



# Automatic Sign Language Translation: dos and dreams.

Author: Giacomo Inches, Martel Innovate



### A brief intro





COST ACTION

2020-2024



Leading Platform for European Citizens, Industries, Academia and Policymakers in Media Accessibility

H2020-ICT19-2017





2017-2021





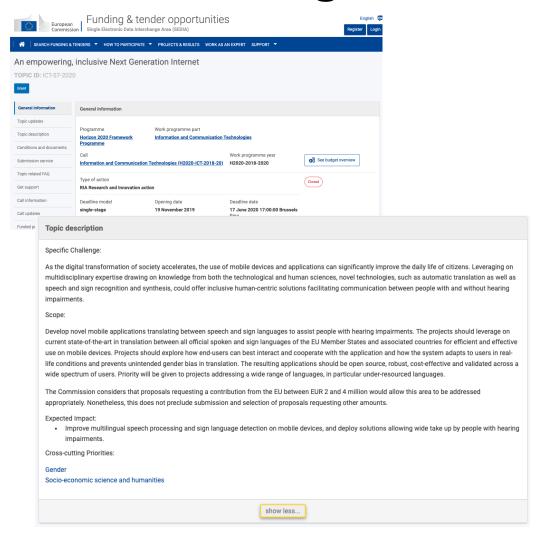
H2020-ICT57-2020

2021-2023

https://www.project-easier.eu/



# **EASIER- Background**





H2020 Call for Proposal: ICT57 Projects that aims to **design**, **develop**, and validate a complete multilingual machine translation system that will act as a framework for barrier-free communication among deaf hearing individuals and for a variety of specific contexts and languages

### **EASIER- Goals**



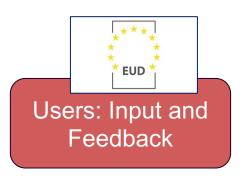
- Translation between sign languages and spoken languages (speech or writing)
- Multiple languages (e.g. DGS, DSGS, LSF, LIS, BSL, GSL, NGT,...)
- Automatic (near-realtime) and semi-automatic (human post-editing)
- Mobile app demonstrating the automatic route
- Goals shared with the SignOn project (both funded within EU ICT-57)

# Lessons learnt from earlier projects





- Producing comprehensible and easy-to-read signing is a must:
  - Producing a signed utterance sign after sign simply is not enough!
  - Communication is more than content-passing.
  - Co-design with final beneficiaries of the innovation!

