



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable Task 6

ELRC Workshop Report for Belgium



Author(s): Ayla Rigouts Terryn

Dissemination Level: Public

Version No.: <V01.0>

Date: 13-04-2016



Contents

1	<u>Executive Summary</u>	3
2	<u>Agenda</u>	4
3	<u>Summary of Content of Sessions</u>	5
3.1	Welcome, Aims and Objectives of the workshop	5
3.2	The EU and Multilingualism	5
3.3	Languages and Language Technologies in Belgium	5
3.4	Automated Translation: how does it work?	5
3.5	How can Public Institutions Benefit from the CEF-AT Platform?	6
3.6	What Data is needed?	6
3.7	Legal Framework for Contributing Data	6
3.8	How can I Contribute: Practical and Technical Aspects	7
3.9	Wrap-up, On-site Conclusions and Commitments	7
4	<u>Synthesis of Workshop Discussions</u>	8
4.1	Panel 1 - Multilingual Services in Belgium: Functioning, Data and Linguistic Resources	8
4.2	Panel 2 - Data and Language Resources in Belgium: where to find them?	8
5	<u>Workshop Presentation Materials</u>	9

ELRC Workshop Report for Germany

1 Executive Summary

This document reports on the ELRC Workshop in Belgium, which took place in Ghent, on the 13^h of April 2016. 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 51 participants, spanning a wide range of ministries and public organisations. The dedicated event webpage can be found at <http://lr-coordination.eu/belgium>.

ELRC Workshop Report for Germany**2 Agenda**

08:30 – 09:00	Registration + coffee	
09:00 – 09:15	Welcome, Aims and Objectives of the workshop	Véronique Hoste, Ghent University Khalid Choukri, ELRA & ELDA
09:15 – 09:45	The EU and Multilingualism	Jeroen Aspeslagh, European Commission
09:45 – 10:15	Languages and Language Technologies in Belgium	Walter Daelemans, University of Antwerp
10:15 – 10:45	Coffee break + Networking	
10:45 – 11:30	Panel 1: Multilingual Services in Belgium: Functioning, Data and Linguistic Resources	Stijn De Smeytere, Chancellery of the Prime Minister
11:30 – 12:00	Automated Translation: how does it work?	Lieve Macken, Ghent University
12:00 – 12:30	How can Public Institutions benefit from the CEF-AT Platform?	Kimmo Rossi, European Commission
12:30 – 13:30	Lunch	
13:30 – 14:15	Panel 2: Data and Language Resources in Belgium: where to find them?	Moderator: Véronique Hoste, Ghent University
14:15 – 14:45	What Data is Needed?	Khalid Choukri, ELRA & ELDA
14:45 – 15:15	Legal Framework for Contributing Data	Joris Deene, Everest Law)
15:15 – 15:45	Coffee break + Networking	
15:45 – 16:15	How can I Contribute: Practical and Technical Aspects	Khalid Choukri, ELRA & ELDA
16:15 – 16:30	Wrap-up, On-site Conclusions and Commitments	Véronique Hoste, Ghent University Stijn De Smeytere, Chancellery of the Prime Minister Khalid Choukri, ELRA & ELDA
16:30	Reception	

3 Summary of Content of Sessions

3.1 Welcome, Aims and Objectives of the workshop

Véronique Hoste, the National Anchor Point in Belgium, and Khalid Choukri, as 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.

They stressed the importance of multilingualism in the EU and emphasised that, despite the difficulties multilingualism can create, it should be seen as an asset, an inherent part of a rich European culture. That is where translation comes into play, by indiscriminately allowing people with different native languages to communicate. But since human translation can no longer keep up with demand, linguistic resources are needed to improve machine translation, which is why this Workshop has been organised: to implore national public services to share their data for this worthy cause. As a concrete example, Khalid Choukri explained the concept of the multilingual European Single Digital Market (DSM).

After this introduction, the agenda for the day was presented and the stage was ceded to Jeroen Aspeslagh.

3.2 The EU and Multilingualism

Jeroen Aspeslagh, Main Terminologist at the Dutch department of the DG Translation of the European Commission, set the European scene with its 24 official languages and 60 major regional/minority language pointing out that all languages are equal in Europe and that Europe is committed to multilingualism.

3.3 Languages and Language Technologies in Belgium

After highlighting the importance of translation within the EU, Walter Daelemans, Professor of computational linguistics at the University of Antwerp, took the stage and showed the audience where language technology fits into this bigger picture. With some examples from everyday life, he explained that language technology is omnipresent and that it has already come a long way. However, he also clarified that “we are not nearly there yet”. Machine translation for English can already be considered “good”, French and Spanish offer “moderate” support, but all other European languages lag behind with “fragmentary” or even “weak/no support”. Once again, it became clear that data is pivotal for progress in this area.

After concluding with a short overview of language technology industry in Belgium and the Netherlands, it was time for a coffee break and some networking.

