

# Facilitating the Spread of New Sign Language Technologies across Europe

Hope E. Morgan<sup>1</sup>, Onno Crasborn<sup>1</sup>, Maria Kopf<sup>2</sup>, Marc Schulder<sup>2</sup>, Thomas Hanke<sup>2</sup>  
<sup>1</sup> Radboud University, Netherlands; <sup>2</sup> University of Hamburg, Institute of German Sign Language and Communication of the Deaf, Germany

The EASIER project for automatic translation of sign languages adopts the view that **annotated corpora associated to rich lexical databases** have an important role to play in developing and improving language technologies for the automatic translation of sign languages. Doing so requires several steps, and should also include a pathway for under-resourced sign languages to be included. This aspect of the larger EASIER project is outlined here around three main questions...

## Q1. What is the current status of sign language datasets across Europe?



Overview of the Datasets for the Sign Languages of Europe (Kopf, Schulder, & Hanke 2021)

This report describes all the known sign language resources in the EU that would meet the needs of the EASIER pipeline. It identifies and describes **26 corpora** and **41 lexical resources**, covering 24 European sign languages

## Q2. Given variation in annotation and coding across datasets, how can they be aligned for the same pipeline?



Specification for the Harmonization of Sign Language Annotations (Kopf, Schulder, Hanke, & Bigeard 2022)

This technical report contains two key contributions:

- (1) Analysis of 20+ aspects of differences in annotation standards between the existing corpora (e.g., segmentation, compounds, directional verbs)
- (2) Proposes a basic single unified **interchange format in JSON container** to encode information relevant to EASIER translation pipeline

## Q3. How to prepare under-resourced languages for integration into this pipeline?



### Identify new language documentation projects

Outreach to both academia and deaf associations, especially to countries with gaps in documentation (yellow in chart & map) to find new, unknown, or upcoming documentation projects



### Definition of minimal contents of dataset for participation

Overview of what would be minimally necessary to create quality datasets usable for language technologies used in automatic translation of sign languages



### Workflow documents

Materials that will be a guide to those starting language documentation projects. These materials will cover both linguistic and technological questions about annotation and coding. These will also support training sessions



### Training sessions for new documentation projects

Training workshop offered to language documentation teams, prioritizing those with current, new, and upcoming projects of less-resourced languages



### Infrastructure to automatically analyze other datasets

Video processing services in the form of an infrastructure running on high-performance clusters will be offered to less-resourced languages; i.e., for the use of state-of-the-art 2D pose estimation techniques used to feed translation pipelines



Lexical resources	Corpora	Country	Sign-Spoken pairing
■	■	Germany	DGS - German*
■	■	Netherlands	NGT - Dutch*
■	■	United Kingdom	BSL - English*
■	■	Finland	FinSL/SVK - Finnish
■	■	Sweden	STS / SSL - Swedish
■	■	Belgium	LSFB - French
■	■	Belgium	VGT - Dutch
■	■	Greece	GSL - Greek*
■	■	Poland	PJM - Polish

\* Main language pairs in EASIER project (6)

