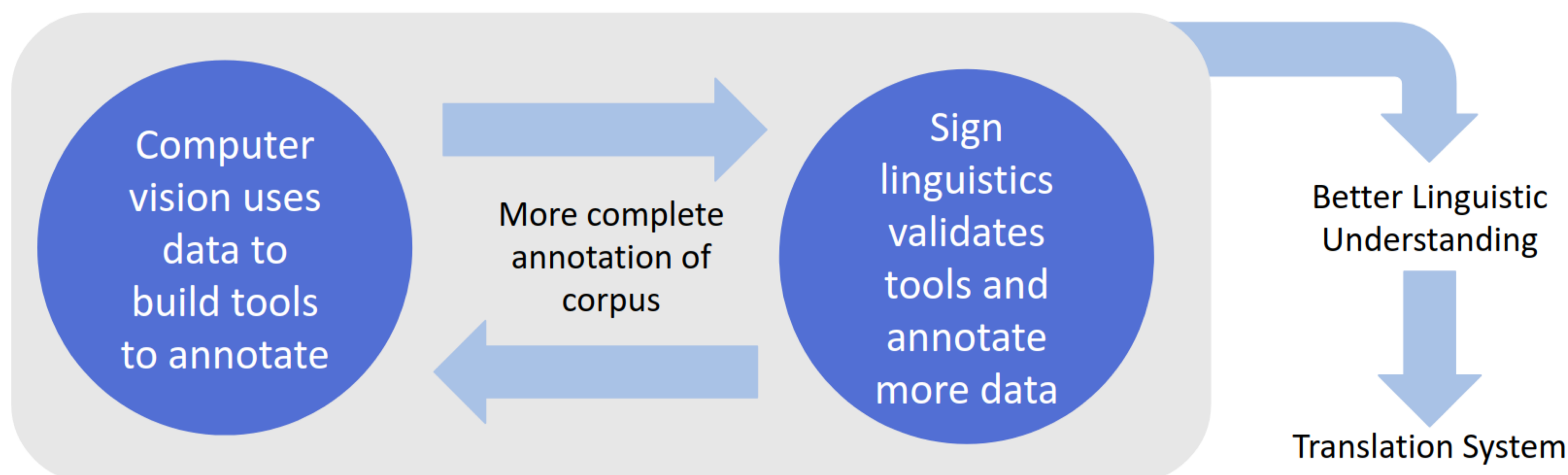


ExTOL: End to End Translation of BSL

A partnership between sign linguists and computer scientists

- Computer vision is undergoing a revolution in terms of deep learning, breaking new ground
- Linguists want to better understand the language, its structure and use – achieved by recording and studying language data
- However, annotation of that data is extremely slow and laborious. Linguists need tools to help them automate the annotation process