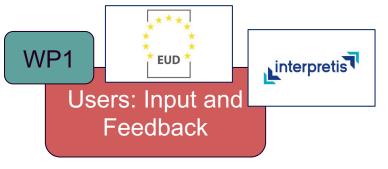






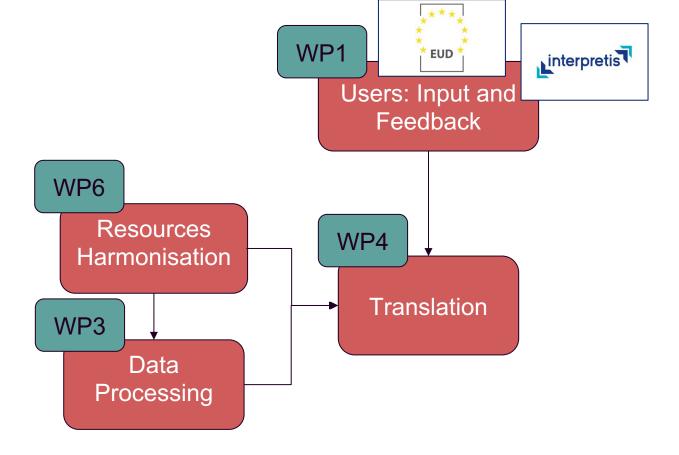






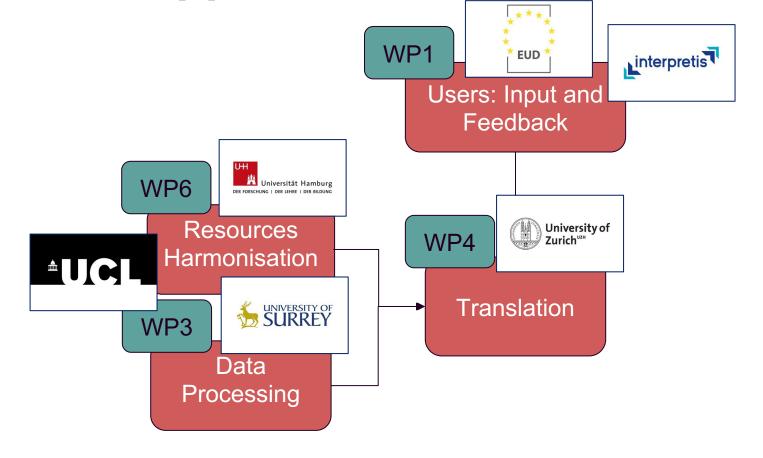


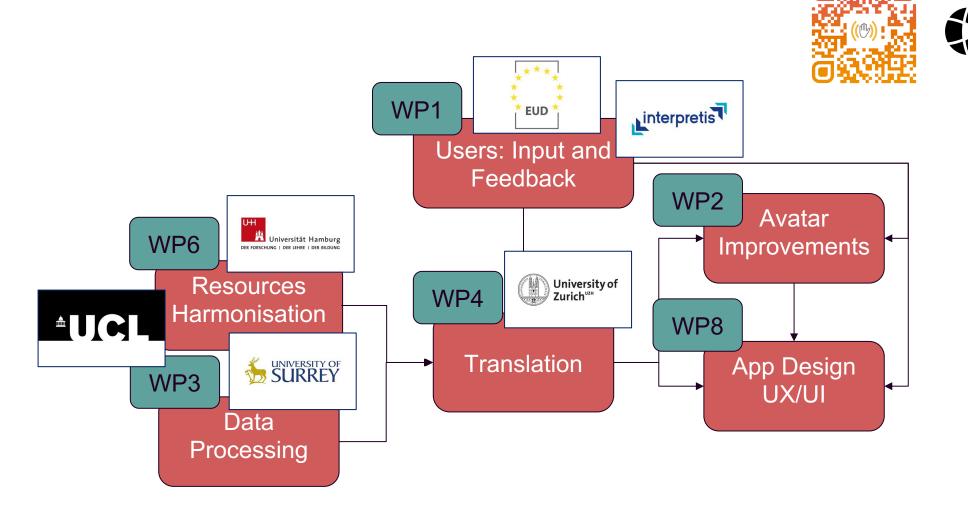






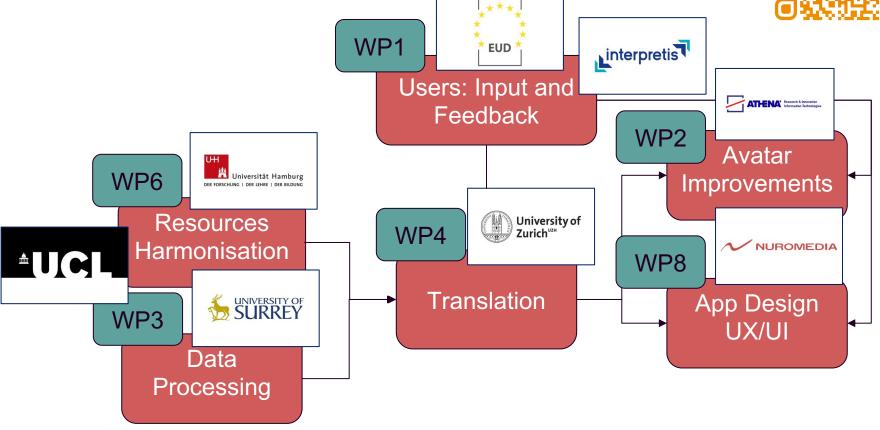




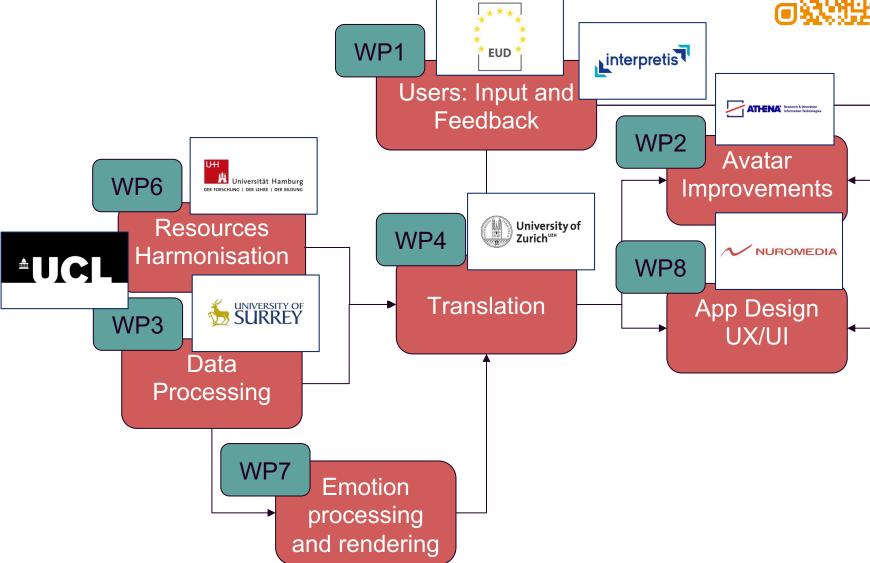


**LEAD-ME** 

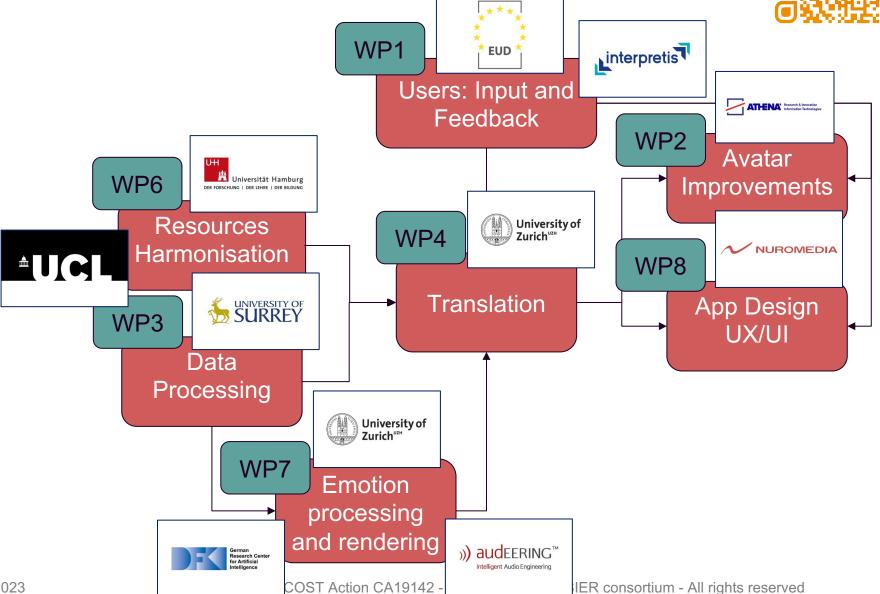




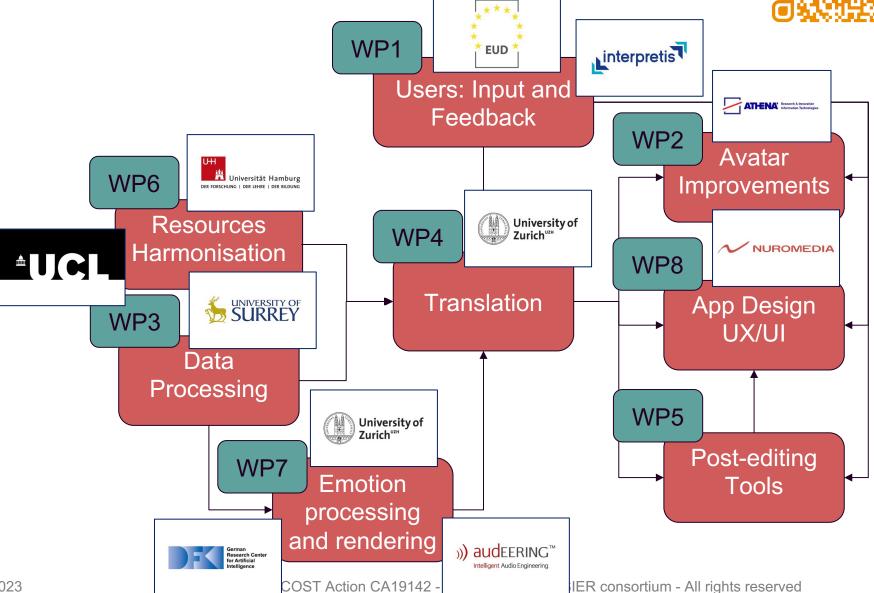




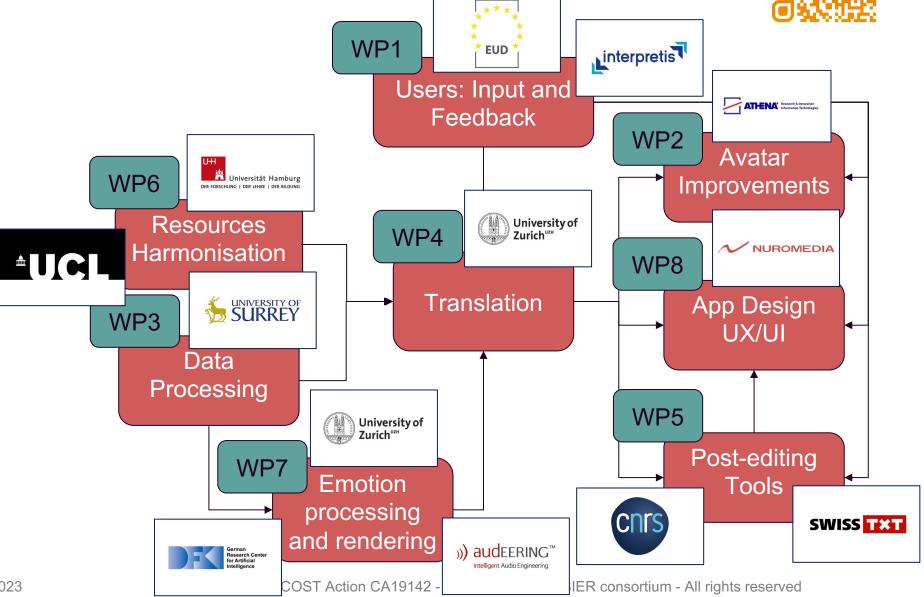




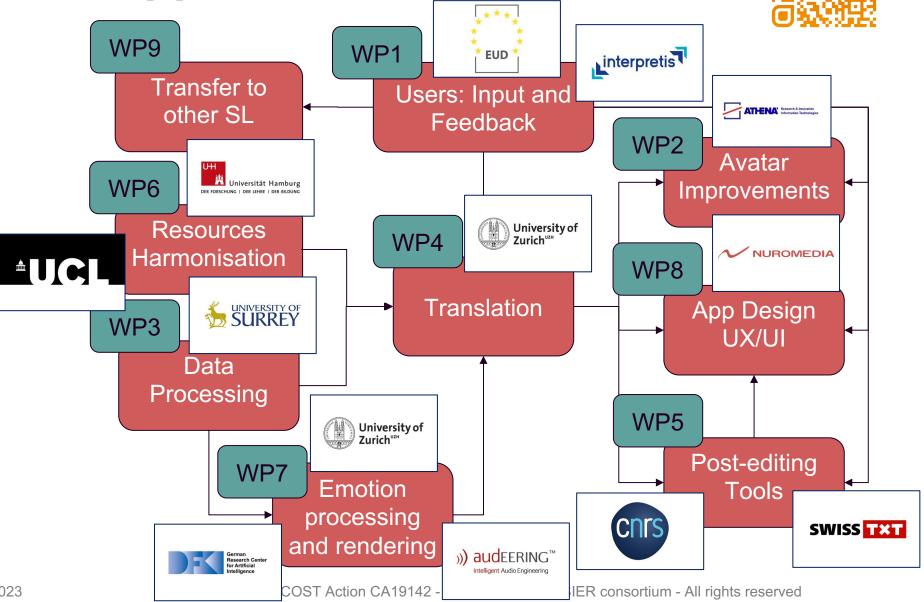