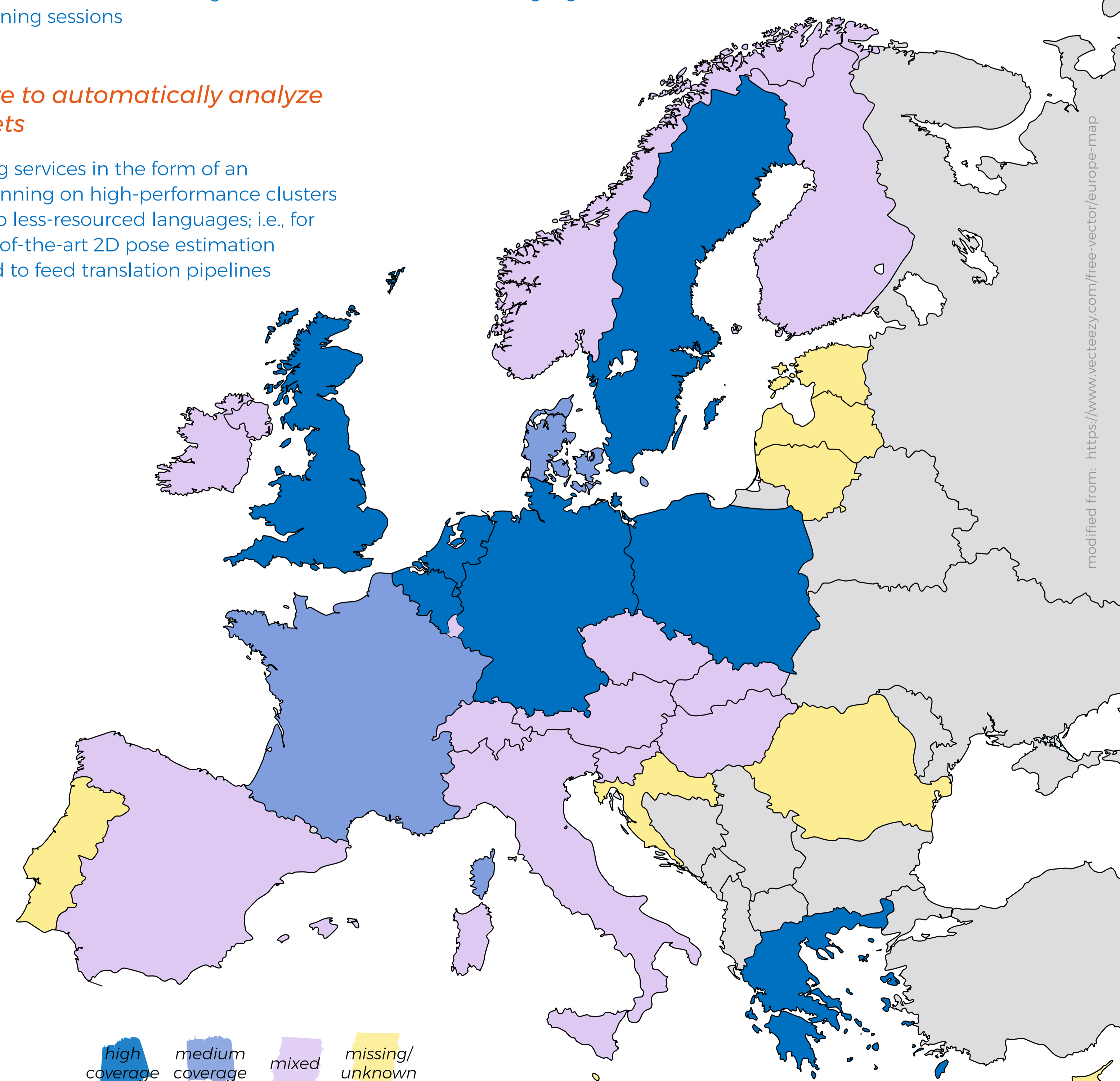
■	■	Denmark	DTS - Danish
■	■	France	LSF - French*
■	■	Switzerland	DSGS - Swiss German
■	■	Ireland	ISL - English
■	■	Italy	LIS - Italian*
■	■	Slovenia	SZJ - Slovene
■	■	Spain	LSE - Spanish
■	■	Austria	ÖGS - German
■	■	Czech Republic	CZJ - Czech
■	■	Hungary	HSL - Hungarian
■	■	Finland	FinSSL - Finnish
■	■	Norway	NTS - Norwegian
■	■	Slovakia	SPJ / SPR - Slovak

■	■	Bulgaria	BŽE - Bulgarian
■	■	Croatia	CSQ - Croatian
■	■	Cyprus	ASL/GSL - Greek
■	■	Cyprus	TID - Turkish
■	■	Estonia	EVK - Estonian
■	■	Latvia	LZV - Latvian
■	■	Liechtenstein	DSGS - German
■	■	Lithuania	LGK - Lithuanian
■	■	Luxembourg	DGS - French
■	■	Luxembourg	DGS - German
■	■	Malta	LSM - Maltese
■	■	Malta	LSM - English
■	■	Portugal	LGP - Portuguese
■	■	Romania	LMGR - Romanian
■	■	Romania	HSL - Hungarian
■	■	Spain	LSC - Catalan / Spanish
■	■	Spain	LSCV (Valencia) - Spanish
■	■	Switzerland	LSF-SR - French

■ high coverage  
 ■ some coverage  
 ■ data, but amount not known  
 ■ unknown!

Help us find new sign language documentation projects!

contact:  
[hope.morgan@ru.nl](mailto:hope.morgan@ru.nl)



Map showing the extent of sign language documentation in SLs in EU countries

modified from: <https://www.vecteezy.com/free-vector/europe-map>