

WORKSHOP CONTENTS

MARIA GIAGKOU, STELIOS PIPERIDIS (ILSP / ATHENA R.C.)

Virtual LRB meeting, 29.04.2020



MASTER WORKSHOP AGENDA

08:00 – 09:00 Registration

09:00 – 09:10 Welcome and introduction

09:10 – 09:35 The potential of Language Technology and AI – where we are, where we should be heading

09:35 – 10:20 Language Technologies in <Country> / for <Language> - Panel session

10:20 – 10:50 *Coffee Break*

10:50 – 11:15 The CEF AT Platform

11:15 – 12:00 Language technologies by/for the public sector - Panel session

12:00 – 12:30 LT requirements and offerings: do they converge? – Q/A

12:30 – 13:30 *Lunch Break*

13:30 – 13:45 The value of data for the development of LT

13.45 – 14:30 Language data creation, management and sharing: existing practices and challenges - Panel session

14:30 – 14:35 Conclusions

14:35 – 15:30 Language technologies demonstrations and networking

MASTER PRESENTATIONS



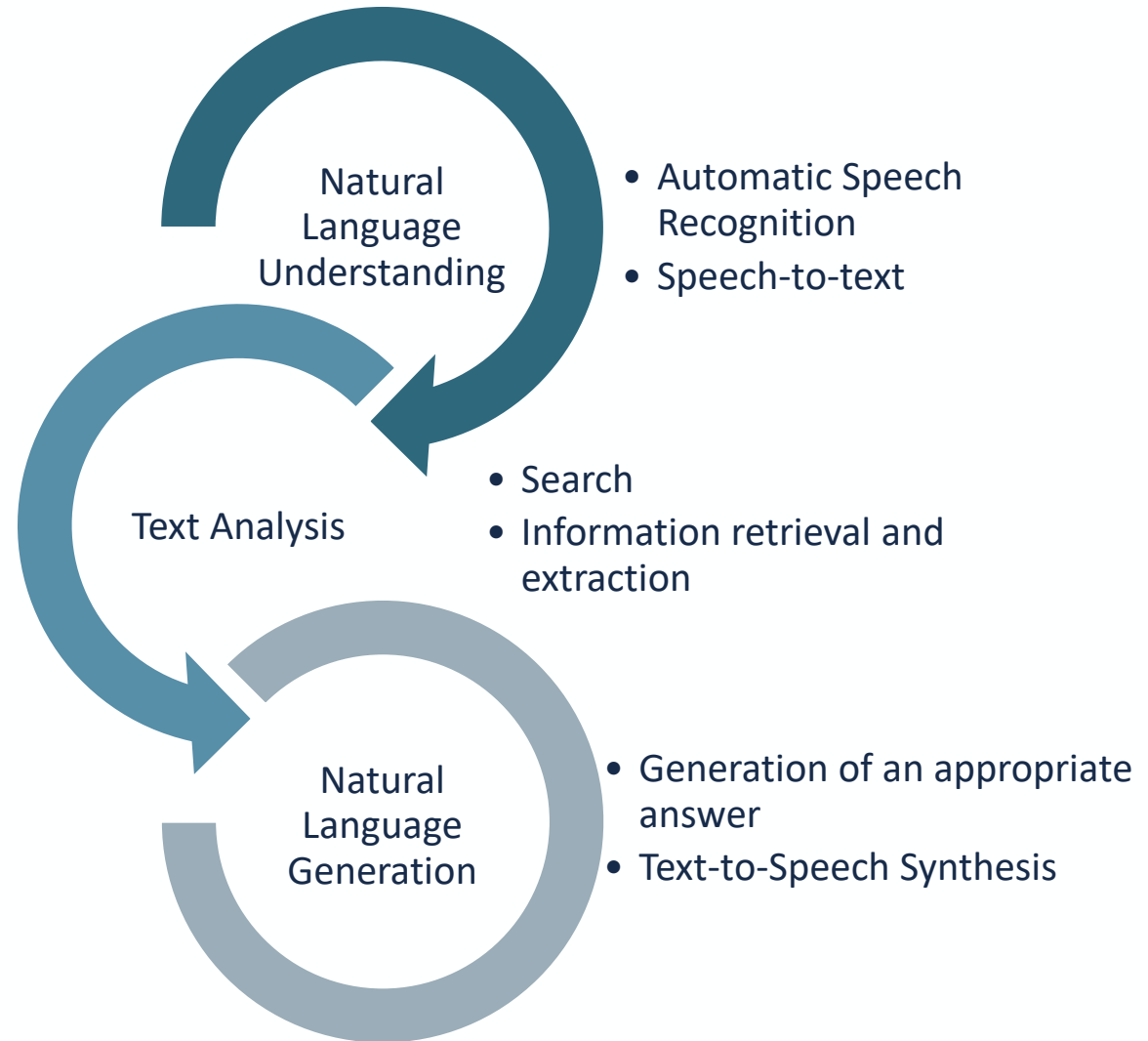
**S1. Welcome and
introduction**

- **Speaker:** local organiser
- **Est. duration:** 10 min.
- **Contents:**
 - Topic of the workshop
 - Context and background of ELRC
 - Program and practical notes



