



Deliverable D3.2.3 Task 8

ELRC Workshop Report for Poland



Author(s):	Maciej Ogrodniczuk (Institute of Computer Science, Polish Academy of Sciences) Anna Kotarska (National Health Fund, Central Office in Warsaw)
Dissemination Level:	Public
Version No.:	<V1.0>
Date:	2018-03-02



Contents

1	Executive Summary	3
2	Workshop Agenda	4
3	Summary of Content of Sessions	5
3.1	Welcome and introduction / Welcome by the EC	5
3.2	Connecting public services across Europe: ambition and results so far	5
3.3	National initiatives for digital public services and (open) data	5
3.4	CEF in Poland: an outlook into current and future challenges – Panel session	5
3.5	The CEF eTranslation platform @ work	6
3.6	The European Language Resource Coordination (ELRC) action	7
3.7	ELRC in Poland	7
3.8	Can language data be shared and how?	7
3.9	Preparing and sharing data with the ELRC repository – and what happens next	7
3.10	Identifying and managing your data: Questions, Answers and Conclusions	7
4	Synthesis of Workshop Discussions	9
4.1	ELRC and Open language Data in Poland	9
4.2	Success stories and lessons learnt	10
5	Workshop Presentation Materials	11

1 Executive Summary

This document reports on the ELRC+ Workshop in Poland, which took place in Warsaw on the 18th of December 2017 at the premises of The European Commission Representation in Poland (ul. Jasna 14/16a). It includes the agenda of the event (section 2) and briefly informs about the content of each individual, interactive and panel workshop session (section 3).

Some of the main issues raised at the workshop are the following:

- Need for a public-private sector initiative for developing guidelines and good practices for contracting translation services in the public sector, to be approved by the Public Procurement Office
- Need for awareness-raising actions on the purpose of ELRC in relevant public sector events and focused meetings
- The lack of technological expertise in the public sector especially with regard to the use of CAT tools. As a result sharing of language resources by public entities is limited mostly to non-TMX formats.
- Polish public institutions are usually unaware of the technical problems of data delivery and this is why the role of ELRC is critical: taking over the technicalities is absolutely necessary to make data delivery a success
- The degree of aggregation of the different channels used by the EC for collecting language data and their relation, e.g. IATE, ELRC, generic services projects.

The dedicated event webpage together with all presentations can be found at <http://lr-coordination.eu/l2poland>.

2 Workshop Agenda

08:00 – 09:00 Registration

09:00 – 09:15 Welcome and introduction / Welcome by the EC

Maciej Ogrodniczuk – Institute of Computer Science, Polish Academy of Sciences, ELRC Technical NAP

Marek Prawda – Head of Representation of European Commission in Poland

09:15 – 09:35 Connecting public services across Europe: ambition and results so far

Susan Fraser – European Commission, DG CONNECT

09:35 – 09:55 National initiatives for digital public services and (open) data

Grzegorz Zajączkowski – Ministry of Digital Affairs

09:55 – 10:40 CEF in Poland: an outlook into current and future challenges – Panel session

Anna Kotarska – National Health Fund, ELRC Public Services NAP (moderator)

Małgorzata Ławicka – National Health Fund / EESSI

Artur Miękina – Polish Chamber of Information Technology and Telecommunications

Wojciech Wołoszyk – IURIDICO Legal & Financial Translations

Grzegorz Zajączkowski – Ministry of Digital Affairs

10:40 – 11:00 The CEF eTranslation platform @ work

Szymon Klocek – European Commission

11:00 – 11:30 Coffee break

11:30 – 11:55 The European Language Resource Coordination (ELRC) action

Stelios Piperidis – ELRC

11:55 – 12:15 ELRC in Poland – Maciej Ogrodniczuk

12:15 – 12:40 Can language data be shared and how?

