easier

intElligent Automatic SIgn languagE tRanslation

The EASIER project at a glance

Dr. Eleni Efthimiou
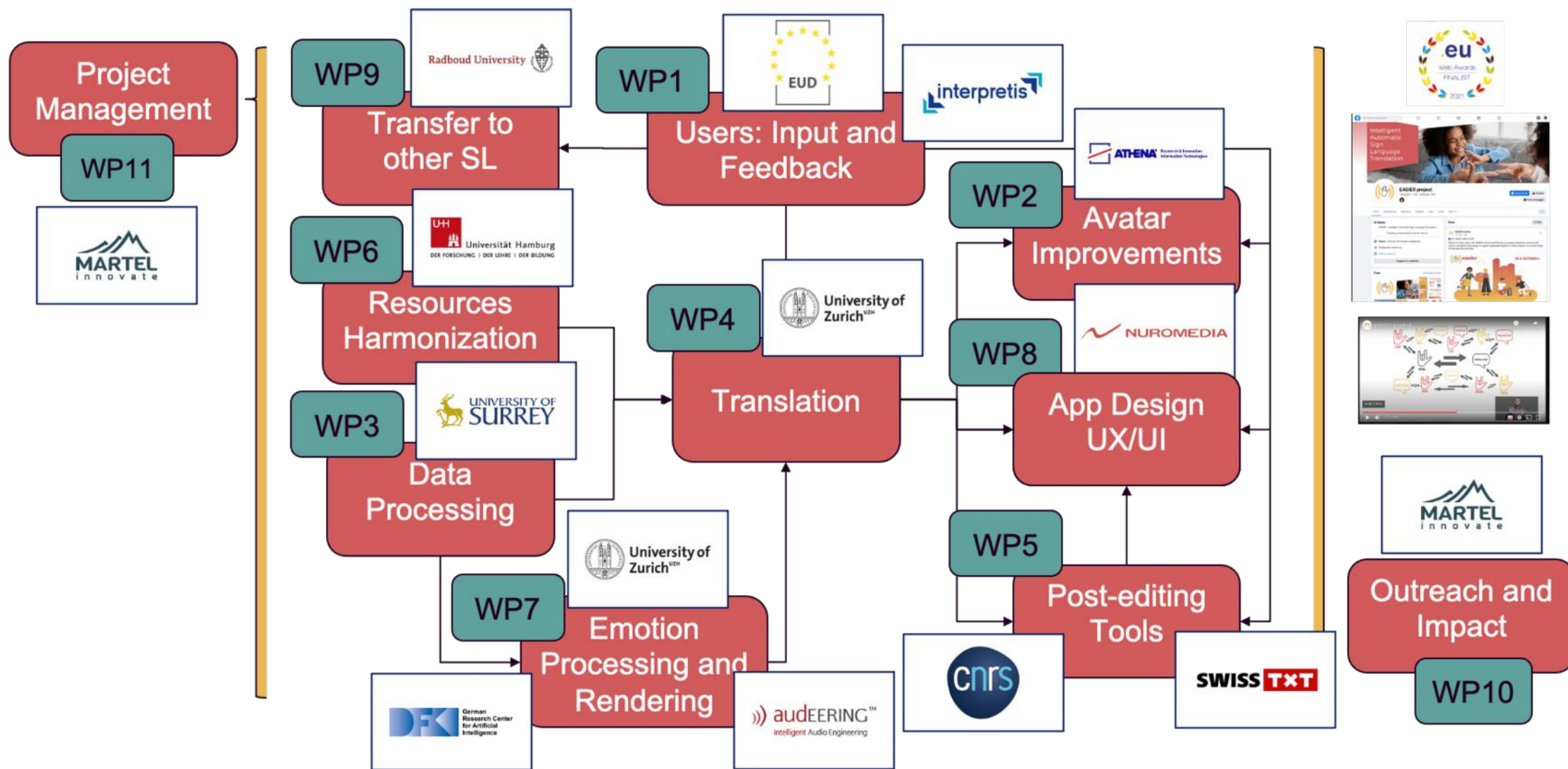
EASIER Project Scientific Coordinator