Hey there!
What's the
weather going
to be like today?

Language Technologies behind the scenes



Indicative slide
from S1 master
presentation

Connecting Europe Facility (CEF)

EU funding
instrument

trans-European
infrastructures

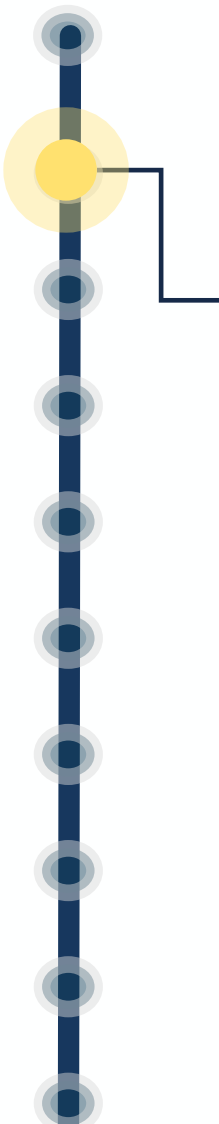
CEF Telecom

pan-European
digital services

Automated Translation (CEF.AT) /
eTranslation

European Language Resource
Coordination (ELRC)

Indicative slide
from S1 master
presentation



S2. The potential of Language Technology and AI – where we are, where we should be heading

- **Speaker:** Distinguished academic/researcher in the field of AI / NLP
- **Est. duration:** 25 min.
- **Contents:**
 - AI represents one of the greatest opportunities for global societal and economic progress
 - Zoom into Language Processing; how machines learn to process human languages
 - Status quo: Depending on the country, different sources of information can be referenced, for instance one or more countries' digital agendas/AI plans, the European Commission's National Interoperability Framework Observatory (NIFO), the Digital Economy and Society Index, the OECD Observatory of Public Sector Innovation

Indicative slide
from S2 master
presentation

WHAT IS AI? LANGUAGE AND INTELLIGENCE

- “If a conversation with the device cannot be differentiated from a similar conversation with a human being then the device can be called intelligent” (Alan Turing)
- Problem: Human languages are incredibly elegant, efficient, flexible, complex!
 - One word may mean many things
 - Many ways of saying the same thing
 - Meaning depends on context
 - Literal and figurative language (metaphor)

Indicative slide
from S2 master
presentation

DEEP LEARNING IN LANGUAGE TECHNOLOGIES

Deep learning applies to many language technologies, e.g.

- Acoustic Speech Recognition (ASR)
- Text to Speech Systems (TTS)
- Dialog Systems / Chat Bots
- Question-Answering Systems (QA)
- Named Entity Recognition (NER)
- Relation Extraction (RE)
- Text Summarisation
- Sentiment Analysis

Indicative slide
from S2 master
presentation

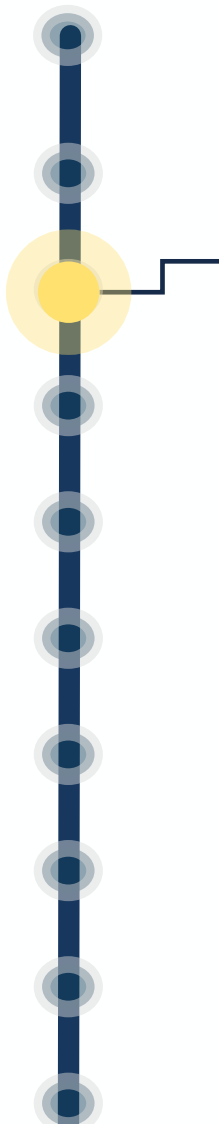
AI: WHERE WE SHOULD BE HEADING

Example: AI for Humanity (France)

AI FOR

HUMANITY

- **Developing an aggressive data policy**
 - ➔ Encourage companies to pool and share their data
 - ➔ Create data that is in the public interests
 - ➔ Support the right to data portability



**S3. Language
Technologies in
<Country> / for
<Language> (Panel
session)**

- **Panellists:** ≥ 3 local LT industry representatives
- **Moderator:** ELRC T-NAP or ELG national representative or Researcher active in the field of LT for the local language
- **Est. duration:** 45 min.
- **Contents:**
 - LT supply side at the national level
 - Introduction by moderator: technological readiness of local language, the risk of digital extinction

Indicative slide
from S3 master
presentation

Will the Internet Make Most Languages Go Extinct?

