Machine Translation for Broadcast Journalism

Susanne Weber

Language Technology Producer, BBC News Labs

In this presentation…

Automated MT for Media Processing

CAT for Video Translation

SUMMA – H2020 EU-funded collaboration

Scalable Understanding of Multilingual Media

summa-project.eu

SUMMA – Objectives

- Scalable and extensible monitoring platform

- Higher quality and richer tools for monitors

- Extensible automated knowledge base construction

- Multilingual and cross-lingual capabilities

SUMMA – Consortium

SUMMA – 3 Use Cases

1 - External Media Monitoring (BBC)
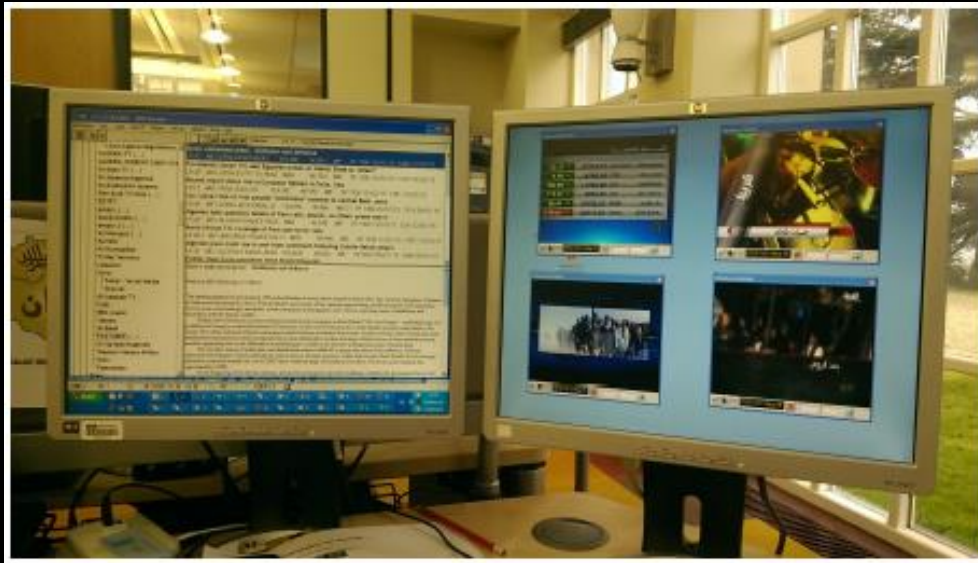
2 - Internal Media Montoring (DW)

