

## Deliverable D3.2.27 Task 3

# ELRC Workshop Report for Denmark



<b>Author(s):</b>	Carl Frederik Bach Kirchmeier (Agency for Digital Government), Kristīne Vīgante (Tilde), Aivars Bērziņš (Tilde)
<b>Dissemination Level:</b>	Public
<b>Version No.:</b>	V3.0
<b>Date:</b>	2023-01-19



## Contents

<b><u>1</u></b>	<b><u>Executive Summary</u></b>	<b><u>3</u></b>
<b><u>2</u></b>	<b><u>Workshop Agenda</u></b>	<b><u>4</u></b>
2.1.1	ELRC WORKSHOP	5
2.1.2	ELG WORKSHOP	5
<b><u>3</u></b>	<b><u>Summary of Content of Sessions</u></b>	<b><u>6</u></b>
<b>3.1</b>	<b>Session 1: “Welcome and introduction”</b>	<b>6</b>
<b>3.2</b>	<b>Session 2: “On Danish chat bots and a Danish open source initiative”</b>	<b>6</b>
<b>3.3</b>	<b>Session 3: “Natural language processing: Recent advances and challenges”</b>	<b>7</b>
<b>3.4</b>	<b>Session 4: “What eye movements can teach us about language processing”</b>	<b>7</b>
<b>3.5</b>	<b>Session 5: “Speech Recognition for Digital Inclusion and Support”</b>	<b>7</b>
<b>3.6</b>	<b>Session 6: “Sprogteknologi.dk – A Common Danish Language Resource”</b>	<b>8</b>
<b>3.7</b>	<b>Session 7: “COR – Central Word Register for Danish for Language Technology purposes”</b>	<b>8</b>
<b>3.8</b>	<b>Session 8: “eTranslation”</b>	<b>8</b>
<b>3.9</b>	<b>Session 9: “European Language Resource Coordination (ELRC) and beyond”</b>	<b>9</b>
<b><u>4</u></b>	<b><u>Country Profile: Language data creation, management and sharing</u></b>	<b><u>10</u></b>
<b><u>5</u></b>	<b><u>Workshop Participants</u></b>	<b><u>11</u></b>
<b><u>6</u></b>	<b><u>Workshop Evaluation – Participant Feedback Analysis</u></b>	<b><u>Error! Bookmark not defined.</u></b>
<b><u>7</u></b>	<b><u>Appendix</u></b>	<b><u>12</u></b>
<b>7.1</b>	<b>Localised agenda</b>	<b>12</b>
<b>7.2</b>	<b>Localised press release</b>	<b>12</b>
<b>7.3</b>	<b>Workshop Presentations</b>	<b>13</b>

## 1 Executive Summary

The third ELRC workshop in Denmark took place on 16 November 2021 at 13.20-15.00. The ELRC workshop was part of the broader programme of the day. The Centre for Language Technology at the Department of Nordic Studies and Linguistics, UCPH, and the Danish Agency for Digitisation, hosted the conference “Language Technology conference 2021” with a focus on the language technology developments and status in Denmark and in the European Union. It was decided by the organisers to schedule the workshop within a larger conference, since both organisations had planned a few language technology conferences/seminars etc. soon. Both parties, therefore, agreed that it would be better to hold one larger event, since the community for language technology in Denmark isn’t big enough to host multiple events in a short span of time.

The morning session’s focus was on research and development in Denmark; the afternoon part was dedicated to two workshops on the European Language Resource Coordination (ELRC) with a focus on EU machine translation (eTranslation), and the European Language Grid (ELG), which is developing a pan-European language technology portal.

## 2 Workshop Agenda

The final workshop programme in English

### **Welcome and introduction**

09:00 – 09:10 Bolette S. Pedersen, Centre for Language Technology, UCPH, and Christian Plaschke, The Danish Agency for Digitisation

### **On Danish chat bots and a Danish open source initiative**

09:10 – 09:40 Henrik Fabrin, Certainly

### Presentation

### **Natural language processing: Recent advances and challenges**

09:40 – 10:10 Barbara Planck, IT University of Copenhagen

### Presentation

### **What eye movements can teach us about language processing**

10:00 – 10:20 Nora Hollenstein, Centre for Language Technology, University of Copenhagen

### Presentation

10:20 – 10:30 **Discussion/questions: Speakers from the first morning session in the panel**

