European Language Resource Coordination
Connecting Europe Facility

Deliverable D3.2.10
Task 8

ELRC Workshop Report for Austria

Author(s):	Barbara Heinisch, Jürgen Kotzian
Dissemination Level:	Public
Version No.:	V1.0, Final
Date:	2018-10-18

Contents

1 Executive Summary

This document reports on the second ELRC Workshop in Austria, which took place on 25 September 2018 at the Austrian National Defence Academy in Vienna and was hosted by the Austrian Armed Forces Language Institute.

The objective of the event was to raise awareness of ELRC, eTranslation and the importance of language resource contributions to ELRC to adapt eTranslation to the needs of public administrations.
The event addressed topics such as Austrian digital public services and open data, multilingualism of digital infrastructures in Austria, the CEF eTranslation platform, ELRC (in Austria), legal issues when sharing language resources and preparing and sharing data with ELRC. Finally, a questions and answers session and a final statement by the Public Services NAP concluded the second ELRC Workshop in Austria.

The results of Austria's engagement with and participation in the ELRC network were made apparent by highlighting (a) the language resources from Austrian bodies already contributed to ELRC and eTranslation and (b) the fostering of the ARG GUT network (established at the Austrian Armed Forces Language Institute in 2012) consisting of translators and terminologists in the public sector, with more than 20 members. This network consists of potential language resource providers from different bodies and covering different domains.

One of the major issues raised by the participants is the treatment of Austrian German as a variety of German, which is in many ways different from the German language used in Germany. This becomes apparent, for example, in administrative terminology. Although, eTranslation or any machine translation system has not yet addressed this issue, all necessary measures have been adopted on the ELRC-SHARE repository side, to ensure the use of appropriate metadata descriptions of language resources from Austrian bodies.

Additionally, the participants requested that eTranslation becomes publicly available, instead of limiting access to public sector bodies, and that its performance is enhanced in a wider range of text types and domains. Moreover, translators requested advanced features such as providing alternative translations or providing the source of the machine translation output.

Sharing data and the processes associated with it was highlighted as a challenge. The main problems are related to copyright issues, confidentiality and privacy and the uncertainty regarding the right to re-use and re-purpose the translations, including approval for sharing by the hierarchy.

Currently, there does not seem to be a need for multilingual digital public services, except for those areas that require international data exchange or interfaces with other systems. The workshop demonstrated that many participants either do not know eTranslation or have not used it before.

The workshop was attended primarily by representatives from Austrian public bodies and academia.

Further information on the event as well as the presentations can be found at: http://www.lr-coordination.eu/l2austria.

2 Workshop Agenda

	Agenda **ELRC workshop in Austria** **25 September 2018** Landesverteidigungsakademie Vienna Stiftgasse 2a 1070 Vienna	**Programm Österreichischer** **ELRC-Workshop** **25. September 2018 in Wien** Landesverteidigungsakademie Wien Stiftgasse 2a 1070 Wien
09:00 – 09:30	**Registration**	**Registrierung**
09:30 – 09:40	**Welcome and introduction** *Bruno Nestler (Head of Language Institute of the Austrian Armed Forces)*	**Begrüßung und Einführung** *Dr. Bruno Nestler (Leiter Landesverteidigungsakademie/Sprachinstitut des Bundesheeres)*
09:40 – 09:45	**Welcome by the EC** *Achim Braun (Directorate-General for Translation, European Commission)*	**Grußworte der Europäischen Kommission** *Achim Braun (Generaldirektion Übersetzung, Europäische Kommission)*
	Session 1. Connecting a multilingual Europe: European context and local needs	**Teil 1. Das mehrsprachige Europa verbinden: Europäischer Kontext und nationale Bedürfnisse**
09:45 – 10:10	**Connecting public services across Europe: ambition and results so far** *Aleksandra Wesołowska (Project Officer, Directorate-General for Communications Networks, Content and Technology, European Commission) - video presentation*	**Öffentliche Dienste in ganz Europa miteinander verbinden: Zielsetzung und erste Ergebnisse** *Aleksandra Wesołowska (Projektverantwortliche, Generaldirektion Kommunikationsnetze, Inhalte und Technologien, European Commission) - Videobotschaft*
10:10 – 10:30	**National initiatives for digital public services and (open) data** *Reinhard Posch (CIO AUT, Graz University of Technology)*	**Initiativen zur Digitalisierung öffentlicher Verwaltungsservices und Open Data in Österreich** *Prof. Reinhard Posch (CIO AUT, TU Graz)*
10:30 – 11:00	**CEF in Austria: an outlook into current and future challenges – Panel session**	**Diskussionsrunde: Mehrsprachigkeit von digitaler Infrastruktur im öffentlichen Dienst in**

