Deliverable D3.1.1

Report on 3rd ELRC Conference

Author(s):	Andrea Lösch (DFKI)
	Lilli Smal (DFKI)
Dissemination Level:	Public
Date:	2018-02-09
Copyright:	yes

Contents

1 Introduction

This report provides a summary of the 3rd ELRC conference. The conference took place in Brussels, Brussels BluePoint (www.bluepoint.be), on the 7th and 8th of November 2017 as a subsequent satellite event to the Translating Europe Forum (6th – 7th of November 2017).

The report is structured as follows: First we describe the aims and objectives of the conference, followed by an overview of the thematic structure and organisation of the conference (see 2 Focus and contents of the Conference). This also includes details on the contents of the demonstrations and presentations for each day (see section 2.3.1 for Day 1: Arena of Opportunities, and section 2.3.2 for Day 2: Plenary Sessions). Most importantly, an overview of the major discussion points is given (see section 3 Major Discussion Points). Last but not least, the Annex provides the conference programme. All conference presentation materials including videos are available through the ELRC website (http://www.lr-coordination.eu/ELRC-conference-2017).

2 Focus and contents of the Conference

2.1 Context

Europe's linguistic diversity can turn into an obstacle for public service administrations and for the provision of public services in Europe, for example when it hinders the free exchange of information or when it prevents unbiased access to public services, business opportunities, job opportunities, and support. Through the **European Language Resource Coordination (ELRC)** (www.lr-coordination.eu), the European Commission is undertaking an unprecedented effort to support the further development and adaptation of the automated translation platform (CEF eTranslation) to the needs of public services across all EU Member States, Iceland and Norway. In doing so, the ELRC directly supports the provision of multilingual services for Europe's citizens, administrations and businesses. CEF eTranslation will power Europe's public online services (see, for instance, the Online Dispute Resolution Platform, the eJustice Portal, Electronic Exchange of Social Security Information (EESSI), the European Data Portal, etc.)

2.2 Target audience

The ELRC Conference targets **public service administrators and representatives of public sector bodies in all CEF affiliated countries**. In particular, this includes the responsible translation services, as well as representatives for digitisation, representatives of information offices and/or public relation offices, and people responsible for data and data management. The 3rd ELRC Conference was organised as an invitation-based event to ensure that the target audience is reached and present at the event. All invitations were issued by the European Commission and the ELRC coordinator.

Overall, there were 150 registered participants from which 142 participants attended the 3rd ELRC Conference; all countries were represented with the exception of Cyprus. Overall, the audience was composed of the following participants:

- 74 Representatives of Public Service Administrations
- 27 Representatives from Industry
- 41 Representatives from Research / Academia

The list of participants is available in Annex 2. The higher percentage of representatives from research/academia (including also Technology NAPs) and from industry in comparison to the last conference stems from the fact that there were more than 20 presenters, exhibitors and other contributors from research/academia and industry extended their stay to also take part as participants.

2.3 Focus and contents

The central aim of the 3rd ELRC Conference was two-fold: (i) to illustrate the multi-lingual needs of public services across Europe (and how to achieve sufficient quality of the translations in these domains) and (ii) to kick off the discussion on how to

achieve a sustainable supply of language resources (LR) for such public services across Europe.

This latter is a key concern to the ELRC whose goal it is to enable the sharing of language resources across and among all CEF-affiliated countries in order to provide multi-lingual public online services (in particular Digital Service Infrastructures (DSIs), but also others). As such, the sharing of LR is not and cannot be a one-time effort,; it must be a continuous activity, which is undertaken even beyond the life-time of the ELRC. The collection and sharing of LR hence cannot stop with the end of the ELRC project which is currently supported from the Connecting Europe Facility Programme. Rather, it is essential to find and establish a **sustainable** model for the sharing and for the provision of relevant language resources in order to ensure the functioning of European public online services in the future. In order to achieve this goal, the sharing of language resources must be anchored in and carried by all of the participating countries. This means that the future model needs to acknowledge the diversity of the participating countries: Each country has different processes and structures and hence, the modalities and conditions for the sharing of language resources may differ from country to country. It is the mission of the ELRC to identify the modus operandi for each individual country and to determine a modus operandi that is suitable for all countries together.

The detailed agenda of the conference is available in Annex 1. The following sessions (2.3.1 and 2.3.2) provide the details of the contents covered and discussed on each of the conference days.

2.3.1 Day 1: Arena of Opportunities

The first day featured the **Arena of Opportunities** where participants were able to meet different experts and language technology service providers and to discuss existing solutions face-to-face. During the past year, many technological advances have been made with regard to multi-lingual technologies, especially in the area of machine translation, but also with regard to other tools created to enable fast high-quality translations. The Arena of Opportunities hence included demonstrations of the latest tools available and services for computer-assisted translation, machine translation, and for the processing of language resources. Moreover, the representatives of all collaborative projects funded under CEF Automated Translation which started recently were also present and illustrated their activities with regard to the enhancement of CEF eTranslation. Overall, the Arena of Opportunities included the following exhibitors:

1. **Demonstrations of Neural Machine Translation (NMT):**

eTranslation (upgraded version of the EC's earlier MT@EC service): This demonstration referred to the European Commission's machine translation service. The service produces translations into and from any official EU language. It is currently available to and in use by all EU institutions and agencies, and is also available free of charge to public administrations in all EU countries, as well as Iceland, Norway , which is why the ELRC Conference's participants were highly interested in that topic. eTranslation / MT@EC are trained by DGT using the vast Euramis translation memories, comprising over one billion sentences in the 24 official EU languages produced by the translators of the EU institutions over the past

decades. The systems are particularly suited to the needs of EU policy documents. The demonstration gave a preview of the Neural MT engines released in November 2017, showing the significant improvements in translation quality that can be achieved with the help of the new systems. Further information on eTranslation / MT@EC is available online at
https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Machine+Translation.

Open Source Neural Machine Translation System and Pure Neural™ Machine Translation (PNMT): This demonstration illustrated how the first ready-to-use Neural Machine Translation Software, developed by Systran, works. PNMT presents an inherently fluent, accurate, context sensitive, end-to-end, high-speed translator. Moreover, the exhibitor included a demonstration of OpenNMT (www.opennmt.net), an open source initiative for neural machine translation and neural sequence modelling which can be used by parties that are interested in it. For further information please contact Djamel Mostefa (djamel.mostefa@systrangroup.com).