Natalia Mileszyk – Centrum Cyfrowe

12:40 – 13:40 Lunch break

13:40 – 14:15 Preparing and sharing data with the ELRC repository – and what happens next

Maria Giagkou – ELRC

14:15 – 14:45 Identifying and managing your data: Questions & Answers

Maciej Ogrodniczuk, Stelios Piperidis

14:45 – 15:00 Conclusions

Maciej Ogrodniczuk, Stelios Piperidis

15:00 – 15:30 Coffee break and networking

3 Summary of Content of Sessions

3.1 Welcome and introduction / Welcome by the EC

Maciej Ogrodniczuk from the Institute of Computer Science, Polish Academy of Sciences, the local ELRC representative and the Polish Technical National Anchor Point, opened the event by welcoming the audience and introducing the key persons in conceiving and organizing the event, the ELRC consortium, the EC/DGT representatives, speakers and panelists. Marek Prawda, the Head of European Commission Representation in Poland welcomed the audience on behalf of the EC.

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

Susan Fraser, the Project Officer at the European Commission – DG Connect presented via video link the Connecting Europe Facility and CEF Digital, including the position of Poland in Europe according to the Digital Economy and Society Index 2017, sectoral platforms of the digital single market (DSIs), the eTranslation building block and how it is reused within the CEF programme. The presentation ended with a call for involvement to use the MT@EC/eTranslation web interface, integrate eTranslation API in Polish online public services and to continue the contacts/dialogue with CEF and the ELRC.

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

Grzegorz Zajączkowski, the Digital Champion of the EC for Poland and Advisor in the Political Cabinet of the Polish Minister of Digital Affairs presented the characteristics of several open data projects carried out by the Ministry of Digital Affairs (primarily danepubliczne.gov.pl but also obywatel.gov.pl, biznes.gov.pl or widok.gov.pl) and strategic guidelines of the Ministry towards the digital public services in Poland. The presentation showed the rising adoption of open data portals by businesses and citizens, increasing quality of access and quality of public data available as well as new methods of monitoring e-services and their use by the citizens. Mr. Zajączkowski also presented a new initiative of a social helpdesk targeted at users experiencing problems with access to public services and how the government can react by tracking social media activity and responding using the same channel.

3.4 CEF in Poland: an outlook into current and future challenges – Panel session

The panel was moderated by Anna Kotarska from the Central Office of the National Health Fund, ELRC Public Services National Anchor Point for Poland and it focused on presenting the challenges linked with adopting the CEF building blocks in the panelists' institutions in the areas related to translation. The panelists were: Małgorzata Ławicka-van der Veen from the Central Office of the National Health Fund (NHF), an EESSI SPOC and expert, a member of the project steering committee; Artur Miękina from the Polish Chamber of Information Technology and Telecommunications/Asseco, trust services expert (eSignature); Wojciech Wołoszyk, a lawyer-linguist from IURIDICO Legal & Financial Translations, providing translation services for Polish and European public institutions (industry); and Grzegorz Zajączkowski from the Ministry of Digital Affairs.

Małgorzata Ławicka-van der Veen presented the EESSI project organization in Poland and its needs with regard to translation services. The three major Polish public institutions involved in the project implementation include: the Polish Social Insurance Institution (<http://www.zus.pl/>), the Agricultural Social Insurance Fund (<https://www.krus.gov.pl/en/>), and the NHF (<http://www.nfz.gov.pl/>). At present the stakeholders of the project are not fully satisfied with the quality and cohesion of the terminology used in the SED forms that are used for the electronic exchange of social security

ELRC Workshop Report for Poland

information and, in their opinion, it should be re-verified. They see a possibility of employing the automated translation for the translation of the numerous and broad procedures/ descriptions related to the SED documents, constituting the guidelines for the users - the administrative staff of the specified institutions.

Artur Miękina presented the use of eSignature in the context of communication between both individuals (i.e. freelance and sworn translators), translation agencies and public authorities, regarding delivery of translation services. He emphasized the crucial importance of security of information (translation) delivered as well as time effectiveness of the solution. Although the number of eSignature users has been growing, it is still far from being popular with the translators, who prefer traditional means of delivery, identification and authentication. CEF leaders should raise the awareness that translation (eTranslation) is a significant factor facilitating the development of the Digital Single Market, and that along with the eID, eDelivery, and eInvoicing, it is an inherent component of a comprehensive strategy for the development of pan-European digital services. It should be stressed the concurrent use of the building blocks would bring the synergistic benefits.