Research Director, ATHENA RC/ILSP

eleni_e@athenarc.gr

EASIER goals aligned with ICT57 call

- 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 automated route

    Goals shared with the SignOn project (also funded within EU ICT-57)

Translation directions
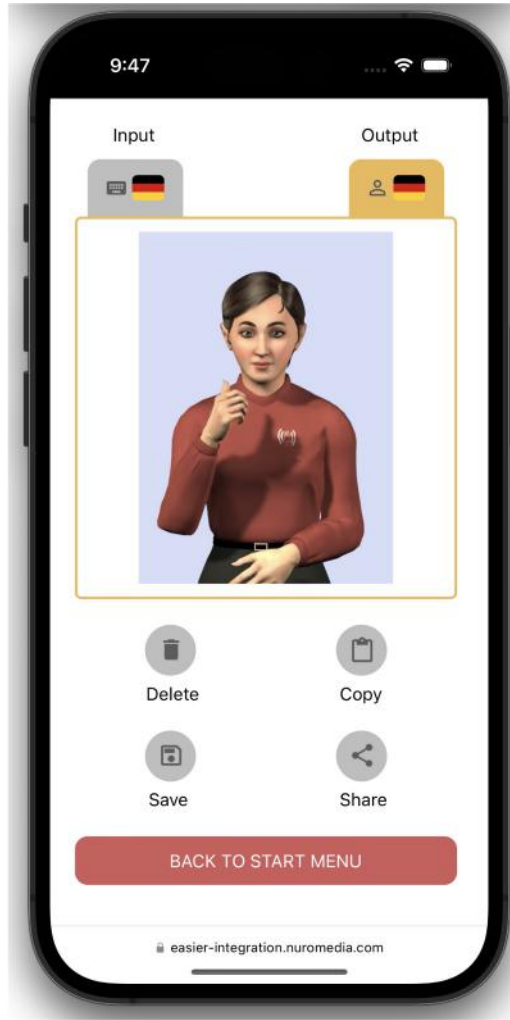
- **Speech-to-text to sign**

