European Language Resource Coordination

Connecting Europe Facility

Deliverable Task 6

ELRC Workshop Report for Greece

Author(s):	Maria Koutsombogera (ILSP)
	Stelios Piperidis (ILSP)
	Kanella Pouli (ILSP)
	Maria Gavrilidou (ILSP)
Dissemination Level:	Public
Version No.:	<V1.1>
Date:	2015-10-12

Contents

1 Executive Summary

This document reports on the ELRC Workshop in Greece, which took place in Athens, on the 24th of September 2015 at the National Hellenic Research Foundation. It includes the agenda of the event (section 2) and briefly informs about the content of each individual, interactive and panel workshop session (sections 3 & 4). The event was attended by 84 participants spanning a wide range of ministries and public organisations, as well as independent data experts and freelance translators. The dedicated event webpage can be found at http://lr-coordination.eu/greece.

2   Workshop Agenda

Agenda for CEF.AT - ELRC Training Workshop
(European Language Resource Coordination)
http://lr-coordination.eu/greece
National Hellenic Research Foundation, "Leonidas Zervas" Amphitheatre
Athens, September 24 2015

08:00 – 09:00	Registration
09:00 – 09:10	Opening by ELRC (**S. Piperidis** – ILSP/ELRC)
09:10 – 09:20	Welcome by the EC Representation in Greece (**P. Karvounis** – Head of EC Representation in Greece)
09:20 – 09:30	Welcome by the EC (**S. Pilos** – EC, DG Translation, Head of Language Applications)
09:30 – 09:40	Workshop Objectives and Programme (**S. Piperidis** – ILSP/ELRC)
09:40 – 09:50	The EU and Multilingualism (**P. Alevantis** - Head of DGT Athens)
09:50 – 10:10	Language(s) and related Language Technologies in Greece (**M. Gavrilidou** – ILSP, ELRC Technical NAP)
10:10 – 11:00	Panel 1: Multilingual Services in Greece (Moderator: **P. Alevantis** – Head of DGT Athens)
	Panelists: P. Benekos (Director of translation services, Ministry of Foreign Affairs), E. Antonoglou (Deputy Head of Publications and Translation Section, Bank of Greece), D. Gioutikas (Ministry of Justice), S. Tsoukalas (President of the Greek National School of Public Administration Alumni Association), A. Katerinopoulos (Ministry of Health)
11:00 – 11:30	*Coffee Break, Networking, Demos*
11:30 – 12:00	Automated Translation: How does it work and what is it used for? (**S. Piperidis** - ILSP/ELRC)
12:00 – 12:30	Machine Translation, CEF.AT and how can Public Institutions benefit from the CEF.AT Platform? (**S. Pilos**, EC, DG Translation, Head of Language Applications)
12:30 – 13:30	*Lunch Break, Networking, Demos*
13:30 – 14:00	What data is needed? (**P. Prokopidis** - ILSP)
14:00 – 14:30	Legal framework for Contributing Data (**P. Tsiavos** – The Media Institute, UCL)
14:30 – 15:15	Panel 2: The data and Language Resources in Greece: Where to look for what? (Moderator: **M. Gavriilidou**),
	Panelists: K. Aggeletopoulou (Ministry of Interior-Administrative Reform and e-Governance), E. Papavasileiou (Minstry of Foreign Affairs), G. Veikopoulos (Ministry of Foreign Affairs), K. Nadalis (TMServe), A. Routzouni (Ministry of Interior-Administrative Reform and e-Governance) L. Tossounidis (Cedefop, Ministry of Interior-Administrative Reform and e-Governance), Th. Karounos (Open Data- Greek Free/Open Source Software Society)
15:15 – 15:45	*Coffee Break, Networking, Demos*
15:45 – 16:15	Sharing Data and Language Resources: Technical and Practical aspects (**P. Labropoulou** - ILSP)
16:15 – 16:45	Interactive session: Can We Engage? (**S. Piperidis** - ILSP/ELRC)
16:45 – 17:15	Best Practice for the Future: Capitalize on your Valuable Data (**M. Koutsombogera** - ILSP)
17:15 – 17:30	Wrap-up, On site Conclusions and Commitments (Moderator: **S. Piperidis** - ILSP/ELRC)

3   Summary of Content of Sessions

3.1   Session 1: "Opening by ELRC"

Stelios Piperidis, the local ELRC representative, opened the event by welcoming the audience and introducing the key persons in conceiving and organizing the event, namely the ELRC consortium and the EC/DGT representatives.

3.2   Session 2: "Welcome by the EC Representation in Greece"

The welcome speech of Panos Karvounis, Head of the EC Representation in Greece, stressed the importance of the CEF instrument in supporting the DSIs and, specifically, the investment in machine translation. He pointed out the potential for improvement offered to MT@EC provided the system is trained on national administrations texts; he furthermore acknowledged the fact that improving the performance of the machine translation system will contribute to further protect the Greek language in the audio-visual interconnected ecosystem of the 21st century. He expressed the intent to help the Greek public administration to perfect the EU machine translation tool for the Greek language to ensure a better service to citizens and concluded his welcome speech with wishes for a successful workshop.