By Alex Mayyasi



Like 6



If you ever find yourself at an Internet cafe in the Middle East, you may be surprised to find

Many European languages in danger of digital extinction



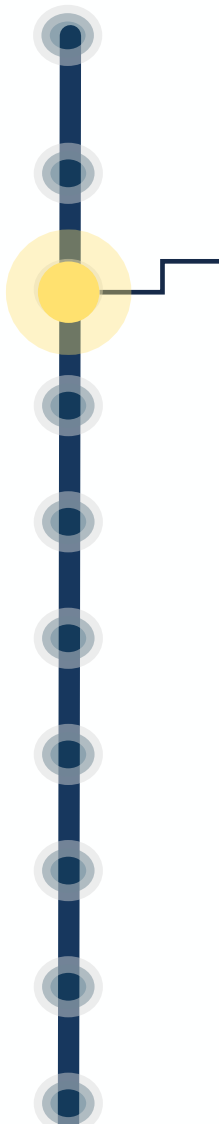
The alphabets at risk of extinction

LANGUAGE



It isn't just languages that are endangered: dozens of alphabets around the world are at risk. And they could have even more to tell us.

Linguists predict that out of the 6000+ languages spoken today, 50 to 90 per cent face extinction this century



**S3. Language
Technologies in
<Country> / for
<Language> (Panel
session)**

- **Contents (continued):**

- Panellists discuss the exploitation potential of LT, use (especially by the public administration), needs and future plans for deployments
- The moderator presents the CEF AT tools and services catalogue (<https://cef-at-service-catalogue.eu/>) as a source for further information on available language technologies for the local language

<https://cef-at-service-catalogue.eu/>

Indicative slide
from S3 master
presentation

eTranslation services Home Browse Services Manage Services Administration Account Logout About

Results 68 (1/4)

Filter By:

Function

- Translation (68)
- Language Technologies (7)
- Data And Data Management (1)

Task

- Machine Translation (68)
- Computer-aided Translation (5)
- Speech Recognition (5)

MORE

Language Coverage

- English (39)
- French (11)
- German (28)

MORE

Service Ownership Status

- External (68)
- Managed Internally By Ec (0)

SHOW ADVANCED FILTERS

Clear Filters

Vilnius University machine translation
Function: Translation
Task: Term Extraction, Machine Translation
Provided by: Vilnius University, TILDE

DeepL Translator
Function: Translation
Task: Machine Translation
Provided by: DeepL GmbH

Translexy: The TraMOOC platform
Function: Translation
Task: Machine Translation
Provided by: TraMOOC consortium

Personal Translator 20
Function: Translation
Task: Machine Translation
Provided by: Linguattec

ESTeam Translator
Function: Translation
Task: Machine Translation
Provided by: ESTeam AB

AMPLEXOR MachineTranslationExpert
Function: Translation
Task: Machine Translation
Provided by: AMPLEXOR International S.A.

Tildes Birojs
Function: Language Technologies
Task: Speech Recognition, Text To Speech Synthesis, Machine Translation, spelling and grammar checker, optical character recognition
Provided by: TILDE

CE5: Translation and synthesis of



S4. The CEF AT platform

- **Speaker:** EC representative, if possible, or ELRC consortium representative
- **Est. duration:** 25 min.
- **Contents:**
 - target audiences
 - use scenarios
 - languages coverage
 - translation output quality with examples
 - how to register
 - hands-on eTranslation web user interface

Indicative slide
from S4 master
presentation



I'm a SME. Is eTranslation for me?

YES!

eTranslation is open to European SMEs since March 2020
Specific conditions and limitations apply.

A SECOND EXAMPLE: eTRANSLATION (EU DOMAIN ENGINE)

Indicative slide
from S4 master
presentation

EN

Come gather round people
wherever you **roam**, and admit
that the waters around you have
grown.

The Times They Are A-Changin', Bob Dylan

FR

Venez rassembler des
personnes **où que vous**
utilisez l'itinérance aux tarifs
nationaux, et admettre que
les eaux autour de vous ont
grandi.