  State-of-the-art machine translation

  Avatar presenting signed output

- **Sign to text (to speech)**

  Robust data-driven video recognition

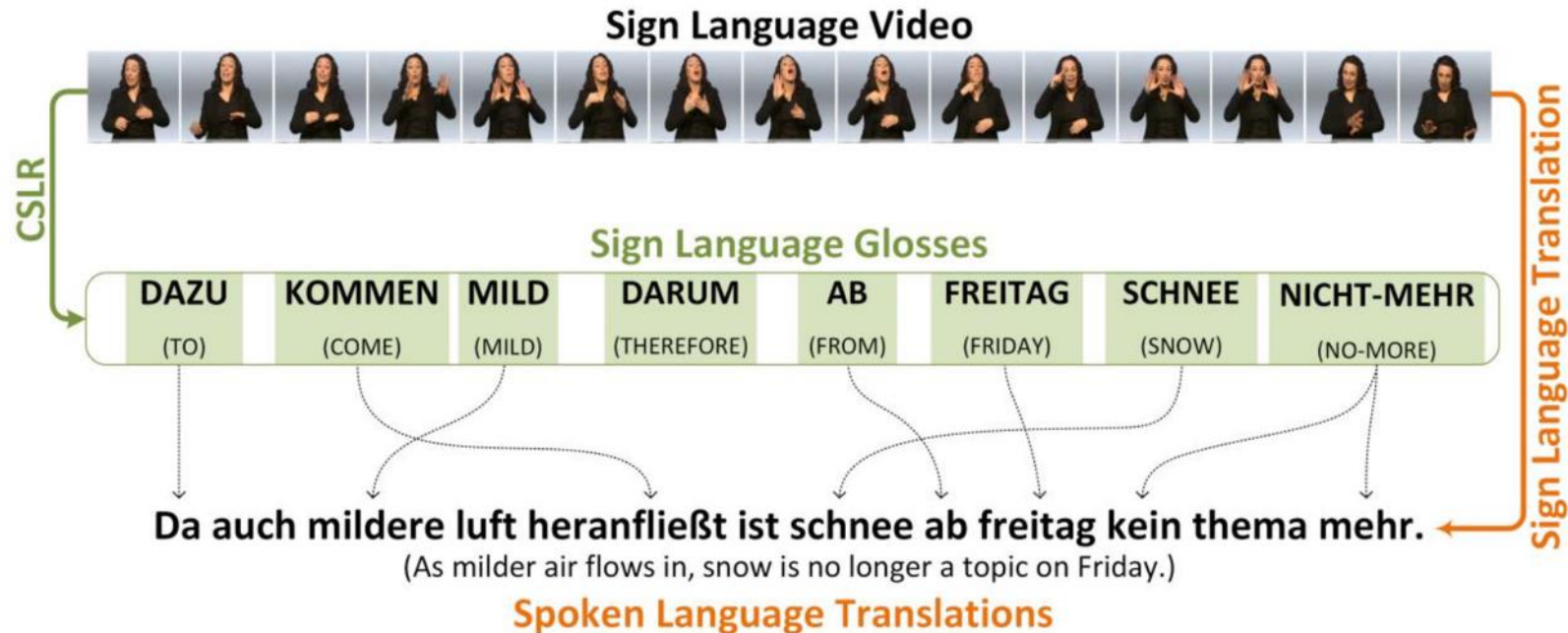 State-of-the-art machine translation

  Output in text

Main challenges

- 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 vs rule based)

- Presentation of automatic translation results
  - in a Mobile interface
  - via Avatar (virtual human)

- User acceptance and quality of translation
  - Post-editing
  - End users always in the loop
  - Continuous evaluation

Data: Orders of magnitude

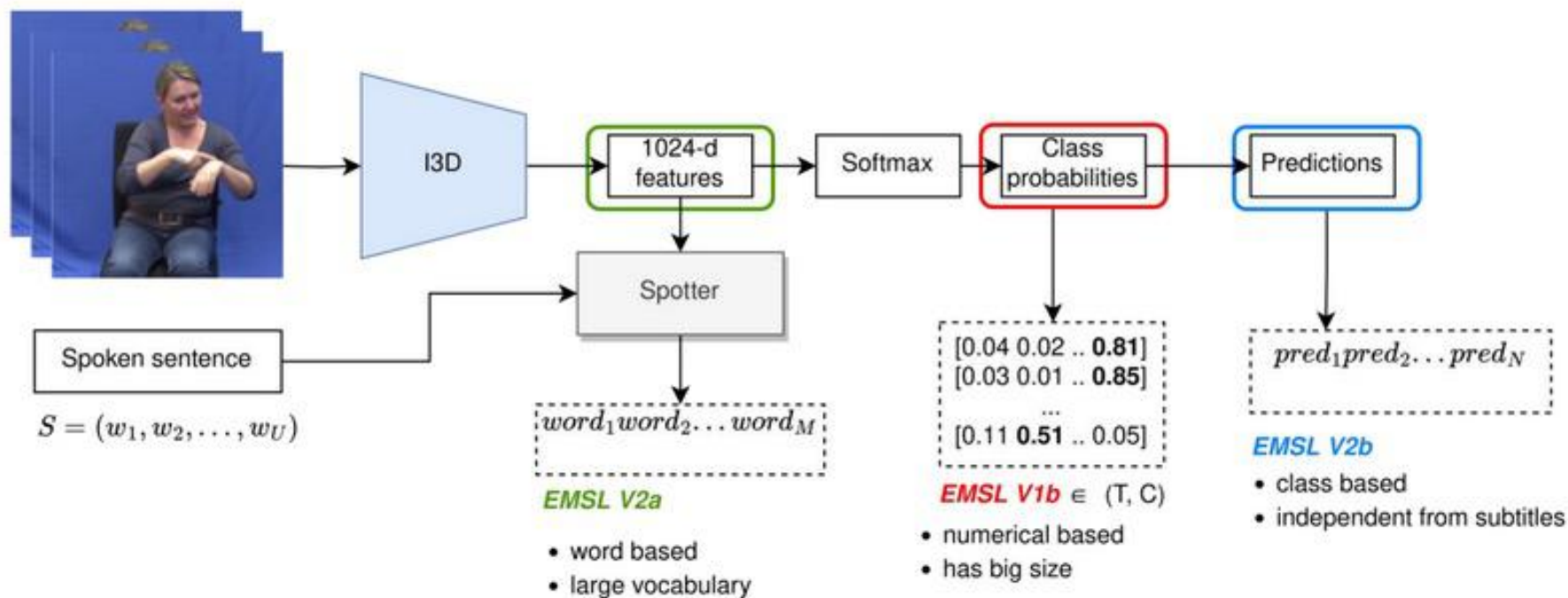| Language	Number of sentence pairs in corpus*	Number of sentence pairs in broadcast data**
DGS	64 000	1 130 000
BSL	6 000	1 150 000
Std corpus for English-German	150 000 000	150 000 000

