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 Monitoring (DW)
3 - Data Journalism
SUMMA – Use Case 1 External Monitoring

External Monitoring
SUMMA – Use Case 2 Internal Monitoring

Internal Monitoring
SUMMA – The Pipeline
SUMMA – UI Concept

UI Concept 1

- Channel ID & native language
- Segment Unique timestamp
- Player (SD? HDR?)
- Player controller - tag instances marked
- Tools - screen grab, snip video, save, attach
- Semantic Tag word cloud - size indicates current frequency across region/group
- "Now playing" text highlighted
- Translated transcript
- Tags shown underlined
- Add new tag - click pencil to 'underline', and enter text
- Translation Confidence: 74%
- Semantic Confidence: 83%

In the eastern Ukrainian city of Donetsk, a group of activists have declared their region independent from Kyiv. This comes after protests stormed a local government building last night.

Mass demonstrations against the country's new leadership started peacefully on Sunday, but the situation quickly escalated.

Pro-Russian protesters in Donetsk have seized the local power building, including the headquarters of the Security Service of Ukraine and proclaimed the creation of a 'People's Republic of Donetsk.'

Ukraine's police and security services have not interfered.

Segment machine analysis confidence (possibly better represented graphically?)
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

1. Translate Script
2. Record Voice-over tracks
3. Align Audio & Video
4. Edit Audio
5. Balance Audio Tracks
Typical Workflow for Video Translation

- **Translate Script**: Use MT (Machine Translation) to translate the script.
- **Record Voice-over tracks**: Record the voice-over tracks.
- **Align Audio & Video**: Use FFmpeg for alignment and editing of audio and video tracks.
- **Edit Audio**: Balance and align audio tracks.

Diagram elements include:
- Blue arrows indicating the flow of processes.
- MT (Machine Translation) icon.
- TTS (Text to Speech) icon.
- FFmpeg icon.
Off-the-shelf products

- Google Translate
- Computer-Assisted Translation

Text-To-Speech Voice Synthesis
ALTO – Translation Window

A dress for every occasion - on display together for the first time. From her christening gown, wedding dress, coronation gown to the outfit her Majesty wore at the recent Royal wedding.

They're all here. This collection the focal point as nineteen state rooms at Buckingham Palace are opened to the public all summer.

FEMALE SOT “With a milestone of her Majesty's birthday this was a real opportunity to look at something that's very personal to the Queen, to look across the ten decades of her life and see how clothes have supported her in her various different roles as head of state of the commonwealth, of the armed forces and so on and also to glimpse some of the finest pieces of British couture of the twentieth century.”

A dress for every occasion - on display together for the first time. From her christening gown, wedding dress, coronation gown to the outfit her Majesty wore at the recent Royal wedding.

They're all here. This collection the focal point as nineteen state rooms at Buckingham Palace are opened to the public all summer.

FEMALE SOT “With a milestone of her Majesty's birthday this was a real opportunity to look at something that's very personal to the Queen, to look across the ten decades of her life and see how clothes have supported her in her various different roles as head of state of the commonwealth, of the armed forces and so on and also to glimpse some of the finest pieces of British couture of the twentieth century.”

A dress for every occasion - on display together for the first time. From her christening gown, wedding dress, coronation gown to the outfit her Majesty wore at the recent Royal wedding.

They're all here. This collection the focal point as nineteen state rooms at Buckingham Palace are opened to the public all summer.

FEMALE SOT “With a milestone of her Majesty's birthday this was a real opportunity to look at something that's very personal to the Queen, to look across the ten decades of her life and see how clothes have supported her in her various different roles as head of state of the commonwealth, of the armed forces and so on and also to glimpse some of the finest pieces of British couture of the twentieth century.”
ALTO - Audio Editing is Text Editing
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
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.”
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.”
Examples: 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
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)