A SECOND EXAMPLE: eTRANSLATION (GENERAL TEXT DOMAIN ENGINE)

Indicative slide
from S4 master
presentation

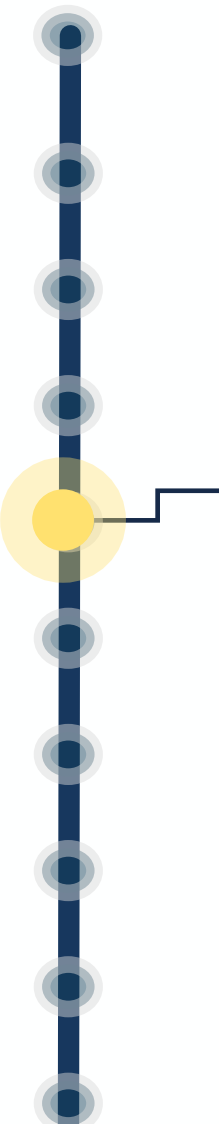
EN

Come gather round people
wherever you **roam**, and admit
that the waters around you have
grown.

The Times They Are A-Changin', Bob Dylan

FR

Venez rassembler les gens
partout où vous errez, et
admettez que les eaux autour
de vous ont grandi.



S5. Language technologies by/for the public sector (Panel session)

- **Panellists:** ≥ 3 local public sector representatives
- **Moderator:** ELRC P-NAP
- **Est. duration:** 45 min.
- **Contents:**
 - The LT demand side, especially demand by the public sector
 - Introduction by moderator on the roles of the governments in the development of LT; examples of LT-based digital services by/for the European public sector.
 - Discussion with panellists: how LT can address their needs; what are their requirements

ROLES OF GOVERNMENTS WITH REGARD TO ARTIFICIAL INTELLIGENCE (AI) AND LANGUAGE TECHNOLOGY (LT)

Indicative slide
from S5 master
presentation

- financier or direct investor
- regulator
- convener and standards-setter
- data steward
- smart buyer and co-developer
- user and service provider


Indicative slide
from S5 master
presentation

EXAMPLES OF LT-BASED SERVICES IN EUROPE


Maahanmuuttovirasto
Migrationsverket
Finnish Immigration Service

Starting up smoothly


chatbot service for foreign entrepreneurs in Finland



Kamu
helps you with residence permits to live in Finland
[TALK TO KAMU](#)




PatRek
advises you how to set up your company in Finland
[TALK TO PATREK](#)



VeroBot
tells you about business and work-related taxes in Finland
[TALK TO VEROBOT](#)

Kamu chatbot ? X




Heil I'm Kamu, your robot assistant. You can ask me about immigration or citizenship.

For security reasons, don't use any personal details such as your personal identity code in the chat. We save these conversations to help you better in the future.

How can I help you?

Write your message here. 0 / 110





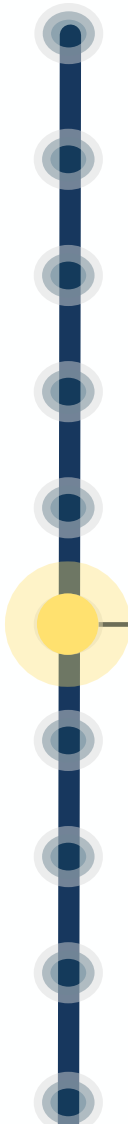


Latvijas Republikas Uzņēmumu reģistrs

I'm UNA. Don't hesitate to ask - I'll help you!








**S6. LT
requirements and
offerings: do they
converge? (Q/A
session)**

- **Moderator:** ELRC P-NAP or T-NAP
- **Est. duration:** 30 min.
- **Contents:**
 - Questions from audience
 - Can be skipped if many questions have already risen during the previous sessions
 - The participants will be able to use an event app to raise questions
 - Indicative conversation triggers are provided in the master presentation and in the guidelines for NAPs