10:30 – 11:00 *Coffee Break*

### **Speech Recognition for Digital Inclusion and Support**

11:00 – 11:30 Klaus Bjørn Larsen, Roskilde Kommune

### Presentation

### **Sprogteknologi.dk – A Common Danish Language Resource**

11:30 – 11:50 Danish Agency for Digitisation

### Presentation

### **COR – Central Word Register for Danish for Language Technology purposes**

11:50 – 12:10 Peter Juel-Henrichsen, The Danish Language Council  
Sanni Nimb, Det Danske Sprog- og Litteraturselskab  
Bolette Sandford Pedersen, Centre for Language Technology University of Copenhagen

### Presentation pt. 1 Presentation pt. 2

12:10 – 12:20 **Discussion/questions. Speakers from the second morning session in the panel**

12:20 – 13:20 *Lunch Break*

## 2.1.1 ELRC WORKSHOP

### **eTranslation**

13:20 – 13:45 Claus Thornby Larsen, the European Commission

### Presentation

13:45 – 14:10 **Update from the ELRC and new tools on the CEF platform**  
Thierry Declerck, European Language Resource Coordination

14:10 – 14:30 Panel on ELRC and eTranslation seen in the Danish Context

14:30 – 15:00 **Break**

## 2.1.2 ELG WORKSHOP

15:00 – 15:30 **The European Language Grid and European Language Equality  
Projects**  
Georg Rehm, German Research Center for Artificial Intelligence, Germany

15:30 – 15:50 **Panel on ELG and Synergies with the Danish LT community as  
presented in the morning sessions**

15:50 – 16:00 **Conclusions**

### 3 Summary of Content of Sessions

The presentations at the conference were given by 9 different speakers, representing researchers, public and private organisations. The first 6 presentations were focused on development, outlook and use-cases of Danish language technology. It was discussed how Danish language technology could help businesses, what the major obstacles for further developments are, and how to improve the development and use of Danish language technology.

The last 3 presentations were given during the ELG and ELRC workshops.

Claus Thornby Larsen from the European Commission presented eTranslation, with a focus on the possibilities and current limitations of the tool.

This was followed by a presentation from Thierry Declerck (representing the ELRC) with the title “Update on ELRC and new tools on the CEF platform”. It focused on the importance of gathering language resources, so that tools, such as eTranslation, could be further improved.

In the Panel discussion, the participants Thierry Declerck, Thornby Larsen and Sabine Kirchmeier from EFNIL discussed how European politics could benefit from Danish language technology and how Denmark could further support the initiatives taken by the EU.

Lastly, Georg Rehm held a presentation on the ELG (European Language Grid) platform. It was discussed how the ELG has contributed to the European LT thus far and its effects on the upcoming Language Data Space and the future of AI-related solutions in LT.

#### 3.1 Session 1: “Welcome and introduction”

Bolette S. Pedersen, (Centre for Language Technology) and Christian Plaschke (The Danish Agency for Digitisation), the local ELRC representatives, opened the event by welcoming the audience and introducing the key persons in conceiving and organising the event, namely the ELRC consortium and the EC/DGT representatives.

#### 3.2 Session 2: “On Danish chat bots and a Danish open source initiative”

This session was presented by Henrik Fabrin, CEO and Co-founder of Certainly. The presentation was launched with an introduction to the company. He briefly described Certainly, explaining that it is a Conversational AI Platform made for e-commerce and customer service and was first developed 10 years ago using open-source technologies. A few sentences were also devoted to the history of the company, goals, vision and mission, and main export markets as well aspects of e-commerce and future developments. Henrik Fabrin reported on the growth in conversations within the platform and overall usage of chatbots from January 2019 to October 2021, as well as e-commerce all together.

Henrik Fabrin explained that their choice was to use open source, mainly because it is a perfect alignment between commercial strategy and social responsibility. He mentioned that the company contributed to the open-source community by opening four language models (Danish,

## ELRC Workshop Report for Denmark

Swedish, Norwegian and Finnish), datasets of questions for Coronavirus in English and Danish as well as Unity test libraries. All listed technologies are part of their chatbots. He emphasised the benefits for companies, and that the use of the platform increases the average order value by 20 percent. The presentation was rounded up with conclusions about the changes that e-commerce has experienced over recent years.

