



European Language
Resource Coordination



Norsk Språkteknologi og internasjonalt samarbeid: ELRC, eTranslation og ELG

Kan roboten egentlig snakke?

Tredje ELRC-seminar i Norge, 3. Mars 2021

kristine.eide@sprakradet.no

Språkrådet

- Internasjonalt samarbeid om språkteknologi



● ● Hvorfor?

Norge er ikke det eneste landet som

- Har en ikke ubegrunnet frykt for de store språkernes dominans.
- Har særheter vi (tror vi) er alene om (Dansk tale er uforståelig – selv for en robot)
- Deltar i felles europeisk digital infrastruktur (CEF Digital)



• Hva vil vi ha?

- Språkteknologi på eget språk som fungerer like godt som på større språk.
- Språkteknologi som er uavhengig av de store aktørene.
- Språkteknologi som gjør det mulig å kommunisere på tvers av språk.



• Hva vil vi ha?

- Språkteknologi på eget språk som fungerer like godt som på større språk.
- Språkteknologi som er uavhengig av de store aktørene.
- Språkteknologi som gjør det mulig å kommunisere på tvers av språk.

• DATA



• CEF Digital



CEF – Connecting Europe Facility

Europeiske infrastrukturer
(Digital Service Infrastructures (DSIs))

- eHealth
- Online Dispute Resolution
- eProcurement

Felleseuropeiske plattformer og portaler

- European Platform for Desinformation

Europeiske fellesløsninger (Byggeklosser)

- eID
- eSignature
- eTranslation

ELRC

European Language
Resource Coordination



Nasjonale Ankerfester (NAP) i Norge:

Jon Arild Olsen, Nasjonalbiblioteket (teknisk)

Kristine Eide, Språkrådet (offentlig sektor)



- ELRC og eTranslation
- (European Language Resource Coordination)



ELRC WHITE PAPER

**Sustainable Language Data Sharing
to Support Language Equality
in Multilingual Europe**

WHY LANGUAGE DATA MATTERS

https://lr-coordination.eu/sites/default/files/ELRC_Conference/ELRCWhitePaper.pdf

ELRC og eTranslation (European Language Resource Coordination)

Norsk handlingsplan:

- Øke mengden tospråklige ressurser (engelsk-norsk og bokmål-nynorsk)
- Parallellkorpus bokmål-nynorsk
- Øke oppmerksomhet rundt språkdata som åpne data
- Øke interessen for automatisk oversettelse i offentlige tjenester
- God praksis for datahåndtering i offentlige tjenester
- Identifisere og håndtere juridiske hensyn