3 - Data Journalism

SUMMA – Use Case 1 External Monitoring

External Monitoring
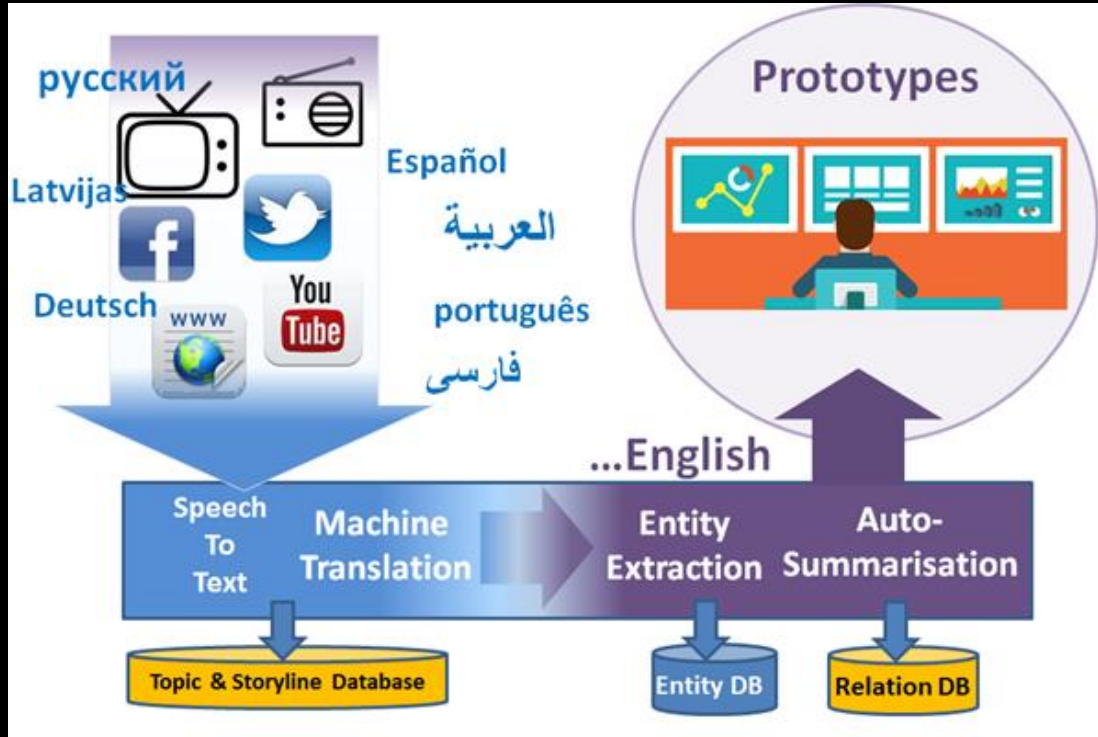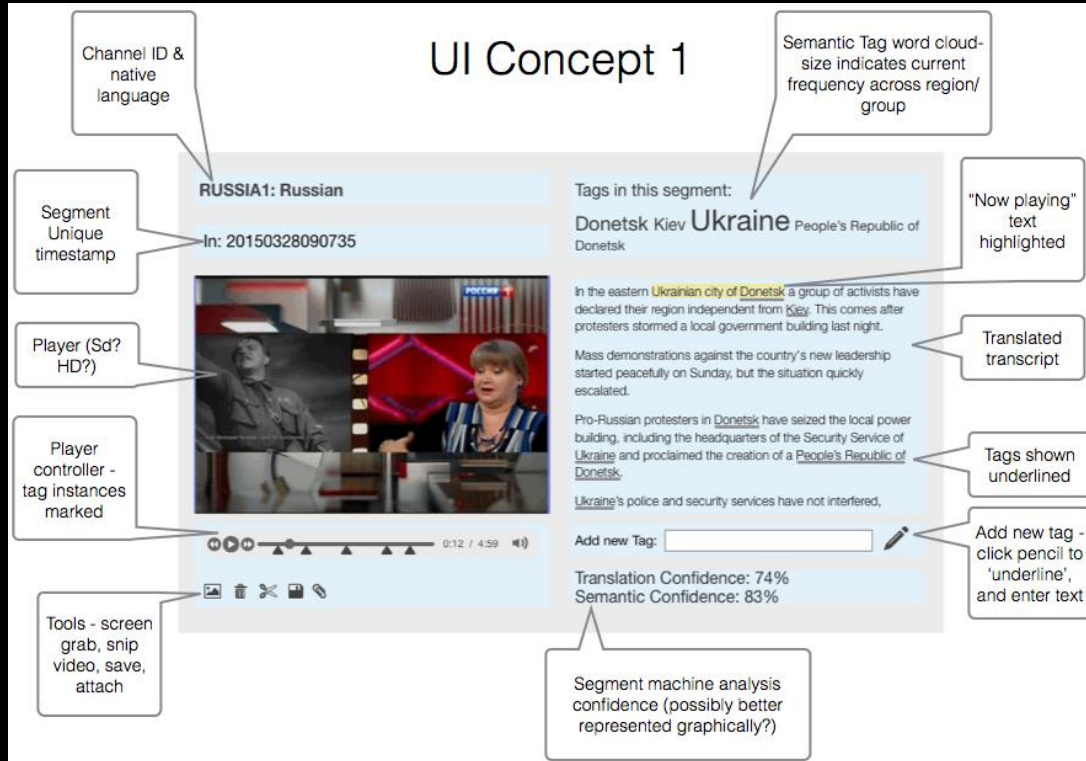
SUMMA – BBC Monitoring

