



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable D3.2.21

Task 8

ELRC Workshop Report for Hungary

Author(s):	Tamás Váradi, ELRC/Research Institute for Linguistics of the Hungarian Academy of Sciences
Dissemination Level:	Public
Version No.:	<V1.0>
Date:	2019-01-16



Contents

1	Executive Summary	3
2	Workshop Agenda	3
3	Summary of Content of Sessions	5
3.1	Welcome and introduction	5
3.2	Welcome by the EC	5
3.3	Connecting public services across Europe: ambition and results so far	5
3.4	National initiatives for digital public services and (open) data	6
3.5	OFFI in the Service of e-government	6
3.6	CEF in Hungary: an outlook into current and future challenges	6
3.7	The implementation of the CEF Programme in Hungary: Panel session	7
3.8	The CEF eTranslation platform @ work	7
3.9	The European Language Resource Coordination (ELRC) action	8
3.10	ELRC in Hungary	8
3.11	Can language data be shared and how? National and European legal framework?	9
3.12	Identifying and managing your data: Questions & Answers	9
3.13	Conclusions	10
4	Synthesis of Workshop Discussions	10
4.1	ELRC and Open Language Data in Hungary	10
4.2	Success stories and lessons learnt	11
5	Workshop Presentations	11
6	Workshop photo gallery	11

1 Executive Summary

This document reports on the second ELRC Workshop in Hungary, which took place in Budapest, on the 12th of November 2018 at the Research Institute for Linguistics, Hungarian Academy of Sciences. It includes the agenda of the event and briefly reviews the content of each individual, interactive and panel workshop session. The participants spanning a wide range of stakeholders, including ministries, the language industry, libraries, public organisations and language technology partners in total 21 organisations attended or showed interest.

The contributions to the workshop provided an overview of EU initiatives for the realization of the Digital Single Market as well as the national policies and actions in Hungary for Digital Public Services and Digital Economy and Society. Showcases of how Machine Translation and/or other language technologies enhance multilingual public digital services were provided both at the European and national level. The workshop especially focused on the European Commission's eTranslation system and its role in cross-border exchanges and how language data created by Hungarian public bodies is the key factor in enhancing its quality and extending its applicability to new domains.

The dedicated event webpage can be found at <http://lr-coordination.eu/l2hungary>.

2 Workshop Agenda

- 08:00 – 09:00 **Registration**
- 09:00 – 09:10 **Welcome and introduction**
Tamás Váradi, ELRC/Research Institute for Linguistics of the Hungarian Academy of Sciences
- 09:10 – 09:15 **Welcome by the EC**
István Peger, The European Commission Representation in Hungary
- Session 1. Connecting a multilingual Europe: European context and local needs**
- 09:15 – 09:35 **Connecting public services across Europe: ambition and results so far**
Alexandra Wesolowska, DG CONNECT
- 09:35 – 09:55 **National initiatives for digital public services and (open) data**
Presenter: Barna Gaál, Ministry of Interior
- 09:55 – 10:05 **OFFI in the Service of e-government**
Presenter: Ákos Kántor, Hungarian Office for Translation and Attestation
- 10:05 – 10:25 **CEF in Hungary: an outlook into current and future challenges**
Presenter: Ferenc Kneisz, Ministry of Innovation and Technology
-