* Public DGS Corpus, BSL Corpus

** Broadcast data prepared for processing within EASIER

Sign Language Video

CSLR

Sign Language Glosses

DAZU	KOMMEN	MILD	DARUM	AB	FREITAG	SCHNEE	NICHT-MEHR
(TO)	(COME)	(MILD)	(THEREFORE)	(FROM)	(FRIDAY)	(SNOW)	(NO-MORE)

Sign Language Translation

Da auch mildere luft heranfließt ist schnee ab freitag kein thema mehr.
(As milder air flows in, snow is no longer a topic on Friday.)

Spoken Language Translations

- CSLR: **C**ontinuous **S**ign **L**anguage **R**ecognition
- We use "Recognition" to refer to identifying specific signs within a continuous video sequence.
- Recognizing the sign is similar to "glossing" the data, which is what a linguist would do to annotate a sign language video.
- However, **translation means converting the underlying message into the equivalent spoken language sentence** (sign->spoken).
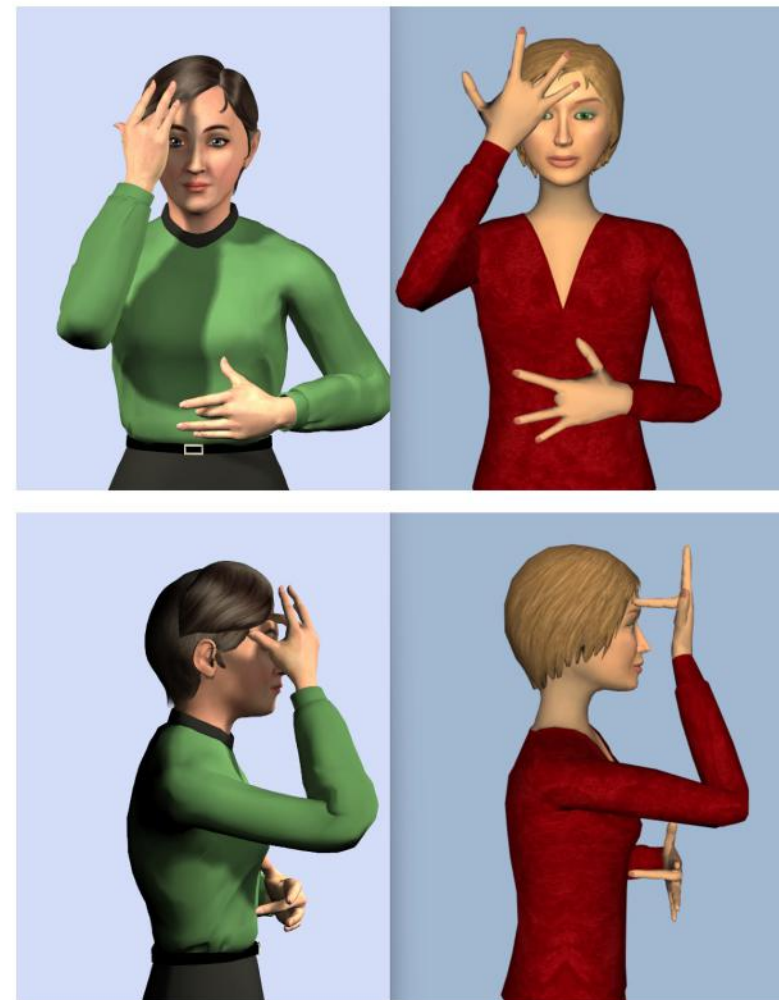
- The idea is to describe sign language video representation for all European SLs.