	Moderator: Jürgen Kotzian (PS NAP, Language Institute of the Austrian Armed Forces) *Panellists: Elsa Lischka (Federal Computing Centre), Robert Scharinger (Federal Ministry of Labour, Social Affairs, Health and Consumer Protection)*	**Österreich: Herausforderungen der Gegenwart und Zukunft** *Moderation: Jürgen Kotzian (PS NAP, Sprachinstitut des Bundesheeres)* *Panel: Elsa Lischka (Bundesrechenzentrum)* *Robert Scharinger (Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz, eHealth)*
11:00 – 11:30	***Coffee break***	***Kaffeepause***
11:30 – 11:50	**The CEF eTranslation platform @ work** *Achim Braun (Directorate-General for Translation, European Commission)*	**CEF eTranslation-Plattform @ work** *Achim Braun (Generaldirektion Übersetzung, Europäische Kommission)*
11:50 – 12:10	**The European Language Resource Coordination (ELRC) action** *Stelios Piperidis (ILSP/"Athena" R.C., ELRC)*	**Die Initiative European Language Resource Coordination (ELRC)** *Stelios Piperidis (ILSP/"Athena" R.C., ELRC) – auf Englisch*
12:10 – 12:30	**ELRC in Austria: Overcoming language barriers – how and why?** *Barbara Heinisch (Centre for Translation Studies, University of Vienna)*	**ELRC in Österreich: Sprachbarrieren überwinden – wie und warum?** *Barbara Heinisch (Zentrum für Translationswissenschaft, Universität Wien)*
12.30 – 13:30	***Lunch break***	***Mittagessen***
	Session 2. Engage: hands-on data	Teil 2. Interaktiver Teil: Daten in der Praxis
13:30 – 14:00	**Sharing language resources: Legal issues** *Christian Galinski (Infoterm)*	**Sprachressourcen teilen: Rechtliche Fragen** *Dr. Christian Galinski (Infoterm)*
14:00 – 14:20	**Preparing and sharing data with the ELRC repository – and what happens next** *Maria Giagkou (ILSP/"Athena" R.C., ELRC)*	**Datenaufbereitung und -bereitstellung mit Hilfe des ELRC-Repository – sowie nächste Schritte** *Maria Giagkou (ILSP/"Athena" R.C., ELRC) – auf Englisch*

14:20 – 14:50	**Identifying and managing your data: Questions & Answers** *Moderator: Barbara Heinisch (Centre for Translation Studies, University of Vienna)*	**Diskussionsrunde: Ausfindigmachen und Management von Sprachdaten: Fragen und Antworten** *Moderation: Barbara Heinisch (Zentrum für Translationswissenschaft, Universität Wien)*
14:50 – 15:00	**Conclusions** *Jürgen Kotzian (PS NAP, Language Institute of the Austrian Armed Forces)*	**Schlussfolgerungen** *Jürgen Kotzian (PS NAP, Sprachinstitut des Bundesheeres)*
15:00 – 15:30	***Coffee Break and networking***	***Kaffeepause und Networking***

3 Summary of Content of Sessions

3.1 Welcome and introduction

Bruno Nestler, head of the Language Institute of the Austrian Armed Forces and former National Anchor Point, welcomed the participants at the Austrian National Defence Academy. It was his personal wish to have Austria's second ELRC workshop in the Sala Terrena of the National Defence Academy and he was grateful that a lot of people followed this invitation.

The Language Institute (SIB) is Austria's largest public language service provider. Therefore, it is justified that the public National Anchor Point (NAP) is a SIB member. Jürgen Kotzian is the current National Anchor Point Public Service, a renowned position accompanied by a plethora of tasks.

Dr. Nestler emphasised the cooperation with the Centre for Translation Studies at the University of Vienna, which is the local organiser, whereas the SIB provided the framework for this event. Bruno Nestler expressed his gratitude to Prof. Gerhard Budin, Technology NAP and his staff member Barbara Heinisch.

Bruno Nestler welcomed the ELRC representatives, Stelios Piperidis and Maria Giagkou as well as Achim Braun from the DGT Local Field Office in Vienna.

According to Dr. Nestler, ELRC and eTranslation are a matter of importance. The aim is not to replace translators with machines or to generate fluent, but wrong translations. High-quality human translation should be the benchmark. However, machines can do a lot of things and they can do it faster and faster. Therefore, they are an indispensable tool for future translators. Whether we use them or not, we should help design the future of translation proactively. Moreover, the Austrian variety of German should be positioned in Europe and should benefit from eTranslation.

Bruno Nestler went on to emphasise the commitment of SIB to ELRC and to supporting the vision of connecting a multilingual Europe, upon request of the Federal Chancellery. This commitment is in line with the leading role of SIB as a public language services provider in Austria that has already produced and made available a number of glossaries and specialised dictionaries.

The speaker concluded by expressing his gratitude to all people participating in the organisation of the event and inviting the audience to the European Day of Languages taking place in the same venue the following day.

3.2 Welcome by the EC

Achim Braun, representative of the DGT in Austria, welcomed all workshop participants. He always experienced the importance of language since he comes from Germany and does not speak the Austrian variety of German. He stressed that language is part of a person's identity and that the European Commission's machine translation system, i.e. eTranslation, also supports him in his professional life as a translator.