Wojciech Wołoszyk shared his opinion and recommendations regarding the procurement of translation services by public sector entities and its implications for the ELRC project. While giving a critical overview of the tender specifications drawn up in the recent years, frequently leading to insufficient contract execution (poor quality translation), and imperfect application of the provisions of the amended public procurement act, he highlighted the need for informative and awareness raising actions, including drafting relevant guidelines for contracting of translation services. Those should be approved and distributed among public institutions in collaboration with the Public Procurement Office. Most importantly, W. Wołoszyk emphasized the role and the legal status of the Translation Memories (TMs) generated in the process of translation with the use of CAT tools. Due to general lack of technological expertise in the public sector concerning these tools and the proper translation process itself, sharing language resources by public entities is limited mostly to non-TMX formats. Firstly, the contracts concluded in the public sector almost never include the requirement to deliver the TMs to the contracting party. Secondly, while including such provisions, particular attention should be paid to securing respective rights of the institution that would allow it to submit the TMs as „ready to use” language resources to the ELRC. Including the described solution in the guidelines and good practices for the industry combined with the widespread implementation in the contracts of a clause regulating TMs delivery and status could be a highly efficient and sustainable solution for continuous delivery of language resources for the eTranslation project. It would also be in line with current technological trends and improve the overall quality of translations in the long-term. Grzegorz Zajązkowski rounded up the discussion by expressing optimism on the uptake of digital services in Poland. He agreed that awareness raising and educational actions combined with other forms of support are necessary for the growth of digital society.

3.5 The CEF eTranslation platform @ work

Szymon Klocek from the European Commission presented via video link the major features of the eTranslation platform started in November 2017, the differences between statistical and neural translation and the agenda of the adoption of the system for new languages (including Polish, to be available in April 2018). Mr. Klocek also presented the interface of the new system, an English-German translation example and target audiences of the system (in Poland: all users representing public sector institutions, with e-mail addresses ending with .gov.pl).

3.6 The European Language Resource Coordination (ELRC) action

Stelios Piperidis presented the overview of the ELRC project within CEF, the ELRC consortium and ELRC tasks. The talk underlined the necessity of combining in-domain and out-of-domain data, complementing the Polish contribution so far, the largest of all South- and East-European countries. The presentation ended with the call for action to use the ELRC Share repository, ELRC helpdesk and on-site assistance.

3.7 ELRC in Poland

Maciej Ogrodniczuk presented the achievements of ELRC in Poland with respect to data collection (18 data sets collected so far, mostly bilingual Polish-English corpora but also two terminological resources and the monolingual Polish Court Rulings Corpus), basic statistics of data collected (10 domains, 2M words in the multilingual part and 178M words in the monolingual corpus) and the process of data conversion (all data delivered in TMX format after their in-house conversion from PDF). The presentation ended with the summary of main lessons learned and several successes of the process (see section 4.2).

3.8 Can language data be shared and how?

Natalia Mileszyk from Centrum Cyfrowe presented different legal considerations regarding data, resources and databases with respect to data reuse, differences between public information and public sector information and the transposition of the PSI directive into national regulations in Poland. The focus of the presentation was also on the Programme of Opening the Polish Public Data adopted in September 2016 and advocated by the Ministry of Digital Affairs.

3.9 Preparing and sharing data with the ELRC repository – and what happens next

Maria Giagkou from ELRC presented how the data is shared using the ELRC repository using several Polish examples from existing data, recommended types of data, file formats for parallel texts as well as dos and don'ts of data preparation process. The ELRC Share repository was presented step-by-step and ELRC services for data processing and on-site assistance, available free-of-charge to all Polish public organisations, were presented to workshop participants.

3.10 Identifying and managing your data: Questions, Answers and Conclusions

Maciej Ogrodniczuk, Stelios Piperidis and Anna Kotarska moderated the concluding discussion with the following three main points raised:

1. almost all participants agreed that they are willing to use Polish-English eTranslation when ready,
2. Bartłomiej Czerkowski from the Civil Aviation Authority, who already contributed terminological data, expressed his willingness to share corpus data as well and asked about the level of integration of multilingual data by the EC in their different systems (IATE, eTranslation databases etc.). He also raised a concern about the different channels through which the EC/EU requests data for reuse. For example, the Civil Aviation Authority already share their terminologies with IATE on a regular basis, but they were also approached for the same purpose by ELRC.
3. Grażyna Hadjiraftis from the National Fund for Environmental Protection and Water Management thanked for the hints they will be able to use in the contracts with translating companies and raised the issue of translating tables, drawings and images in EU documents

ELRC Workshop Report for Poland

which they frequently translate – the Polish documents are much larger than English which makes the documents difficult to typeset; the cost of post-editing is higher than the cost of translation (e.g. translation – 7K PLN, post-editing – 10K PLN).

