



**European Language
Resource Coordination**
Connecting Europe Facility

ELRC Workshop Report for Cyprus



Author(s): Georgios Floros (University of Cyprus)
Dissemination Level: Public
Version No.: <V1.1>
Date: 2015-12-08



Contents

1	<u>Executive Summary</u>	3
2	<u>Workshop Agenda</u>	4
3	<u>Summary of Content of Sessions</u>	5
3.1	Session 1: “Opening and welcome (ELRC)”	5
3.2	Session 2: “Local welcome”	5
3.3	Session 3: “Welcome by the European Commission”	5
3.4	Session 4: “Aims and objectives”	5
3.5	Session 5: “The EU and multilingualism”	5
3.6	Session 6: “Language and language technology in Cyprus”	6
3.7	Session 7: “Round table 1: Multilingual public services in Cyprus”	6
3.8	Session 8: “Automated translation: How does it work?”	6
3.9	Session 9: “Machine translation: How can public institutions benefit from the CEF.AT platform?”	6
3.10	Session 10: “What data are needed?”	6
3.11	Session 11: “Legal framework for data contribution”	7
3.12	Session 12: “Round-table 2: Language data of the Cypriot public sector”	7
3.13	Session 13: “Data and language resources: Technical and practical aspects”	7
3.14	Session 14: “Discussion with the audience: How can we engage?”	7
3.15	Session 15: “Wrap-up and Conclusions”	7
4	<u>Synthesis of Workshop Discussions</u>	9
4.1	Panel 1 (Session 7): “Round table 1: Multilingual public services in Cyprus”	9
4.2	Panel 2 (Session 12): “Round-table 2: Language data of the Cypriot public sector”	9
4.3	Panel 3 (Session 14): “Discussion with the audience: How can we engage?”	10
5	<u>Workshop Presentation Materials</u>	11

ELRC Workshop Report for Cyprus

1 Executive Summary

This document reports on the ELRC Workshop in Cyprus, which took place in Nicosia on 1 December 2015 at the premises of the Representation of the EC in Cyprus (EU House, 30 Byron Ave., 1096 Nicosia, Cyprus, +357 22 81 77 70). 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 36 participants from public organizations, as well as independent data experts and freelance translators. The dedicated event webpage can be found at <http://lr-coordination.eu/cyprus> .

ELRC Workshop Report for Cyprus

2 Workshop Agenda

The numbers in red indicate the Sessions

08:00 – 09:00		Registration
09:00 – 09:10	1	Opening and welcome (ELRC) (S. Piperidis – ILSP, G. Floros – University of Cyprus)
09:10 – 09:20	2	Local welcome (M. Neokleous , on behalf of the Head of the Representation of the EC in Cyprus, Mr. G. Markopoulitiotis)
09:20 – 09:30	3	Welcome by the European Commission (A. Vassiliou , former EU-Commissioner for Education, Culture, Multilingualism and Youth)
09:30 – 09:40	4	Aims and objectives (S. Piperidis – ILSP/ELRC)
09:40 – 09:50	5	The EU and multilingualism (M. Neokleous – DGT, Nicosia Office)
09:50 – 10:10	6	Language and language technology in Cyprus (G. Floros – University of Cyprus, NAP for Cyprus)
		Round table 1: Multilingual public services in Cyprus (Moderation: M. Neokleous – DGT, Nicosia Office)
10:10 – 11:00	7	Presentations: M. Adamidou (House of Representatives), M. Gavriilides (Government Printing Office), I. Soulos (Ministry of Interior), Ch. Christodoulou (Ministry of Transport & Communications)
11:00 – 11:30		Coffee break
11:30 – 12:00	8	Automated translation: How does it work? (S. Piperidis – ILSP/ELRC)
12:00 – 12:30	9	Machine translation: How can public institutions benefit from the CEF.AT platform? (S. Pilos – DGT, European Commission, <i>through teleconferencing</i>)
12:30 – 13:30		Lunch break
13:30 – 14:00	10	What data are needed? (M. Koutsombogera – ILSP)
14:00 – 14:30	11	Legal framework for data contribution (T. Synodinou – University of Cyprus)
		Round-table 2: Language data of the Cypriot public sector (Moderation: G. Floros – University of Cyprus/ELRC)
14:30 – 15:15	12	Presentations: A. Xyda (Cyprus Police), P. Charalambous (Ministry of Energy, Commerce, Industry and Tourism), I. Soulos (Ministry of Interior), A. Stylianou (Central Bank of Cyprus)
15:15 – 15:45		Coffee break
15:45 – 16:15	13	Data and language resources: Technical and practical aspects (M. Koutsombogera – ILSP)
16:15 – 16:45	14	Discussion with the audience: How can we engage? (Moderation: S. Piperidis – ILSP/ELRC, G. Floros – University of Cyprus/ELRC, M. Neokleous – DGT, Nicosia Office)
16.45 – 17:00	15	Wrap-up and Conclusions

