5ᵀᴴ ELRC CONFERENCE WELCOME AND INTRODUCTION

Andrea Lösch (DFKI)

ELRC AT A GLANCE …

THE ELRC CONSORTIUM

STARTED
in April 2015

END ?

GOAL
To provide language resources for building better MT systems in CEF eTranslation to overcome language barriers in public services

LANGUAGE RESOURCES (LR) COLLECTED WITHIN ELRC

ELRC collects in particular the following types of language data:

- Corpora, i.e. a set of documents or a text in one or more languages:
  - official documents / translations in the official administration (decisions, legal acts etc.);
  - articles, reports, magazines, newsletters, etc.;
  - Translation Memories (i.e. aligned text segments in the source and target language)
- Language or translation models
- Lexical and Conceptual resources, e.g.
  - terminologies,
  - glossaries,
  - thesauri,
  - wordnets,
  - ontologies.

WHY ELRC...? WHY LANGUAGE DATA ...?

WHY LANGUAGE DATA...?

- **Slator Language Industry Market Report**[1]:
  - <u>global</u> language services and technology industry market: 23.2 billion USD in 2018
  - expected growth up to 28.2 billion USD by 2022 with Europe being among the key players
- **CEF Market Study on European LT-Market**[2]:
  - Expected growth rate for the <u>European</u> Language Technology (LT) market about 10% annual growth!
  - Especially Natural Language Understanding (NLU) (e.g. chatbot applications) as well as domain- and application-specific machine translation (MT) are expected to be of increasing importance
- **European Parliament resolution / Report on "Language Equality in the Digital Age"**[3]:
  - ➔ "make as a priority of language technology those Member States which are small in size and have their own language"
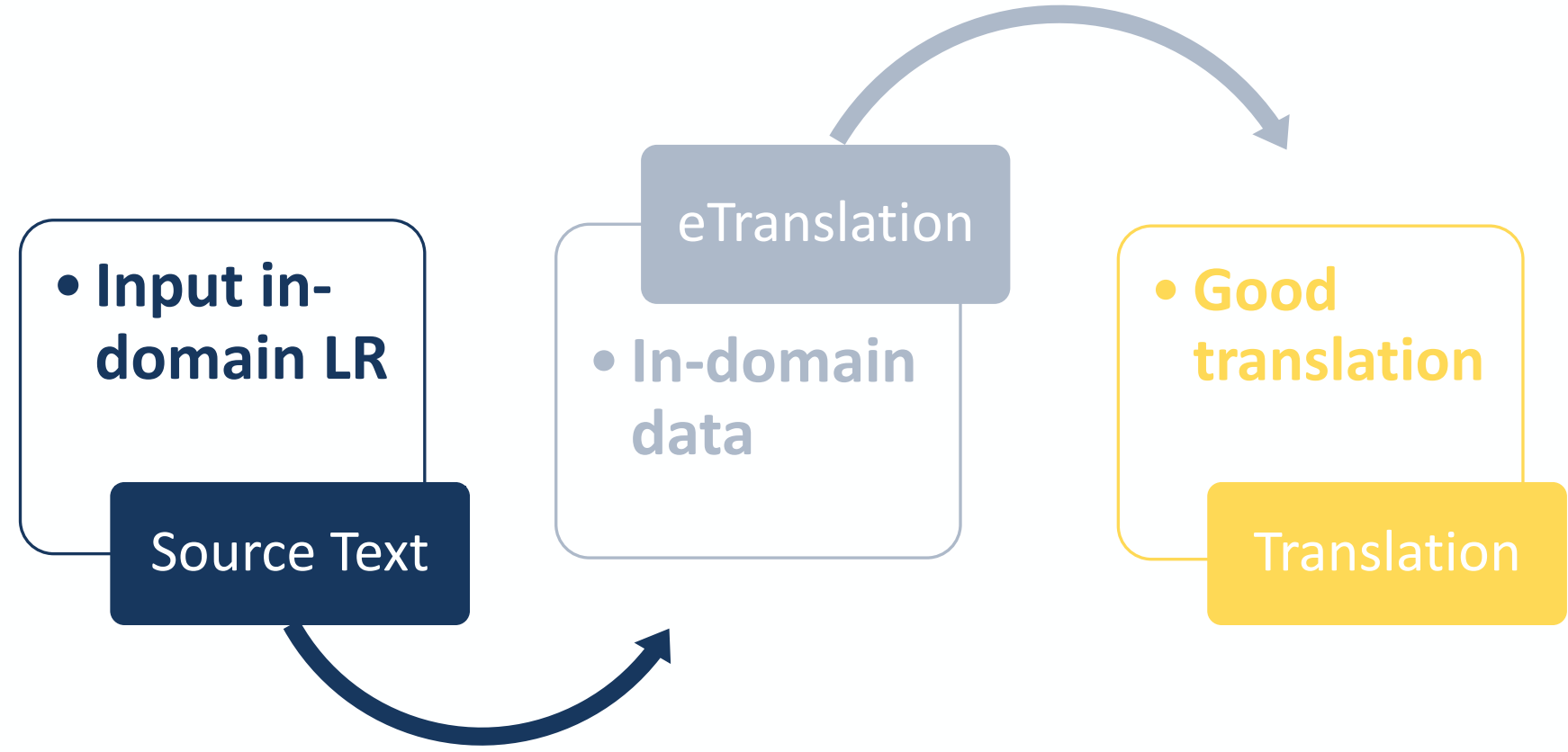
Why Language Data…?
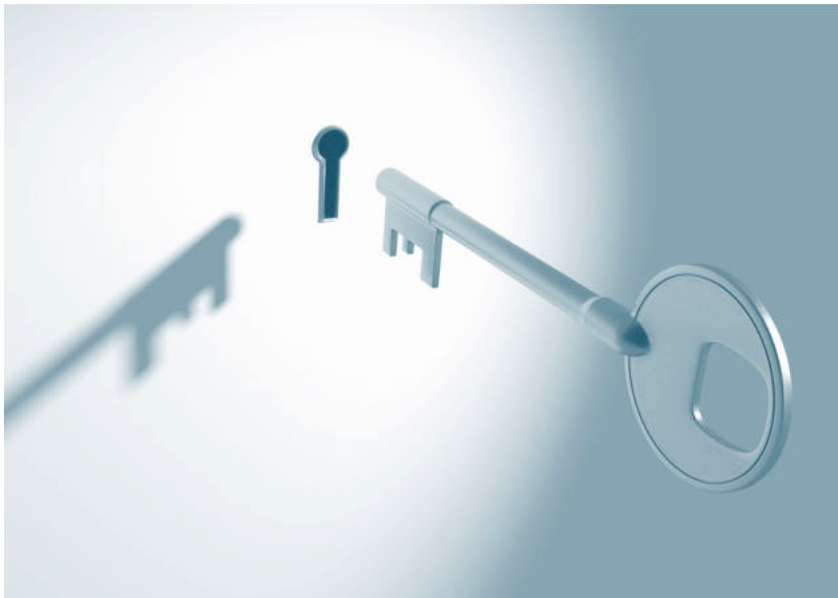
LT is a Key Market for Europe and Language Data is the Key for LT Development!

