SignGPT and the Visual Language Toolkit

Matt Brown[1], Oline Ranum[2], Edward Fish[2], Heidi Proctor[1], Richard Bowden[2], Bencie Woll[1], Kearsy Cormier[1]

[1] DCAL, University College London
[2] CVSSP, University of Surrey

The data problem

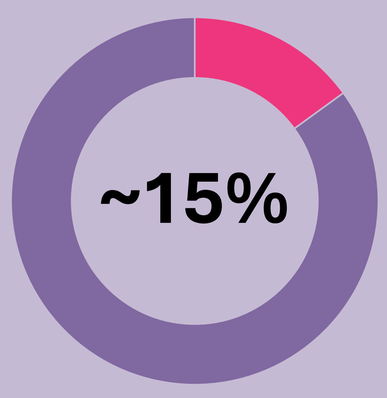
Many potential sources of sign language video data have **ecological validity issues:**

Interpreted media
Errors, omissions, propositional flattening

Broadcast captioning
Non-verbatim, errors, omissions, substitutions

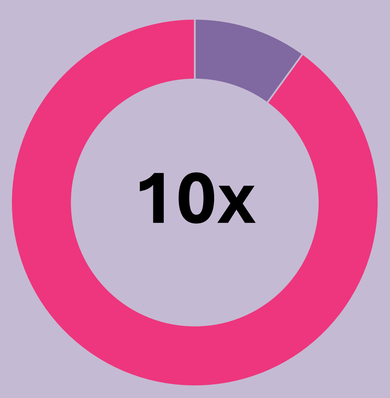
Social media
Consent, licensing, provenance unknown

Instead, we should be using carefully annotated corpora, grounded in sign language linguistics. But their creation is slow and resource-heavy. **Human annotators need better AI tools.**



BSL Corpus video data annotated so far

Proportion of non-lexical BSL constituents, weakly captured by glossing

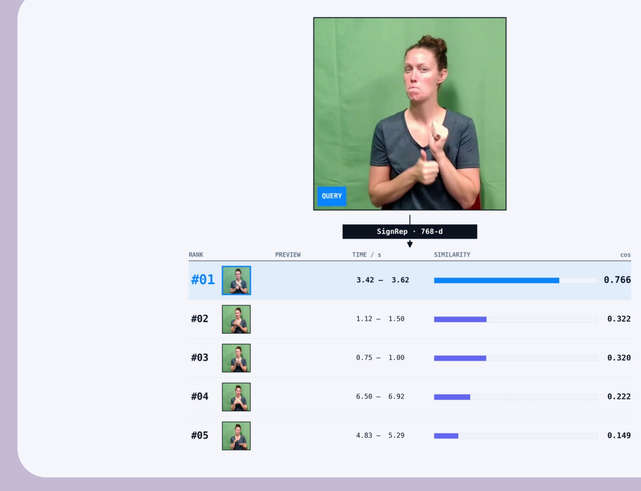


Target expansion of annotated BSL resources

The Visual Language Toolkit (VLTK): work in progress

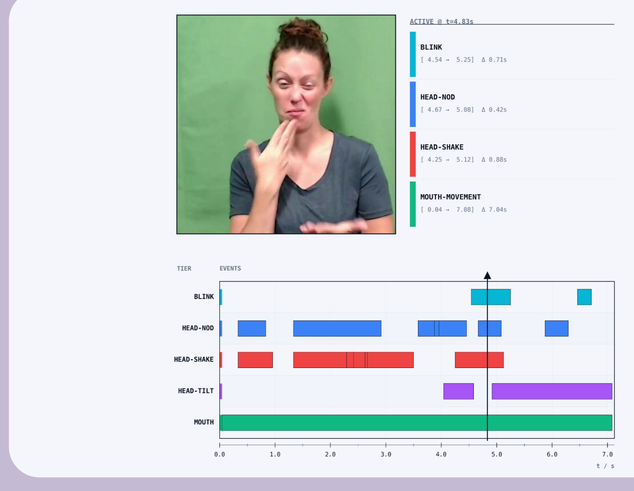
Segmentation

BIO self-attention model predicts sign boundaries from skeletal pose and hand data



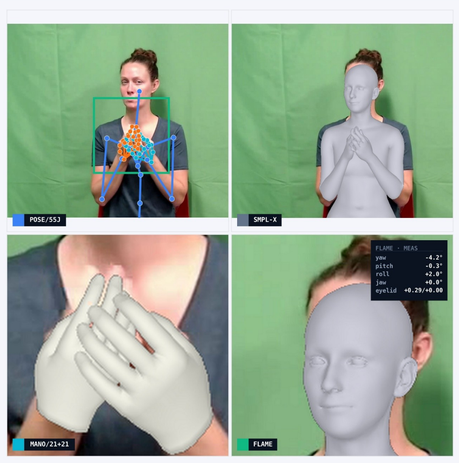
Lexical sign spotting

SignRep masked autoencoder learns pose-based embeddings, compared against a reference source



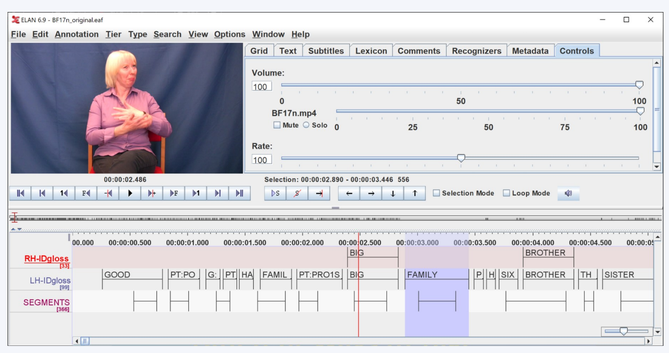
Non-manual features

Confidence estimates for blinks, head tilts, nods, gaze, and brow movement (Teaser face estimation)



Pose extraction pipeline

Keypoints and pose estimations derived from various models (SMPL-X, WiLoR, FLAME)



Export to ELAN

Each tool is interoperable with ELAN, e.g. for ease of annotator review or comparison with existing data

What else do you need?
Let us know!

For more information please contact:
Matt Brown (m.brown.13@ucl.ac.uk)
DCAL, 49 Gordon Square, London WC1H 0PD, UK

SignGPT is a collaboration between CVSSP (Surrey University), VGG (Oxford University), and DCAL (University College London), funded by the EPSRC (EP/Z535370/1).