Neural Machine Translation with Quality Translation 21: Quality Translation 21 (www.qt21.eu) is a neural machine translation project which has received funding from the European Union's Horizon 2020 research and innovation programme. The project focusses on languages which are currently not adequately supported by machine translation. QT21 therefore provides solutions for high-quality translation for these languages. For further information please contact Christian Dugast (christian.dugast@dfki.de).

2. Tools for translators:

translationQ: This demonstration featured translationQ which was developed by televic (www.televic-education.com) and which presents a unique computer-assisted tool for revision, particularly for the evaluation and review of the quality of translations. translationQ will also automatically detect errors across documents and provides automatic feedback. This does not only save significant time and effort, but also ensures that all translations are corrected following exactly the same criteria. For further information please contact Dirk Verbeke (d.verbeke@televic.com).

3. Tools for the processing / generation of language resources:

ILSP Focused Crawler (ILSP-FC): The ILSP Focused Crawler (http://nlp.ilsp.gr/redmine/projects/ilsp-fc/, contact: prokopis@ilsp.gr and vpapa@ilsp.gr) is a research prototype for acquiring domain-specific monolingual and bilingual corpora. The input required from the user consists of a list of seed URLs pointing to relevant web pages and a list of terms that describe a topic. ILSP-FC integrates modules for text normalization, language identification, document clean-up, text classification, bilingual document alignment (i.e. identification of pairs of documents that are translations of each other) and sentence alignment. As such, it is a valuable tool to create language resources in domains and areas where there are no resources readily available. In addition to that, the software can be used as a general crawler if the user does not provide a list of terms.

ParaCrawl: The ParaCrawl project focusses on the generation of large-scale parallel web crawls in order to generate corresponding large-scale language resources for various language pairs. These can serve as a good basis for training MT systems although/despite the fact that content adaptation is not possible. Further information on ParaCrawl see http://statmt.org/paracrawl/ or contact: Kenneth Heafield – kheafiel@inf.ed.ac.uk)

Anonymisation: Regarding the sharing of language resources, there may be two parallel sets of issues involved preventing their use for machine translation: (i) copyright and related rights issues and (ii) privacy related issues. The latter frequently presents a problem for the sharing and use of resources, since data that contains for example personal or confidential information usually cannot be shared. The automated anonymization of language resources hence can be considered a useful tool or service to overcome such problems and was therefore included as exhibit for the Arena of Opportunities. For further information please contact Victoria Arranz (arranz@elda.org) or Pawel Kamocki (Kamocki@elda.org).

4. CEF-TC-2016-3 Projects for Automated Translation:

CEF Automated Translation for the EU Council Presidency: In order to reach the goal of an EU Council Presidency without any language barriers, the translation tool uses Neural Machine Translation and the CEF eTranslation platform to provide secure, instant translation of texts, full documents, and websites. By the use of AI-based machine learning, Neural MT produces highly fluent, readable, and almost human-like translations. The EU Council Presidency in Estonia was the first to use the ground-breaking new tool, which also features the world's very first Neural MT systems for Estonian. Besides journalists and delegates, the tool's users include Estonian public sector translators and staff members at the EU Council Presidency secretariat. The EU Council Presidency Translator is available online through www.translate2017.eu (contact: Rihards Kalnins - rihards.kalnins@tilde.com).

eTranslation TermBank: The objective of the eTranslation Termbank action is to identify and collect terminology resources relevant for national public services, administrations, and governmental institutions across European countries. Terminology resources will be collected within domains including health, business legislation, and consumer protection. All data collected within this initiative will be used to develop high-quality automated translation systems for EU languages in the CEF eTranslation platform which, as indicated above, is open for public institutions. For further information please contact Tajana Gornostaja (tatjana.gornostaja@tilde.com).

IADAATPA (Intelligent, Automatic Domain Adapted Automated Translation for Public Administrations): IADAATPA will deliver a scalable and highly secure platform for the provision of automated translation services capable of integrating, using and extending the EU's current Automated Translation platform. It will help to foster the adoption of the Automated Translation Building Block by easing the integration of automated translation into cross-border European digital services and serve as a successful use case for future implementations at local, regional and national levels

and by different administrations. The action will incorporate a series of state-of-the-art and domain-adapted Machine Translation engines based on statistical and neural networks. They will be deployed and tested by the IADAATPA Public Administrations early adopters in Spain, the Republic of Ireland and Latvia. For further information please contact Manuel Herranz (m.herranz@pangeanic.es).

European Language Resources Infrastructure (ELRI): The objective of the ELRI project is to deploy a network that enables accessible, secure and reliable sharing of Language Resources by different stakeholders at EU level and promotes its uptake and growth. The focus will be on the language resources which are produced by translation centers and public institutions in Europe, aiming to render the data available for the DGT. As such, the project targets the deployment of a centralized Language Resource sharing infrastructure (based on ELRC-SHARE technology) for all the Translation Centres providing translations and derived materials such as translation memories to the respective national administrations. The main actions to be taken by the project are (i) the establishment of local LR relay stations for the different participating countries and (ii) the connection of each node to the centralised sharing centre. For further information please contact Thierry Etchegoyhen (tetchegoyhen@vicomtech.org).

Cross-border eProcurement notifications: The goal of the project is to enable the exchange of tender notices across three different procurement systems (Croatia, Slovakia, Slovenia) and to distribute such notices from other countries to potential bidders. In particular, this includes the provision of basic tender information in the user's preferred language. Accordingly, the project also establishes the necessary shared functions for achieving these objectives, which include: (i) the steering of the translation workflows, (ii) the temporary storage of notices, and (iii) the creation of a keywords/tagging database. For further information please contact Michal Ohrablo (michal.ohrablo@anasoft.com).

Provision of web-scale parallel corpora for official EU languages: see above, 3. Tools for the processing / generation of language resources: ParaCrawl

5. Other

Public Multilingual Knowledge Infrastructure (PMKI): The main objective of this project is to help SMEs and public administrations to save resources by creating and maintaining their own multi-lingual tools. The public multilingual knowledge management infrastructure (PMKI) will support e-commerce solutions such as machine translation, localisation and multilingual search by creating interoperable multilingual classifications and terminologies that will be easily reusable by small and medium-sized enterprises (SMEs) and public administrations. Compared to big companies, SMEs are currently at a disadvantage due to the high costs for providing multilingual services. In 2017, core data model for multilingual taxonomies/terminologies has been established and will be provided via PMKI. For further information, please visit https://ec.europa.eu/isa2/actions/overcoming-language-barriers_en or contact Najeh Hajlaoui (najeh.hajlaoui@ext.publications.europa.eu).