### 3.3 Session 3: “Natural language processing: Recent advances and challenges”

Professor Barbara Plank (ITU, Copenhagen) began her presentation by raising what she sees as the big goal of Natural Language Processing, that is for AI to understand and produce language, just as humans do. She briefly described the Sources of Bias and how they are not only in the data. She further presented a short historical insight and future growth prospects for natural language processing (NLP). The market size is expected to triple from USD 11.6 billion in 2020 to USD 35.1 billion by 2026. She also briefly touched on the development of language models and the importance of data to sustain the development of large language models.

Barbara Plank introduced various NLP methods for Danish: Dependency Parsing, Coreference Resolution, Nested Named Entities, Lexical Normalisation and provided some examples for these methods. She also presented results for Nested NER: MTL versus STL. The main conclusion of the presentation is that NLP has grown tremendously, biases are everywhere, and the direction should be towards more inclusive & robust NLP.

### 3.4 Session 4: “What eye movements can teach us about language processing”

Nora Hollenstein (Centre for Language Technology, University of Copenhagen) briefly presented the basics of eye-tracking technology, introducing features such as: First fixation duration, Total reading time, Landing position and Mean fixation duration. She continued with a brief overview of Psycholinguistics.

The speaker then elaborated on the potential of eye-tracking in NLP, and how NLP can be improved on various levels with eye-tracking data. She emphasised that eye-tracking data reflects human text processing and forms Multi-modal NLP models; the compiled data provides modest but consistent improvements on various levels of text processing.

She closed her presentation with two statements:

- Eye-tracking provides online metrics of language comprehension on multiple levels of analysis.
- With eye-tracking technology, a better understanding of the similarities and differences between human and machine language processing is acquired.

### 3.5 Session 5: “Speech Recognition for Digital Inclusion and Support”

Klaus Bjørn Larsen gave a very informative presentation about a project on intelligent speech recognition and the benefits for the public and businesses. Opening his presentation with the project purpose and benefits, he introduced the project’s focus groups: 18-29 year-old, Elderly 70+, Disabled. He also briefly touched on the usage scenarios and explained why the public

## ELRC Workshop Report for Denmark

services area is good for data collecting and testing. Considering the fact that one of the most important aspects is data protection, he emphasised that the use of the chat bot is non-binding, no personal data is retained, and contact is anonymous.

Klaus Bjørn Larsen concluded his presentation with the statement that business benefits from voice-bots in service availability and compliance (distinctly standardised responses), and that full automation of speech recognition is the new goal.

### 3.6 Session 6: “Sprogteknologi.dk – A Common Danish Language Resource”

This session was presented by Christian Plaschke (Danish Agency for Digitisation). He began his presentation by introducing what the Sprogteknologi.dk initiative is and its purpose – to strengthen the development of solutions for the Danish language using AI based technologies. He then continued by describing why this initiative is important to language technology companies as well as to end users, whether they be public administrations, researchers, companies or citizens. He concluded his presentation with a recommendation on how to ensure that the Danish language thrives in the digital age: share Danish language resources, Open-source as much as possible and think about the future collection of language resources.

### 3.7 Session 7: “COR – Central Word Register for Danish for Language Technology purposes”

Bolette Sandford Pedersen presented COR – Central Word Register, and explained how Danish companies now cooperate to a large extent to create Danish-orientated AI. She explained and gave examples of why COR must facilitate language-centred AI systems for Danish and is therefore based on locally rooted knowledge of the Danish language and society rather than simply adapting applications from English. She continued by explaining how COR develops basic guidelines for the amalgamation of meanings from the Danish Dictionary and deletion of obsolete or overly specialised meanings.

### 3.8 Session 8: “eTranslation”

The presentation began by introducing the Directorate-General for Translation in numbers and facts. Claus Thornby Larsen (Language Technology Coordinator Directorate-General for Translation) compared machine translation from the past to nowadays. He continued his presentation with a deeper introduction into eTranslation, emphasising the main disadvantages of machine translation, such as inconsistent terminology, omissions or additions, providing several examples. He gave some examples of how MT mistakes can be politically sensitive. He emphasised that MT quality depends on the available data and data consistency. He continued by explaining how each language pair is trained and what amount time is necessary to do so. He gave as an example of fields such as legislation, where a legal act is often composed of bits and pieces from other legal acts and defined terms must be used consistently.

He concluded his presentation with the statement that AI is of great help to professionals.