- 10:25 – 11:10 **The implementation of the CEF Programme in Hungary: Panel session**
Moderator: Bódi Zoltán, Institute for Hungarian Language Strategy
Panelists:
- *Lucza Vilmos, National Health Insurance Fund of Hungary*
 - *Ferenc Kneisz, Ministry of Innovation and Technology*
 - *Zsuzsa Kurkó, International Children's Safety Service*
 - *Réka Sárközy, National Széchényi Library*
 - *Szabolcs Sebő, Hungarian Official Gazette*
- 11.10 – 11:30 **Coffee Break**
- 11:30 – 11:50 **The CEF eTranslation platform @ work**
MICHAEL Jellinghaus, eTranslation Expert at European Commission
- Session 2. Engage: hands-on data
- 11:50 – 12:30 **The European Language Resource Coordination (ELRC) action: why and how you can engage and benefit**
Stelios Piperidis, ILSP/ELRC
- 12:30 – 13:30 **Lunch Break**
- 13:30 – 13:50 **ELRC in Hungary**
Tamás Váradi, ELRC/Research Institute for Linguistics of the Hungarian Academy of Sciences
- 13:50 – 14:15 **Can language data be shared and how? National and European legal framework**
Nóra Takács, Hungarian Intellectual Property Office
- 14:15 – 14:45 **Identifying and managing your data: Questions & Answers**
Moderator: *Tamás Váradi Tamás Váradi, ELRC/ Research Institute for Linguistics of the Hungarian Academy of Sciences & Stelios Piperidis, ILSP/ELRC*
- 14:45 – 15:00 **Conclusions**
Moderator: *Tamás Váradi Tamás Váradi, ELRC/ Research Institute for Linguistics of the Hungarian Academy of Sciences & Stelios Piperidis, ILSP/ELRC*
- 15:00 – 15:30 **Coffee Break and networking**

3 Summary of Content of Sessions

3.1 Welcome and introduction

Tamás Váradi, the National Technical Anchor Point and local organizer of the workshop and the Hungarian coordinator of the ELRC in his welcome message greeted the audience and thanked particularly the speakers, the panelists and the potential data providers.

He stressed the importance of breaking down the barriers of linguistic isolation. Language technology and automated translation in particular are central components in the efforts to achieve this goal.

At the heart of automated translation lies the recycling of human translation expertise that serves to train translation systems through machine learning. High quality can only be achieved by feeding large (preferably parallel) corpora of annotated texts that come from the specific domain. The central objective of the ELRC project is to make open data from public administration available to enhance the quality of automated translation. He also described the objectives of the workshop i.e. to identify language resources and tools that could enhance the performance of machine translation and to investigate and address the multilingual needs of Hungarian public services.

He was pleased to note the large number of participants from the public sector and hoped for a fruitful collaboration in the workshop and later on.

3.2 Welcome by the EC

The video address by Rytis Martikonis, Director-General of DG Translation and Claire Bury, the Deputy Director-General at DG CONNECT, was presented to the audience. The two high-level officials presented an excellent and concise overview of the objectives of the ELRC and how it fits into the big picture. While this welcome video message obviously required extra time, its benefits far compensated for its effect on the timetable of the Workshop, which was restored by the end of the event.

István Perger, Deputy Head of the the European Commission Representation in Hungary, emphasized the importance of languages in the rich cultural history and diversity of Europe. From 2004 the volumes of text to translate has multiplied, therefore efficient and quick translations are more needed than ever. EU texts are translated everyday with the use of tools that improve translators' efficiency. Mr. Perger talked about the Digital Services Infrastructures (DSIs), among them the machine translation system that the European Commission is offering to public services, to facilitate the Digital Single Market.

Finally he thanked for the opportunity to be at the workshop and wished a successful and fruitful workshop.

3.3 Connecting public services across Europe: ambition and results so far

Alexandra Wesolowska, from EC DG CONNECT gave a very informative presentation about connecting public services across Europe via video recording. Those who were not familiar with

the Digital Services Infrastructure (DSIs) and Digital Single Market sectorial platforms/websites could get an overview of them and their characteristics. Ms. Wesolowska described the connection between DSIs and eTranslation building block and the reuse of building blocks within the CEF Programme.

3.4 National initiatives for digital public services and (open) data

The speech of **Barna Gaál** from the Ministry of Interior was about the national strategies for eGovernment. As one of the most informative and authoritative presentations of the workshop, it was a great start to the first session.

Mr. Gaál showed diagrams of the position of Hungary in the field of Digital Public Services and Digital Economy and Society. He concluded that while there is a lot of room for improvement in many specific areas, Hungary fortunately has exhibited significant improvements and successes such as eID cards, Electronic Health Cooperation Service Space and Hungarian Municipality ASP.

The audience could get also an excellent overview of the multilingual public digital services as well as some examples of multilingual public services enabled by MT and/or other language technologies.