2.3.2 Day 2: Plenary Sessions

The second day of the ELRC Conference was dedicated to the plenary sessions:
- Session 1: Illustrated the needs of public services (in particular CEF Digital Services) with regard to multilinguality and language resources
- Session 2: Demonstrated the benefits and possibilities offered by new tools and services to support the translation process
- Session 3: Illustrated successful approaches for sustainable sharing of language resources in different public service scenarios

1. Session 1: Multi-lingual needs of public services

As part of this session an introduction to different online services (e.g. eJustice, eProcurement, European Refugee Information and Communication System ERICS), their underlying translation process, and corresponding needs and requirements for language resources was given.

The **eJustice** portal was presented by Agnieszka Jelnicka from the Directorate for Translation (DGT) of the European Commission. Following a short introduction to the eJustice Portal as a "one-stop shop" which was released in 2010 and currently covers 150 justice-related topics in 23 languages (including 28 judicial systems), Ms. Jelnicka provided an insight into the different contents. She also demonstrated that the use of the eJustice portal has increased significantly and constantly from 2010 (less than 200.000 visits) until 2016 (almost 4 Mio. visits). The eJustice translation process was explained and major constraints were illustrated, including, for instance, difficulties of translating texts based on provision from national legal acts (i.e. not EU legal acts), the large number of eJustice pages that need to be translated, the difficulty to ensure consistency of the translations per theme, or the obligation to outsource translations. Finally, the conclusions were drawn on the usability of MT output which, due to the non-existence of domain specific engines for the eJustice portal still has limited usability.

The multi-lingual requirements of the **eProcurement** DSI were presented by Marc-Christopher Schmidt from the European Commission's DG GROW. Following an overview of the aim and coverage of the eProcurement platform, Mr. Schmidt explained the different tools included in the eProcurement DSI, namely
- TED (tenders electronic daily - http://ted.europa.eu/TED/main/HomePage.do): a tool which makes all tender notifications available
- eESPD (the European Single Procurement Document - http://ec.europa.eu/growth/single-market/public-procurement/e-procurement/espd/): a tool for self-declaration which is used as preliminary evidence, replacing the certificates issued by public authorities or third parties to confirm that the tenderer fulfils the exclusion or selection criteria of the tender, and
- eCertis (https://ec.europa.eu/growth/tools-databases/ecertis/): a mapping tool that helps public buyers and bidders to identify certificates and attestations

> requested as evidence of eligibility in procurement procedures across all EU countries.

His presentation then focussed on different use cases for eTranslation in the context of eProcurement. For suppliers, this includes in particular the multi-lingual support (i) for the identification of business opportunities (notifications), (ii) for the understanding of the requirements (eCertis and eESPD), and (iii) for the understanding of the procurement documents (which, however, are often unstructured). For buyers, multi-lingual support is needed in all cases where it is not obligational to have the documents in a particular language.

Last but not least, Mika Nieminen from University Aalto presented the **European Refugee Communication and Information Service** (ERICS). Following the overview of the system's technical architecture, Mr. Nieminen showed how Artificial Intelligence is used in the crowd-sourced information workflow. The AI functions employed by ERICS particularly include the use of Neural Machine Translation (https://github.com/tensorflow/nmt). The collaborative translation workflow is an iterative one, starting with the translation of a particular sentence (for which MT is being used), the re-writing/correction of the sentence by crowd-workers and the improvement of the language and syntax by the crowd-workers. This workflow was used to translate an entire book within an hour. Moreover, it outperformed the sole use of MT both in terms of adequacy, as well as fluency. However, the NMT system used is not adapted to the particular domain which should be improved in the future.

2. Session 2: Tools and services for high-quality translation

As part of this session, experts from the language technology industry, as well as users of language technologies and Language Service Providers (LSPs) provided clear and hands-on accounts from their day-to-day practice of novel tools and services used for translation. Furthermore, the effects on the translation process, the benefits and, above all, the quality of the translation were demonstrated. A panel on "How to ensure maximum quality of translations" concluded this session.

The session was opened with a presentation on **translation quality, neural machine translation and language resource**s by Kenneth Heafield from the University of Edinburgh. By using lively examples, he illustrated that neural machine translation achieves better results with regard to agreement and fluency (word selection), and that generally, a more appropriate style is used. On the other hand, he pointed out that rare words may present problems even for NMT. As regards to the importance of data, Kenneth Heafield indicated the data sizes required to ensure that machine translation works and showed the positive effects of data size (in particular with regard to in-domain data) on the translation quality. The case study on the translation of patents illustrated that it is usually better for the translation quality to build in-domain engines with fewer data rather than giving the same data to a generic system (where it will not have the desired positive impact on translation quality). The end of his talk gave an overview of the creation of large-scale language resources as part of the ParaCrawl project – and also the importance of such LR as baseline data for MT engines.

Teresa Lynn from the ADAPT Centre of Dublin City University in Ireland presented ADAPT's experience in providing **machine translation for the Irish Government**, and in particular on how to achieve good translation results even when the language resources are scarce. Her presentation started with insights into Irish as minority language which did not only include a history of Ireland and the Irish Language, but also insights into recent legislative developments. She then illustrated the translation needs of the Irish government (Department for Culture, Heritage and the Gaeltacht), their former translation practices, and the milestones when introducing them to machine translation. Details were given about the development of the MT Engine (including the collection and management of language resources in particular), the incremental improvement of the quality of the MT output, and the take-up of the MT system by the public services in Ireland. The factors for success were explained and the virtuous circle of sharing of language resources was illustrated. The presentation concluded with an outlook to the future plans of the Irish Government departments and the vision of a Shared Translation Service. Last but not least, Teresa Lynn showed how the development of the MT for the Irish government feeds into CEF activities.

The concluding **panel** of this session focused on the **quality of translation**. The panel was moderated by Rihards Kalniš (Tilde) and included the following panelists:
- Anna Kotarska – Senior Specialist, Department of Analysis and Strategy, National Health Fund, Department of Analysis and Strategy, Poland
- Markus Foti – Project Manager, European Commission, DG Translation
- Shanna Ni Rabhartaigh – Translator, Department of Culture, Heritage and the Gaeltracht, Ireland)
- Jānis Ziediš – Project Manager, Culture Information Systems Centre, Latvia

The central aim of the panel was to underline the importance of language resources (LR), even if MT is not used and to show that LR are of fundamental importance for an efficient translation process and for high-quality translations. The questions addressed included the following:
- Quality management:
  - How do you manage translation quality in your organisation (process)?
  - How do you evaluate the quality or accuracy of a translation (criteria)?
- Technologies and quality:
  - How important do you think the human factor is (i.e. technical fluency in the target language; localisation expertise, subject matter experience etc.)? And to what extent do you think that language technologies can help?
  - Do you think there has been a change in the importance and the value of machine translation with regard to the emergence of neural machine translation?
  - What are your expectations for the new paradigm? How do you see the benefits?
  - Which other tools and technologies do you consider useful for optimizing the translation process and outputs?