3.3   Session 3: "Welcome by the EC"

Spyros Pilos, Head of Language Applications in DG Translation, EC, expressed his welcome message to the audience by setting the CEF scene from the perspective of the multilingual Digital Single Market strategy. He identified current multilingual challenges in the European public services and the support of the EC to digital multilingualism. He reported on the nature and the objectives of the CEF Digital and explained the rationale behind the CEF.AT platform and the expected benefits for the Greek public services.

3.4   Session 4: "Workshop Objectives and Programme"

Stelios Piperidis went through the workshop objectives and its logistics. He introduced ELRC and explained its relation to CEF and CEF.AT, while he briefly presented the main stakeholders, principles and goals of this endeavor. He stressed the main points of potential collaboration between ELRC and the Greek public sector in view of the multilingualism support within the EU.

3.5   Session 5: "The EU and Multilingualism"

Panagiotis Alevantis, Head of DGT Athens Office, presented 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. He also reported on the challenges that the Greek language encounters in terms of machine translation and in view of the multilingual Digital Single Market, stressing the need for multilingual support and securing digital inclusion.

3.6   Session 6: "Language(s) and related Language Technologies in Greece"

Maria Gavriilidou, the Greek Technology NAP presented the aspects of multilingual Greece, i.e. domains and cross-cultural topics requiring communication and collaboration in several languages, by giving concrete examples. She also described the current state-of-play in Greece with regard to the Language Technology, in terms of both monolingual and multilingual processing. She pointed out the key players in Greece in the domains of Language Technology and specifically Machine Translation, and she concluded with problems,

challenges and improvement perspectives that the Greek language encounters with respect to Machine Translation.

3.7 Session 7: "Automated Translation: How does it work and what it is used for?"

Stelios Piperidis elaborated on issues related to Machine Translation, i.e. the large volumes of quality translation needed within the EU and the associated cost, as well as the language peculiarities that make MT a complex task. He explained how modern statistical machine translation technologies work by providing simple examples and stressing the importance of abundant and appropriate data. In this respect, he briefly listed the type of data that are ideally required for the CEF.AT platform and what would be possible sources for this data.

3.8 Session 8: "Machine Translation, CEF.AT and how can Public Institutions benefit from the CEF.AT Platform?"

Spiros Pilos 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 stressed that the CEF.AT platform will build upon MT@EC and aims to address the CEF DSIs. Most importantly, he presented the benefits that CEF.AT will bring to its users in terms of translation of better quality and security, adapted to the user needs and specific domains, among others.

The audience questions focused on issues of accessing MT@EC, the restrictions in terms of number of users, the dissemination of information regarding the use of the MT@EC system, and requests to access the system by the local administration organisations. An important issue that was raised was that of freelance translators (subcontractors of the official Translation Service) not being able to use MT@EC.

3.9 Session 9: "What data is needed?"

Prokopis Prokopidis (ILSP/ELRC) focused on the data that need to be collected, the required format of the data so that they can be fed into MT systems and the importance of the data that the public sector holds and with which it can contribute to the CEF.AT platform. The main data types mentioned were texts (i.e. technical reports, webpages, speeches, etc.), ideally translated in one or more EU languages, but also glossaries, terminological databases or lists of words in one or more languages. It is desirable that such data cover specific domains addressed by the DSIs, and that are aligned (parallel) or comparable. The presenter encouraged the participants to help ELRC by providing or identifying the required textual content or pointing out key persons/ organisations with which they collaborate. The questions that followed were related to the type of documents that are desirable (digital vs. non-digital) and issues of translation validation in the web content that is crawled and further processed.

3.10 Session 10: "Legal framework for Contributing Data"

Prodromos Tsiavos (The Media Institute, UCL) stressed the need for a clear and simple to follow legal framework for the public sector information and data reuse throughout the EU. He briefly presented the PSI directive and explained the structure of the data rights. Most importantly, he listed the step-by-step process to be followed to release data under the PSI Directive and presented actual case studies from various EU countries. The distinct stages in the process refer to: exclusion of confidential information; consent for, anonymization or exclusion of personal data; take care of 3rd party copyrights; follow the national PSI transposition rules; use a standard Open Government Licence, Open public licence or re-use licence, and follow the national or organisational PSI re-use policy,

3.11 Session 10: "Sharing Data and Language Resources: Technical and Practical aspects"

Penny Labropoulou (ILSP/ELRC) presented in detail the workflow for the collection, processing and sharing of language resources. Most of the stages of this process, e.g. the identification of the data sources and datasets, the basic metadata documentation, data cleaning and privacy and ethics management are tasks in which the public sector providers will collaborate with ELRC. The presenter encouraged the audience to participate in these activities and work together with ELRC, and she showcased the mechanisms with which ELRC will fully support the providers throughout the whole process, i.e. the helpdesk and user forum mechanism, the ELRC repository and the ELRC website.