Regarding open data policies, Mr. Gaál urged for a paradigm shift and stronger coordination between the Ministry of Interior and the Ministry for Innovation and Technology. As an initiative Mr. Gaál mentioned the Hungarian Central Governmental Service Bus (KKSZB) which is an Interoperability Platform that aims to ensure a service-oriented and standardized connection between the national base registries and the different specific public administration information systems.

3.5 OFFI in the Service of e-government

Ákos Kántor, chief consultant to the Director General of the Hungarian Office for Translation and Attestation (OFFI) gave a talk about the role of OFFI in eGovernment and about the adopted solution for e-attestation.

The Hungarian Office for Translation and Attestation is a unique institution in Europe in the field of certified (or attested) translations with a history of over 150 years. It has also a unique situation as from 1994 it is a 100% state-owned shareholding company.

Mr. Kántor described the organization and work of OFFI with some figures, as well as the connection with the Ministry of Justice and the duties of the Office to provide public services. Finally, he presented the steps taken by OFFI to modernise the translation certification processes through e-attestation.

3.6 CEF in Hungary: an outlook into current and future challenges

As a preface of the panel session **Ferenc Kneisz**, senior adviser in the Ministry of Innovation and Technology, national contact point for CEF Digital and as such a key stakeholder in the CEF Programme in Hungary, gave a talk about the CEF Programme current and future challenges in Hungary.

In the Digital Economy and Society Index from 2017 Hungary is the 8th from the bottom, obviously a position that calls for improvement.

Mr. Kneisz presented a highly informative first-hand overview of the CEF Programme in Hungary. He provided some practical advice as to how to prepare a successful CEF proposal. He showed an overview of the Work Programme 2019 and the proposed draft budget distribution and pointed out the fields where Hungary could improve such as eProcurement, eHealth, eSignatures etc.

3.7 The implementation of the CEF Programme in Hungary: Panel session

The moderator of the panel session was **Bódi Zoltán** from the Institute for Hungarian Language Strategy. The panelists were as follows: **Lucza Vilmos** (National Health Insurance Fund of Hungary), **Ferenc Kneisz** (Ministry of Innovation and Technology), **Zsuzsa Kurkó** (International Children's Safety Service), **Réka Sárközy** (National Széchényi Library), **Szabolcs Sebő** (Hungarian Official Gazette).

In the panel discussion four outstanding projects were introduced. The representative of National Health Insurance Fund of Hungary reported the implementation of Electronic Exchange of Social Security Information (EESSI) Programme between Hungarian and EU (social security) institutions. The implementation of the Programme takes place in Hungary according to the EU regulations. The Programme seeks to eliminate the paper-based exchange of social security information and to replace it by electronic exchanges.

The International Children's Safety Service performs as the realizer of the Safer Internet Programme among the implementers of CEF. The aim of the Safer Internet Programme launched by the European Commission is to enable internet users to browse safely/securely and to know whom they can ask for help in case of problem.

The National Széchényi Library is the realizer of the Europeana project in Hungary. The Europeana Collections make European cultural heritage content available to a global audience, making it easier for people to use for work, learning or pleasure.

The Hungarian Official Gazette shared professional experiences connected to integration of Hungarian Business Register into EU Business Registers Interconnection System (BRIS) project. Participants were informed about the BRIS platform that provides EU-wide company information by linking central, commercial and business registers.

The panelists agreed that the eTranslation platform can be a useful tool to overcome the language barrier in the implementation of each of the above mentioned projects.

3.8 The CEF eTranslation platform @ work

Michael Jellinghaus, eTranslation Engines Project Manager, DGT, EC

Mr Jellinghaus is a member of the DGT development team, where the latest machine translation system, aka eTranslation, is being developed. This presentation was unique in that it came from the most authentic source and also it came in a live video link. Mr. Jellinghaus gave an overview of the history of the machine translation system developed at DGT, discussing the technological aspects of the succeeding generations of MT systems. He dwelled on the recent inauguration of the eTranslation system, which now has engines for all EU languages in the neural network

paradigm. He introduced workshop participants to the eTranslation website where translation of texts in a variety of formats and sources can be requested. Mr. Jellinghaus also described the process of the gradual deployment of the eTranslation system Europe-wide, both in terms of increasing accessibility to a wider range of users and also of covering an increasing number of domains. He also confirmed, along with many other presenters during the day that data, and domain specific data at that, is the key to the success of the eTranslation system and ELRC is working to provide the necessary data sets.