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

Due to technical problems, the corresponding pre-recorded video from Aleksandra Wesołowska (DG CONNECT) was not presented. However, the video will be made available via the workshop website.

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

Reinhard Posch, Chief Information Officer of Austria and professor at the Graz University of Technology, gave a presentation on "Digital public services and Open Data". He presented current trends in the populations' usage of digital services, which the providers have to address when developing new services.

He started with current digital practices that are commonly accepted, such as the concept of "digital by default", which means that more and more transactions are carried out online, and the providing of data according to the data once-only principle", that citizens and businesses provide diverse data only once in contact with public administrations, while public administration bodies take actions to internally share and reuse these data – even across borders. At the same time, data protection and security will be an important issue of the future. He also pointed out that more and more people are using their mobile phones instead of their computers, and therefore, it is crucial for digital administration to offer services that are up to date.

Prof. Posch discussed the difference between using a mobile phone and using a computer. For example, mobile phones are "always on" – people can always be contacted, which creates a new way to communicate. In addition, we identify more strongly with our mobile phones than we do with our computers – while the same computer can be used by several people, a mobile phone usually only belongs to one person, which also means that we do not need to "log in" to our phones.

On 29 September 2018, the EU regulation eIDAS went into force, which provides new, improved safety aspects in electronic traffic, and Prof. Posch pointed out that especially in the administrative sector, this kind of safety is indispensable. eIDAS also recognised the need for addressing mobile structures.

Another aspect that is currently changing in the digital world is that we are moving from session communication to single transactions. On the computer, we work in sessions, since we can turn it on and off, and, for example, the information we deliver in one singular email message is usually more complete. However, if we work on mobile phones, we often send shorter messages, and the communication can easily be interrupted and then resumed later.

Prof. Posch also mentioned the concept of remote eID. In the past, we used to have chip cards where our security-relevant information was saved. Today, we cannot put chip cards into our mobile phones anymore. The concept of remote eID enables us to deposit the safety element at the provider, which is safer, since a small chip card can easily be lost.

He ended his talk by mentioning a surprising fact: every second, 57,000 new devices are going online. Of course, this makes it even more important for developers to provide the necessary safety aspects for identification and data security.

After his presentation, Prof. Posch answered a number of questions on multilingualism and machine translation. He pointed out that machine translation already works quite well for domain-specific, technical texts (for example, dealing with law or medicine), but it has not yet reached the same performance level for literary texts. For texts that are ambiguous and allow for various interpretations – which often applies to literary texts – machine translation still does not produce satisfactory results. Prof. Posch argued that our progress in machine translation will also depend on our research advances in the field of artificial intelligence, as this forms the basis for various machine translation services.

At the same time, Prof. Posch has confidence in the future advances of machine translation. He considers domain specification very important to increase the quality of machine translation systems, since the same word order is often being interpreted differently in various domains. In chess, researchers believed for a long time that no computer will ever reach human skills, but then, computers were developed that could even defeat the best chess players. Prof. Posch believes that if researchers keep working on the improvement of machine translation, also machines will eventually be able to produce human-like translations.

3.5   CEF in Austria: an outlook into current and future challenges – Panel session

In the panel session, Jürgen Kotzian, Public Services NAP, had a discussion with Elsa Lischka from the Federal Computing Centre and Robert Scharinger from the Federal Ministry of Labour, Social Affairs, Health and Consumer Protection on multilingualism of public digital infrastructures in Austria.

The aim of the panel was to present the institutions and to discuss the interoperability and multilingualism of digital public services in Austria.
Elsa Lischka works as a Digital Transformation Consultant at Austria's Federal Computing Centre, which is responsible for IT solutions for all public services in Austria. Robert Scharinger is head of the department for Health Information Management and Health Informatics at the Federal Ministry of Labour, Social Affairs, Health and Consumer Protection. He represented eHealth and related services, such as the electronic health record (ELGA).

Jürgen Kotzian, who moderated the discussion, asked the following questions:

Who is the target group for your service?

Robert Scharinger mentioned doctors and patients as his target group. Elsa Lischka stated that her target group includes public administrators, ministries and citizens.

Which digital services do you use?

Elsa Lischka replied that the Federal Computing Centre is responsible for providing almost all IT services, among others: access management, electronic legal system, citizen service portals, mobile device management, open data portal, personnel management, online finance system (tax, etc.), licence management, procurement platform, cloud storage, file management, digital workplace, ELGA, e-invoice, e-participation, ELAK (electronic files), etc.
(https://www.brz.gv.at/en/what_we_do.html). She and her colleagues also test many new digital services. Their task is the implementation of a "Digital Austria".

Robert Scharinger explained the task of the electronic health record. It tries to connect all health professionals, and it also takes care of particularly sensitive data. These data must remain where they are and where they are supposed to be: in hospitals and doctors' offices. To secure the data, also the different devices in hospitals need to be connected to each other, and Scharinger stated that he also collaborates with the Federal Computing Centre to safely conduct all these transactions.