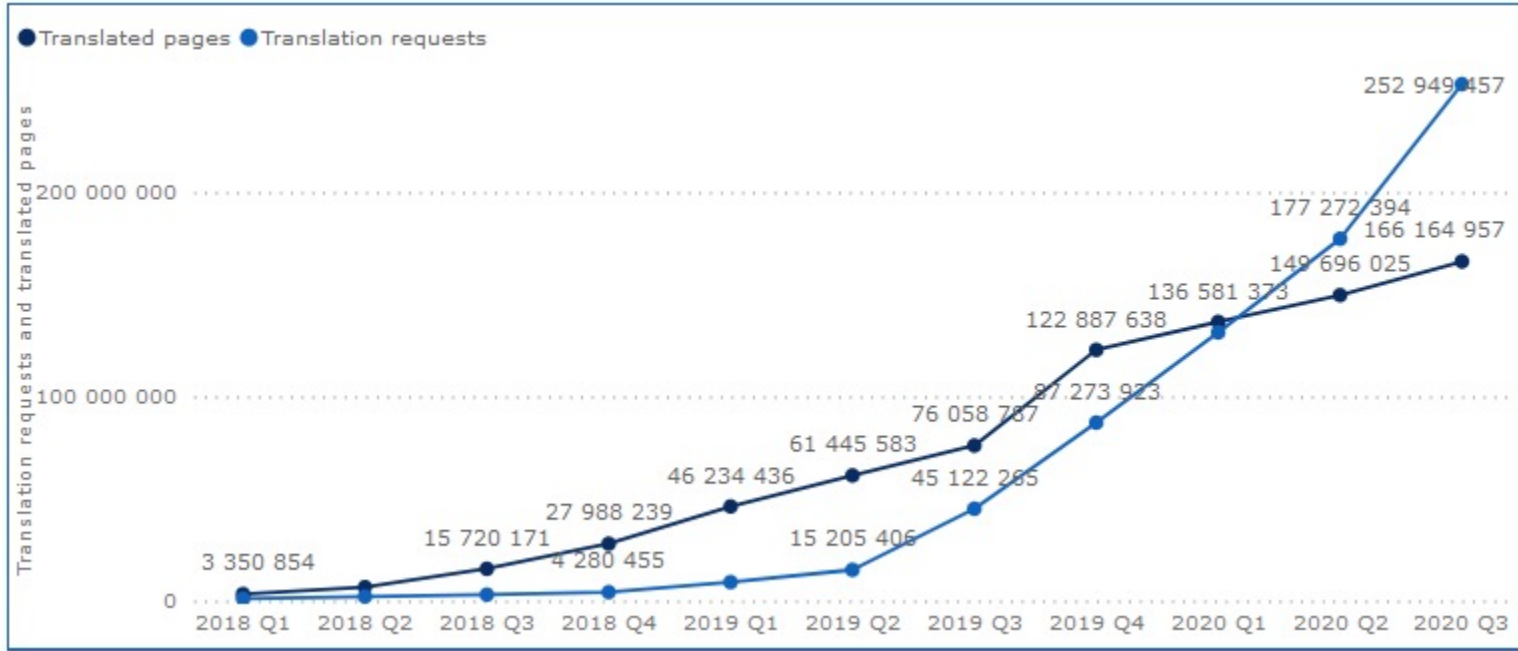
ELRC WHITE PAPER

Sustainable Language Data Sharing
to Support Language Equality
in Multilingual Europe

WHY LANGUAGE DATA MATTERS

https://lr-coordination.eu/sites/default/files/ELRC_Conference/ELRCWhitePaper.pdf

• Bruk av eTranslation



<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation+dashboard>



Two new and exciting upgrades have just been launched – eTranslation now handles **Arabic**, and **Speech-to-Text** joins the family of CEF Language Tools. Try it [here](#).



eTranslation vient d'élargir sa gamme de services - eTranslation supporte désormais l'**arabe** et le « **Speech-to-Text** » rejoint la famille des outils « [CEF Language Tools](#) ».

Zwei neue spannende Upgrades: eTranslation übersetzt jetzt auch **Arabisch**, und die [CEF-Language-Tools-Familie](#) wurde um **Speech-to-Text** erweitert.

Language Technology is shaping our multilingual future.

55 / 2500





Two new and exciting upgrades have just been launched – eTranslation now handles **Arabic**, and **Speech-to-Text** joins the family of CEF Language Tools. Try it [here](#).

eTranslation vient d'élargir sa gamme de services - eTranslation supporte désormais l'**arabe** et le « **Speech-to-Text** » rejoint la famille des outils « [CEF Language Tools](#) ».

Zwei neue spannende Upgrades: eTranslation übersetzt jetzt auch **Arabisch**, und die [CEF-Language-Tools-Familie](#) wurde um **Speech-to-Text** erweitert.

Language Technology is shaping our multilingual future.

Forskningsteknologien er i stand til å utforme vår flerspråklige fremtid.

● DATA, DATA!

Language Technology is shaping our multilingual future.

Forskningsteknologien er i stand til å utforme vår flerspråklige fremtid.

Språkteknologi formar vår flerspråkiga framtid.

● DATA, DATA!

Language Technology is shaping our multilingual future.

It has already been transforming the way we interact with our devices and with each other, the way we shop, work and travel.

Forskningsteknologien er i stand til å utforme vår flerspråklige fremtid.

Og det er allerede en del av hvordan vi samarbeider med våre innretninger og med hverandre, hvordan arbeidslivet, arbeidet og reiser, arbeid og reiser.

Språkteknologi formar vår flerspråkiga framtid.