# WHAT DOES ELRC DO …?

WHY ELRC?

ELRC LR SUPPORT THE DEVELOPMENT OF CEF eTRANSLATION!

- <u>CEF eTranslation helps public services being multilingual:</u>
  - Neural machine translation tool provided by the European Commission to all EU bodies but also public services and public administrations across Europe
  - for 24 official languages of the EU (24 <-> 24), Chinese, Russian, Turkish, Arabic, …
  - Secure & scalable
  - Facts and figures:
    - 70+ systems connected to eTranslation
    - 7.000 SMEs registered
    - More than 40 Mio. pages translated in 2020
- <u>Additional upcoming services:</u>
  - Multilingual tweets
  - Speech-to-text
  - Anonymsiation etc.
  - ➔ **https://language-tools.ec.europa.eu/**
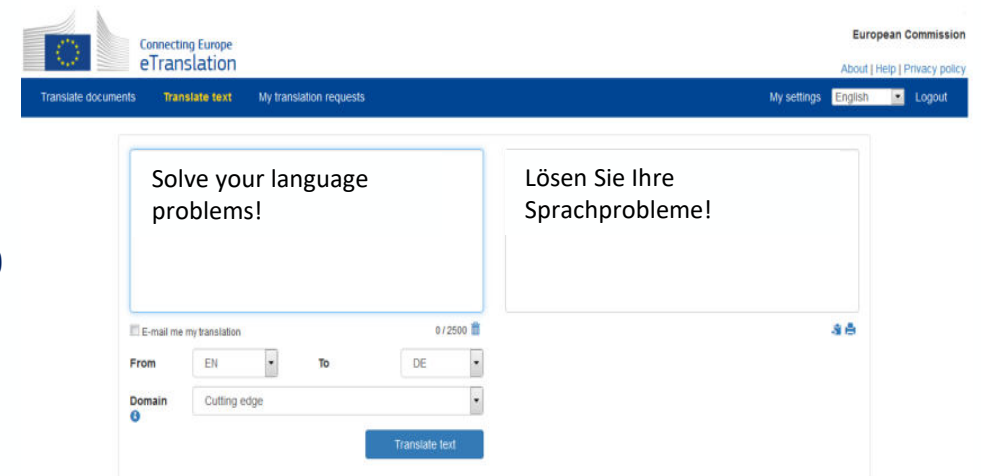
CEF eTranslation supports Public Services