3 Summary of Content of Sessions

3.1 Session 1: “Opening and welcome (ELRC)”

Welcome addresses by Dr Georgios Floros from the University of Cyprus in Nicosia and by Mr. Stelios Piperidis from the ILSP in Athens. Dr Floros thanked all participants for accepting the invitation to the workshop and stressed that their participation is much appreciated. The presence of Ms. Androulla Vassiliou was of particular importance and a great honor for the workshop organizers. Dr Floros then informed the participants of the feedback forms and asked them to submit the filled-out forms to the conference secretary at the end of the workshop. He also informed the participants that the certificates of attendance would be distributed at the end of the workshop.

3.2 Session 2: “Local welcome”

Welcome address by Ms. Martha Neokleous on behalf of Mr. Georgios Markopouliotis, Head of the Representation of the EU in Cyprus. Mr. Markopouliotis thanked the organizers and the participants and stressed the importance of the CEF instrument in supporting the DSIs and, specifically, the investment in machine translation. He pointed out that the MT@EC and its enrichment through the CEF.AT platform 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 Cypriot public administration to contribute to the machine translation tool for the Greek language and concluded his welcome speech with wishes for a successful workshop.

3.3 Session 3: “Welcome by the European Commission”

Welcome address by Ms. Androulla Vassiliou, former EU-Commissioner for Education, Culture, Multilingualism and Youth. Ms. Vassiliou congratulated the organizers on the event and thanked the participants for accepting to come to the workshop. She reminded the audience that, in her role as Commissioner, she had the chance to inaugurate the MT@EC machine translation system some years ago and stressed the importance of the contribution on the part of the Cypriot public services, since the success of the system depends on the volume of language data contributed. She further stressed the importance of language technology for the Greek language and culture and concluded with wishes for a successful workshop.

3.4 Session 4: “Aims and objectives”

Stelios Piperidis analyzed the multilingual character of the EU and the role of translation in achieving communication within Europe. He then focused on the CEF and CEF.AT, and briefly presented the main stakeholders, principles and goals of this endeavor, as well as its value towards serving the needs of European citizens. He stressed the fact that machine translation depends on language data which need to be provided by the users in order for the benefits to return to them. He then went through the workshop objectives and the agenda and concluded by stressing the importance of the project to the creation of a multilingual Digital Single Market.

3.5 Session 5: “The EU and multilingualism”

Martha Neokleous, Head of DGT Nicosia 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

ELRC Workshop Report for Cyprus

work. She also focused on the importance of multilingualism and translation as pillars of safeguarding democracy within the EU. She then 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 and, finally, briefly informed the audience about the latest trends and achievements in supporting the Greek language through networks and tools.