3.12 Session 11: "Interactive session: Can We Engage?"

In this interactive session, Stelios Piperidis discussed a set of key questions with the audience:

- Does your organisation produce data that you think as appropriate for CEF.AT?
- Do you think that your organisation can share this data for the CEF.AT purposes?
- Are there practical difficulties that could affect your contribution?
- What do you think are the benefits from your participation in CEF.AT?
- Do you have any suggestions on how to move on?

The audience raised the following points:
- There are organisations that have and produce data that are useful to CEF and find it possible to share them. Practical difficulties and quality issues are to be examined.
- Institutional actions and measures need to be further taken, i.e. official correspondence, political decisions, etc. as well as a series of dissemination efforts, i.e. short presentations per ministry or public organisation.
- The institutional and political framework is already there, what is perhaps missing is a contact point/link to the public organisations, as well as a better infrastructural organisation of the translation work, so that the data are further exploited and reused.
- Some freelance translators see no motivation in sharing data, because they see no direct benefit. The first step would be for them to gain access to MT@EC and then discuss possible ways of collaboration.

3.13 Session 12: "Best Practice for the Future: Capitalize on your Valuable Data"

Maria Koutsombogera (ILSP/ELRC) recommended a set of best practices to be optionally followed by data providers, to maximise the potential and reusability of their data, either existing or to be acquired in the future. The best practices were structured in the form of a Data Management Plan (DMP) that defines the strategy according to which the data are collected, described, stored and preserved. The presenter explained the distinct aspects in the construction of a DMP that the audience may undertake, in addition to its own internal processes, to ensure the exploitation and sustainability of its data. The audience expressed the opinion that it would be useful to transfer the knowledge about research projects DMPs to the public domain data and that the public domain data would greatly profit from a DMP.

3.14 Session 13: "Wrap-up, on site Conclusions and Commitments"

Stelios Piperidis concluded the Workshop by pointing out that the event was a first step towards the collaboration of public sector with ELRC, in view of the contribution to the high quality service that the CEF.AT platform is expected to bring. In this respect, he stressed the need for a continuous data contribution as an investment to technology solutions that have a clear effect on the European languages sustainability and thus the digital presence of the Greek language.

4 Synthesis of Workshop Discussions

4.1 Panel 1: Multilingual Services in Greece

The panel was moderated by P. Alevantis, the Head of DGT Athens. The panellists were representatives from the Ministry of Foreign Affairs, the Ministry of Justice, the Ministry of Health and the Bank of Greece. The moderator addressed three rounds of questions to the members of the panel; these rounds were about:

- Languages and text types used in translation services in each organization (important and difficult languages; type of translated documents; problems encountered)
- Translation workflows (who translates, e.g. in-house, outsourcing, other; quality control systems; digital archives of translations)
- Changes attested / evolution (language priorities; future changes; experience with combinations of human and machine translation; incoming or outgoing documents).

The answers of the panelists displayed most vividly the diversity that exists in the different services of the public sector: from the digitized workflows, the translation memories and the term lists of the Bank of Greece to the surprisingly "paper-only" procedure of the Official Translation Service of the Ministry of Foreign Affairs. The translation needs were more convergent: English is the language most used, with the addition of non-EU languages due to immigration. Quality control is also non-existent in the Greek public sector, with the exception of the Bank of Greece.

4.2 Panel 2: The data and Language Resources in Greece: Where to look for what?

The Panel was moderated by M. Gavriilidou, the Greek ELRC Technology NAP. The panellists were representatives from the Ministry of Interior and Administrative Reform, the Ministry of Foreign Affairs, a private translation company and the Open Data- Greek Free/Open Source Software Society. The focus of this panel was on the:

- Availability of language data in the Greek public sector
- Appropriate data identification procedures
- Sensitization and mobilization of the public sector for this cause.

Following a short presentation of the key issues (language data, bi-texts, parallel texts, available data, open data, structured data) by the moderator, the panelists presented the situation in the country in relation to these issues. It became evident that the legal framework for open public data is in place (transposition of the PSI Directive), and that the public sector is currently obliged to upload their data on the public data central inventory (http://www.data.gov.gr/). However, this initiative is very recent and still needs reinforcement.

The panelists proposed tips and ideas as best practices for approaching the public sector and acquiring the necessary data, which will be implemented as follow-up actions after the workshop.

5   Workshop Presentation Materials

The workshop presentations and videos can be accessed at the event webpage, at http://lr-coordination.eu/greece_agenda.