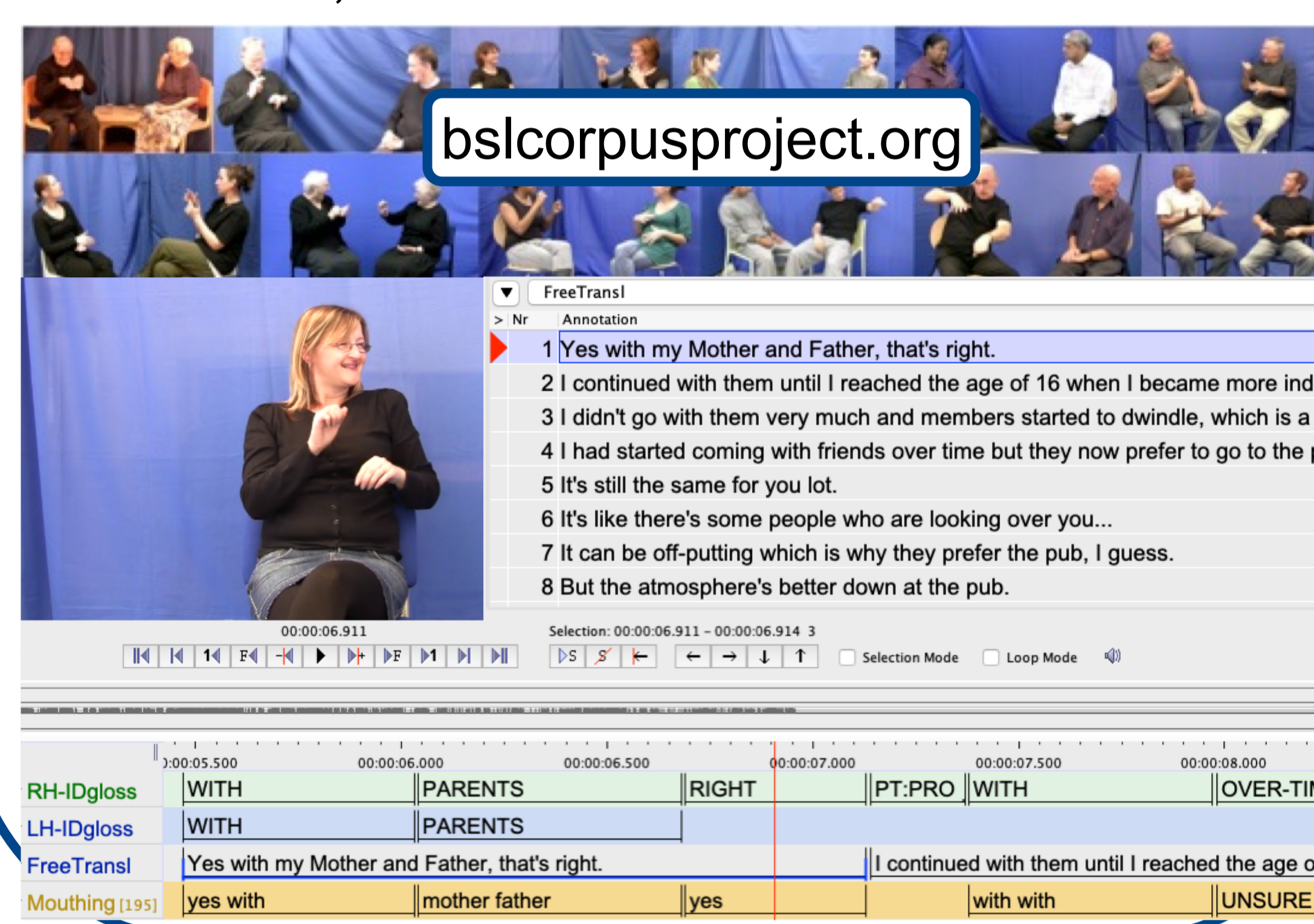
Research Questions

1. What manual/non-manual linguistic annotation is needed for reliable machine translation from sign to spoken language?
2. What automatic annotation tools will provide the most utility to linguistic study?
3. What can machines do better than humans, and vice-versa?

Here focus on: English mouthings and baseline translation tests

BSL Corpus

- Collection of ~125 hours of videos & background info from 249 deaf signers from 8 UK regions in conversation, telling stories, being interviewed and producing their signs for 102 concepts
- Videos, metadata & annotations online



Automatic Recognition of English mouthings

The Oxford-BBC Lip Reading Sentences 2 Dataset

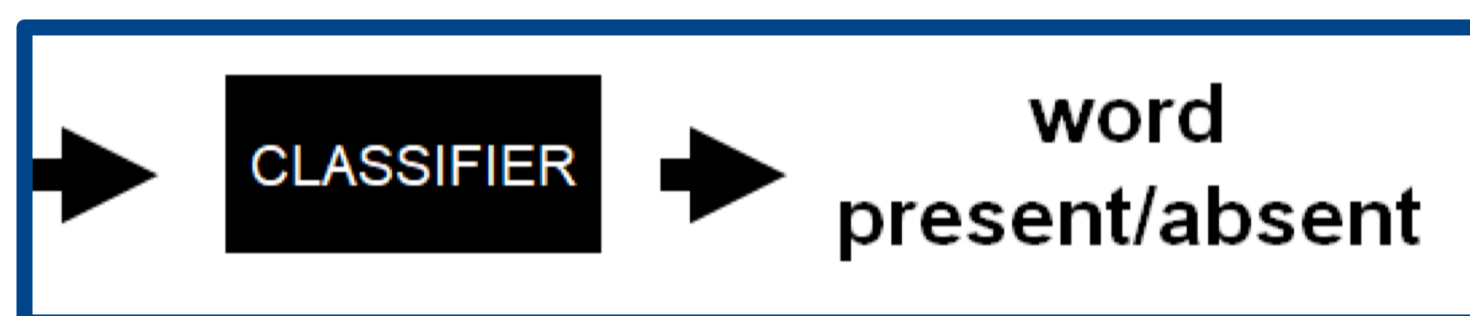
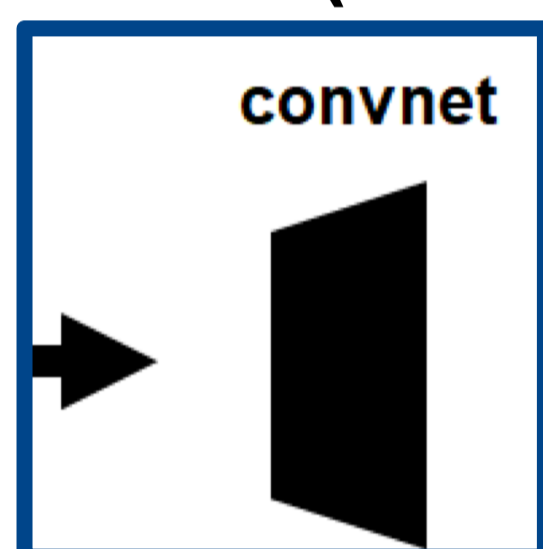
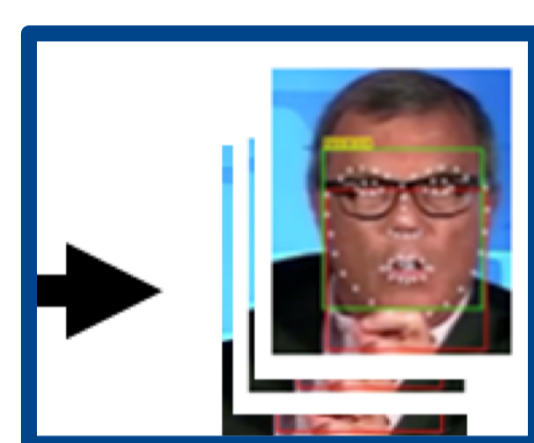
- 200 hours of video
- Word level alignment between text and video