Internal Monitoring

SUMMA – The Pipeline

SUMMA – UI Concept

SUMMA – MT – The Languages

High resourced

Arabic
German
Spanish
English

Medium resourced

Portuguese
Russian
Farsi

Low resourced

Latvian
Ukrainian

SUMMA – Machine Translation

Sources:

- from ASR

- from online text

- from social media

SUMMA – Hack Events

BBC #newsHACK

2017 ➡️ Automated monitoring of multi-media sources

2018 ➡️ Focus on Data Journalism

2019 ➡️ Focus on multilingual aspect

ALTO - Video Translation Tool (BBC)

ALTO – video translation for News

- Audience expectations

- Multilingual journalism

- Language technologies

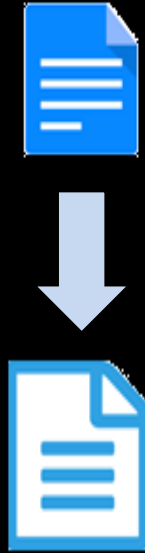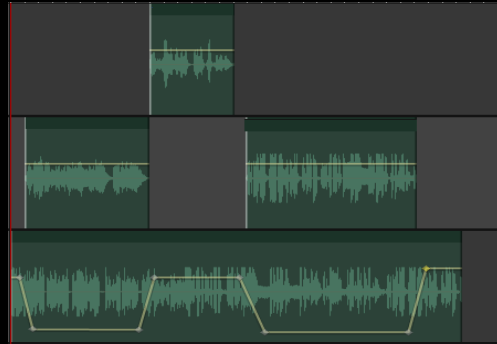# How do we currently translate videos?

