European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2015/1091 programme.

European Language Resource Coordination

Connecting Europe Facility

Deliverable D3.2.17

Task 3

ELRC Workshop Report for Austria

Author(s): Dagmar Gromann, Elisa Schnell, Miloš Milohanović, Felix Funda

Dissemination Level: Public

Version No.: V1.0

Date: 2021-12-22

Contents

1. Executive Summary

This document reports on the third ELRC Workshop in Austria, which took place on 10 November 2021 at the National Defence Academy in Vienna and was hosted by the Austrian Armed Forces Language Institute and the Centre for Translation Studies (CTS), University of Vienna.

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. In accordance with the SMART 2019/1083 tender specifications, the third ELRC workshop shall effectively engage the relevant stakeholders regarding the Connecting Europe Facility — Automated Translation (CEF AT) goals to lower language barriers.

The event addressed numerous language resources related topics. Among others, those topics included Austrian digital public services and open data, multilingualism of digital infrastructures in Austria, the CEF eTranslation platform, and the presence of ELRC in Austria. Furthermore, the event addressed the current state and future developments of language technology developed in Austria as well as for Austrian German, such as a chatbot developed by the City of Vienna for Viennese German. An important part of the workshop were lightning talks, i.e., one minute presentations, by language technology developers that provided an introduction to the variety of language technologies in Austria and a teaser for the software demonstration session. After a question and answering session and a final statement by the Public Services NAP, the third ELRC Workshop in Austria was concluded with the aforementioned software demonstration session to showcase the AI landscape in Austria accompanied by networking and socialising.

This third Austrian ELRC workshop built on the foundation laid by previous workshops. The results of Austria’s engagement with and participation in the ELRC network were made apparent by highlighting the language resources from Austrian bodies already contributed to ELRC and eTranslation. This network consists of potential language resource providers from different organisations and covering different domains. The number of attendees from different companies and areas of expertise are proof of the success of previous ELRC workshops and the endeavours that followed.

One of the major questions raised once again by the participants is how exactly should Austrian German as a specific variety of German be treated, independent of the German language used in Germany. The differences between these two varieties of the pluricentric German language are especially apparent in areas of technical language and jargon, for example, in administrative terminology.

Another key challenge was the topic of sharing data and the processes associated with it. As mentioned during the previous workshop, the main problems are related to copyright issues, confidentiality and privacy and the uncertainty regarding the right to reuse and repurpose the translations, including approval for sharing by the hierarchy.

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: https://lr-coordination.eu/austria3rd.

2. Workshop Agenda

08:00 – 09:00	***Registration***
09:00 – 09:05	**Introduction** Elisa Schnell, BA MA (NAP) and Ass.-Prof. Dr. Dagmar Gromann (NAP)
09:05 – 09:10	**Welcome** Brigadier General Reinhard SCHÖBERL, Deputy Commandant, National Defence Academy
09:10 – 09:35	**Language Technology and Artificial Intelligence:** **Latest trends and perspectives** Prof. Benjamin Roth, Faculty of Computer Science, University of Vienna
09:35 – 10:20	**Language Technology in and for Austria from the industry** **perspective (panel discussion)** Klaus Fleischmann, MA, CEO of Kaleidoscope Dr. Gerhard Backfried, CSO of Hensoldt Analytics Dr. Andreas Rath, CEO of ONDEWO Francisco Webber, CEO of Cortical.io Moderation: Ass.-Prof. Dr. Dagmar Gromann, University of Vienna
10:20 – 10:50	***Coffee break***
10:50 – 11:15	**Advances of the CEF AT platform** Andreas Eisele, Project Manager Machine Translation, Directorate-General for Translation, European Commission
11:15 – 12:00	**Language technologies in and for Austria by/for the public sector** **(panel discussion)** Ass. jur. Matthias Lichtenthaler, Federal Computing Centre Col Ing. Mag. Klaus Mak, Central Documentation and Information/ National Defence Academy Sindre Wimberger, City of Vienna Mag. Elisabeth Taucher, Criminal Intelligence Service Austria Mag. Claudia Lisa, Criminal Intelligence Service Austria Moderation: Felix Funda, BA, Austrian Armed Forces Language Institute/National Defence Academy