Major discussion points covered by this panel are presented in section 3.6 below.

3. Session 3: Approaches for sustainable sharing of language resources

As part of this session, best practice examples and different approaches for sustainable sharing of language resources within prominent national ministries and public service administrations were given.

The session was opened with a presentation on the **use of language resources within eTranslation** by Markus Foti, the eTranslation project manager from the European Commission's DGT. Following an introduction into MT@EC and EURAMIS (DGT's database of translated segments), Markus Foti illustrated the document translation workflow within DGT. It is important to note that because of DGT's outstanding expertise in the area of machine translation and also due to its unprecedented in-house collection of language resources (over one billion sentences currently, growing at 2.6 million a month), eTranslation/MT@EC and EURAMIS present best practice for the use of machine translation world-wide. Markus Foti also highlighted that because of its history, EURAMIS became a treasure trove of Eurospeak. As a consequence, translations with eTranslation/MT@EC for other domains than the ones within the EC are less good in terms of their quality. For instance, the quality of translations for colloquial language by eTranslation/MT@EC is noticably lower. The aim is now to build more domain-specific engines (in particular for CEF Digital Services) so that the quality of translation is improved for these application scenarios.

The session also included the case of **inter-ministerial collaboration in Finland** as a first best practice example on the national level. Taru Virtanen, the Head of the Foreign Languages Unit at the Prime Minister's Office in Finland, presented the history of the translation services in Finland where a remarkable transformation of the services took place: in 2015, the translation specialists from all 12 government ministries were brought together in one single in-house translation service which now provides translation, language and terminology services to all ministries. Ms. Virtanen also illustrated the management of the translation memory system and internal termbases, as well as the maintenance of the government online termbank Valter (www.valter.fi). In summary, several translation-related functions were automated, including especially the submission and handling of translation requests, but also other internal functions. Most importantly, all relevant information on translation and revision assignments was stored in electronic form (Shake) which also collects statistics for operational planning and budgeting. Particular emphasis was placed on the management of language resources.

Subsequently, the case of **eTranslation for the EU Council Presidency** was presented by its project manager Rihards Kalni š from Tilde. He first provided insights into the HUGO translation service – a specialized MT service which had been developed for the 2015 Latvian Presidency of the EU Council. Then, a short explanation was given about eTranslation to set the base to then illustrate how eTranslation could be integrated for the EU Council Presidencies 2017-2018. In particular, the EU Council Presidency Translator 2017 was introduced, followed by an illustration of the use of neural machine translation and the importance of language

resources for the development of the 2017 translator for Estonian. Again, the importance for sharing language resources even across different presidencies was stressed, in order to be able to adapt the MT engines to the particular domain. As such, the EU Council Presidency Translator represents an important example of sharing LR in a particular domain across several presidencies. Last but not least, a short demonstration of the current EU Council Presidency translator was given. The presentation concluded with an outlook for the upcoming EU Council Presidencies (Bulgaria and Austria) and the corresponding work that needs to be done.

Next, Djamel Mostefa, R&D Manager of Systran, illustrated in his presentation the **best practices for managing data for MT**. Following a brief overview of Systran's engagement in machine translation, he pointed out that the major success factors for corpus management include (i) an overview of the types of language resources useful for MT, (ii) an illustration of the corpus management workflow, (iii) the corner stones for corpus specification, and (iv) an insight into the production of parallel corpora. He elaborated in particular the legal framework for the use of language resources for machine translation (including a brief demonstration of key solutions like anonymization). Finally, he summarized the main conclusions, thus providing the best practices for sustainable management of language resources.

The final presentation of this session focused on key questions associated with **a sustainable supply pipeline for language resources among public services in Europe** by Andrea Lösch (DFKI, ELRC Project Manager). Following a brief summary of the aims of the ELRC and its achievements so far, the current structure for the supply of language resources was presented, including key aspects (coordination, technical infrastructure, support services, and communication) at the European level, the national level, and at the institutional level in each country. Based on this analysis, it was shown that currently several key questions will need to be figured out:

- On the **national level**, the key question is about the communication structures: We need to identify and implement appropriate communication structures on national level with regard to the sharing of LR*.*
- Similarly, on **the institutional level** in each country, we need to identify corresponding institutional coordinators and ensure their commitment. Based on this, we can set-up the corresponding institutional coordination and communication structures.
- Last but not least, and directly associated to the questions just posed, with regard to the ELRC, there is one remaining question on the European level: How can we ensure sustainability of the structures on all levels? What is the best way to move forward from here?

The presentation continued looking at initiatives and best practice examples for the sharing of language resources on the national level, best practice on the European level, existing Europe-wide initiatives supporting the sharing of language resources (commercial and non-profit) as well as existing collaboration and exchange of best practice among translation services in Europe. It has been concluded that in order to find the right solution, all relevant stakeholders must be involved. Moreover, the solution(s) must respect the national level and be deeply rooted in it; they should ideally build on already existing initiatives to avoid duplication of efforts and to allow the creation of synergies.

The conference concluded with a **panel on framework and conditions for sustainable supply of language resources** in which the foundations of a sustainable supply pipeline were critically discussed by representatives from the public sector (both national governmental institutions and the EC). The panel was moderated by Josef van Genabith (Scientific Director of Multilingual Technologies Department at DFKI and ELRC Coordinator). The panlists included:

- Thierry Etchegoyhen – Vicomtech; ELRI Coordinator
- Jeffrey Ganellen – Ministry for Foreign Affairs, Sweden
- Djamel Mostefa – R&D Project Manager, Systran
- Marco Marsella – European Commission, DG CNECT, Head of Unit G3 – Learning, Multilingualism and Accessibility
- Taru Virtanen – Head of Unit, Foreign Languages, Prime Minister's Office of Finland

The central aims of the panel were (i) to underline the importance of language resources (LR), even if MT is not used, (ii) to identify best practice and feasible models for sharing of LR, and (iii) to identify existing obstacles / pain points with regard to the sharing of LR. In order to achieve these goals, the panel critically discussed the following key issues:

- Examples and organisation of information exchange on national level (in particular insights by Taru Virtanen and Jeffrey Ganellen), discussing questions such as
    - How do translators collaborate in your organisation? Do you exchange useful information, documents, example translations, language resources, other?
    - How is the collaboration organised on process level and on technological level? (i.e. where is the relevant information stored, how can people access and use it? In general, what kind of tools and technologies do you use to support the translation process?)
    - Is there any collaboration with other translation services / translators outside your organisation? If so, how is it organised and is there any sharing of relevant information, language resources etc.?
- Examples of internal information exchange in an international organisation (in particular insights by Djamel Mostefa):
    - How is the sharing of LR organised in Systran?
    - Are there any differences between the EU and the US?
- Main difficulties regarding the sharing of language resources on the national and on the European level as perceived by the participants
- Possible solutions in order to overcome these obstacles (all)
- A short explication and insights into the work of the Conference of Translation Services of European States (COTSOES) and mechanisms established to exchange best practice, latest developments in key areas for translation services (i.e. terminology, new technologies, human resources).