3.9 The European Language Resource Coordination (ELRC) action

Stelios Piperidis, Institute for Language and Speech Processing / Athena R.C., ELRC

Mr. Piperidis provided a comprehensive overview of the EC's initiatives towards breaking down language barriers in the Digital Single Market. He introduced all the evolving set of components such as MT@EC, CEF.AT and eTranslation and their relationship to each other. One leading idea of implementing the Digital Single Market is through various Digital Service Infrastructures (DSI), which also face the challenge of the language issue. Among the DSIs, those that can be reused by other digital services are called Building Blocks. eTranslation is such a building block, as it can not only be used as a stand-alone service, but it can also be embedded in other DSIs, such as the Online Dispute Resolution.

Mr. Piperidis discussed ELRC from an organisational point of view. ELRC as a project consists of a core consortium that works with a network of technology and public administration contact points. The main activities carried out by ELRC is to identify, locate, prepare and make such language resources (preferably parallel corpora from the public administration, hence data subject to the PSI) available as training data for the Translation system. As present day machine translation systems employ machine learning to train the system, key to the success of such system is clean and relevant (domain-specific) data in as big quantities as possible.

Mr. Piperidis followed with a detailed overview of the workflow of collecting the above kind of data. He laid great emphasis on the services of the central ELRC team, who can come to the aid of data holders whenever they need to resort to outside help. ELRC is ready to give on-site assistance, by mobilising the network of technology experts, to provide advice and help whether on technical, legal or organisational issues.

Finally, Mr. Piperidis explained in very clear terms what the potential benefits of donating data to the ELRC may be. The workshop participants were pleased and intrigued to note that they can benefit from handing over raw data by receiving valuable processed data (clean, aligned bilingual text in some standard format).

3.10 ELRC in Hungary

Tamás Váradi the ELRC National Anchor Point described how the objective of breaking the language barriers in an effort to build the European Digital Single Market rest on several pillars, including the ELRC action, as well as the project work in the framework of CEF Telecom project work.

He reported that, after the administrative reshuffling that took place following the recent elections in April 2018, ELRC could establish and build solid contacts with new relevant government departments including the Cabinet Office and the Department of e-Government at the Ministry of the Interior. As regards the motivations for Government Departments to take

part in the ELRC, he reminded workshop participants of the European Parliament resolution of 11 September 2018 on language equality in the digital age, which defined very specific objectives in this area and called on the European Commission as well as member states to cooperate and implement specific Programmes to ensure that the digital viability of all EU languages.

Finally, Mr. Váradi was very pleased to note that a consortium of seven neighbouring countries (Bulgaria, Croatia, Hungary, Poland, Romania Slovakia, and Slovenia) have recently been awarded funding for the MARCELL project, which aims to compile seven large-scale domain-specific corpora from the national legal domain. The two-year project started on the 1st of October 2018.

3.11 Can language data be shared and how? National and European legal framework?

Nóra Takács is a legal expert from the Hungarian Intellectual Property Office. In her presentation she described the Public Sector Information directive and in more details the 2012 LXIII. law about the Reusage of public data.

She presented the step-by-step procedure to publish data. Ms Takács discussed in detail how to: 1. exclude confidential information. 2. gather the permission and assure the rights. 3. Anonymize or delete the personal data. 4. gather the licences. 5 observe the national PSI law.

The ensuing discussion revolved around what quantity of data use is considered fair and reasonable. She regretted that such issues are not defined *in abstracto* but the answer to such questions can only be given in full consideration to the specific contexts and the circumstances of the case.