Det har redan förändrat hur vi interagerar med våra enheter och med varandra, hur vi handlar, arbetar och reser.

● DATA, DATA!

Language Technology is shaping our multilingual future.

It has already been transforming the way we interact with our devices and with each other, the way we shop, work and travel.

More and more it reshapes our interaction with service providers, either public or private.

Forskningsteknologien er i stand til å utforme vår flerspråklige fremtid.

Og det er allerede en del av hvordan vi samarbeider med våre innretninger og med hverandre, hvordan arbeidslivet, arbeidet og reiser, arbeid og reiser.

Flere og mer omformer vårt samarbeid med tjenesteytere, enten offentlig eller privat.

Språkteknologi formar vår flerspråkiga framtid.

Det har redan förändrat hur vi interagerar med våra enheter och med varandra, hur vi handlar, arbetar och reser.

Mer och mer omformar det vår interaktion med tjänsteleverantörer, antingen offentliga eller privata.

• Veileder for innlevering av data

Samarbeid mellom Digitaliseringsdirektoratet,
Nasjonalbiblioteket og Språkrådet

- Kommer snart
- Kort fortalt sier vi:
 - *Fagtermer/begreper (tilgjengeliggjøres)*
 - *Oversettelser (sendes)*
 - *Offentlig produsert tekst (høstes automatisk)*
 - *Utviklede ressurser (sendes)*

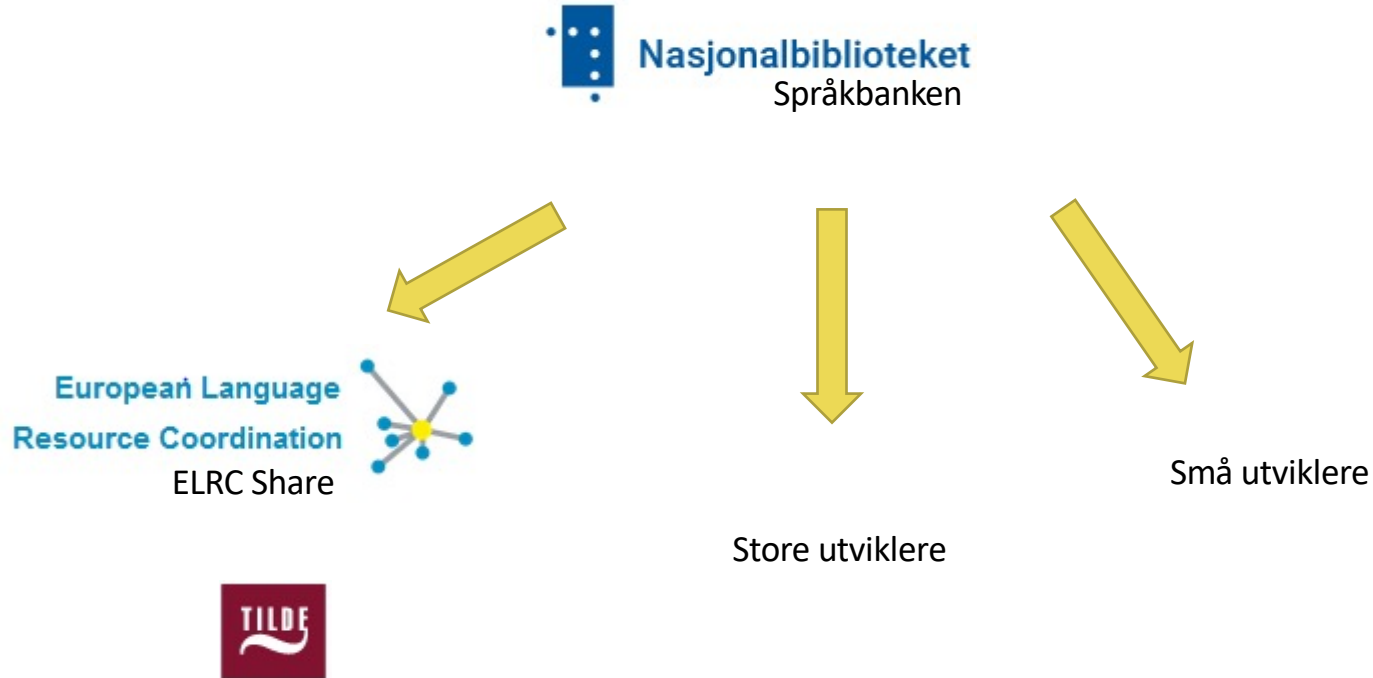


 Felles datakatalog



 **Nasjonalbiblioteket**
Språkbanken

• Hva skjer videre med dataene?