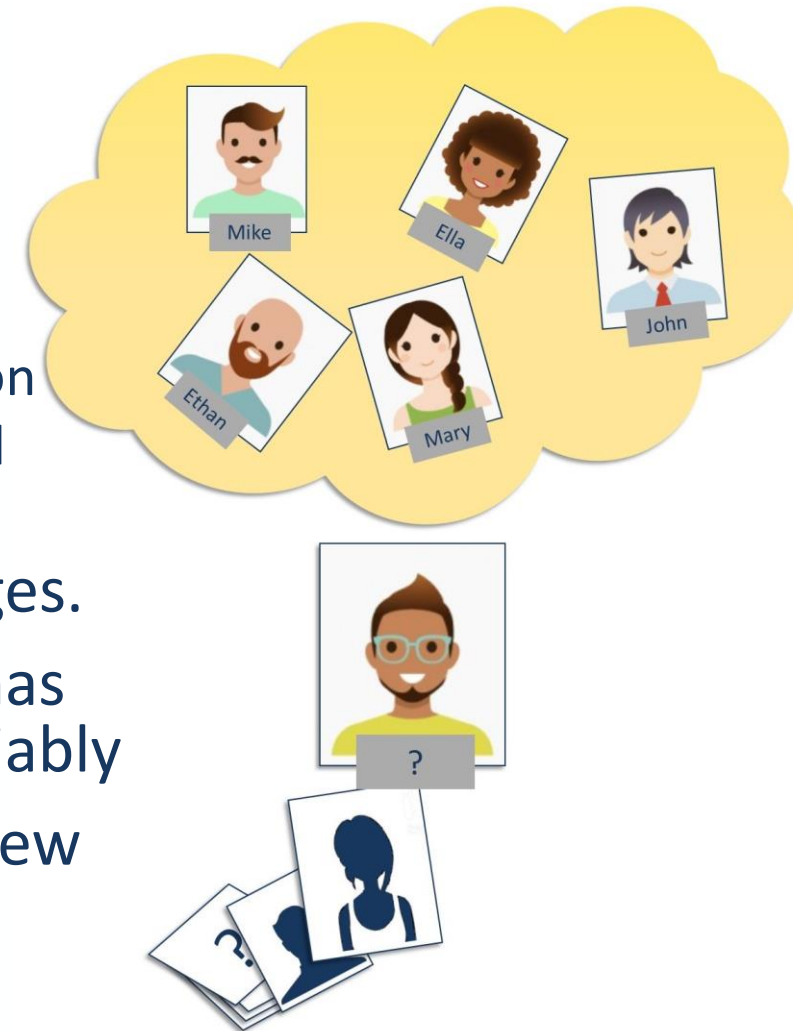
**S7. The value of
data for the
development of
top quality LT**

- **Speaker:** A distinguished local researcher in the field of language resources (e.g. the national language corpus) or a representative of the national Open Data Portal
- **Est. duration:** 15 min.
- **Contents:**
 - link the preceding presentations and discussions on LT to the need for more and better language data
 - how AI applications, and especially LT and machine translation systems, are trained with data
 - initiatives that address the growing need for language data

Indicative slide
from S7 master
presentation

WHY DOES DATA HELP?


- Imagine we have a bunch of corresponding face images
 - All of these images belong to the same person
 - We call this „starting point“ of our work Gold Standard
- The task is now to classify new face images.
- From the Gold Standard, the computer has learned how to compare face images reliably
- As such, it is able to deal properly with new face images



Indicative slide
from S7 master
presentation

HOW MUCH DATA IS ENOUGH DATA?


- Hard to tell. We try it out and train with what is there.
- We evaluate the model. There are various ways.
- One is a similarity measure that indicates closeness to the Gold Standard
- We find a bad translation. Is 85% good enough?
- No absolute value, but relative to other systems
- Measuring own improvements
- You have enough data once the results are satisfying



S8. Language data creation, management and sharing: existing practices and challenges (Panel session)

- **Panellists:** 3-4 representatives of existing infrastructures (e.g. National open data portal, Clarin, ELG etc.) as well as other key national stakeholders as identified in the ELRC European Countries Profile
- **Moderator:** Speaker of S7 or ELRC consortium representative
- **Est. duration:** 45 min.
- **Contents:**
 - existing infrastructures and operations for creating, managing and sharing language data at the national level

- **Speaker:** Local organiser
- **Est. duration:** 5 min.
- **Contents:**
 - overview of the topics discussed



**S9. Take-home
message and
conclusions**

Indicative slide
from S9 master
presentation

TOPICS RAISED

- The potential of LT to transform governance, administration, commerce and society
- The technological readiness of our language in the AI era
- Our country's position with regard to the development and adoption of LT
- Developments at the EU level towards a wider development and adoption of LT

Indicative slide
from S9 master
presentation

LESSONS LEARNT

- Speakers of less spoken languages are facing a digital timebomb
- The LT research and industry can facilitate digital interaction in our multilingual Europe
- LT and AI need more data!
- Share your language data with Europe!

- **Participants:** Local LT industry representatives
- **Est. duration:** ~60 min.
- **Contents:**
 - Demonstrations of LT for the local language, informal setting during coffee and networking



**S10. Demos
session**