Major discussion points covered by this panel are presented in section 3.11 below.

The conference was closed by Josef van Genabith, underlining again the slogan of this conference "Identify – Share – Benefit", with an outlook on the upcoming activities of the ELRC and the planned follow-up of the sustainability discussion within the Language Resource Board.

3   Major Discussion Points

3.1 eJustice

Following Agnieszka Jelnicka's presentation of the multi-lingual needs within the eJustice portal (see above, section 2.3.2), several important discussion points emerged. The first one concerned the perceived benefit of having eTranslation trained for application in the eJustice domain. It was concluded that the biggest benefit of this would be the considerable increase in the efficiency and better support of the internal translation process. The translators could then manage to translate more documents than they are now able to.

The question was raised on how DGT copes with the great translation requests which clearly exceed the capacity of the DGT translators. It was explained that currently, much of the translation work for the eJustice portal is being outsourced, and that unfortunately, DGT does not ask to obtain the tmx-files from the translation agencies/freelance translators. Consequently these documents do not feed into the Euramis data base of DGT and they cannot be used for machine translation training. The parallel texts of the outsourced translations however represent very useful training materials for eTranslation engine in the eJustice domain. It was concluded that it is hence of utmost importance to work towards a policy of obtaining the tmx-files from external translators / translation agencies together with the outsourced translation.

Following a direct question from the audience it was explained that the subcontracted translation agencies do not have access to the Euramis database, which means that they cannot use the language resources in order to guide their translations. The reason for that is that Euramis is an EC-internal system. However, external translation agencies can use EUR-LEX to inform their translations.

Last but not least, it was stressed that even though eTranslation cannot be directly used by non-public service organisations and institutions, it is and will be available within the context of different public services. This means that the users of the eJustice portal can for instance use it as part of the eJustice portal in order to have the contents of the portal translated.

As conclusive statement, it was again highlighted that the current eTranslation engines do not, work very well for the texts published on the e-Justice portal because they are trained mainly on EC data. What is hence most important to have now is a higher amount of translated text on national legislation – not just on EU legislation – in order to make the eTranslation service work better in the eJustice domain.

3.2 eProcurement

One of the key discussion points raised by the presentation of Marc-Christopher Schmidt (see above, section 2.3.2) concerned the benefits of using eTranslation for

the eProcurement DSI. Most interestingly, for eProcurement the most important aspect of using eTranslation instead of using other freely available or commercial solutions (e.g. Google translate) is the IPR (Intellectual Property Right). Marc-Christopher Schmidt pointed out that if someone uses eTranslation, the IPR remains with the one using the eTranslation tool – and not with the translation service provider. Moreover, by applying eTranslation to the eProcurement domain, the benefits are significant in the sense that the entire user group of eProcurement will benefit from this service (i.e. all suppliers and buyers across Europe, businesses etc.)

Regarding the question of who can actually have access to the structured data used in eProcurement, it is interesting to note that within eProcurement, an open source management tool is used that stores all the translation files – and that these files could be shared even with the ELRC and help the further adaptation of CEF eTranslation. He explained that potentially a set of multilingual structured data could be very helpful for improving Machine Translation applied to unstructured documents.

Another interesting suggestion was made with regard to the development of the CEF eTranslation service: it would be good to provide direct links towards the source (country) which would help improve the overall understanding; the more information a translator has, the easier is his job. As such, in the future, more structured information will be provided. So far, numerous code lists are already being provided by the publications office, e.g. code lists for the different countries.

Marc-Christopher Schmidt also pointed to the Metadata Registry (MDR) of the Publication Office of the EU (http://publications.europa.eu/mdr/index.html), in which one can find potentially relevant collections of meta data elements in all European Languages.

3.3 European Refugee Communication and Information Service

The European Refugee Information and Communication Service (ERICS) which was presented by Mika Nieminen from Aalto University aims to support refugees with a mobile translation and search service, including an extremist and fake news filter solution. In 2018, ERICS will be available, also for mobile users. It is the first AI-based translation system on the market that integrates crowd-sourcing with further exploitation of the collected and interpreted data to enhance the machine-learning system. The advantage of the crowd-sourced approach is its efficiency. As indicated earlier (see above, section 2.3.2), the collaborative translation workflow is an iterative one, starting with the translation of a particular sentence (for which MT is being used), the re-writing/correction of the sentence by crowd-workers and the final improvement of the language and syntax by the crowd-workers. The application of this process outperformed the sole use of MT both in terms of adequacy and fluency.

A key question raised in the audience was how freelance translators are selected for the crowd-sourcing in order to ensure high quality. It was explained that freelance translators serving ERICS are evaluated and rated based on simple tests (example translations). In the course of the provision of translations for the service, these ratings are being monitored and can change over time (e.g. better translators will be

given jobs more frequently). Overall, the crowd-sourcing approach to translation represents an interesting usage scenario for all public services in which human translation is not possible for reasons related to time or resources, but in which fast and high quality translation is needed.

3.4 Translation quality, Neural Machine Translation and LR

Following the presentation by Kenneth Heafield from University of Edinburgh (see above, section 2.3.2), it was stressed again that in neural MT it is always best to have in-domain parallel data, which is very similar to statistical machine translation (SMT). As a reference, roughly 10 000 words of in-domain parallel data are enough to test a system. 50 000 words of in-domain parallel data are enough for light customisation. Moreover, Kenneth Heafield pointed out that even monolingual data can improve the translation quality by 1-4 BLEU points.