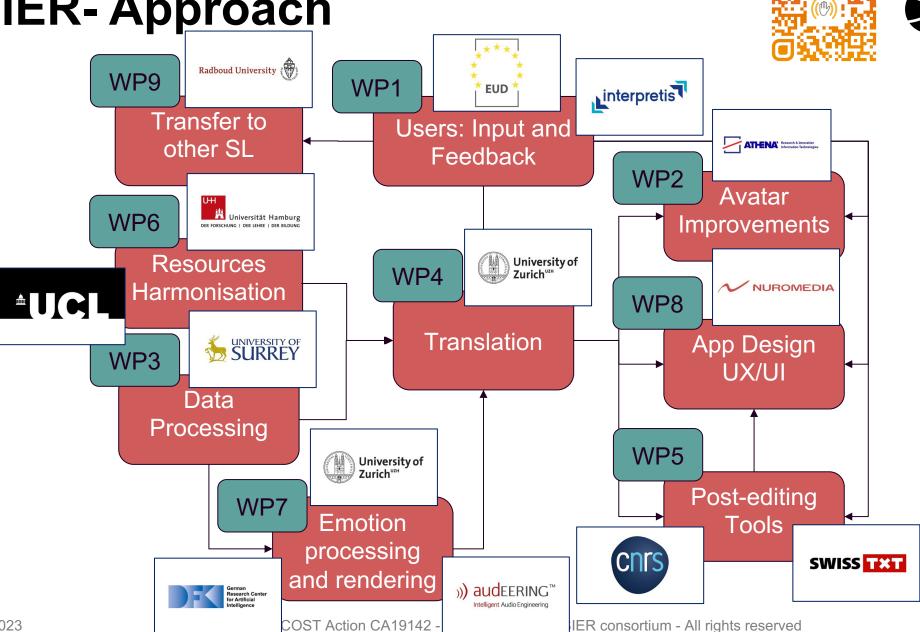




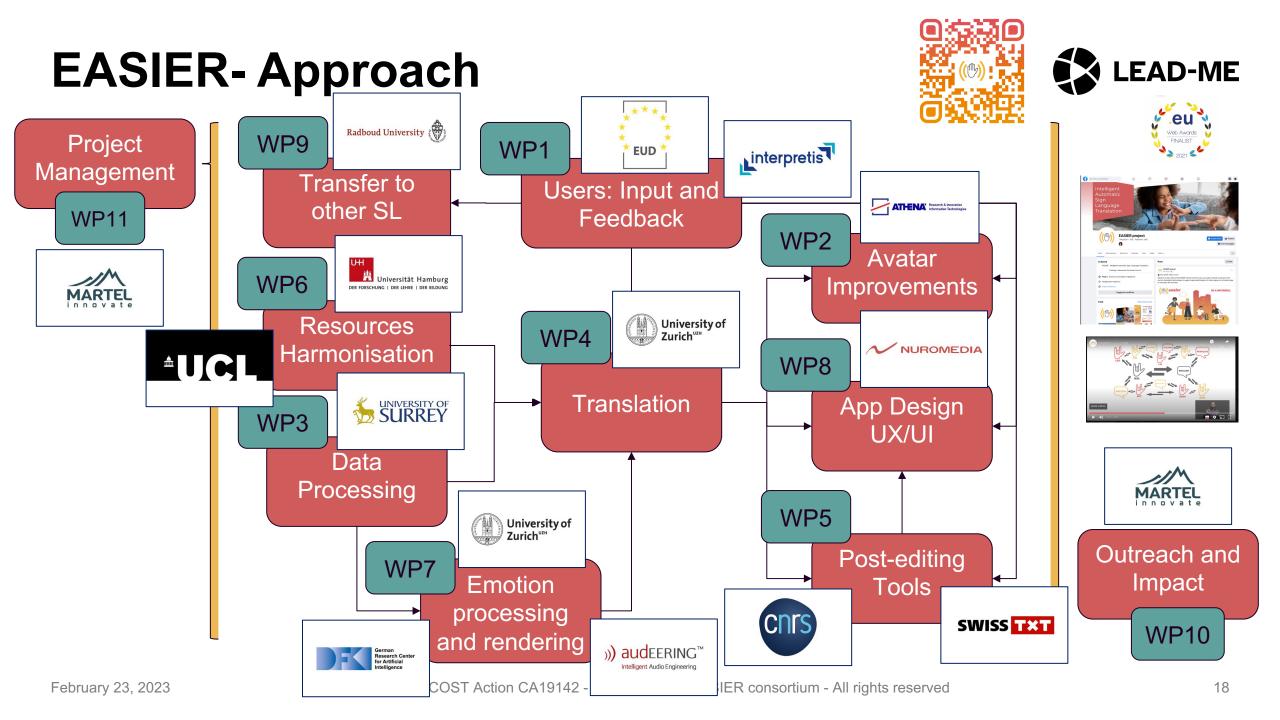












# Main challenges and opportunities





#### Sign Languages are different from Spoken Languages

- No direct mapping between spoken languages and sign languages e.g. grammar, syntax, sequentiality
  vs. signing space
- Multidimensionality and multimodality of sign languages
  - Annotation
  - Data processing
- Data scarcity