BSL Corpus

- 50,000+ BSL IDglosses annotated
- 5,000 English Mouthing annotations

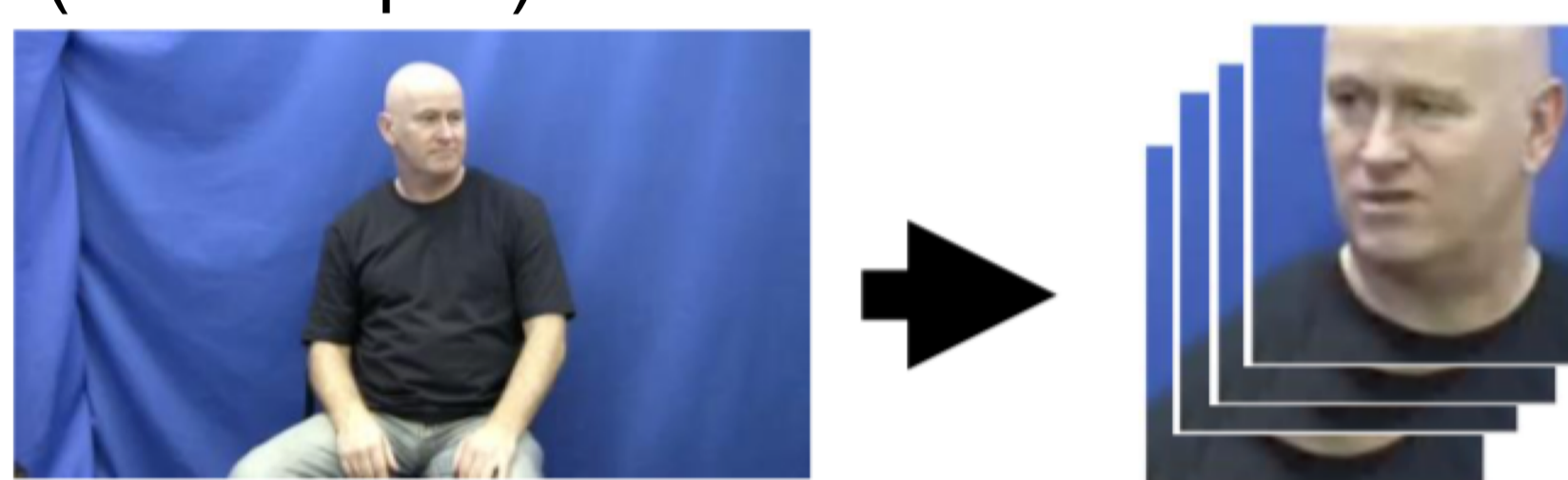


Word spotting using lip movements (no audio)



Early Results (Accuracy)

Good: "weather, insect" (BBC), "remember, hearing" (BSL Corpus)
 Bad: "million" (BBC), "but, london" (BSL Corpus)



What this will be used for

- Automatically labelling mouthings
- Use of mouthing to help identify BSL signs

Baseline Translation Tests

If we had perfect recognition of BSL, how good would the translation to English be from ID glosses alone using the state-of-the-art machine translation approaches?