Typical Workflow for Video Translation

Translate Script

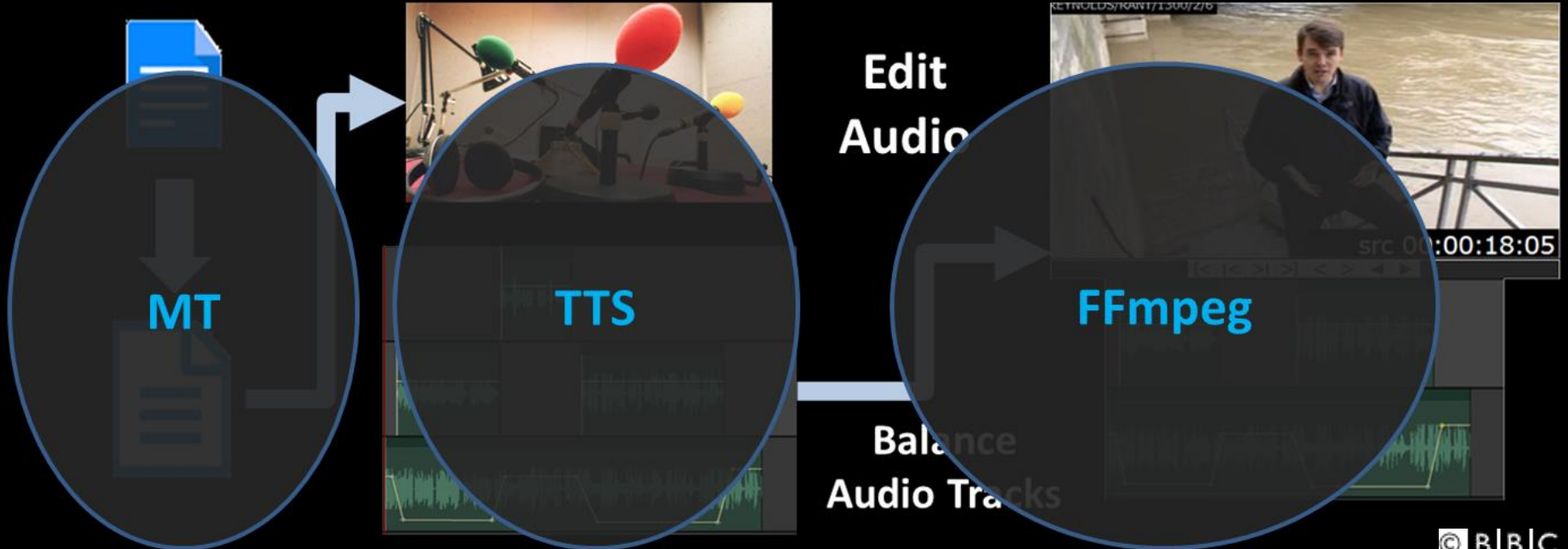**Record Voice-over tracks**

Edit Audio

Align Audio & Video

src 00:00:18:05

Balance Audio Tracks

© BBC

Typical Workflow for Video Translation

Translate Script → **Record Voice-over tracks** → **Edit Audio** → **Align Audio & Video**