$$V = (x_1, x_2, \ldots, x_T)$$

$$S = (w_1, w_2, \ldots, w_U)$$

EMSL V2a
- word based
- large vocabulary

EMSL V1b $\in$ (T, C)
- numerical based
- has big size

EMSL V2b
- class based
- independent from subtitles

Affect representation in avatar signing

Incorporating emotion markers

(Signing in LSF)

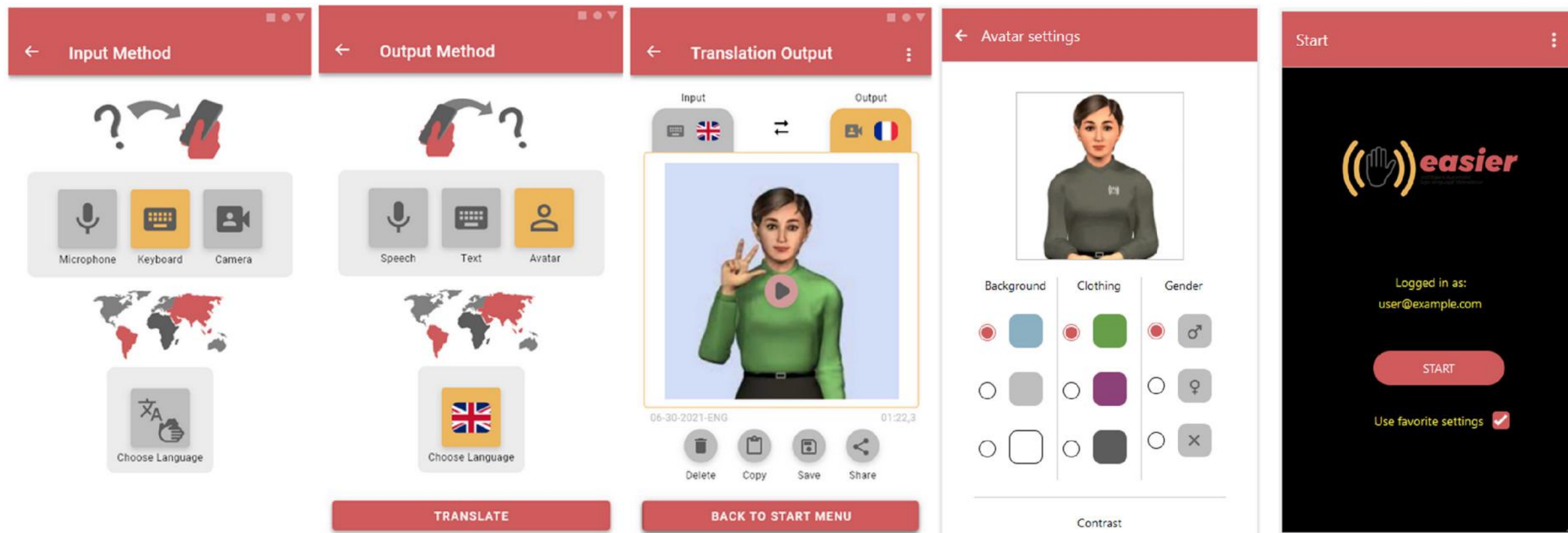incorporating non-manual SL articulators

(Signing in GSL)

incorporating mouthing features

(Signing in DGS)