12:00 – 12:30	**Language Technologies: requirements and offerings (Q&A session)** Moderation: Ass.-Prof. Dr. Dagmar Gromann
12:30 – 13:30	***Lunch break***
13:30 – 13:45	**Value of data for the development of top quality Language Technologies** Mag. Hannes Pirker, Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences
13:45 - 14:30	**Language data creation, management and sharing: existing practises and challenges (panel discussion)** Ing. Brigitte Lutz, MSc, Open Data Portal, City of Vienna Prof. Barbara Schuppler, Graz University of Technology Mag. Hannes Pirker, Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences Mag. Jürgen Kotzian, Austrian Armed Forces Language Institute/National Defence Academy Moderation: Elisa Schnell, BA MA, Austrian Armed Forces Language Institute/National Defence Academy
14:30 – 14:35	**Lightning Talks for Language Technology demonstrations**
14:35 – 14:40	**Concluding remarks** Elisa Schnell, BA MA (NAP) and Ass.-Prof. Dr. Dagmar Gromann (NAP)
14:40 – 15:30	**Demo session of Language Technologies and socialising**

3. Summary of Content of Sessions

3.1. Welcome and introduction

Dagmar Gromann, Assistant Professor at the University of Vienna and current Technology National Anchor Point (NAP) welcomed the participants at the Austrian National Defence Academy. Gromann expressed her gratitude for the respectable number of attendees in spite of the difficulties caused by the ongoing COVID-19 pandemic.

The introduction was continued by Elisa Schnell, the current Public Services NAP and interpreter, translator, language teacher at the Austrian Armed Forces Language Institute. She was especially grateful for the attendance of representatives of the Austrian Armed Forces, who were invited as VIPs.

Schnell delivered a short introduction of ELRC for those unfamiliar with its work, as well as the programme of the third Austrian ELRC workshop. Furthermore, she emphasised the continued cooperation with the Centre for Translation Studies (CTS) of the University of Vienna. The Austrian Armed Forces Language Institute not only provided the location for the event, but its team, including Felix Funda and Stefan Bienert, jointly organised the entire event with the CTS team, including Barbara Heinisch and Miloš Milohanović, to whom Schnell expressed the NAPs gratitude.

Schnell continued by thanking the numerous panelists for their attendance. She went on to emphasise the commitment of the Austrian Armed Forces Language Institute to ELRC and to supporting the vision of connecting a multilingual and multicultural Europe. This commitment is in line with the leading role of the Austrian Armed Forces Language Institute as a public language service provider that has already produced and made available a number of glossaries and specialised dictionaries.

A brief welcome address was delivered by Brigadier General Reinhard Schöberl, Deputy Commandant, National Defence Academy, who thanked Elisa Schnell and Dagmar Gromann for their work in organising the event. Schöberl emphasised the importance of the Austrian Armed Forces Language Institute due to their many years of experience in working with multiple languages, as well as their role as pioneers of digitalisation in Austria. As such, the importance and potentials of AI technology and their integration in language technologies is an area of great interest for the Austrian Armed Forces Language Institute. The speaker concluded by expressing his gratitude to the people attending online via video conference in Zoom, which was simultaneously translated into English by Stefanie Göstl and Yasmina Müller, as well as to all people participating on site.

3.2. The potential of Language Technology and AI – where we are, where we should be heading

Benjamin Roth, Professor at the Faculty of Computer Science, University of Vienna, in the area of deep learning and statistical Natural Language Processing (NLP), provided a first keynote on Language Technology and Artificial Intelligence: Trends and Perspectives, the slides of which are available online. The presentation started with an introduction to machine learning, and Roth emphasised the importance of training with adequate and high quality data. Roth then continued to explain the basics of deep learning, emphasising that neural systems usually use raw input data instead of defined feature representations as is the case in traditional machine learning. He also noted that an important goal is for a system to be able to attain progressively higher levels of abstractions.

Roth continued with a basic explanation of current statistical language models as well as neural language models, such as BERT, GPT-3, and GPTJ. As a use case, he demonstrated how GPT-3 can be utilised to generate human-like text based on a few seed sentences. Current and future applications of neural language models include Neural Machine Translation, Question Answering, and Image Recognition.