MT

TTS

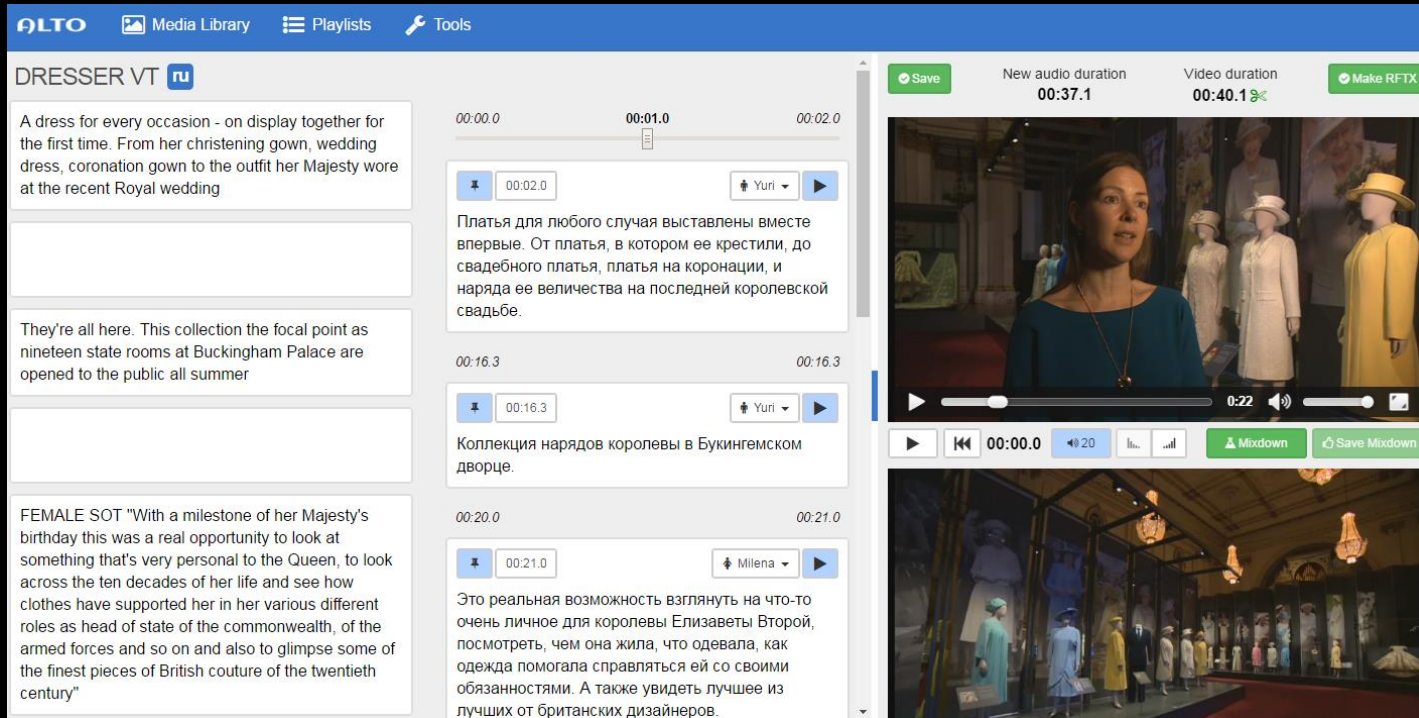**Balance Audio Tracks**

FFmpeg

src 00:00:18:05

©BBC

Off-the-shelf products
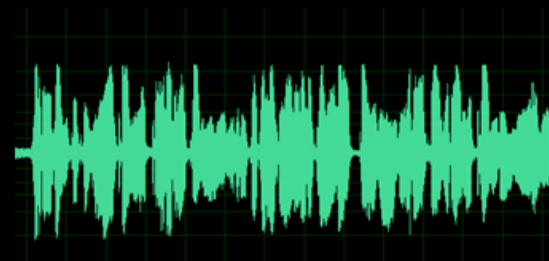
**Text-To-Speech
Voice Synthesis**

➢Google Translate

➢Computer-Assisted Translation

ALTO – Translation Window

ALTO - Audio Editing is Text Editing

BBC World Service – Languages and Styles

Translations from English into other languages:
Africa, Middle East, Asia, Latin America

Language styles

- Journalists: correct, complex, metaphors, idioms

- Interviewees: less correct, sometimes incomplete,

    disfluencies, colloquial

- Headlines: short, factual

© B|B|C

BBC World Service – Language Styles

Examples: Simple, factual, headlines style:

- "Hurricane Newton has reached Mexico's west coast near a popular tourist destination."

- "We'll bring you the latest with BBC Weather."

- "No trip to Myanmar is complete without a visit to Bagan."

- "Two weeks ago the temples were shaken by a strong earthquake."

BBC World Service – Language Styles

Examples: Complex – authored by journalists:

- "Beating the most unpopular Republican Presidential nominee in the history of modern day polling should have been something of an open goal."

- "…But it's with white males that her poll numbers are tanking - local businessman James O. would rather dive into a shark pool than give Hillary Clinton his vote."

- "In a city where some live in peace and prosperity, others are in poverty and peril."

BBC World Service – Language Styles

<u>Examples</u>: interviewees and vox pops:

"you can't just be cruising through no neighbourhood that's in a war with each other - because nine times out of ten - they probably don't recognise the car you in at the time and they can mistake you as a shooter from one of their rival gangs and they can open fire on you… you can't make no mistakes." (Chicago citizen)

"The life here in the camp – it's not bad - like we have more friends we are like family - brothers - but I hope they do not take this camp from here. I am just staying here - I am just waiting." (Calais migrant)

ALTO – User Feedback on CAT

Positive:

- Faster than translating from scratch
- Headlines / summaries are more correct than news packages
- Useful tool for proof-reading, use as dictionary

Negative:

- Syntax
- Declensions (Russian in particular)
- Idioms, metaphors, phrasal verbs

© B|B|C

ALTO – Users' suggestions

- Phrase dictionary

- Translation memory

- Ability to translate honorifics

Conclusion and Next Steps:

- low resourced languages

- Machine Learning - small number of users

- MT - different language styles

- Collaborate with research groups and universities

"Your Japanese language is really well done - thank you. I was impressed that you get an expert to work with the automatic translation…"
(viewer from UK)

"As I can't speak English, I've been waiting for a site like this. "
(viewer from Tokyo)