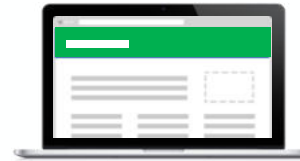- Examples of European Public Services: Digital Service Infrastructures (DSIs)

BRIS

Citizens and business partners need legal certainty when doing business cross-border

ODR

Citizens need to solve disputes **online across borders**

eHealth

Citizens need to have online access to their patient summary when abroad

EESSI

Citizens need to get to enjoy their social security seamlessly and online when abroad

It also includes

e-Justice
Open Data
eProcurement
Cybersecurity
Europeana
Disinformation
And many more...

ELRC OBJECTIVES

Identify	• The needs of public services
Engage	• With the public sector in the identification of language resources
Collect	• Language resources
Help	• With any technical or legal issues
Act	• Observatory for language resources across Europe

ELRC AT A GLANCE

National Anchor Points (NAP) in each country

www.lr-coordination.eu/anchor-points

CONFERENCE AGENDA

09:30 – 09:45 Welcome by the EC *(June Lowery-Kingston, Head of Unit "Accessibility, Multilingualism & Safer Internet", European Commission)*

09:45 – 10:00 Welcome and Introduction *(Andrea Lösch, ELRC Project Manager, DFKI)*

10:00 – 11:00 INSIDE ELRC AND DIGITAL EUROPE

10:00 – 10:15 ELRC Update *(Andrea Lösch, ELRC Project Manager, DFKI)*

10:15 – 10:30 European Language Industry Survey: Tech Tools – Initial Findings and Outlook *(John Anthony O'Shea, Legal Translation Practitioner, Board Member FIT Europe)*

10:30 – 11:00 Inside Digital Europe: LT Funding Opportunities *(Philippe Gelin, Head of Sector Multilingualism, European Commission)*

11:00 – 11:20 Coffee Break

11:20 – 12:20 SPOTLIGHT: QUALITY OF DATA

11:20 – 11:40 Assessing the quality of translations – a practical guide *(Renate Müller, Quality and Language Coordinator, DGT, European Commission)*

11:40 – 12:00 Using crawled data for MT development – promises and pitfalls *(Marcis Pinnis, Chief AI Officer, Tilde)*

12:00 – 12:20 Language data collection and LT development for low-resourced languages *(Andrew Bredenkamp, Chair, Translators without Borders)*

12:20 – 12:30 Wrap-up and outlook *(Andrea Lösch, ELRC Project Manager, DFKI)*

12:30 – 13:30 Lunch Break

13:30 – 14:30 RE-USING LANGUAGE DATA: TECHNICAL POSSIBILITIES AND LEGAL CONSTRAINTS

13:30 – 13:50 Language Data Validation and Curation in ELRC *(Victoria Arranz, Head of Language Resources Projects - R&D, ELDA and Mickaël Rigault, Junior Project Manager & Legal Counsel, ELDA)*

13:50 – 14:10 Anonymising language data within the CEF Data Marketplace *(Amir Kamran, Head of NLP, TAUS)*

14:10 – 14:30 How much anonymisation is needed – a legal perspective *(Andreas Sesing, Research Assistant, Manager of the Institute of Legal Informatics, Saarland University and Jonas Baumann, Research Associate, University of Johannesburg)*

**14:30 – 15:15 DATA CREATION AND SHARING IN CEF GENERIC SERVICES RESOURCE PROJECTS**

14:30 – 14:45 Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages *(Miquel Esplà Gomis, Research Assistant, University of Alicante)*

14:45 – 15:00 FedTerm: Federated eTranslation Termbank Network *(Gabriele Sauberer, Leader of Dissemination Work Package TermNet and Arturs Vasiljevskis, Head of MT solutions, Tilde)*

15:00 – 15:15 PRINCIPLE *(Jane Dunne, EU Research Project Coordinator, ADAPT Centre)*

15:15 – 15:30 Summary and Conclusions *(Andrea Lösch, DFKI)*

… because
#LanguageDataMatters
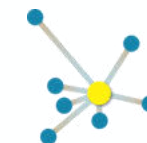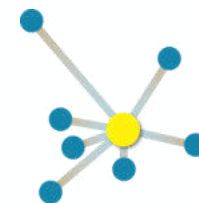
European Language Resource Coordination
Connecting Europe Facility

PRACTICAL HINTS & TIPPS

10 March 2021
9:30 am – 3:30 pm CET
via Zoom

FIFTH ELRC
CONFERENCE

- Please keep your video switched off.

- Please keep your audio switched off.

- Please submit all questions through the chat.

- Share your opinion about the conference in our online survey!

Thank you for your cooperation!