Roth went on to present current challenges and limitations of deep learning, the main ones being the need for long training procedures, adaptation to different domains of pretrained models, the leveraging and integration of human experience and knowledge, and a lack of explainability. A debated question is whether statistical models are an appropriate alternative to and representation of (human) intelligence. Experts of deep learning will need to find answers to relevant questions in regard to expert knowledge, the optimal procedure for annotating data and leveraging annotated data, trust in Artificial Intelligence (AI), as well as ensuring transparency, accountability, and fairness. The arguably most difficult challenge is to understand and properly define what intelligence is.

3.3. Language Technology in and for Austria from the industry perspective (Panel session)

Gromann was the moderator for the first panel discussion that focused on language technologies in and for Austria from an industry perspective. Four highly reputable industry representatives participated in this first panel.

Klaus Fleischmann is founder and CEO of Kaleidoscope GmbH as a system and consulting company for global content and translation technology as well as terminology software. He is additionally Managing Director of eurocom Translation Services GmbH. He studied conference interpretation in Vienna and Monterey, USA, and technical communication at the Donau-Universität Krems. Furthermore, he is a member of the Executive Committee and Board of the International Network for Terminology (TermNet).

Francisco Webber is co-founder and CEO of the Austrian AI company, Cortical.io, that specialises on Natural Language Understanding and provides services for the automated processing and analysis of large text data. In 2005, Webber founded Matrixware Information Services, the first company to develop a global standardised patent database. In 2007, he founded the Information Retrieval Facility, a non-profit research institute with the mission to bridge the gap between academia and industry in the field of information retrieval. He founded Cortical.io in 2011 with the goal to apply the principles of cerebral processing to machine learning and text processing, as formulated in his "Semantic Folding Theory” white paper. Based on this novel approach, Cortical.io has developed a unique natural language understanding technology that solves many challenges related to big text data. Cortical.io solutions are implemented at several Fortune 500 companies, in various contexts, including semantic search and contract analytics.

Andreas S. Rath is an AI expert and creative thinker while at the same time a sharp problem solver and execution specialist. He worked for five years in Artificial Intelligence research and six years in digitising international companies around the world as a top management consultant for McKinsey & Company. Now as CEO and founder of ONDEWO, an Austrian high-tech company based in Vienna, he applies his knowledge and experience to enable machines to engage in natural conversations with humans. With his international team he builds the revolutionary ONDEWO Call Center AI platform including Speech2Text, Natural Language Processing & Understanding and Text2Speech. The on-premise ONDEWO Call Center AI platform supports agents with natural language AI services and automates inbound and outbound phone calls with software-based AI agents.

Gerhard Backfried is one of the co-founders and currently holds the position of Chief Scientific Officer (CSO) at HENSOLDT Analytics. His technical expertise includes acoustic and language modelling as well as speech recognition algorithms. More recently he has been focussing on the combination of traditional and social media, particularly in the context of multilingual and multimedia disaster-communication. He holds a master's degree in computer science (M.Sc.) from the Technical University of Vienna with specialty in Artificial Intelligence and Linguistics and a Ph.D. degree in Computer Science from the University of Vienna. He holds a number of patents, has authored several papers and book chapters, regularly participates in conference program committees and has been contributing to national and international research projects, such as KIRAS/QuOIMA, FP7/M-ECO, FP7/SIIP, H2020/ELG or H2020/MIRROR.

The first topic of discussion was a **general overview of how language technologies are used in Austria**. The participants stated that right now, both Austria and the global community are undergoing a linguistic and technological revolution, both mainly driven by Big Data. They emphasised that Austrian Standard German is threatened by "digital extinction" due to Austria being a comparatively small market and Austrian Standard German rarely being a consideration. In their view, innovation is driven by the free market and the highest chance of preventing the "digital extinction" of Austrian Standard German is to find ways to fine-tune existing systems in order to adapt them for the needs of the Austrian market.

The second topic for the panel discussion regarded the **current state of language technologies** in relation to Austria and Austrian German. According to the panellists, the current situation in Austria is promising, however, it is heavily dependent on market conditions. Internationally, there exists an overwhelming presence of the US market, where machine learning is performed to a much higher degree and with far fewer language combinations than in Europe. To improve this situation, more expertise in language technology and linguistic data science and more higher-education programmes would be needed. A future goal is to move on from probability-based models, and instead offer technologies which are more appropriate for realistic applications, and which could work with much smaller data sets.