Last but not least, the case of the European Patent Office (EPO) was discussed again. This case proved that it is important to understand that specialized engines will always yield higher quality translations than generic systems (even if in-domain data is contained in them). While EPO decided to give his data to Google (a generic system), WIPO (the world patent organization) used the same in-domain data to set up a custom system with the result that the WIPO system significantly outperforms the Google generic system. This means that adding even the best in-domain data to a system which largely consists of out-of-domain data, will barely improve the system. It is better to have a system trained sufficiently on in-domain data.

A last question arose regarding the impact and necessity of the ParaCrawl project and the large-scale parallel corporal obtained from it. Kenneth Heafield explained that large parallel corpora which are out of domain are always helpful (and absolutely necessary) as a baseline, but not sufficient for ensuring adequate translation quality. Quality only derives from in-domain data. However, in cases of languages where the European Commission has very little baseline data, large-scale corpora are being sought in addition to the always necessary in-domain data.

3.5 Machine Translation for the Irish Government

The case of the machine translation developed for the Irish Government as presented by Teresa Lynn from the ADAPT Centre of Dublin City University (see above, section 2.3.2) illustrated how to achieve good translation results even when the language resources are scarce. ADAPT was asked by the Department of Culture, Heritage and the Gaeltacht (DCHG) to develop an MT system, especially in order to support the translation of repetitive translation tasks and to ease the increasing translator workload. Moreover, since language resources were scarce, ADAPT was asked to create more Irish language resources.

All existing resources (i.e. notices, public relations, parliamentary questions, reports, speeches that had already been translated within DCHG) were combined into a baseline system. However, the quality of the automated translation was poor. As such, ADAPT made efforts to clean the existing TM data from DCHG and also crawled public information sources online to cover gaps identified in the cleaned data. Based on this data, the MT system was then fine-tuned and re-trained. It is

important to mention that there was a significant improvement and that the trained system clearly outperformed the freely available online translation services for the DCHG use case.

Teresa Lynn stressed again that unlike other languages (such as English, French, German etc.) the Irish language is an under-resourced language for many application scenarios. This is why the main issue of developing the MT system for DCHG was indeed to find the right data to be able to train and customize the system.

Last but not least, it was also explained that in the translation departments of most public administration today, there is a general issue regarding the appropriate management of language resources. Not many translation services store their translation memories / tmx files. And if they do, in many cases, the translations are not stored in an organized / re-usable way. Given the importance of data (language resources) for the development and the customisation of MT systems, it is hence of utmost importance for all translation departments to manage their language resources adequately, so that they are available for future reference.

3.6 Panel: How to ensure maximum quality of translation?

The central aim of this panel was (i) to underline the importance of language resources (LR), even if MT is not used and (ii) to show that LR are of fundamental importance for an efficient translation process and for high-quality translations. The questions addressed covered both the management of translation quality in translation departments as well as the relation between translation quality and the use of technologies.

As before, the major discussion point concerned the connection between language resources and the quality of MT outputs. As was already apparent in several presentations preceding the panel, the technology experts of the panel (namely Markus Foti, Project Manager for eTranslation/ MT@EC of the European Commission's DGT, and Janis Ziedins, Project Manager of the Culture Information Systems Centre in Latvia) highlighted again that language resources are of utmost importance for high quality translations. By using several examples, they explained and illustrated that without sufficient in-domain and baseline training data, the output of any MT system will be poor.

Another important issue that was raised was the availability of language resources for translation services. Anna Kotarska explained (Senior Specialist at the Department of Analysis and Strategy at the National Health Fund of Poland) explained that many of the translations for ministries are being outsourced. As a consequence, it remains a problem to get hold of the tmx files translated by external agencies. It was concluded that subcontracting needs to be adjusted in this case to include a clause to hand over the tmx files, too.

Following a discussion with the audience, it turned out that several other translation services present at the conference found themselves in the same situation. However, some of them had already included corresponding clauses in their agreements and contracts that would require external agencies to hand over the translation memories

(tmx files) together with the translations. Unfortunately, the inclusion of such clauses seems to be more difficult in some countries (e.g. Poland) than in others (e.g. Ireland).

Last but not least, the discussions within the panel illustrated another important aspect of achieving high quality MT outputs. As explained by Janis Ziedins, in Latvia a memorandum was recently issued by the government demanding that all translations should be made centrally available. Therefore, the sharing of LR on the national level appeared to be a prerequisite to ensure the availability of language resources for MT training – but also as a translator's aid in general (see below, section 3.11, for further discussion points regarding the sharing of LR).

3.7 Using language resources in CEF eTranslation

In the presentation of Markus Foti (Programme Manager of eTranslation/MT@EC of the European Commission's DGT), it was demonstrated that the main goal of eTranslation is to make it usable in different public service scenarios. This includes in particular the building of more domain-specific engines as well as regular updates and incremental training).

With regard to the addition and use of language resources within and for eTranslation, one question was whether DGT creates a separate engine or whether they simply add data to the data that already exists. It was explained that this depends on the situation. For under-resourced languages such as Irish, new data is simply added (because of the existing lack of Irish LR in general), whereas in other cases, where there is sufficient baseline data available, separate domain-specific engines can be built and data will only be added to the relevant engine.

Moreover, it was stressed that eTranslation will move from the existing infrastructure within DIGIT to a cloud-based infrastructure. The reason for this is that DIGIT's infrastructure (hardware) does not offer the hardware required for neural machine translation.

Another interesting question was whether the same training data is actually used for MT@EC (statistical machine translation system) and for eTranslation (neural machine translation system). Markus Foti explained that currently, eTranslation is only trained on a subset of all training data, but this relates to security reasons (i.e. that some data cannot yet be used within a cloud infrastructure).

Regarding the incremental training of the MT system (for instance in analogy to Google or to the ERICS system) that has been planned, Markus Foti pointed out that DGT had only been experimenting with incremental training so far, because for statistical machine translation systems (like MT@EC), the process would turn out to be too time-consuming. However, with eTranslation (neural machine translation), incremental training will be made possible, as neural machine translation is based on an iterative approach.

With regard to the incorporation of direct feedback or corrections in eTranslation / MT@EC, Markus Foti confirmed that currently, there is no system for incorporating

such feedback. There have only been different plug-ins within SDL Trados Studio ensuring that a correction which has been made once will actually carry through the whole document. However, this correction would not change the MT engine unless it is made repeatedly by other translators as well and fed back into the Euramis database. In addition, DGT is currently able to harvest and parse the xliff-file (the file which is generated by SDL Trados Studio in addition to the tmx-file) in order to look for common corrections.

