briefly presented the history of machine translation. The core of the presentation focused on Statistical Machine Translation and the information available in training data sets. Other sources of information useful in translation were also mentioned (bi-/multi-lingual dictionaries, lexical ontologies, terminological databases). Participants were encouraged to become aware of the treasure represented by the data they own and to foresee a management plan for them.

3.7 How can Public Institutions benefit from the CEAF.AT Platform?

After a historical overview of how the language problem became more and more complex with the addition of member states, Mr. Spiros Pilos, Head of Language Applications of DGT elaborated on the reasons why MT was the only viable solution to tackle the multilingualism challenge in Europe. He described in detail the MT@EC system, i.e. the languages it supports, the technologies upon which it is based, its user interface and other technical features, including input data format, delivery of results and security in document transfer. He also presented statistics about its usage and invited the public sector representatives to access it. He concluded emphasizing that the focus will be the coverage of more domains and increasingly in-domain text are needed to implement this move at sufficiently high quality.

3.8 What data are necessary? Why?

Mr. Dan Cristea, from the faculty of Computer Science of the “Al. I. Cuza” University of Iasi, explained concepts such as “web”, “big data” and their connection with translation. He presented the types of data to be collected for machine translation: texts (translations, aligned translations, collections of comparable texts), glossaries, terminological databases, dictionaries and lists of words in one or more languages. He explained how different formats are useful to varying degrees and stressed the importance of metadata. Several examples of language resources already available were shown. Finally, he also encouraged the participants to become aware of the treasure represented by the data they own and find ways of making them available to the community.

3.9 Legal framework for data contribution”

Mr. Mihai Floroiu, from the University “Dunărea de Jos” stressed the need for a clear and simple to follow legal framework for the public sector information and data reuse throughout the EU. He then described how data would be used by ELRC, and indicated the five stages for releasing data: exclude confidential information; obtain prior informed consent, find a legal basis, anonymize or exclude personal data; ensure there is no 3rd party copyrights, that the material is in the Public Domain or that the necessary licenses have been obtained; follow

ELRC Workshop Report for Romania

the national PSI transposition rules (e.g. use the national Open Government licenses or the standard procedure for releasing PSI); use a standard Open Government licence, open public license or reuse license, and follow the national or organisational PSI re-use policy.

3.10 Data and language resources

Stelios Piperidis (ELRC/ILSP) presented in detail the value chain activity consisting of identification and selection of language data, its documentation, cleaning, validation, processing and sharing. The talk addressed several issues related to this process, concerning both legal and technical issues (such as data formatting, anonymization etc.) The session ended with the presentation of the ELRC portal, technical and legal support helpdesk, forum and repository for sharing language resources.

3.11 Open Discussion: How can we engage?

This session was moderated by Stelios Piperidis, Dan Tufiş and Liliana Comănescu and outlined the interest of the participants to get involved in the ELRC initiative and to contribute with different language resources available at the institutions they represented (translation memories RO-FR, specialized-domain dictionaries, various Romanian language corpora, etc.). As the status of some of these resources as far as the IPR and legal status was unclear, things remained to be clarified and the authorized people be informed. One of the issues raised in this session was related to whether the access to the CEF.AT facilities will be opened only to juridical persons or to free-lance translators as well. Mr. Stelios Piperidis clarified the issue.

3.12 Wrap-up, on site conclusions and commitments

The workshop was concluded with very positive comments from the ELRC representative, Mr. Stelios Piperidis, who thanked the organizers of the meeting, the speakers and all participants for their interest and future cooperation. Dan Tufiş, appreciated the declared commitments, expressed his optimism for the Romania's involvement in the fulfilment of the objectives of ELRC and additionally, thanked for the support offered by the ELRC representatives, to the Representation of EC in Romania and especially to the DGT Local Field Officer Liliana Comănescu and to the two interpreters who ensured a very high quality service. Finally, Liliana Comănescu reiterated the full support of DGT-EC Romania for future actions along ELRC work-plan and thanked all participants.

4 Synthesis of Workshop Discussions

4.1 Panel “Multilingual Public services in Romania”

This panel was chaired by Alexandru Ioan (Radio Romania). The panelists were representatives from the Ministry of Communication and for the Information Society, Romanian Parliament, the National Institute of Statistics, and Romanian Academy Library.

The aim of this panel was to have these representatives share with the ELRC workshop participants their experience in problems related to languages and multilinguality.

The participants presented the state of the art with respect to the way ministries deal with translation problems: although translation services are of extreme necessity for each domain of activity, there are no translation departments in the ministries. Terminological inconsistencies are only one of the outcomes of this situation. The need for a close cooperation between translators and domain specialists was emphasized.

4.2 Panel “Data and language resources in Romania”

The panelists were representative of several institutions: Romanian Academy Research Institute for Artificial Intelligence “Mihai Drăgănescu”, the European Institute in Romania, the Institute of Cultural Memory, the Romanian Association of Terminology, and Executive Agency for Higher Education, Research, Development and Innovation Funding.

The panel participants identified internal data that are useful for CEF.AT. They are stored on the servers of the institutions they represent and are subject to various restrictions (confidentiality, property right being only some of them). The institutions, through their representatives, expressed their openness to discussing the way the data they have can be used and become beneficial to others.

ELRC Workshop Report for Romania

5 Workshop Presentation Materials

The presentations of the Romanian ELRC Workshop are available on the Workshop web-site http://lr-coordination.eu/ro/romania_agenda, together with the recordings of the respective sessions.

The pre-workshop press-release was distributed by the Representation of the European Commission in Romania and the Romanian Academy:

http://ec.europa.eu/romania/news/agenda/15032016_atelier_cordonarea_resurselor_lingvistice_europene_bucuresti_ro.htm

<https://academiaromana.wordpress.com/2016/03/15/atelierul-coordonarea-resurselor-lingvistice-europene-organizat-la-bucuresti-comunicat-de-presa/>

The post-workshop press-release has been released by the Romanian Academy:

<http://www.academiaromana.ro/com2016/doc/d0323-ComunicatELRC-siteRCE.pdf>

Also, an interview with Dan Tufiş, on the ELRC topic, was broadcasted by the National Radio program on March 15th (22:46 - Radio România Actualități "Agenda Globală").