  - Combination of datasets (broadcast & linguistic)
  - Combination of language-pairs
  - Combination of approaches for translation (statistical vs lexical vs neural)
- Presentation of automatic translation results
  - Mobile Interface
  - Avatar/Virtual Human
- User acceptance and quality of translation
  - Post-editing
  - End users always in the loop
  - Continuous evaluation

# Challenge 1: Data (quality & quantity)





#### Multidimensionality and multimodality of sign languages

- Annotation
- Data processing

#### Data scarcity

- Combination of datasets (broadcast & linguistic)
- Combination of language-pairs
- Combination of approaches for translation (statistical vs lexical vs neural)

#### Linguistic corpora



High quality

Variety of elicitation tasks

Source: Sign Language

Semi-spontaneous language production

Rich linguistic annotation + translations

#### Broadcasting data



Large quantity
News domain
Source: spoken language
Interpreting under time pressure
Subtitles/Captions

# **Challenge 1: Orders of magnitude**



Language	Number of sentence pairs in corpus*	Number of sentence pairs in broadcast data**
DGS	64 000	1 130 000
BSL	6 000	1 150 000
Std corpus for English-German	150 000 000	150 000 000

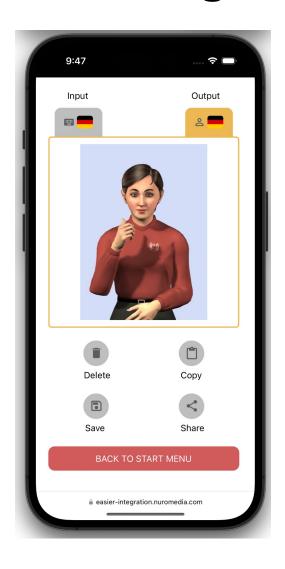
<sup>\*</sup> Public DGS Corpus, BSL Corpus