On a more positive note, it was highlighted that Austrian companies offer a very high degree of expertise and competence, which allows them to stay competitive. It was stated that Austrian companies are exemplary in certain niches and that they are perceived positively outside of Austria, in contrast to a more negative perception within Austria.

Other highlighted needs were the need to educate customers and increase the general technological literacy, as well as the need to offer simpler, user-oriented solutions instead of aiming for minimal improvements of existing technologies.

The last topic of discussion for the first panel session was the **current application scenarios and use cases of language technologies for the general public and administrative bodies** in Austria. These include a currently in-development customer-oriented internal search system that can answer very specific questions that users might ask, an in-development speech recognition and speech synthesis technologies with a special focus on creating trust between users and service providers, and software solutions designed for the Austrian Armed Forces. Aside from these, there are numerous technologies still in use, ranging from very traditional ones up to state-of-the-art solutions.

3.4. The CEF-AT platform

The CEF-AT platform was introduced to the attendees by Andreas Eisele, Project Manager for Machine Translation at the Directorate-General for Translation at the European Commission. The slides of this presentation are available online.

CEF-AT offers machine translation and terminology solutions for multiple groups, including public servants, translators and interpreters in the EU, administrative and public employees, SMEs, and freelance translators. As the main area of application, he mentioned raw translation of texts for fast editing of EU documents in different languages.

An integral part of the CEF-AT project is eTranslation. eTranslation was supposed to offer improved machine translation for EU languages that are not well suited for statistical machine translation systems. Therefore, it was developed based on neural machine translation technology. Furthermore, eTranslation needed to be adapted to new domains. Among other ways in which this adaptation took place, one way was to integrate eTranslation into public online services.

Eisele sees the quality of the eTranslation system output as good, but not flawless, especially in the domains with formulaic texts, and stresses that the texts still need post-editing. Planned derivative technologies include named entity recognition, Computer Assisted Translation (CAT) tools, speech technology, media data, and transcription. Currently, the eTranslation API is only available for authorised users.

In relation to the first panel session, it was noted that eTranslation is suitable for both big companies as well as for SMEs, but it was also stated that it currently does not differentiate between Austrian and German standards varieties.

3.5. Language technologies by/for the public sector (Panel session)

Felix Funda from the Austrian Armed Forces Language Institute moderated this panel. Funda explained that the Austrian Armed Forces frequently uses MT and CAT-Tools, and also implements other language technologies. The second panel was made up of five participants representing the public sector.

Matthias Lichtenthaler is the Head of Digital Government & Innovation of the Federal Computing Centre. He coordinates the measures for implementing digitization and innovation for public administration and, together with his team, develops various initiatives of the administration for a digital Austria, including joint projects with the private sector. Lichtenthaler's professional focus is on cognitive computing, content analytics, process automation, and currently also on the structural and legal applicability of artificial intelligence and blockchain technology in public administration and government-related private business.

Klaus Mak is a career officer working for the Central Documentation Office of the National Defence Academy. He carries out teaching and lecturing activities as well as consulting and evaluation projects at a wide variety of domestic and foreign educational institutions for information profession and knowledge management. Mak studied political science, communication science and journalism at the University of Vienna after graduating from the Technical College for Mechanical Engineering and the Military Academy.

Sindre Wimberger has been developing digital products and services for almost 20 years. He currently works as a service designer for the city of Vienna. As the "botfather" of the WienBot, he ensures that it is constantly learning and becoming smarter. With the WienBot, he makes information more accessible to everyone with the help of artificial intelligence, natural language and direct answers.

Elisabeth Taucher is a translator and terminologist for the Interpreting and Translation Service of the Criminal Intelligence Service, Federal Ministry of the Interior. She is responsible for the TM tool and terminology management system, as well as for terminology work in German, English, French, and Spanish.

Claudia Lisa is Deputy Head of Unit of the Interpreting and Translation Service at the Criminal Intelligence Service, Federal Ministry of the Interior. She is a translator and interpreter, responsible for the interpreting processes from the order acceptance and selection of the appropriate interpreting mode to process handling and post-processing.

The first question addressed the topic of **requirements and demands for language technologies in the public sector** in Austria. The participants have listed a high number of different needs of the public sector, including the demand for voice recognition and voice input technologies, faster and more efficient written and verbal communication (including translation and interpreting), as well as a need to filter out misinformation and false data. Regardless of the specific type of work faced by different parts of the public sector, there is an increase in the dependence on language technologies and all of them need to be developed further.