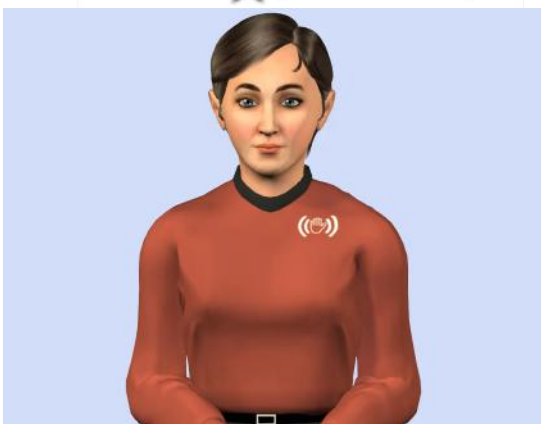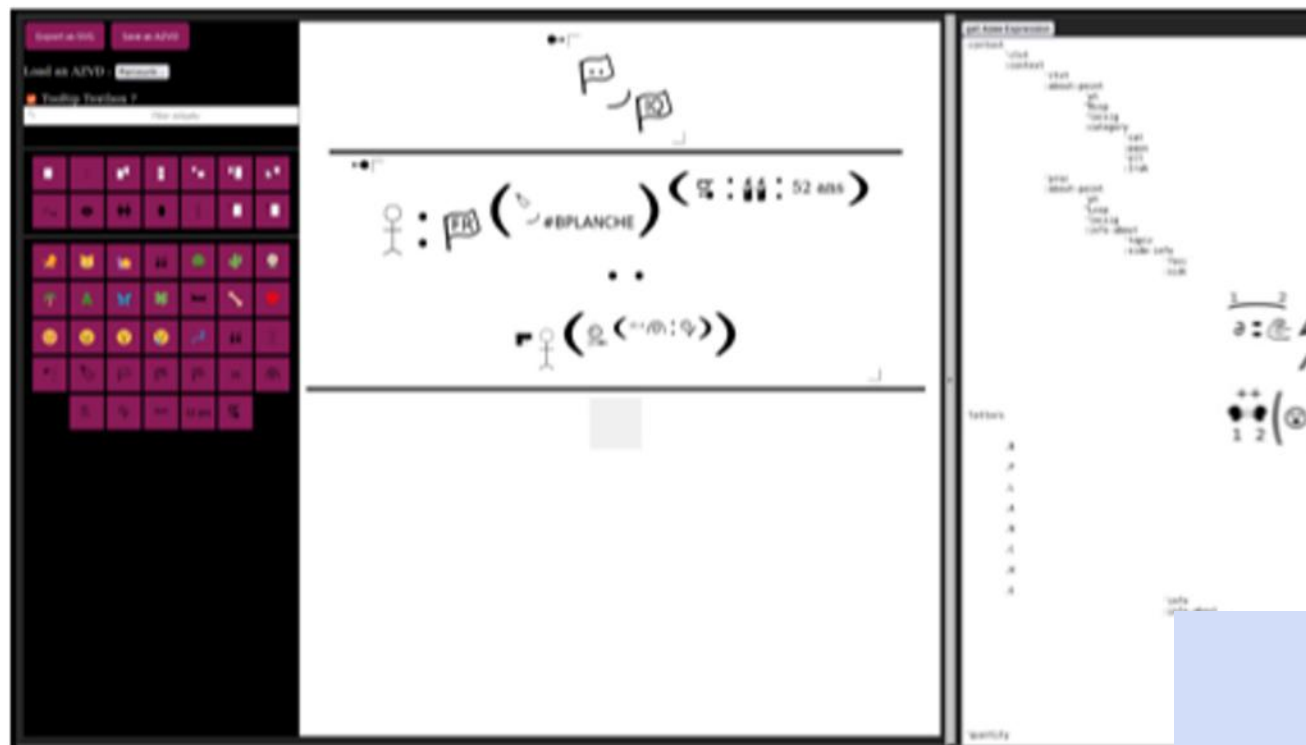# Interaction via a mobile app with low vision support

Post-editing (NERstar)

User involvement

- **User centric design**

- **User evaluation**

Mobile App

Translation

Animation

Country	Sign–Spoken pairing	Lexical resources	Corpora
Germany	DGS - German	high coverage	high coverage
Netherlands	NGT - Dutch	high coverage	high coverage
United Kingdom	BSL - English	high coverage	high coverage
Finland	FinSL/SVK - Finnish	high coverage	high coverage
Sweden	STS / SSL - Swedish	high coverage	high coverage
Belgium	LSFB - French	high coverage	high coverage
Belgium	VGT - Dutch	high coverage	high coverage
Greece	GSL - Greek	high coverage	high coverage
Poland	PJM - Polish	high coverage	high coverage
Denmark	DTS - Danish	high coverage	some coverage
France	LSF - French	some coverage	high coverage
Switzerland	DSGS - Swiss German	high coverage	data, but amount not known
Ireland	ISL - English	some coverage	some coverage
Italy	LIS - Italian	some coverage	some coverage
Slovenia	SZJ - Slovene	some coverage	some coverage
Spain	LSE - Spanish	some coverage	data, but amount not known
Austria	ÖGS - German	some coverage	data, but amount not known
Czech Republic	CZJ - Czech	some coverage	unknown!
Hungary	HSL - Hungarian	unknown!	some coverage
Finland	FinSSL - Finnish	data, but amount not known	data, but amount not known