3.6 Session 6: “Language and language technology in Cyprus”

Georgios Floros, Associate Professor at the University of Cyprus and ELRC-NAP for Cyprus, started by briefly presenting the language situation in Cyprus and the multilingual needs of the Cypriot public sector. He attempted a comparison between Greece and Cyprus as regards such needs and focused on the particularities of the Cypriot context, which is distinguished by the extended use of English and by deviations from the Standard Modern Greek variety spoken in Greece. He then explained in detail the terms *language technology* and *language resources*, and stressed the importance of contributing to machine translation systems with language data from Cyprus, since this will offer the opportunity to include Cypriot terminological and other linguistic particularities in language technological tools.

3.7 Session 7: “Round table 1: Multilingual public services in Cyprus”

Cf. 4.1 below.

3.8 Session 8: “Automated translation: How does it work?”

Stelios Piperidis gave a thorough and succinct presentation on how statistical machine translation tools works. The presentation was tailored to the needs of an audience without previous experience or knowledge of machine translation systems. Mr Piperidis gave examples of how a machine translation system can learn from a suitable corpus of data and explained that the interaction between the translation component (translation model) and the linguistic component (language model) of a machine translation system can fruitfully collaborate to produce a successful translation output. He further talked about how it is possible to make a computer “learn” natural language and vocabulary. Moreover, he explained how it is possible to teach a computer to align segments of different languages on the basis of data which are chosen carefully and in a targeted way. He concluded by calling upon the audience to realize that volume and suitability of data will exponentially reflect on the quality of the translation output by giving further examples of what machine translation systems can afford today.

3.9 Session 9: “Machine translation: How can public institutions benefit from the CEF.AT platform?”

Spyros 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.

3.10 Session 10: “What data are needed?”

Maria Koutsombogera (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

ELRC Workshop Report for Cyprus

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 they are aligned (parallel) or comparable. She then gave an extensive example of the technology used to automatically align data and encouraged the participants to contribute to the project by providing or identifying the required textual content or pointing out key persons/organizations with whom they collaborate.

3.11 Session 11: “Legal framework for data contribution”

Tatiana Synodinou (University of Cyprus) presented the legal framework for data contribution in Cyprus, which actually consists of a combination of a group of laws and directives. She further explained the recent changes regarding the way(s) in which data can be re-used or disseminated, as well as the definitions of the very terms *data*, *information* and *document*. She stressed that the use of language data from the public sector for statistical machine translation systems is a non-commercial use and can therefore be allowed under certain simple conditions. She also informed the audience that the PSI directive for the uploading of public sector data on the data.gov.cy platform will soon be implemented and further explained that the data are separated into four different categories, depending on the degree of accessibility. Moreover, she explained that some data categories are exempted from being allowed to be re-used, due to intellectual property rights, high confidentiality etc. She then listed the step-by-step process to be followed to release data under the PSI Directive and presented actual case studies from various countries. The distinct stages in the process refer to: exclusion of confidential information; consent for, anonymization or exclusion of personal data; taking care of 3rd party copyrights; follow the national PSI transposition rules; use a standard Open Government License, Open public license or re-use license, and follow the national or organizational PSI re-use policy.

3.12 Session 12: “Round-table 2: Language data of the Cypriot public sector”

Cf. 4.2 below.

3.13 Session 13: “Data and language resources: Technical and practical aspects”

Maria Koutsombogera (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. She 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.

3.14 Session 14: “Discussion with the audience: How can we engage?”

Cf. 4.3 below.

3.15 Session 15: “Wrap-up and Conclusions”

Stelios Piperidis, Martha Neokleous and Georgios Floros concluded the workshop by thanking the audience for their participation in the presentation of a very innovative project, and by pointing out that the event was a first step towards the collaboration of the Cypriot public sector with ELRC, in view of the contribution to the high quality service that the

ELRC Workshop Report for Cyprus

CEF.AT platform is expected to bring. In this respect, they all stressed the need for continuous data contribution, as this will have a clear effect on the European languages sustainability and, more specifically, on the digital presence of the Greek language.