The second question concerned **obstacles to a wider adoption of language technologies** in the public sector. The listed obstacles include a sub optimally functioning bureaucracy, lack of trust and accountability, inconsistent translation quality, and the lack of awareness for the need of high-quality translations. Furthermore, there exists a need for more user-friendly, specialised systems, as well as a need to increase general technological and media literacy.

Finally, there is a need for better funding. Due to insufficient funding, there is currently a lack of positions for experts that would oversee and organise further development and implementation of new technologies, and it is impossible to finance state-of-the-art technologies.

Funda summarised some of the most important points of the discussion and agreed that the public sector often develops slower than the private sector, especially due to concerns like accountability and security. This summary concluded the panel discussion and lead to a Question-Answering session with all panellists, first the current panel and then the first industry panel.

3.6. LT requirements and offerings: do they converge? (Q/A session)

The first question in the Q and A session was aimed at the **second panel participants in the public sector** and regarded the **use of Google** in the process of information and knowledge procurement. Mak explained that Google is useful, but not a perfect tool. He elaborated by explaining that in previous studies, people were able to find information equally fast or even faster without Google than with Google, even when not relying on tools such as the deep web. Solutions that do not rely on Google are, therefore, a viable option.

The second question regarded the perceived need for **AI-based products and how marketable** they are. Lisa voiced her opinion that there is a missing awareness for the potential and usefulness of AI-based technology in the realm of language technology. This is intensified by the sub-optimal communication between different departments of the public sector. Lichtenthaler added that the implementation of AI-based products is hindered by a lack of appropriate staff training, by the lack of awareness about the cultural and organisational impact of such an implementation, and by the presence of people who are not qualified enough to perform certain jobs.

A follow-up question regarded **potential solutions to these problems listed**. Wimberger explained that the fact that Vienna and even Austria as a whole are a melting pot of different cultures creates a lot of potential and that this should be used as a starting point. This could take the form of organising summits where representatives of the public sectors and representatives of big players, such as banks and large companies, could come together and find solutions together. He thinks that the third ELRC Workshop might exactly be an example of such an event.

The next questions were aimed at the **first panel discussion participants from industry**. The first of those was about how much people actually care about **trustworthy AI solutions**. Backfried explained that studies show that clients usually care more about the functionality of a product than about where their data is stored and other privacy concerns. Backfried went on to explain that trust is about more than just storing data; the trust between clients and providers is equally important. It is necessary to not promise too much and to keep expectations realistic. Another element of trust is the awareness that with very large data sets, nobody has a complete overview of what the sets contain and how algorithms will make use of data. Trust needs to be built, and this can be done by finding ways to validate data.

In summary, in terms of quality, trust, and security, the bigger challenges lie rather within machine learning than language technology. There exists a need for certified training sets; smaller, curated data sets might be preferable to large, uncontrolled data sets that could introduce noise. There is also a need for more protection mechanisms against bots, spoofing, AI created content, and similar phenomena that exist within what is viewed as a digital arms race.

The next question concerned the **simplification of language** and how the usefulness of AI could be measured. The panellists explained that due to the keyword-based way of searching for information during the times where typing was most common, a simplification of language could be noticed. There is, however, a slight reversal of this trend, and this reversal can be traced back to the increased use of sentence and voice input to search engines.

3.7. The value of data for the development of top-quality LT

Hannes Pirker from the Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences started his presentation, the slides of which are publicly available, on the value of data for the development of LTs with the commonly used phrase that data is the lifeblood of economic development. In his view, the same is true for the development of LTs.

Pirker explained that the first step toward developing proper data sets is to define what kind of data is needed. As an example, he mentioned that facial recognition can be performed in two different ways. Firstly, systems for recognising faces among other objects are trained by using data sets consisting of images with and without faces. Secondly, systems for identification of people are usually trained by using very large data sets of faces from different angles, but do not use images without faces.

Problems that arise are often due to problematic data sets. One example within machine translation systems are data sets with improper translations, but also data sets where the source text is of a lower quality, despite the translation being adequate. An example from image recognition technology is racial bias where systems trained only with, for example, faces of people of European origin, will perform sub-optimally when applied to recognising faces of people of Asian or African origin.