4 Synthesis of Workshop Discussions

Other points raised and ideas mentioned in less formal discussions during coffee breaks and lunch:

- Need for improvement of contracting translation services in the public sector, by widespread use of quality and technical criteria in tender specifications;
- Public-private-sector initiative of developing guidelines and good practices for contracting translation services in the public sector, to be approved by the Public Procurement Office and distributed across the public sector;
- Organizing a workshop in collaboration with Public Procurement Office for raising the awareness of the translation process, technology used, products (including translation memories) in the context of their significance for the ELRC project, re-use under relevant local provisions implementing the PSI Directive;
- Collaboration with the National Centre for Health Information Systems (PL: CSIOZ) in respect of implementing semantic interoperability related to eHealth including the use of international, local, and aggregation terminologies as well as other international classifications (i.e. SNOWMED CT, ICD-10, ICD-9, International Classification of Functioning, Disability and Health – IC, International Classification for Nursing Practice - ICNP®, Anatomic-Therapeutic-Chemical Classification, Logical Observation Identifiers Names and Codes) in transborder exchange of health information (Electronic Health Documentation, ePrescription, eDispensation and other ICT systems). Investigating the use of eTranslation Building Block for translation or codification of the unstructured information included in the above mentioned documents. A meeting with the management of CSIOZ and experts scheduled for 14th February 2018 (ELRC to be represented by Anna Kotarska, PS NAP);
- Delivering presentations of the ELRC programme directly at the public institutions (i.e. the Ministry of Development expressed their interest in the idea);
- Possible collaboration with the University of Warsaw, Institute of Applied Linguistics, in the scope of language resources preparation and their quality assessment, translation engines tests and/or development;
- Joint workshops with the university as part of the Translating Europe Workshops event series, including terminology extraction, corpora. Possibly joint participation in other projects;
- Invitation to present the ELRC /eTranslation at the Translation and Localization Conference in Warsaw (23-24 March 2018, <https://sites.grenadine.co/sites/tlc/en/tlc-2018/schedule>). ELRC to be represented by Anna Kotarska and Rihards Kalnins from Tilde.

4.1 ELRC and Open language Data in Poland

The Polish open data portal Dane Publiczne (<https://danepubliczne.gov.pl/en/>) contains datasets provided by government administration bodies, earmarked funds, Social Insurance Institution, Agricultural Social Security Fund, National Health Fund, national research institutes and other state legal entities. Resources offered by central administration constitute almost 90% of all available data. Most datasets offered by the portal are numerical/tabular, with prevailing share of XLS(X), CSV, XML and JSON formats. PDF, DOC(X), RTF and other textual resources constitute only approx. 15% of all data formats (as of February 2018). All resources are categorized as 1- to 3-star open data. Majority of them are updated in a yearly cycle.

ELRC Workshop Report for Poland

4.2 Success stories and lessons learnt

In the first phase of the ELRC project (2015-2017) Poland managed to be among the best performing countries in terms of the quantity of language data delivered to ELRC. Some of the factors that contributed to this achievement are the following:

- in some institutions access to decision makers, whose approval is required for unlocking the data, is very limited; calling them always brings much better results than mailing. In general, personal contacts (meetings and phone calls) are more effective than written formal inquiries.
- persistence in the process of data acquisition is a very important factor (e.g. the case of the Institute of National Remembrance, ready to sign the contract for data after 12 months of official exchange of documents)
- finding a local ELRC/MT advocate in a public institution helps a lot (e.g. the case of Civil Aviation Authority)
- Polish public institutions are usually unaware of the technical problems of data delivery but this is where ELRC can help; taking over the technicalities is absolutely necessary to make data delivery a success
- introduction of the uniform and stable procedures of data retrieval from Polish public institutions and its delivery to ELRC is still to be introduced in Poland.

5 Workshop Presentation Materials

The presentations are published on the Polish workshop agenda webpage (http://lr-coordination.eu/l2poland_agenda and http://lr-coordination.eu/pl/l2poland_agenda).