Which part does multilingualism play in your work?

Robert Scharinger stated that with the continuous development of science and medicine, structured documentation is becoming increasingly important. In hospitals and doctors' offices, this structured documentation is implemented via code-schemata, such as the ICD-code (which was translated into several languages) or SNOMED. He said that there are so many different terminologies that it would probably take a whole semester to learn them all.

Elsa Lischka replied that in her field of work, one important topic is that of artificial intelligence, which, for example, is employed for the development of chatbots that work with automated translation. Users have to speak clearly and intelligibly when using these chatbots, otherwise the technologies will not be able to translate well. For the further development of these tools, it is important to invest in contextualisation.

Do you know or use any of the CEF building blocks?

Robert Scharinger praised the work of the CEF initiative and mentioned several building blocks used, such as epSOS or STORK. He also talked about CEF eHealth Call, a project by the EU which aims to provide digital service infrastructures including semantic services. ELGA participates in this project by providing ePrescription, which offers translation services for medical prescriptions.

Elsa Lischka stated that there are several experts from the Federal Computing Centre who work on EU projects. She herself is involved in two EU projects, one of which works on providing a "Digital Wallet" for EU citizens, whereby passports or driving licences can be accessed digitally.

Does your service require interaction with (administrative) offices of other EU member states?

Elsa Lischka said that she often works with members of the German counterpart of the Federal Computing Centre, since their technological developments can all be applied in Austria. The Nordic countries are also important cooperation partners, since they play a leading role in technology.

Robert Scharinger stated that the basis for interaction is trust. Every state must provide a safe server infrastructure, and the servers have to trust each other. There must be clear rules for the exchange of data, and it is also important to be aware of national and cultural differences.

Do you notice any differences between the Austrian administrative language and the one used in Germany?

Elsa Lischka replied that in both Germany, Austria, and also German-speaking Switzerland, the difficulty of understanding administrative language is an important topic. Her sector is involved in the simplification of language, and the question of how administrative language can be translated into everyday language. Chatbots and artificial intelligence are important tools in this process.

Jürgen Kotzian then summarised the interesting point that basically, there are three situations where translation is required: intercultural communication between national languages, communication between language varieties (of German in Austria and Germany), and communication between citizens using every language and administrative staff using specific language from various domains.

How do you meet the challenges of multilingualism? Do you employ human translators, do you use automated translation, or did you find other solutions?

Robert Scharinger used a hypothetical example from the sensitive domain of health: A psychotherapeutic or psychiatric report can only be dictated if it contains no reference to the patient. German law does not permit a third party to translate the interaction between a doctor and a patient. If the doctor were to use automated speech recognition and automated translation to dictate and translate the report, he/she would have to speak clearly, otherwise the machine will not understand the speech. However, he stated that machine translation is continuously improving, so this is a hypotherical scenario that may come true in the near future.

Elsa Lischka stated that she and her team work with various sub-contractors who support them in the implementation of these new technologies for customers, but not all of these technologies are of good quality. However, she also added that she is surprised by pace at which automated translation is improving. But they still also work with human translators who review the automated translations.

Which other services that you do not use yet could further facilitate your life?

Elsa Lischka wished for more standardisation in the IT field. The topic of IT consolidation is very important.

Robert Scharinger once again summarised the main purpose of his service, which is to provide health-related information, thereby preventing medical negligence. What he would wish for is a better semantic interoperability framework.

Questions raised by the audience were:

- Multilingual digital services: How can non-German speakers use services in the health care sector? Answer: no integrated service provided at the moment.
- If the same technology is used throughout Europe, is there a monoculture which is more susceptible to hackers because there are similar interfaces and similar weaknesses? Answer Lischka: Security is a major issue, also addressed in the governmental programme.

Not everybody should re-invent the wheel and some services should be consolidated. An access portal should be at the BRZ where informatics specialists can be found.

- A court interpreter welcomes the unification of medical terminology since there are many medical reports and records to be translated. The international codes help both translators and patients coming from another country. Doctors can use these codes if no translator is available at the moment.

3.6   The CEF eTranslation platform @ work

Achim Braun, representative of the DGT in Austria, a translator himself, gave a brief introduction into the translation workflows at the EC offices in Vienna. He stressed that he uses eTranslation regularly, not only as a translator but also to understand texts in foreign languages. Exemplified by an automatic translation of a text about Pinocchio from Portuguese into German with the eTranslation system, Achim Braun demonstrated that the system delivers good quality despite problems with ambiguous words or proper names. He said that his background knowledge of various languages in combination with the machine translation output helps him understand the original text. Moreover, he explained the workflow when translating press releases from English into German for an Austrian audience. First, he receives the English source text from the European Commission. These texts usually contain a lot of references to the European Commission. Second, he uses eTranslation to pre-translate the text and have a German draft version. Third, he adapts the content and the title to the interests of an Austrian audience. Fourth, he sends his partly re-written press release to the communications department. They make further changes and have the final say before sending out the press release. Achim Braun showed that eTranslation helps him translate EC-related press releases. However, he never uses the raw machine output but always has his target audience in mind when he edits the text.