Pirker continued with a brief explanation of how machine translation systems and data sets are connected. He explained that while some systems are meant to be universal or cross-domain, others are designed to be domain-specific or genre-specific. This is why, for example, EuroParl would not be ideal for translating online chats. He continued to describe potential problems regarding data sets. As an example, he described the possibility of a system performing inadequately despite being trained with a good data set; one possible explanation could be that the data was good but not sufficient in quantity. In other words, both the quality and the quantity of data are important.

Pirker then moved on to the broader field of language technology. He explained that data is collected from different sources and that gold standard data is usually used for evaluation only - not for training. He suggested that small amounts of gold standard data could potentially be used during training to improve results. In regard to Austrian Standard German, he explained that data is often insufficient and that the notion that the differences between Austrian Standard German and Standard German is limited to the famous EU Protocol No. 10[1] is simply false; the possibly biggest difference in the terminology of the two standards lies within the administrative and legal terminology. Pirker concluded his presentation with a short summary and explained that the challenges are bigger than most think and that more work needs to be put into solving the described problems.

3.8. Language data creation, management and sharing: existing practises and challenges (Panel session)

The third panel session was moderated by Elisa Schnell. This session was made up of 4 participants.

Brigitte Lutz works as a Data Governance Coordinator in the Municipal Department of the City of Vienna. She is an IT expert for various areas, project manager, senior process manager (SPcM) and eGovernment expert. Other areas of responsibility include IT strategy, blockchain and eGovernment building blocks and services. Lutz is a founding member and currently spokesperson of Cooperation Open Government Data Austria and responsible for related national and international co-operations.

Barbara Schuppler is Assistant Professor at the Signal Processing and Speech Communication Laboratory of the Graz University of Technology. She is a speech scientist with an interdisciplinary educational background and research focus. The central topic of her doctorate thesis was the analysis of conditions for acoustic reduction in large conversational speech corpora using ASR technology. Schuppler has been leading three FWF projects with the focus of building cross-layer models for conversational speech, which also included creating the first large-scale Austrian German conversational speech corpus (GRASS).

Hannes Pirker studied general linguistics at the University of Klagenfurt and computational linguistics at the University of Saarbrücken. He worked for many years at the Austrian Research Institute for AI in the research area of language technology. At the Austrian Centre for Digital Humanities and Cultural Heritage, he is mainly involved in making text corpora, such as the Austrian Media Corpus (AMC), available. He acts as both developer and consultant for applications in the fields of linguistics and digital humanities.

Jürgen Kotzian is Head of Unit at the Language Institute of the Austrian Armed Forces at the National Defence Academy. He served as ELRC Public Services National Anchor Point between 2017 and 2020

[1] Protocol No. 10 refers to a part of EU document 11994N/ACT, in which 23 Austrian Standard German mostly food-related terms are acknowledged as language variety-specific.

and spearheaded the implementation of the Austrian Language Resource Portal (sprachressourcen.at).

The first topic for the discussion was the **situation of Open Data in Austria** and the creation of the Open Data Project in Austria. It was explained that the project started around 10 years ago. The aim was for different parts of public administration to come together, cooperate and exchange data sets. The data also includes machine-readable data and metadata. The Open Data Project now includes over 300 applications and an English-German glossary. Lutz is of the opinion that this glossary should be made available on the Austrian Open Data Portal.

Different concerns about the situation of the Open Data in Austria project were voiced. The biggest concern is the management of data in regard to user privacy and sensitive data. For example, written data is, generally speaking, less sensitive than voice recordings. Another concern are oversights in the creation of data bases. If certain data is missing from a database, it can lead to the unnecessary production of redundant translations, for example. The ways to solve these problems all involve experts who would be responsible for the overseeing of the project, and this requires additional funding.

The next topic of the discussion was the **potential ways to promote the exchange of data**. First and foremost, funding is needed so that institutions can afford to hire people who would oversee the sharing of data. The currently present data stewards and data experts are insufficient in this regard. It was stated that companies will find ways to share data even if civil servants and legal experts are against it, so it stands to reason that it would be better to share the data in the first place and to do so in an organised and legally compliant way.