4 Synthesis of Workshop Discussions

4.1 Panel 1 (Session 7): “Round table 1: Multilingual public services in Cyprus”

The panel was moderated by M. Neokleous, the Head of DGT Nicosia. The panelists were representatives from the Press and Information Office (Ministry of Interior), the Ministry of Transport and Communications, the House of Representatives, and the Government Printing Office. 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 due to multilingual needs).
- 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: The Press and Information Office has a more elaborate workflow as well as electronic archives, while the translation work at the Ministry of Transport and Communications is mainly based on its higher officials, while they do not keep a central electronic archive. The situation looks similar at the International Relations Service of the House of Representatives, while the Government Printing Office rarely translates, but has a large corpus of texts. The experience with machine translation is very limited, apart from the use of wide-spread, online (but of insufficient quality) systems. The translation needs were more convergent: English is the language most used, with the addition of non-EU languages due to immigration. All panelists pointed out the need for translation to and from Turkish, due to the specific local needs. Quality control is also non-existent in the Cypriot public sector.

4.2 Panel 2 (Session 12): “Round-table 2: Language data of the Cypriot public sector”

The panel was moderated by G. Floros, Associate Professor at the University of Cyprus. The panelists were representatives from the Press and Information Office (Ministry of Interior), the Ministry of Energy, Commerce, Industry and Tourism, the Cyprus Police, and the Central Bank of Cyprus. The moderator addressed the following questions to the members of the panel:

- Language data (text types, languages, topics) used in each organization.
- Specific translation needs (text types and source/target languages).
- Translated data already available (languages, volume of data, type of availability).
- Language data which can be contributed to the CEF.AT platform (languages, problems and legal or other difficulties).

As in the first panel, the answers of the panelists displayed diversity among the different services of the public sector: The Press and Information Office has data in many languages, of many different text types, and the only possible problem would be personal data issues. The Ministry of Energy, Commerce, Industry and Tourism has a very large corpus of data in Greek, and immense translation needs. The same holds for the Cyprus Police, where the translation needs are high, but the availability is very restricted, due to confidentiality and personal data issues. The Central Bank of Cyprus has achieved a very high level of parallel texts (translated material) which is already available through the bilingual website. English is

ELRC Workshop Report for Cyprus

the language most used in all public sectors represented in the panel, with the addition of non-EU languages due to immigration, especially for the Press and Information Office. All panelists pointed out the need for translation to and from Turkish, due to the specific local needs and in view of a possible solution to the Cyprus issue, as bilingualism will then be officially reactivated. It was concluded that although some public sectors cannot contribute language data, some others could probably compensate for such shortcoming.

4.3 Panel 3 (Session 14): “Discussion with the audience: How can we engage?”

In this interactive session, Stelios Piperidis, Martha Neokleous and Georgios Floros discussed a set of key questions with the audience:

- Does your organization produce data which you think are appropriate for the CEF.AT platform?
- Do you think that your organization can share this data for the CEF.AT purposes?
- Are there practical difficulties that could affect your contribution?

The audience raised the following points, which started a vivid discussion among all participants:

- There are organizations 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. Some organizations will not be able to contribute much, due to confidentiality issues (e.g. Cyprus Police, National Guard).
- 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 organization. Further to this, many representatives of the public sector pointed out that it would be very useful if they were offered targeted training in translation technology and tools, to combat problems in their everyday work. Another idea would be the use of existing lifelong learning schemes.
- The institutional and political framework is already there, what is perhaps missing is a contact point/link to the public organizations, as well as a better infrastructural organization of the translation work, so that the data are further exploited and reused.
- Any machine translation system designed for public services should necessarily include Turkish, as well as further non-EU languages such as Russian, due to the specific local needs and in view of a possible solution to the Cyprus issue, as bilingualism will then be officially reactivated.
- The quality of the translation output of machine translation systems. Do such systems provide a reliable and ready-to-use output?
- The issue of dialectal differentiations in terminology, since Cypriot Greek terminology often differs from the terminology used in mainland Greece.
- Access to the presentations was required.

5 Workshop Presentation Materials

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