3.12 Identifying and managing your data: Questions & Answers

Moderator: Tamás Váradi, ELRC/MTA-NYTI & Stelios Piperidis (ILSP/ELRC)

By the time we reached this point of the Agenda some participants had to leave the workshop to attend to other pressing engagements. Fortunately, however, this did not mean that the audience sat through the session in silence. On the contrary, a very interesting and lively debate ensued with the representatives of the language services industry. Ms Ágnes Szabó, from the Secretariat of the Association of Hungarian Translators and Interpreters asked how their members could serve the objectives of the ELRC. The challenge, in her view, was that their members are typically individual freelancers, sometimes working at home, and that they may not necessarily be well equipped. Tamás Váradi welcomed the remark and suggested that they start negotiating possible cooperation with Hungarian ELRC and the Association. Ákos Kántor, OFFI, remarked that they would be interested to try to test the system but so far as he understood, the eTranslation is not available to stakeholders in the private sector. Mr. Piperidis acknowledged that this was indeed the case and there may be legal constraints involved. Tamás Váradi offered OFFI the chance to have a hands-on experience of the e-Translation system on the premises of the Research Institute of the Hungarian Academy of Sciences.

3.13 Conclusions

Moderator: Tamás Váradi (ELRC/MTA-NYTI)

In the concluding session Tamás Váradi gave a brief summary of the main issues on the agenda of the workshop ranging from the CEF Programme and how ELRC is organised to support its objectives, the role of eTranslation as a building block and how data is the key factor in enhancing its quality and extending its applicability to new domains. He hoped that participants have now appreciated the value of translation data that may be residing in their organisations and appealed to them to share their data with ELRC.

4 Synthesis of Workshop Discussions

In this section we would like to confine ourselves only to some recurring issues that came up in the discussion. One issue concerned was the need to run the eTranslation system as a local installation. It was argued very cogently by Mr. Barna Gaál, deputy head of the eGovernment department of the Ministry of Interior, that they would consider using the system but it would be far more useful if they could use it for all sorts of documents without constraints, which is far from being the case unless they can create their own installation. The participants discussed this issue from many viewpoints: from a language technology point of view, a local installation would amount to what is called in software technology a fork, a variant of the original software that begins to live a separate life of its own thereafter, independent of the original version, with all the responsibility for maintenance and training of the system falling on the party opening up such a fork. It appeared from Mr. Jellinghaus' position on this issue that the deployment plans for the eTranslation system did not consider such need a high-priority issue. On the other hand, one can argue that the request for local installation did not reflect awareness of all the duties and responsibilities that a software fork may imply.

Another issue at the center of debate was the accessibility of the eTranslation system for commercial partners. OFFI, the biggest and venerable 150 years-old LSP would be interested in using the system, but currently they don't even have the right to test it. Although the question of at least testing was informally arranged, this question is still open at the policy level.

4.1 ELRC and Open Language Data in Hungary

Mr. Barna Gaál (Ministry of Interior) is one of the key officials in charge of eGovernment policy in Hungary. The implementation of the PSI directives is the joint responsibility of the Ministry of Interior and the Ministry for Innovation and Technology. Mr. Gaál argued for a paradigm shift but also pointed out that in order to make public data accessible by default requires strong coordination. Open data bring economic opportunities and represent the foundation of emerging technologies. Recently, there have been marked developments in the field of eGovernment in Hungary, which do not show up in the DESI (the Digital Economy and Society Index). Hungary ranks quite low in the DESI Open Data Component, it ranks among the last 4, in fact. The usage of digital public services is relatively high, with 3.6 million regular users of the Public Services Client Portal. Multilingual support for public digital services is still patchy. At the national level the Public Administration and Public Service Development Operational

Programme provides for measures for Programmes for developing tools and services for multilingual public administration. ICT work is the easier part, the challenge being the content provision and the updating of forms as well as the operative administration of the systems.

4.2 Success stories and lessons learnt

- The workshop was a clear success in terms of the quality and relevance of stakeholders that it could engage. It attracted interest from a much wider range of institutions from the public administration compared to the first ELRC workshop. Even those who finally did not turn up replied positively to our follow-up letters.
- We think that success in the outreach of the workshop was due to investing great effort in the preparatory work: to find a key player in public administration (the department of eGovernment of the Ministry of Interior in our case), to hold several preparatory sessions and to rely on their channels of communication to invite participants.

5 Workshop Presentations

The local contributions to the workshop in the form of presentations are available on the ELRC website http://lr-coordination.eu/l2Hungary_agenda.

6 Workshop photo gallery