### 3.9 Session 9: “European Language Resource Coordination (ELRC) and beyond”

Thierry Declerck gave an in-depth overview of the European Language Resource Coordination activities, starting his presentation with a brief history, a description of the consortium members, and ELRC objectives - to provide language resources to overcome language barriers in public services and the European Digital Single Market in general. He characterised various types of language data which are under the focus of the ELRC, as well as described the progress achieved through several phases of ELRC activities. He paid special attention to the importance of achievements in numbers of unique Language Resources (LR) by type, especially large multi-language multilingual corpora. Although much has been achieved since 2015, he emphasised that this is only the tip of the iceberg and there is still much to be done. As he continued his presentation, he highlighted that there are now more than 676 language tools/services offered by CEF eTranslation and third-party services. He explained that besides activities on language data collection and processing, ELRC also offers advice on legal questions and assists in improving data management processes. He encouraged the participants to contact the Helpdesk with any questions on technical or legal aspects related to the use, production, collection, processing and sharing of LR.

To conclude the presentation he highlighted opportunities, challenges, and the future of ELRC and how this has been impacted by the increasing demand for new languages, new domains and modalities.

## 4 Country Profile: Language data creation, management and sharing

An updated profile for Denmark with respect to language data creation and management has recently been published in the 2022 ELRC White Paper (see: <https://lr-coordination.eu/sites/default/files/LRB/LRB-12/ELRC-White-Paper.pdf>)

## **5 Workshop Participants**

There were 143 participants who completed the online registration form maintained by the Danish Agency for Digitisation. This reflects the number of participants as well, as some came without registration and some people who had registered didn't attend.

ELRC Workshop Report for Denmark

## 6 Appendix

### 6.1 Localised agenda



The image shows a screenshot of a conference agenda for 'Program for Sprogteknologisk konference 2021'. The agenda is listed on the left side of a slide, with a background image of a modern interior with wooden beams and hanging lights. The agenda items include:

- Kl. 9.00 Velkomst
- Kl. 9.10 Om danske chatbots og et dansk open source initiativ v. Henrik Fabre, Certainty
- Kl. 9.40 Natural language processing: Recent advances and challenges v. Barbara Frank, ITU
- Kl. 10.00 What eye movements can teach us about language processing v. Nora Hollenstein, CST (KU)
- Kl. 10.20 Paneldebat med spørgsmål fra salen
- Kl. 10.30 Pause og netværk
- Kl. 11.00 Talegenkendelse til digital support v. Klaus Bjørn Larsen, Roskilde kommune
- Kl. 11.30 Sprogteknologi.dk – En fælles dansk sprogpresse v. Christian Plassche, Digitaliseringsstyrelsen
- Kl. 11.50 COR – Central Ordregister for dansk til sprogteknologiske formål v. Peter Jørn Henriksen, DSN, Sanni Nimb, DSL og Bolette S. Pedersen, CST (KU)
- Kl. 12.10 Paneldebat med spørgsmål fra salen
- Kl. 12.20 Pause, frokost og posters
- Kl. 13.20 EU-kommissionens maskinoversættelsessystem: eTranslation v. Claus Thornby Larsen, EU-kommissionen
- Kl. 13.45 Update from ELRC and new tools on the CEF platform v. Thierry Declercq, European Language Resource Coordination
- Kl. 14.10 Panel on ELRC and eTranslation seen in the Danish context
- Kl. 14.30 Pause og posters
- Kl. 15.00 The European Language Grid and European Language Equality project v. Georg Rehm, Deutsches Forschungszentrum für Künstliche Intelligenz
- Kl. 15.30 Panel on ELG and synergies with the Danish LT community
- Kl. 15.50 Afrunding og tak for i dag

### 6.2 Localised press release



The image shows a screenshot of a press release page from the Center for Sprogteknologi (CST). The page features the CST logo and name, a search bar, and a navigation menu. The main content is a press release titled 'Sprogteknologisk Konference 2021'.

**Center for Sprogteknologi** [CST](#) > [Kalender](#) > Sprogteknologisk Konfe...

Om centret  
Ansatte  
Forskning  
Projekter  
Uddannelser  
Omverdensrelationer  
Online-værktøjer  
Sproressourcer  
Kontakt

Institut for Nordiske Studier og Sprogvidenskab  
Det Humanistiske Fakultet

## Sprogteknologisk Konference 2021

Center for Sprogteknologi ved Institut for Nordiske Studier og Sprogvidenskab, KU, er sammen med Digitaliseringsstyrelsen vært for denne konference om sprogteknologi som det udvikler sig lige nu i Danmark og i det europæiske miljø.