- CEFs platform for automatisk oversettelse
- (The CEF Automated Translation Platform)

MT@EC

- Juridisk språk
- Statistisk maskinoversettelse

eTranslation

- Nevrale nettverk
- domenetilpasset
- Integret i offentlige tjenester

CEF.AT

- Flere språkteknologiske løsninger
- ELRC
- Prosjektstøtte



- extending domain coverage (e.g. general; scientific texts; social media)
- extending language coverage (non-EU languages of social & economic importance)
- additional language technologies
 - speech-to-text
 - anonymization
 - named-entity recognition
 - basic CAT tool



-
-

Hvem kan bruke platformen?

- Offentlige virksomheter og ansatte i offentlige virksomheter i EU, Norge og Island
- Små og middels store bedrifter

CEF eTranslation



• ELG – European Language Grid



Forløper: META-NET

- Språkteknologisk forskning og utvikling
- Forskere og teknologibedrifter
- Privatpersoner og næringslivsaktører



• ELG – European Language Grid



ELG

- Tilgang til kommersiell og ikke-kommersiell språkteknologi for alle europeiske språk
- Man kan laste opp programvare og ressurser og bruke dem gjennom plattformen, sammen med andre ressurser som er der.

NCCs (National Competence Centres)

- Samle lokal informasjon om språkteknologiske produkter og tjenester, prosjekter, forskningsentre, eksperter, selskaper og andre tiltak.
- Informere om ELG (seminar)
- Forberede et europeisk språkteknologiråd

norwegian

Search



Norwegian SpeechDat(II) FDB-1000

0 views

The Norwegian SpeechDat(II) FDB-1000 comprises 1016 Norwegian speakers (517 males, 499 females) recorded over the Norwegian fixed telephone network. The FDB-1000 database is partitioned into 4 CDs. The speech databases m...

Keyword: corpus

Language: Norwegian

Licences: ELRA-VAR-ACADEMIC-MEMBER-COMMERCIALUSE-1.0

ELRA-END-USER-ACADEMIC-MEMBER-NONCOMMERCIALUSE-1.0 ...



Collins Multilingual database (MLD) - PhraseBank

2 views

The Collins Multilingual database covers Real Life Daily vocabulary. It is composed of a multilingual lexicon in 32 languages (the WordBank, distributed separately under reference ELRA-T0376) and a multilingual set of se ...

Keyword: lexicalconceptualresource



Search

Other EU/European languages 🔍

- + Albanian (1)
- + Basque (1)
- + Norwegian (1)
- + Turkish (1)

Licences

- + Apache License 2.0 (4)
- + Creative Commons Attribution 4.0 International (1)
- + GNU Affero General Public License v3.0 only (2)
- + GNU General Public License v3.0 only (1)
- + GNU Lesser General Public License v3.0 or later (4)

Show more



Elhuyar Basque ASR

Automatic Speech Recognition service for Basque

Keywords: automatic speech recognition Basque

Language: Basque

Licence: Apache License 2.0

38 views

work in progress



GATE: French Named Entity Recognizer

A named entity recognition service for documents in French. Based on ANNIE, it identifies names of persons, locations, and organizations.

Keywords: Named Entity Recognition French Français

Languages: French English

Licence: Apache License 2.0

34 views

31 times used

Functional service

● Internasjonalt samarbeid om språkteknologi



Særlig rettet mot offentlig
sektor



Særlig rettet mot forskning og
utvikling

Takk for
oppmerksomheten
og husk å svare på

<https://ec.europa.eu/eusurvey/runner/ELRCCountryWorkshops>