3.7   The European Language Resource Coordination (ELRC) action

Stelios Piperidis, member of the ELRC consortium, presented the ELRC initiative in a broader context of EC initiatives. The broader context of ELRC is the Connecting Europe Facility and its three pillars, i.e. transport, telecom and energy. eTranslation is part of Telecom and CEF Digital, Digital Service Infrastructures. He provided an overview of the heads and organisations behind ELRC, i.e. German Research Centre for Artificial Intelligence, ELDA, ILSP and Tilde. Stelios Piperidis explained the role of the ELRC Technological and Public Services NAPs. Gerhard Budin is Austria's Technological NAP and Jürgen Kotzian is Austria's Public Services NAP.

He moved on to show the objectives of ELRC, including the collection of language resources, identifying the needs of public services, engaging with the public sector in the identification of LRs, helping with technical or legal issues and acting as an observatory for LRs across Europe. ELRC was initiated to facilitate multilingual communication and the exchange of information in key public services. Here, he emphasised the importance of feeding eTranslation with domain-specific language resources so that it can work for commerce, open data, social security, consumer rights or eJustice. He mentioned that the European Parliament voted in favour of the resolution on "Language Equality in the Digital Age" on 11 September 2018. The resolution aims at facilitating multilingualism in a digital market and defining minimum language resources that European languages should possess to

prevent digital extinction and preserve languages in the digital world. Moreover, it emphasises the need for language technologies, including databases and translation technologies for all EU languages, to tackle linguistic challenges. ELRC plays a key role in this respect since translation and knowledge management problems of the EU should be solved within the EU. Therefore, language technology for Europe made in Europe is crucial. ELRC collects these language resources and makes them available to CEF.AT (Connecting Europe Facility – Automated Translation), especially in-domain data to provide good machine translations. However, every language resource, both in-domain data and out-of-domain data can make a difference. Therefore, Stelios Piperidis encouraged the participants to contribute their data to ELRC to train the eTranslation system.

Stelios Piperidis provided an overview of the ELRC network, consisting of the ELRC consortium, 60 NAPs and more than 400 identified data holders. So far, 430 LRs have been provided by ELRC, whereas some languages are better covered than others and the majority of the LRs are licenced under CC.

He specified the type of LRs ELRC is looking for, i.e. translation memories, parallel documents, large monolingual texts and terminologies. He presented the ELRC-SHARE repository which enables sharing and access to LRs. The ELRC Helpdesk supports data contributors with technical and legal issues. He concluded his talk with referring to the ELRC on-site assistance.

3.8 ELRC in Austria

Barbara Heinisch from the Centre for Translation Studies at the University of Vienna and representing the Technology NAP, presented the achievements of ELRC in Austria.

She presented the National Anchor Points of ELRC in Austria: Jürgen Kotzian as the Public Services NAP and Gerhard Budin from the Centre for Translation Studies of the University of Vienna in his capacity of Technological NAP. High-level representatives playing a crucial role for the ELRC initiative are Ferdinand Scheidbach from the Federal Ministry for Digital and Economic Affairs as well as Alfred Ruzicka from the Federal Ministry for Transport, Innovation and Technology, who is responsible for the CEF agenda in Austria. She also emphasised the importance of the informal working group ARG GUT (Arbeitsgruppe Gouvernementaler Uebersetzungs- und Terminologiedienste). This working group consisting of translators and terminologists from the public sector are not only creators of language data, i.e. translations at their institutions, but also a group of language experts who create new resources on their own, e.g. the Verwaltungsglossar (administration glossary).

She summarised the collection of language resources in Austria and the contributions to ELRC. Major contributors have been the Language Institute, the Centre for Translation Studies, the Ministry of the Interior as well as the City of Vienna. The contributed datasets cover a variety of different text types and formats, ranging from glossaries, documents concerning constitutional law, transcribed audio guides, parallel data for the Presidency as well as named entities, mainly in German and English. The LRs were provided by the Language Institute and the Centre for Translation Studies, the Ministry of the Interior as well as the City of Vienna. The formats include plain text, TMX, XML, Microsoft Word or Microsoft Excel.

Barbara Heinisch also presented related CEF-funded activities in Austria, i.e. the EU Council Presidency Translator and the CEF eTranslation TermBank project. The EU Council Presidency Translator is a machine translation engine based on eTranslation and a system geared towards the news domain in German and English. It is available at translate2018.eu. It is freely available and was trained with EU Council Presidency-related domains and news data in cooperation with the Centre for Translation Studies.