Aside from a lack of funds, there is also a lack of time as well. Projects are usually short, 3-5 years, and instead, they should be 10-15 years long. An example given were chemists, who are given every piece of state-of-the-art technology for their laboratory so that their work can be up to date, along with longer project times; this kind of support is usually not awarded to data and language personnel at universities. Aside from these problems, work on language and data related projects is challenging because often, projects require very specific data sets that cannot be used for other projects, and conversely, the acquired data from older projects can rarely be used for new projects.

In addition to what was stated above, it was noted that the maintenance of online services needs to be handled by professionals. Transparency and visibility are also crucial, which is why events like the ELRC Workshops should become public.

3.9. Take-home message and conclusions

During the workshop, numerous topics have been touched upon, including the potential of language technologies, politics, administration, trade, society, technological maturity of Austria in regard to the combination of culture and language with AI, the position of Austria in the development of different technologies, and the development across the whole EU.

While a lot has been done already, the workshop has shown that a lot more work lies ahead. The many problems and challenges can be grouped together in different ways, but most can be solved either with better funding, with the hiring of additional experts, with more higher-education programmes to increase the number of local experts, with better communication between different parts of the public and private sectors, with better legislation with less bureaucracy, or a combination of those.

AI and language technology open up the doorway to many dangerous things, such as voice spoofing and identity theft, propaganda and information warfare, and leaks of personal or classified data. The benefits that they offer, however, outweigh the danger, especially considering the fact that those dangers can be worked around if experts from different fields work together. However, to this end Austria requires a higher number of LT experts and programmes for training national LT experts to avoid having to rely on an organisation's ability to attract talents from abroad. Additionally, funding programmes for fundamental LT research and practical LT applications is urgently required.

3.10. Demo session

To conclude the workshop, a socialising and networking event was combined with a demo session. Each organisation in the demo session provided a very brief lightning talk during the workshop, the slides of which are available online. The following organisations presented their software solutions during the demo session of the third ELRC workshop in Austria:

Cortical.io: the company delivers AI-based solutions that help businesses search, extract and analyse information from unstructured text more effectively. These meaning-based solutions, including Contract Intelligence and Message Intelligence, cover a wide spectrum of use cases with proven implementations in Fortune 500 companies. They leverage the company's patented Natural Language Understanding (NLU) technology to solve business problems with complex NLU challenges.

Kaleidoscope GmbH: the company works on customised solutions for successful global content. The company develops and manages different types of software, including software for terminology, quality assurance, reviews, query management, translations and technical documentation.

HENSOLDT Analytics: a global leading provider of Open Source Intelligence (OSINT) systems and Natural Language Processing technologies, such as Automatic Speech Recognition, which are key elements for media monitoring and analysis. The company focuses on end-to-end systems and tools that can efficiently extract and analyse information from open sources (TV, radio, blogs, social media, etc.) and turn them into actionable intelligence, employing cutting-edge technologies across multiple languages, geographies and sources that have been developed with a focus on the requirements of situational awareness.

ONDEWO: an award-winning Austrian high-tech company developing advanced AI algorithms for enabling machines to engage with humans in natural conversations. Their self-developed novel AI approach is a significant innovation leap for the human-machine communication of the future. It is based on numerous in-house developed Deep Learning and Machine Learning algorithms for Speech-to-Text (S2T), Natural Language Processing (NLP), Natural Language Understanding (NLU), and Text-to-Speech (T2S).

Knodle: The WWTF funded project "A framework for KNOwledge supervised Deep LEarning" (Knodle; VRG19-008) at the Faculty of Computer Science, University of Vienna is a Python framework for improved weakly supervised deep learning. Knodle makes use of modularisation, which gives the training process access to fine-grained information such as data set characteristics, matches of heuristic rules, or elements of the deep learning model ultimately used for prediction. Hence, their framework can encompass a wide range of training methods for improving weak supervision, ranging from methods that only look at correlations of rules and output classes (independently of the machine learning model trained with the resulting labels), to those that harness the interplay of neural networks and weakly labelled data.

4. Synthesis of Workshop Discussions

The third ELRC workshop in Austria was attended by a large number of guests from different areas who work with language and technology. The presence and participation of representatives from a large number of private companies and public service providers has been crucial in identifying currently existing problems and communicating the needs of the different parties mentioned above.