3.8 Inter-ministerial collaboration in Finland

As illustrated by Taru Virtanen from the Finnish Prime Minister's Office, the centralization of translation tasks in Finland has proved to be a good possibility to deal with the high increase in translation assignments (i.e. from 2.300 assignments in 2014 up to 17.000 assignments now per year although there was no additional personnel). The internal term bank (and central collection of tmx) proved to be very valuable in providing translation services of a higher quality to all ministries, as the data is now shared and not available in separate silos.

It was underlined that there is a clear goal and wish to provide more language resources on the national open data portal (opendata.fi). Glossaries are currently already made available there, but hopefully, more language resources will be provided by the Finish translation service in the future.

Taru Virtanen stressed that unfortunately, it will not be possible to share politically sensitive data outside the governmental institutions. However, there are already attempts of filtering out such data and providing more language resources to the outside as well. It was highlighted that ELRC can directly help cleaning and processing / preparing such data.

3.9 CEF Automated Translation for the EU Council Presidency

In his presentation of the EU Presidency Translator, its Project Manager Rihards Kalnins (Tilde), illustrated the development of the automated translation tool. Within the Latvian EU Council Presidency, Hugo.lv was adapted in terms of terminologies and also made available to all people attending the event through various tools (mobile app, translation kiosk). As such, the Council Presidency Translator serves various user groups including staff, journalists, delegates, and translators. For 2017/2018, eTranslation was integrated and combined with neural machine translation.

The first stop in the adaptation of the engines to Estonian, Bulgarian, and Austrian was and still remains the gathering of the language resources. The training of the system then takes several months. The training phase for Bulgaria is approaching right now.

Finland expressed their explicit interest in the use of this tool for the Finish EU Council Presidency in 2019. However, the key question is who would be in charge to grant them permission for the use and development of the EU Council Presidency Translator. The approving authority varies from country to country, for example the

Secretariat in the Ministry of Foreign Affairs for Latvia, the Prime Minister's Office in Estonia etc.

3.10 Best practices for managing your data

Concluding his presentation, Djamel Mostefa, R&D Manager of Systran, pointed out that the best practices for managing your data include, above all, the use of existing standards and classifications, a clear and reliable description of the data through the meta-data, the assurance of the quality of the data through a corresponding validation procedure, and ideally, the sharing of data either publicly or directly through dedicated channels (e.g. inter-ministerial collaboration, ELRC-SHARE) etc. In the case of legal questions, Djamel Mostefa once more pointed out that the ELRC Helpdesk can advise and provide support for negotiating the right user licences to protect the data according to the needs.

One important aspect of the presentation was the illustration of different legal frameworks for sharing language resources. The framework available in the US makes it easier to use data for the training of MT systems. Even any data or translated texts available on the Internet (no matter who is the creator of the content) can be used for research purposes / training MT.

In Europe, the sharing of LR is also made easier at least for public service administrations and for data generated by public services thanks to the Public Sector Information (PSI) Directive. Moreover, through its Helpdesk and its services, the ELRC can help with any cleaning and processing of language resources, if required, as well as with clearing any legal issues (see above). The existence of commercial networks for data sharing (e.g. TAUS) shows that even in private settings and industry, it is possible to share data.

3.11 Panel/Presentation: Sustainable LR supply across Europe

In the presentation preceding the panel, it was shown that currently, several key questions regarding the European Language Resource Coordination and the sharing of language resources across Europe will need to be addressed:

- On the **national level**, the key question is about the communication structures: We need to identify and implement the appropriate communication structures on national level with regard to the sharing of LR.
- Similarly, on **the institutional level** in each country, we need to identify corresponding institutional coordinators and ensure their commitment. Based on this, we can set up the corresponding institutional coordination and communication structures.
- Last but not least, and directly associated with the questions just posed, there is one open question on the European level, with regard to the ELRC: How can we ensure sustainability of the structures on all levels? What is the best way to be able to make progress / to be able to move forwards from here?

Following the question on his future vision of the ELRC and the sustainable network for the sharing of LR, Marco Marsella, Head of the Unit on Learing, Multilingualism

and Accessibility at DG CNECT, explained that the infrastructure is the key component to serve the purpose of the DSIs. This does not only include the technical infrastructure for automated translation (i.e. CEF eTranslation), but also the infrastructure to gather and manage the language resources required. According to Marco Marsella, the most important expectation is a difference in culture, i.e. the fostering of a common understanding that the sharing of language resources is valuable and can be associated with benefits. He also stressed that it is of utmost importance to build the framework on people who can engage other people, thus making the network sustainable. As such, the overall vision for the future is to one day deliver high quality translations to a vast majority of citizens and public services.

Building this network is not an easy task because of the greatly varying structures and infrastructures in each country. This was underlined not only by the panellists from national public service administrations (Taru Virtanen and Jeffrey Ganellen) but also by the coordinator of the European Language Resource Infrastructure (ELRI), Thierry Etchegoyhen. For instance, while in Finland all translation services have been centralized and while they store their language resources in a central manner, in Sweden it is not possible yet to even access a translation / tmx file outside the particular ministry, simply for security reasons (i.e. the intranet boundaries of the particular ministry). As such, it would be impossible to share data even among the ministries in Sweden. Therefore, the first step to be taken is to enable sharing of LR on the national level. This will be attempted through the ELRI project (see above, section 2.3.2 for details) in four EU Member States.

It was again stressed by Djamel Mostefa from Systran that it is often not necessary to have access to all possible kinds of data. Instead, it is important to find the relevant data in your particular domain to be able to build an MT engine which meets your requirements. The case of the World Intellectual Property Organisation WIPO (see above, section 3.4) confirmed this. So even within the ELRC, it is important to make data available (and retrievable) for relevant domains.

Fragmentation was seen as one of the main challenges that need to be overcome in the future. Fragmentation refers to the current situation where data is available across various different structures (e.g. TAUS network, ELDA catalogue, different national repositories etc.). Instead, the aim should be to have a common repository through which data across all categories can be shared. This, however, requires the necessary national and local structures (and approval processes) that allow to share data beyond the institutional or national infrastructures.

Such a new framework should therefore be built on the existing national structures, it should be created both bottom-up and top-down, involving key decision makers in each country since they are the only ones to approve such operations. Without their consent and support, activities for sharing language resources will not be sustainable and it is the objective of the ELRC to promote the establishment of the infrastructures necessary in each country.