Another result of the cooperation between the SIB and the Centre for Translation Studies is the Austrian language resource portal (Sprachressourcenportal Österreichs) available at

sprachressourcen.at. It consists of the Verwaltungsglossar developed under the aegis of Jürgen Kotzian by ARG GUT containing German-English terminology in the field of public administration, an English language guide for the EU Council Presidency in Austria, the EU Council Presidency Translator and further resources.

Barbara Heinisch additionally touched upon the obstacles that need to be overcome in collecting language resources. These include internal workflows requiring the authorisation by superiors who may or may not understand the relevance of LRs for different purposes. Other obstacles are related to the lack of resources within an organisation to identify and process the data before sharing it with ELRC. Legal and licencing issues relate to copyright clearance, privacy and confidentiality, personal data, and furthermore, insecure legal situations make data owners reluctant to share their data. In addition, the translation workflows within organisations may not consider data management and data centralisation. In Austria, some ministries and public bodies have in-house translators who may have their own translation memories, but other public bodies outsource their translations without asking for translation memories from the language service providers and without centralising translation request management. Another issue is the lack of awareness for the importance of LRs and the re-use of LRs for different purposes, including MT.

Despite these obstacles, ELRC in Austria has already been successful in collecting a critical size of Austrian German language data, in raising awareness among ICT networks and networks related to the Austrian state, provinces, cities and municipalities at the governmental level, and in creating a network, i.e. ARG GUT.

What worked well? Barbara Heinisch explained that the cooperation within the ARG GUT working group is a good basis as well as the contact with public bodies with the help of Ferdinand Scheidbach. Currently, the LRs in ELRC-SHARE provided by Austria mainly cover the area of military and EU Council Presidency-related topics. Contributions from other areas are more than welcome.

The lessons learned are that the cooperation across organisations on a national level could be improved as well as the cooperation between language service providers, public service translators, information and communication technology departments, etc. Not many people are aware of eTranslation and therefore, more public services could use eTranslation and help to adapt it to their needs. Barbara Heinisch identified further potential improvements in the development of translation processes to identify and share LRs.

Finally, she pointed out why and how language barriers should be overcome. Barbara Heinisch listed the benefits for data contributors, including the improvement of machine translation and the adaptation of the eTranslation system to the peculiarities of the German language in Austria. Moreover, data contributors can use resources provided by others. She wanted to show that LRs are valuable and that every LR counts.

3.9   Can language data be shared and how?

In the session about legal aspects of data re-use, Christian Galinski from Infoterm started with the definition of language resources, i.e. microcontent and semantically structured content. When sharing these data between persons or platforms, e.g. translation memories, there are different legal issues to be considered. According to Christian Galinski, every content may be multilingual and multimedia. However, only structured content (microcontent) can be automatically accessed. Copyright protection depends on the types of content and should be considered when describing resources.

Legal aspects that are crucial include the GDPR and accessibility and the protection of intellectual property.

Areas affected by legal regulations are among others terminological databases and their content, locales/variants, non-verbal descriptions or defining representations such as pictures, icons or formulas.

Christian Galinski listed a number of issues to be considered when it comes to intellectual property rights and copyright, including trademarks, exploitation rights, competition law, patents or brand names. Translations are also protected and there are special regulations for computer programmes or database creators. Different examples show that it is quite easy to infringe copyright, e.g. by using pictures in terminological databases, citing diagrams in academic theses, or sharing pictures on social media. Software helps to find copyright infringements on the internet.

Obstacles preventing data exchange include differences in classification systems or metadata and different data models. However, there are standardisation processes in this respect, e.g. TBX or TMX. Copyright of text and non-linguistic representations as well as database protection are potential risks for data contributors and data receivers.

Christian Galinski explained that works that required creativity and/or effort, i.e. ideas, methods, systems or knowledge, are protected. The author has contributed his/her personality to the work. Works may refer to literature, computer programs, theatre, scientific works, pictures, films, video games, sound, monographies, etc. As an example, he used entries in a terminological database. Texts are protected by copyright. However, definitions are excluded. Moreover, films, photographs, brand names and formulas as well as computer programs may be protected by copyright.

Editors and translators create a new copyright, but they need the permission of the copyright holder of the source text to make translations.

In addition, there is the sui generis database protection in accordance with the Database Directive in the EU.

Finally, Christian Galinski addressed the copyright of linguistic works and the re-use of translations. If a translation memory is considered as a database, then the task of selecting and structuring the elements of the database should be treated as a personal intellectual creation that constitutes a significant investment. The final question raised by Christian Galinski is whether we infringe any rights if we store data in a translation memory. Generally, there is no express right that protects a copyrighted work from being added to a database. However, the right of reproduction or right to edit these works may be infringed.