3.4 Automated Translation: how does it work?

Since the aim of the Workshop was to collect data to improve Machine Translation (MT), some explanation of the subject was in order. This explanation was provided by Professor Lieve Macken of Ghent University. She gave an interactive and comprehensible overview of the technology behind MT. She had the audience decipher Chinese menus and had everyone competing to be the first to solve the puzzle. Then, she linked the strategies that we, as humans, used to solve this riddle, to the strategies used by MT. The presentation led to a better understanding of why data is needed and was met with enthusiasm.

ELRC Workshop Report for Germany

3.5 How can Public Institutions Benefit from the CEF-AT Platform?

Kimmo Rossi (Head of Sector, DG CONNECT, Unit G3 – Data Value Chain, European Commission) began his presentation by illustrating the typical interactions between administrations, citizens, and businesses as much within a Member State as between Member States or between Member States and the European Union. Is it possible to access any service or business in any Member State using your own language?

Based on the illustration that just translating the europa.eu website in all 24 European languages with all 2.000 translators in the EU services would take 100 years, Kimmo Rossi stated that MT is clearly the only viable solution to provide cross-lingual access to information in all Member States. Furthermore, the example showed how MT can be used with respect to Human Translation, for example by giving a rough idea of the content of a document written in a foreign language, allowing the user to decide whether this document is relevant or not and if it needs to be translated accurately (by a human translator).

The MT@EC platform has been launched on 26th of June 2013. It has a web user interface in 24 languages for a human-to-machine use case, or can be used as a web service in a machine-to-machine scenario. CET.AT will augment the service provided by MT@EC by allowing for everyday language, by being able to be more knowledgeable of domains and by being a multilingual enabler not only based on MT.

3.6 What Data is needed?

Khalid Choukri emphasised that ELRC is not only looking at parallel data but at dictionaries, terminologies, ontologies, etc. as well. So-called comparable texts, where texts in two languages share similar features without being direct translations are mentioned as well. A list of particular data formats is presented (*.tmx, *.xliff, *.txt, *.doc, *.docx, *.odt, *.ppt), with a note that PDF format is not really useful. Data identifiers or meta-data sets are important and should belong to the data if possible.

He concludes by saying that most of the needed data exists and that it is mostly a question of collecting everything and that there is no real restriction on the quantity: the more data, the better.

3.7 Legal Framework for Contributing Data

Joris Deene (Everest Law), gave a clear and detailed presentation of the legal difficulties of providing and sharing data. The highlight of the presentation was authorship: who “owns” the content? The writer of the original texts? The translator? The company the translator or the author work for? Joris Deene clarified that the answers to these questions can vary greatly according to the context. The key to the answers is deciding whether or not a text is “original creation”, which can be very subjective.

With his presentation, he raised awareness on the importance of treating authorship and ownership with caution. He raised interesting questions and the audience was keen to know more, which will be remembered for future events. The presentation was further discussed during the following coffee-break.

ELRC Workshop Report for Germany

3.8 How can I Contribute: Practical and Technical Aspects

Khalid Choukri provided some practical information on how to go about sharing data. He explained that this can be done either through the website or via e-mail. He described the process for both options and was happy to answer some further questions by the audience.

3.9 Wrap-up, On-site Conclusions and Commitments

After a successful workshop, Véronique Hoste and Khalid Choukri took the time to thank all of the speakers, organisers and participants. They summed up some of the main subjects of the day, focussing on collaboration and concluded by inviting everyone to the reception.

4 Synthesis of Workshop Discussions

4.1 Panel 1 - Multilingual Services in Belgium: Functioning, Data and Linguistic Resources

Moderator: Stijn De Smeytere (Chancellery of the Prime Minister)

Panellists: Joachim Van den Bogaert (CrossLang), Albert Henn (SPF Intérieur, SCTA), Céline Dubois (ACV BIE), Hugo Zwaenepoel (Vlaamse Overheid)

This panel presentation provided the audience with a general idea of the linguistic practices of several organisations. Some time was allotted to the discussion on data management and the use of translation technology. It was interesting to hear the differences regarding these subjects between different organisations.

4.2 Panel 2 - Data and Language Resources in Belgium: where to find them?

Moderator: Véronique Hoste (Ghent University)

Panellists: Cedric Magnifico (federale pensioendienst), Jean-Francois Bauduin (kamer van volksvertegenwoordigers), Peter Lemmens (NMBS), Serge Lefevre (RIZIV (rijksinstituut voor ziekte en invaliditeitsverzekering))

After giving a short overview of the types of data that are needed and the potential sources of data, Véronique Hoste addressed her questions to the panel. How does the translation process work within your company? What is the volume and what are the languages involved? Who is the target audience? What are the linguistic resources formats and how are the LRs maintained? Hearing the different answers to these questions from the panellists provided us with a good overview of different practices and the impact this would have on the initiative.

5 Workshop Presentation Materials

All presentations are available online on the ELRC website:

http://www.lr-coordination.eu/belgium_agenda