4 Annex

4.1 Annex 1: Conference Programme

7th of November 2017: Arena of Opportunities

14:00 – 14:30: Opening of the conference and Welcoming address (Aleksandra Wesolowska – European Commission, Programme Officer, DG CNECT, Unit G3 – Learning, Multilingualism and Accessibility)

Welcome and introduction (Josef van Genabith – DFKI, ELRC Coordinator)

14:30 – 15:00: Opportunities at a glance (Andrea Lösch – DFKI, ELRC Project Manager)

14:00 – 17:00: Meet the experts and check out live solutions (ongoing), including

- MT@EC / eTranslation (Daniel Prou - Business Manager for Machine Translation, European Commission Directorate-General for Translation)

- Automated evaluation and revision with translationQ (Hendrik Kockaert – Associate Professor, RG Translation and Technology, KU Leuven, translationQ)

- Demonstration of Open Source Neural Machine Translation System and Pure Neural™ Machine Translation (Djamel Mostefa - R&D Project Manager, Systran SA)

    - Neural Machine Translation - Quality Translation 21 (Kenneth Heafield - Assistant Professor, University of Edinburgh)

    - Generating language resources with the ILSP Focused Crawler (Vassilis Papavassiliou, Research Associate, ILSP / R.C. Athena)

    - Anonymisation of Language Resources (Matthieu Camus, Privacy Impact; Pawel Kamocki, ELDA)

    - eTranslation Term Bank (Tatjana Gornostoja – Project Manager, Tilde)

    - iADAATPA - Secure and intelligent integration of eTranslation in public administrations (Manuel Herranz - CEO Pangeanic, iADAATPA Project Manager)

    - European Language Resources Infrastructure (ELRI) (Thierry Etchegoyhen – Principal Researcher, vicomtech)

    - Public Multilingual Knowledge Infrastructure (PMKI) (Najeh Hajlaoui –Publications Office of the European Union)

    - Kantan Neural (Tony O'Dowd – CEO KantanMT)

    - Cross-border eProcurement notifications (Michal Ohrablo – ANASOFT APR)

From 17:00   Cocktail reception and networking

8th of November 2017: Plenary Sessions

09:00 – 09:15: Welcome and introduction (Josef van Genabith – DFKI, ELRC Coordinator)

09:15 – 10:30: Multi-lingual needs of public services in Europe

- 09:15 – 09:40: eJustice: Translation needs and the value of language resources (Agnieszka Jelnicka – European Commission, DG Translation)

- 09:40 – 10:05: Multi-lingual requirements within the eProcurement DSI (Marc-Christopher Schmidt – European Commission, DG GROW)

- 10:05 – 10:30: Translation workflows in the European Refugee Information and Communication Service ERICS (Mika Nieminen - University Aalto, Finland)

10:30 – 11:00: Coffee break

11:00 – 12:30: Effective high-quality translation – new technologies for translators

- 11:00 – 11:30: Translation quality, neural machine translation and language resources (Kenneth Heafield, University of Edinburgh)

- 11:30 – 12:00: Machine Translation for the Irish Government: Achieving good translation results with MT (Teresa Lynn – Dublin City University, ADAPT Centre)

- 12:00 – 12:30: Panel: How to ensure maximum quality of translations? (Moderator: Rihards Kalniš, Tilde; Panelists:
    - Anna Kotarska – Senior Specialist, Department of Analysis and Strategy, National Health Fund, Poland
    - Markus Foti – Project Manager, European Commission, DG Translation
    - Jānis Ziediš – Project Manager, Culture Information Systems Centre, Latvia
    - Shanna Ní Rabhartaigh – Translator, Department of Culture, Heritage and the Gaeltacht, Ireland

12:30 – 13:30: Lunch break

13:30 – 15:00: Approaches for sustainable data sharing:

- 13:30 – 14:00: Using language resources in eTranslation (Markus Foti – Project Manager, DG Translation)

- 14:00 – 14:30 Case study: Inter-ministerial collaboration in Finland (Taru Virtanen – Head of Unit, Foreign Languages, Prime Minister's Office of Finland)

- 14:30 – 15:00: CEF Automated Translation for the EU Council Presidency (Rihards Kalnins – Project Manager, Tilde)

15:00 – 15:30: Coffee break

8th of November 2017: Plenary Sessions (cont.)

15:30 – 16:30: Outlook: Ensuring sustainable data supply across Europe

- 15:30 – 16:00: Best practices for managing your data (Djamel Mostefa – R&D Project Manager, Systran)

- 16:00 – 16:30: A sustainable supply pipeline for public services across Europe? (Josef van Genabith – DFKI, ELRC Coordinator)

- 16:30 – 17:00: Panel: Framework and conditions for ensuring sustainable data supply through and beyond ELRC (Moderator: Josef van Genabith, DFKI; Panelists:
  - Djamel Mostefa – R&D Project Manager, Systran
  - Marco Marsella – European Commission, DG CNECT, Head of Unit G3 – Learning, Multilingualism and Accessibility
  - Taru Virtanen – Head of Unit, Foreign Languages, Prime Minister's Office, Finland
  - Jeffrey Ganellen – Ministry for Foreign Affairs, Sweden

17:00 – 17:15: Summary and conclusions (Marco Marsella – European Commission, DG CNECT, Head of Unit G3 – Learning, Multilingualism and Accessibility; Josef van Genabith – DFKI, ELRC Coordinator)

From 17:15   Cocktail reception and networking

4.2 Annex 2: Conference Presentations

All conference presentation materials including videos are available through the ELRC website (http://www.lr-coordination.eu/ELRC-conference-2017). They include in particular:

1. Day 1 – Welcome and Introduction (Josef van Genabith)

2. Day 1 – Opportunities at a glance (Andrea Lösch)

3. Day 2 – Welcome and Introduction (Josef van Genabith)

4. Day 2 – eJustice: Translation needs and value of LR (Agnieszka Jelnicka)

5. Day 2 – Multi-lingual requirements within eProcurement (Marc-Christopher Schmidt)

6. Day 2 – Translation workflows in ERICS (Mika Nieminen)

7. Day 2 – Translation quality, neural machine translation and LR (Kenneth Heafield)

8. Day 2 – MT for the Irish Government: Achieving good translation results (Teresa Lynn)

9. Day 2 – Panel: How to ensure maximum quality of translations?

10. Day 2 – Using language resources in eTranslation (Markus Foti)

11. Day 2 – Case study: Inter-ministerial collaboration in Finland (Taru Virtanen)

12. Day 2 – CEF eTranslation for the EU Council Presidency (Rihards Kalnins)

13. Day 2 – Best practices for managing your data (Djamel Mostefa)

14. Day 2 – A sustainable LR supply pipeline for public services across Europe (Josef van Genabith)

15. Day 2 – Panel: Framework and conditions for ensuring sustainable supply of LR through and beyond ELRC

16. Day 2 – Summary and conclusions