Formålet med dagen er at samle relevante udviklere, brugere, forskere og studerende til en spændende dag, hvor vi udveksler viden om den aktuelle status på dansk sprogteknologi og samtidig får indblik i de nyeste ressourcer, værktøjer og muligheder, også i en bredere europæisk kontekst.

Hvor formiddagen fokuserer på forskning og udvikling i Danmark, udgøres eftermiddagen af to mindre workshops om hhv. European Language Resource Coordination (ELRC) med fokus på EU's maskinoversættelse (eTranslation), som kan anvendes helt gratis af alle offentlige institutioner og af små og mellemstore virksomheder og endelig European Language Grid (ELG) som udvikler en fælleseuropæisk sprogteknologiportal. Dagen rummer desuden to poster-sessioner hvor der er mulighed for at drøfte folks igangværende projekter.

- Program og præsentationer +
- Accepterede poster +
- Tilmelding +

## ELRC Workshop Report for Denmark

<https://cst.ku.dk/kalender/sprogteknologisk-konference-2021/>



Program for Sprogteknologisk konference d. 16. november 2021

26. oktober 2021

Center for Sprogteknologi ved Institut for Nordiske Studier og Sprogvidenskab, KU og Digitaliseringsstyrelsen er værter for konferencen om sprogteknologi som det udvikler sig lige nu i Danmark og i det europæiske miljø.

Formålet med dagen er at samle relevante udviklere, brugere, forskere og studerende til en spændende dag, hvor vi udveksler viden om den aktuelle status på dansk sprogteknologi og samtidig får indblik i de nyeste ressourcer, værktøjer og muligheder, også i en bredere europæisk kontekst.

Hvor formiddagen fokuserer på forskning og udvikling i Danmark, udgøres eftermiddagen af to mindre workshops om hhv. European Language Resource Coordination (ELRC) med fokus på ELR's maskinoversættelse «Translation, som kan anvendes helt gratis af alle offentlige institutioner og af små og mellemstore virksomheder og endelig European Language Grid (ELG) som udvikler en fælles europæisk sprogteknologiportal.

Dagen rummer desuden to posteressioner hvor der er mulighed for at drøfte folks igangværende projekter.

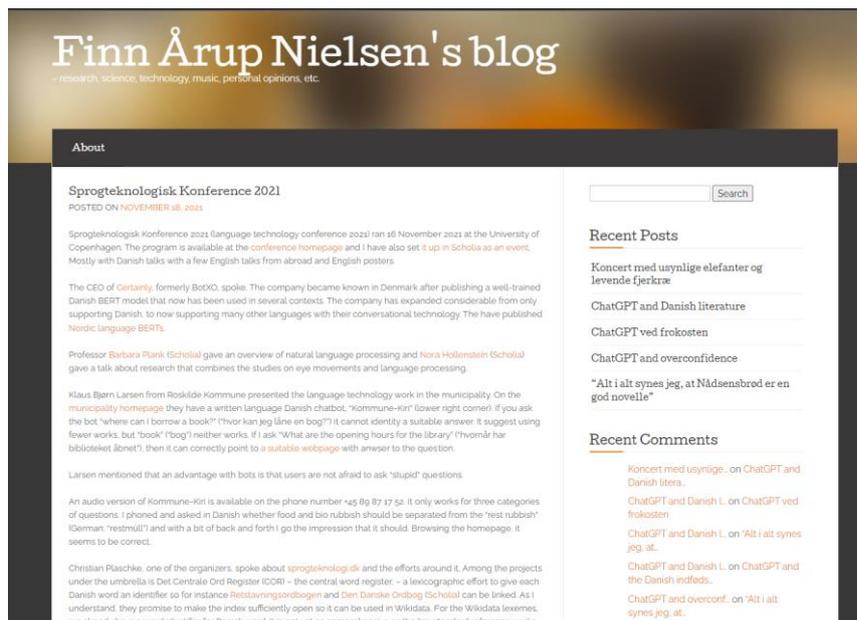
Sted: KU, Søndre Campus, lokale 4A.D.69

[SE PROGRAM](#)

[TILMELDING VIA CENTER FOR SPROGTEKNOLOGI](#)

<https://sprogteknologi.dk/blog/program-for-sprogteknologisk-konference>

## Blogpost about conference



Fin Årup Nielsen's blog  
- research, science, technology, music, personal opinions, etc.