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

Maria Giagkou, member of the ELRC consortium, defined basic concepts addressed by ELRC, ranging from data, language resources, metadata, etc. She pointed out that data is the oil of the 21st century. Data may be any piece of electronically stored content and language data may be any piece of electronically stored text. eTranslation uses a wide range of data, including the DGT translation memories, comprising translations produced by EU bodies and institutions. Although this is a huge amount of data, additional data are needed to improve eTranslation. These types of text include translations or terminology in machine-readable formats, such as bilingual or multilingual parallel texts in .tmx, .xliff, .xlsx format, terminologies in .tbx, .txt., .csv., .xlsx format or monolingual texts as plain text.
Although ELRC welcomes contributions of language data in any format, Maria Giagkou provided some hints and tips for managing translation data in order to ensure their reusability, for example storing the source and target texts in different files, using identical file names and language identifiers, etc.

Maria Giagkou pointed out that when compiling datasets into meaningful language resources to be contributed to ELRC, two criteria should be taken into account. These are language(s) and domain.

The preferred domains are the administrative or regulatory domain and topics relevant to the CEF DSIs, e.g. online dispute resolution, eProcurement, Safer Internet, Public Open Data, Europeana, e-Justice, Business Registers Interconnection System or Electronic Exchange of Social Security Information.

She moved on to present the ELRC-SHARE portal and a step-by-step guide to contribute data to CEF eTranslation. After the data contributor uploads LRs to ELRC-SHARE, ELRC processes these data, uploads them again to ELRC-SHARE, making the processed version also available to the data contributor. Depending on the format of the original dataset, processing may involve data extraction, anonymisation, cleaning, reformatting, data conversion, tag removal, alignment and adding metadata.

She concluded by presenting the ELRC on-site assistance service for data contributors which is free of charge and which might also be used to improve the data management processes within the public sector organisation.

3.11 Identifying and managing your data: Questions & Answers

Barbara Heinisch from the Centre for Translation Studies at the University of Vienna briefly demonstrated how to access the eTranslation web user interface. She highlighted that eTranslation is currently only available to public administrations and universities that are part of the European Masters in Translation Network (EMT).

During the discussion that followed a number of questions were raised by the audiences. The questions and the answers provided by the organisers were as follows:

- Q: What is the difference between eTranslation and the EU Council Presidency Translator?
  A: The EU Council Presidency Translator is openly accessible since users just need to type in the URL and can use the system. Moreover, it has a focus on the Council Presidency domain and news articles in German and English. It has been trained for this respect. eTranslation was developed for translations within the EU and can also be used by public administrations with a focus on EU texts. The EU Council Presidency Translator is based on eTranslation.
- Q: How can ELRC support us if we want to share language resources? Is there financial support for this?
  A: CEF has launched a series of calls for proposals that fund not only the collection of LRs on a national level, but also the re-use of the eTranslation system in public digital services. If you represent a public body in Austria that wants to re-use the eTranslation system for a public digital service, there is funding available.
- Q: Will the translation memories created by ELRC be fed into the central DGT translation memory? Or will the LRs collected by ELRC be separated from DGT translation memory?
- A: The resources collected by ELRC may eventually be merged with the Euramis database. However, some data may be held apart due to legal and technical reasons.
- Q: You talked about resources fed into the system. Will they be checked for quality? Will also bad translations be used as LRs?
  A: The correct translation may never exist. We collect those resources that are of high quality, developed by professional translators. eTranslation will be trained with a high

amount of data and if there is one file that is badly translated, then it does not affect the system.

- Q: Can eTranslation be used offline? Can machine translation systems be used offline?
  A: Neural machine translation systems cannot be used offline. In the 1990s proprietary systems could be used offline. Most of the neural MT systems are used online.

- Q: Can freelance translators share their translations with ELRC? What copyright issues do they have to keep in mind?
  A: If freelance translators share their data, the client's (i.e. the source text owner's) consent is necessary before submitting data to ELRC. It is important to clarify with the client (in written form) that the data may also be used for additional purposes, such as MT. ELRC prefers to work with CC licences, but you can also upload resources that are not publicly available. Since eTranslation is mainly targeting the public sector, the legal framework on which we try to rely is the PSI directive. According to this directive, most of the data generated by public authorities are free to use. In case your client is a ministry and the content does not contain sensitive or personal information, it may can be shareable under the PSI directive.

- Q: Can students use eTranslation without providing resources?
  A: As long as you are a member of the European Masters in Translation Network, lecturers or directorates of studies can apply for accounts to get access to eTranslation.

- Q: What do the EU Council Presidency Translator or eTranslation do when there is linguistic ambivalence, e.g. "province" or "federal state" for "Bundesland" in German? Does the law of the strongest apply, i.e. will the systems use the terminology of those who contribute the majority of the resources?
  A: That might be a good reason to convince your superior to contribute resources to the system. In addition, there are terminological resources and named entity lists that cater for linguistic ambivalence. Current technology is intelligent enough to store and represent all the different meanings of a word in the most appropriate way. If you have ambiguous words in a context, there are mechanisms built in the model to discern the different senses and choose appropriate lexical transfer. But you need to have examples of the different senses and different lexical transfers. Therefore, the ELRC searches for LRs and for domains relevant to your field of work. In this respect, CEF Automated Translation (AT) facilitates cross-border communication in the member states. eTranslation will be the core of CEF.AT.