Country	Sign–Spoken pairing	Lexical resources	Corpora
Norway	NTS - Norwegian	data, but amount not known	data, but amount not known
Slovakia	SPJ / SPR - Slovak	data, but amount not known	unknown!
Bulgaria	BŽE - Bulgarian	unknown!	unknown!
Croatia	CSQ - Croatian	unknown!	unknown!
Cyprus	ASL/GSL - Greek	unknown!	unknown!
Cyprus	TID - Turkish	unknown!	unknown!
Estonia	EVK - Estonian	unknown!	unknown!
Latvia	LZV - Latvian	unknown!	unknown!
Liechtenstein	DSGS - German	unknown!	unknown!
Lithuania	LGK - Lithuanian	unknown!	unknown!
Luxembourg	DGS - French	unknown!	unknown!
Luxembourg	DGS - German	unknown!	unknown!
Malta	LSM - Maltese	unknown!	unknown!
Malta	LSM - English	unknown!	unknown!
Portugal	LGP - Portuguese	unknown!	unknown!
Romania	LMGR - Romanian	unknown!	unknown!
Romania	HSL - Hungarian	unknown!	unknown!
Spain	LSC - Catalan / Spanish	unknown!	unknown!
Spain	LSCV (Valencia) - Spanish	unknown!	unknown!
Switzerland	LSF-SR - French	unknown!	unknown!

Legend:
- high coverage
- some coverage
- data, but amount not known
- unknown!

Https://www.Project-easier.Eu/deliverables/
D9.2: An overview of resources in the making

FIGURE 1: : COVERAGE OF LANGUAGE RESOURCES (LEXICAL, CORPUS) IN LANGUAGES OF THE EUROPEAN UNION, LISTED BY LANGUAGE PAIRING (SIGN LANGUAGE - SPOKEN LANGUAGE PAIR)

Compendium

https://www.sign-lang.uni-hamburg.de/lr/compendium

D6.1

new : available as

single PDF

42
Corpora

84
Lexical Resources

27
Collection Tasks

82
Sign Languages

Wordnet

http://sign-net.meine-dgs.de

- D6.3 : index for DGS and GSL
- D6.4 : index for all core languages of the project
- D6.5 : more languages outside of the project

EASIER AUTUMN SCHOOL

« All Events

This event has passed.

EASIER Autumn School

September 25 – September 26

Autumn School: Sign language data meets data science – data science meets sign linguistics

The goal of the Autumn School is to generate expertise for under-resourced sign languages to extend the scope of EASIER to more European sign languages. The idea is not only to support people who already work with sign languages and train them in technological approaches, but also to train people who already work with in data science, language technologies, etc. in the handling of sign languages. Therefore, both groups will have a general input in the beginning on the subject that is new to them.
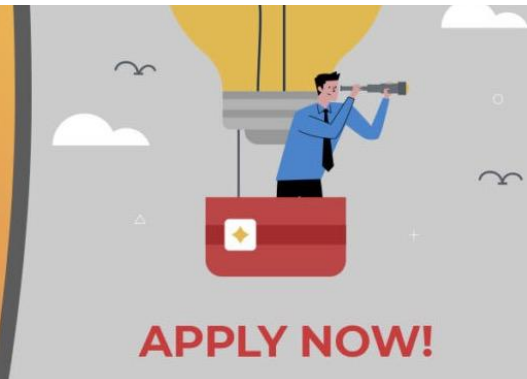
THANKS

EASIER project is funded by the EU's Horizon2020 Programme under Grant Agreement number 101016982

PROJECT-EASIER.EU **@EASIERPROJECT**