About

Sprogteknologisk Conference 2021  
POSTED ON NOVEMBER 18, 2021

Sprogteknologisk Conference 2021 (language technology conference 2021) ran 16-17 November 2021 at the University of Copenhagen. The program is available at the conference homepage and I have also set it up in Scholia as an event. Mostly with Danish talks with a few English talks from abroad and English posters.

The CEO of Certainty, formerly BoixO, spoke. The company became known in Denmark after publishing a well-trained Danish BERT model that now has been used in several contexts. The company has expanded considerably from only supporting Danish, to now supporting many other languages with their conversational technology. They have published Nordic language BERTs.

Professor Barbara Plank (Scholia) gave an overview of natural language processing and Nora Hollenstein (Scholia) gave a talk about research that combines the studies on eye movements and language processing.

Klaus Børn Larsen from Roskilde Kommune presented the language technology work in the municipality. On the municipality homepage they have a written language Danish chatbot, "kommune-Kit" (lower right corner), if you ask the bot "where can I borrow a book?" ("hvor kan jeg låne en bog?") it cannot identify a suitable answer, it suggest using fewer words, but "book" ("bog") neither works. If I ask "What are the opening hours for the library?" ("hvornår har biblioteket åbnet?"), then it can correctly point to a suitable webpage with answer to the question.

Larsen mentioned that an advantage with bots is that users are not afraid to ask "stupid" questions.

An audio version of kommune-Kit is available on the phone number +45 89 87 37 52. It only works for three categories of questions: I phoned and asked in Danish whether food and bio rubbish should be separated from the "rest rubbish" (German "restmüll") and with a bit of back and forth I go the impression that it should. Browsing the homepage, it seems to be correct.

Christian Pläschke, one of the organizers, spoke about sprogteknologi.dk and the efforts around it. Among the projects under the umbrella is Det Centrale Ord Register (ICORI) - the central word register. - a lexicographic effort to give each Danish word an identifier so for instance Restafføringsordbogen and Den Danske Ord bog (Scholia) can be linked. As I understand, they promise to make the index sufficiently open so it can be used in Wikidata. For the Wikidata lexemes, we already have a word identifier for Danish word. It is not yet so comprehensive as the big standard reference works.

Search

Recent Posts

- Koncert med usynlige elefanter og levende fjerkæb
- ChatGPT and Danish literature
- ChatGPT ved frokosten
- ChatGPT and overconfidence
- "Alt i alt synes jeg, at Nådsensbrød er en god novelle"

Recent Comments

- Koncert med usynlige... on ChatGPT and Danish litera...
- ChatGPT and Danish L... on ChatGPT ved frokosten
- ChatGPT and Danish L... on 'Alt i alt synes jeg, at...
- ChatGPT and Danish L... on ChatGPT and the Danish inflects...
- ChatGPT and overconf... on 'Alt i alt synes jeg, at...

<https://finnaarupnielsen.wordpress.com/2021/11/18/sprogteknologisk-konference-2021/>

### 6.3 Workshop Presentations

The presentations are available on the ELRC website: <https://lr-coordination.eu/denmark3rd>.

All the presentations are also available on the local website (via Centre for Language Technology, Department of Nordic Studies and Linguistics): <https://cst.ku.dk/kalender/sprogteknologisk-konference-2021/>.

## ELRC Workshop Report for Denmark

(Facilitator Manex Agirrezabal)

### ELRC-workshop

Ordstyrer Carl Frederik Bach Kirchmeier

- |             |  |
|-------------|--|
| 13.20-13.45 | Claus Thornby Larsen, EU-kommissionen: <b>EU-Kommissionens maskinoversættelsessystem: eTranslation.</b><br><a href="#">Download præsentation.</a>                          |
| 13.45-14.10 | Thierry Declerck, European Language Resource Coordination (ELRC): <i>Update from the ELRC and new tools on the CEF platform.</i><br><a href="#">Download præsentation.</a> |
| 14.10-14.30 | Panel on ELRC and eTranslation seen in the Danish context  |
| 14.30-15.00 | <b>Pause</b><br>Kaffe<br>Posters fra forskere, brugere, udviklere og studerende (Facilitator Manex Agirrezabal)  |