- Q: Is it planned to extend the features of eTranslation with DeepL features, e.g. additional suggestions when clicking on a word?
  A: The CEF eTranslation (or CEF.AT) ecosystem will comprise a set of language technologies and multilingual tools, for example named entity recognisers. The core of CEF.AT is always eTranslation. DeepL has been trained on high-quality translations as a result of translation activity in the European Union in the Directorate-General for Translation. DeepL is related to Linguee. Content-wise, not technology-wise, the user experience might be similar. It is not clear yet, whether such functionalities will be developed by the MT team at the EC.

3.12 Conclusions

Jürgen Kotzian, Public Service NAP in Austria, seized the opportunity to thank all participants, organisers, speakers and interpreters for their time and efforts. He emphasised the importance of feeding eTranslation with data from Austria. Only this way, eTranslation can learn Austrian German and also provide good translations for texts and documents from Austria. eTranslation is a tool that can help translators in their professional life.

4 Synthesis of Workshop Discussions

One major issue raised by the participants during the discussion sessions but also in informal conversations was the demand that eTranslation becomes publicly available and provides better-quality translations for a wider range of texts. Moreover, translators would wish for advanced features such as providing alternative translations or providing the source of the MT output, i.e. whether it is based on the training of a text from Germany or from Austria. Both for the NAPs and the participants, the differentiation between Austrian German and German German is important. In addition, the integration of terminology in eTranslation was also raised as an issue.

Another challenge is the sharing of data. The main problems are related to copyright issues, confidentiality and privacy and the uncertainty regarding the right to re-use and re-purpose the translations.

Participants were also interested in funding opportunities for the processing and sharing of data with the EC.

Quality issues were also mentioned by the participants, i.e. people do not want to provide translations of bad quality to be judged by others. Moreover, the participants wondered about the internal data processing steps at ELRC, i.e. copyright clearance or controlling the quality of the translations provided.

During informal discussions participants working in the legal domain said that all barriers mentioned in the presentations apply in their field. A huge amount of translations from a variety of languages, e.g. different types of certificates, etc. would be available, but sharing these data is too risky.

Currently, there does not seem to be a need for multilingual digital public services in Austria, except for those areas that require international data exchange or interfaces with other systems.

4.1 ELRC and Open language Data in Austria

Austria's open data portal https://www.data.gv.at/ aims at being the central Austrian catalogue. It collects and makes accessible metadata of decentral data catalogues of Austria's administrative bodies. Data.gv.at is the single point of contact to the European Data Portal. The data sets do not only include open government information but also documents related to the PSI directive. Recently, metadata have also been translated into other languages.

Data.gv.at does not only contain numerical data but also various reports and applications. Data are put into different categories: work, population, education and research, finance, geography and planning, society, health, art and culture, agriculture and forestry, sport and leisure, environment, transport and technology, administration and politics and economics and tourism. The portal differentiates between data (numerical data), documents and applications. In the documents section about 1,900 so-called metadata records are available. Contributing organisations include the ministry responsible for social affairs and ministry responsible for transport and innovation, provinces, national parks, the Federal Chancellery and the City of Vienna. The predominant file format is PDF. The majority of the LRs are in German, but there are also some in other languages, e.g. English. The portal contains 1,858 documents, 20,350 data sets and 445 applications.

4.2 Success stories and lessons learnt

Achievements in Austria include:

- The establishment of the ARG GUT network consisting of translators and terminologists in the public sector, with more than 20 members. This working group is beneficial to ELRC in

two ways. First, this group may provide LRs to ELRC and second, they create LRs in their capacity as a working group aimed at standardising language use in Austria.

- The NAPs in Austria were able to establish Austrian German as a variety of German which is in many ways different from the German language used in Germany, e.g. when it comes to administrative terminology.

Recommendations and lessons learned:

- Recommendations for the subsequent ELRC workshops are, among others, that someone from a higher position (e.g. from the EC) should contact the CEF coordinator and ask for their active contribution. This may include co-organising the workshop with the national CEF coordinator, sending invitations or asking for LRs. Having a high-level official who may ask for LRs would make data collection easier and would also allow to create a sustainable data pipeline. Moreover, stressing the benefits of eTranslation would be important in other workshops.
- Moreover, one workshop per year is not enough to reach potential data contributors. There might be other channels to reach the target group more effectively.
- In Austria, public administrations are reluctant to share their data due to confidentiality/privacy issues, the PSI directive as well as internal processes, e.g. approval from the superior.
- The establishment of a centralised translation service in Austria would make it easier to consolidate the LRs in Austria. Currently, the decentralised situation leads to data silos which can hardly be accessed. Moreover, if organisations outsource their translations, they are usually not asking for TMs from the LSPs or re-purposing their translations.
- There are a lot of resources available in Austria but gaining access to these resources is a difficult task. It is unclear whether those participants who already knew about candidate relevant resources owned by their institutions or by affiliated bodies would eventually share their data.