<sup>\*\*</sup> Broadcast data prepared for processing within EASIER

# **Challenge 2: Translation**





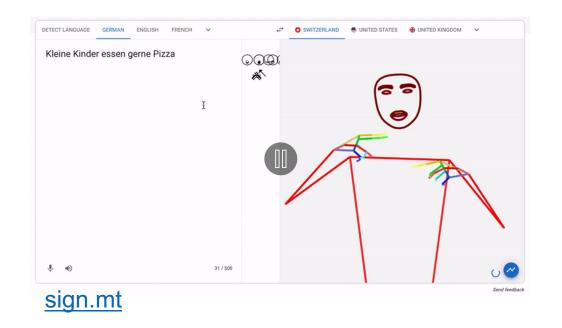


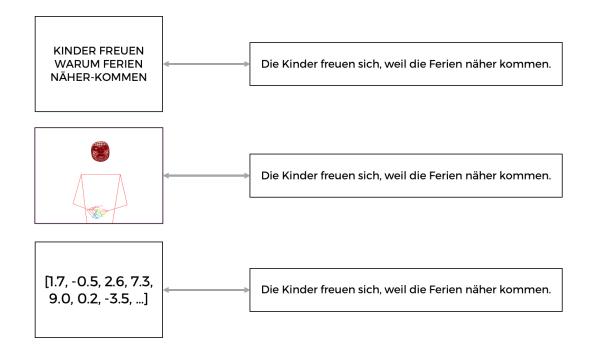
- Speech-to-text to Sign
  - → State-of-the-art machine translation
    - → Avatar presenting signed output
- Sign to text (to speech)
  - → Robust data-driven video recognition
    - → State-of-the-art machine translation
      - → Output in text