The attendees represented a number of industries and fields of study, among others: language technology and machine translation, public administration and political science, economy, and sociology. The discussion was also focused on Austria's technological maturity in terms of combining culture and language with AI, Austria's position in the development of various technologies, and the development across the EU. The different perspectives of the panel participants and the audience were essential for understanding the present issues, as well as for offering suggestions for potential solutions.

Language Technology in and for Austria from the industry perspective

As far as the industry sector is concerned, the biggest challenges to overcome are the status of Austrian Standard German as a unique language variety, and the trust between customers and companies. The Austrian market is significantly smaller than the German market, but the panel participants were of the opinion that with more expert input, this challenge can be addressed adequately. Building trust with current and potential future customers requires a deeper understanding of what that trust entails, but a first necessary step would be to offer better, more accurate descriptions of services and products that the companies provide, without overpromising or misrepresenting the possibilities that the services or products offer. These two challenges go hand in hand and can hardly be solved separately one from another. If Austrian Standard German can become a viable option for companies to focus on, then the use of Austrian Standard German can make language technologies more appealing to the Austrian Market, which in turn can create trust and enable further improvements of the language technologies.

Language technologies in and for Austria by/for the public sector

Unlike with the private sector and industry, where the existing problems within different organisations are similar in nature, the problems faced by the different parts of the public sector are all very distinct. Three big types of challenges can be identified: challenges regarding speech technology and interpreting, challenges regarding written communication and translation, and challenges regarding privacy and security. While the challenges are unique and require individual solutions, a general approach to solving them might exist. In general, more expert knowledge is needed, as well as improvements of technological and media literacy of laypeople. With better funding, it would be possible to employ more experts and implement newer technologies. In combination with efforts to make the work and general use of technologies easier for people who are not necessarily experts for language technology (such as by improving technological and media literacy or offering specialised software with optimised user interfaces), it would be possible to overcome a large number of issues raised by the panel participants. These issues include the speed at which communication needs to occur, the accuracy of translation and information transfer, and privacy and domestic security.

Language data creation, management and sharing: existing practises and challenges

Currently, the biggest challenge concerning data sharing and the Austrian Open Data project is finding the appropriate ways to share data. This is particularly difficult because different types of data cannot be shared in the same way, and some cannot be shared at all. This is mostly influenced by privacy concerns. As an example, written texts that do not contain personal data could be shared more easily and without concerns regarding privacy, while voice recordings need to be handled much more carefully. Once again, expert knowledge is needed for the sake of challenging these issues. Only if experts, companies, and regulatory bodies work together can it be possible to find ways for different entities to share data securely. Unless this cooperation can be developed, it is certain that some companies will find ways to share data with others if it benefits them, regardless of how sensitive the data in question is.

5. Country Profile: Language data creation, management and sharing

The panel session on language data creation, management, and sharing has shown that while a lot has been done already, there is a lot of work to be done in the future, as well. The Open Data in Austria project is currently progressing very well, however, some concerns were raised. There is a lack of overview into which data is being preserved and which data is lacking. There seem to have been oversights in the past and data that should have been stored already is currently missing.

There is a need to employ people who would be responsible for overseeing the data creation, management, and sharing process. And maybe more importantly, they would need to identify the blind spots in the communication channels which lead to the oversights in the first place and offer possible solutions for removing those blind spots.

The panel participants agreed that there is a lack of funding which would allow companies and institutions to hire professionals who would be able to deal with the challenges described above.

Another challenge that was pointed out is the fact that there are a lot of printed documents in use, many of which are outdated. In fact, it seems like any type of printed media becomes outdated as soon as it is printed due to the rapid technological developments that influence everyday life. There exists a need to move toward online media as much as possible.

From a survey on the data resource situation in Austria conducted among workshop participants, it could be gathered that organisations utilise eTranslation as well as machine translation, Computer Assisted Translation (CAT) technologies and online dictionaries. Participating organisations generally provide terminologies, glossaries, parallel texts, and translation memories.

In terms of data sharing, the main inhibiting factors were legal issues, especially confidential information within the data and copyright issues. Furthermore, inadequate practices for the management of data was attributed to a lack of a centralised initiate and a lack of data reliability and interoperability. Finally, a lack of a national policy was identified as complicating the sharing of data, which, as indicated in the answers of one participant, was repeatedly emphasised in the workshop.