# **Challenge 2: Translation**





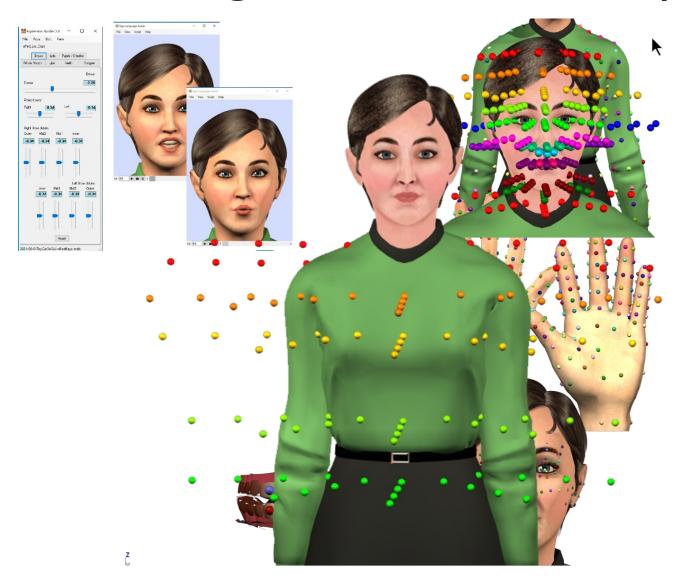




# **Challenge 3: Presentation (Avatar)**







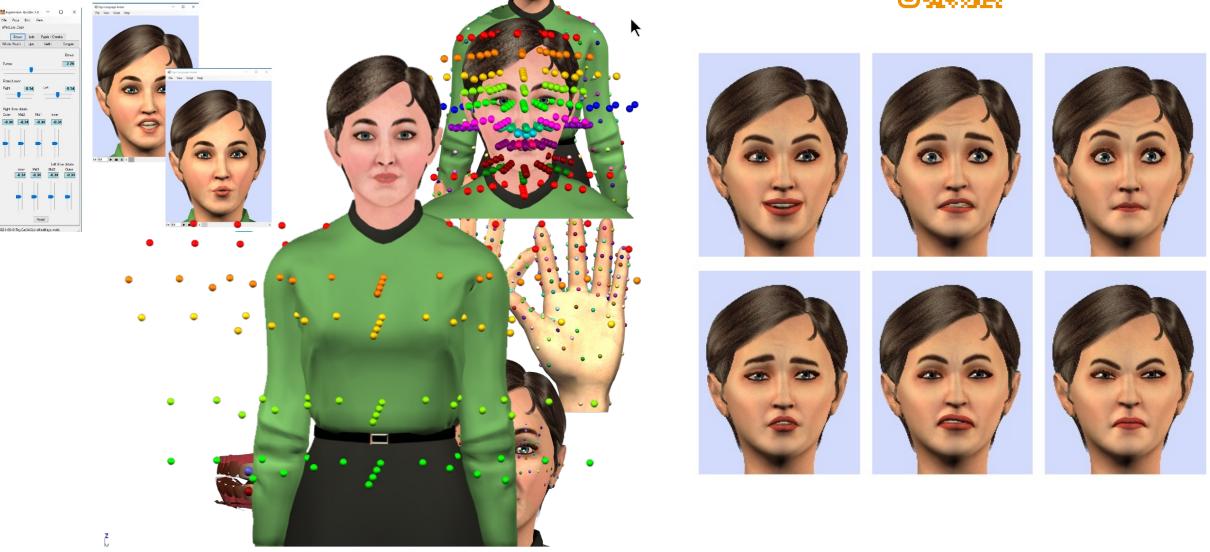




# **Challenge 3: Presentation with affect**





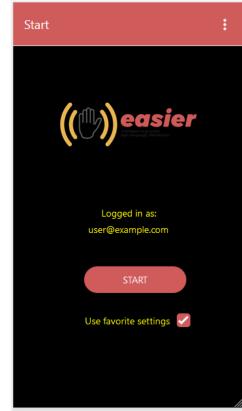


# **Challenge 4: Interaction**





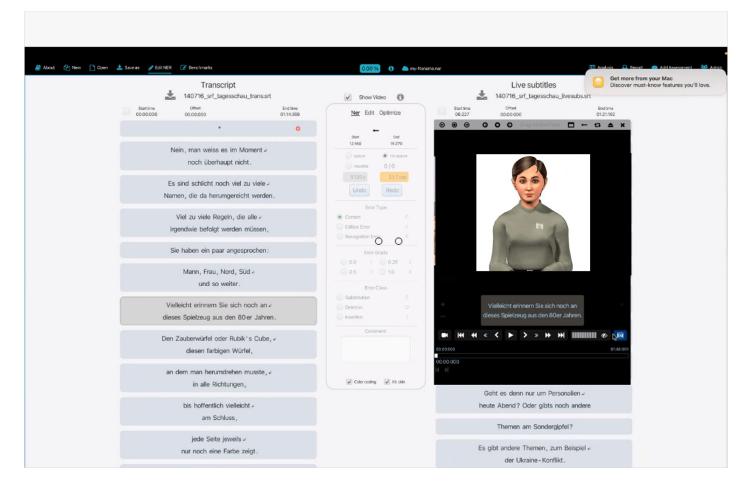




# **Challenge 5: Post-editing (NERstar)**



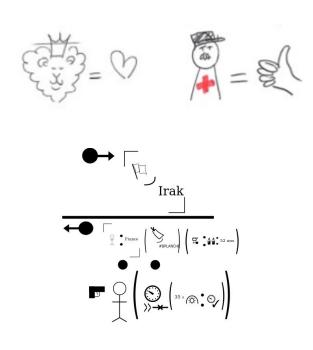


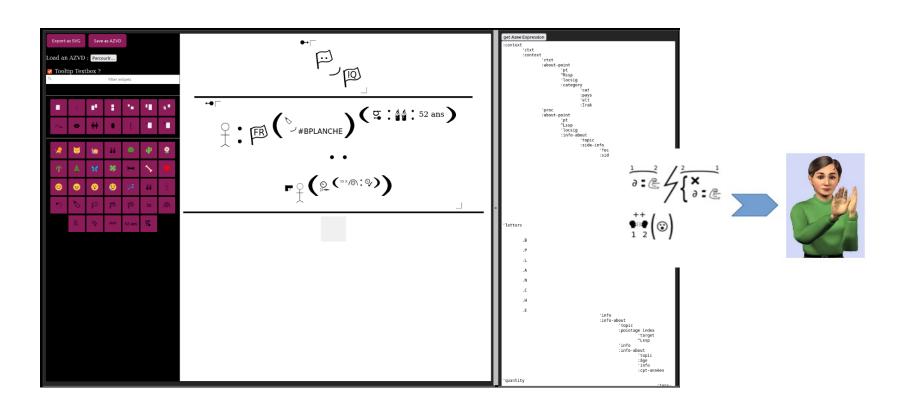


# **Challenge 5: Post-editing (Azee)**









# **Challenge 6: Sign Neologisms**







## Challenge 7: Raise awarness and disseminate





#### **Meet the EASIER Consortium**

















































# Thanks for your attention!

Giacomo Inches giacomo.inches@martel-innovate.com