Dataset:

- Parallel ID-Gloss and Free Translation
- pairs from BSL Corpus.

Approach:

- Transformers (Vaswani et al. 2017)
- Base of nearly all of the state-of-the-art NLP models including NMT.

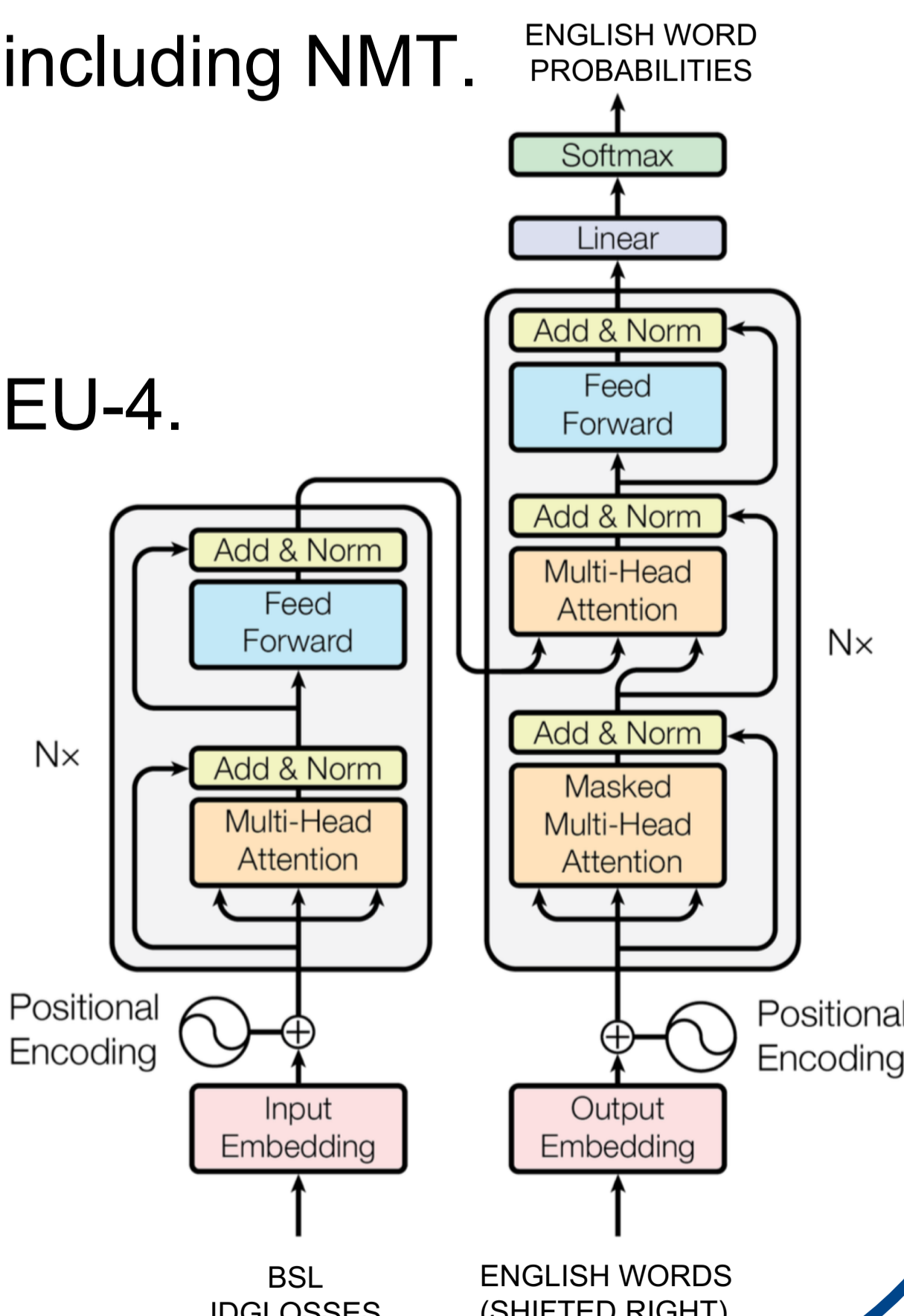
Initial Results:

- BLUE-4 Scores of 4.58 on Dev & 4.00 on Test sets.
- Baselines in spoken languages range from 20 to 40 BLEU-4.

Why is performance so low?

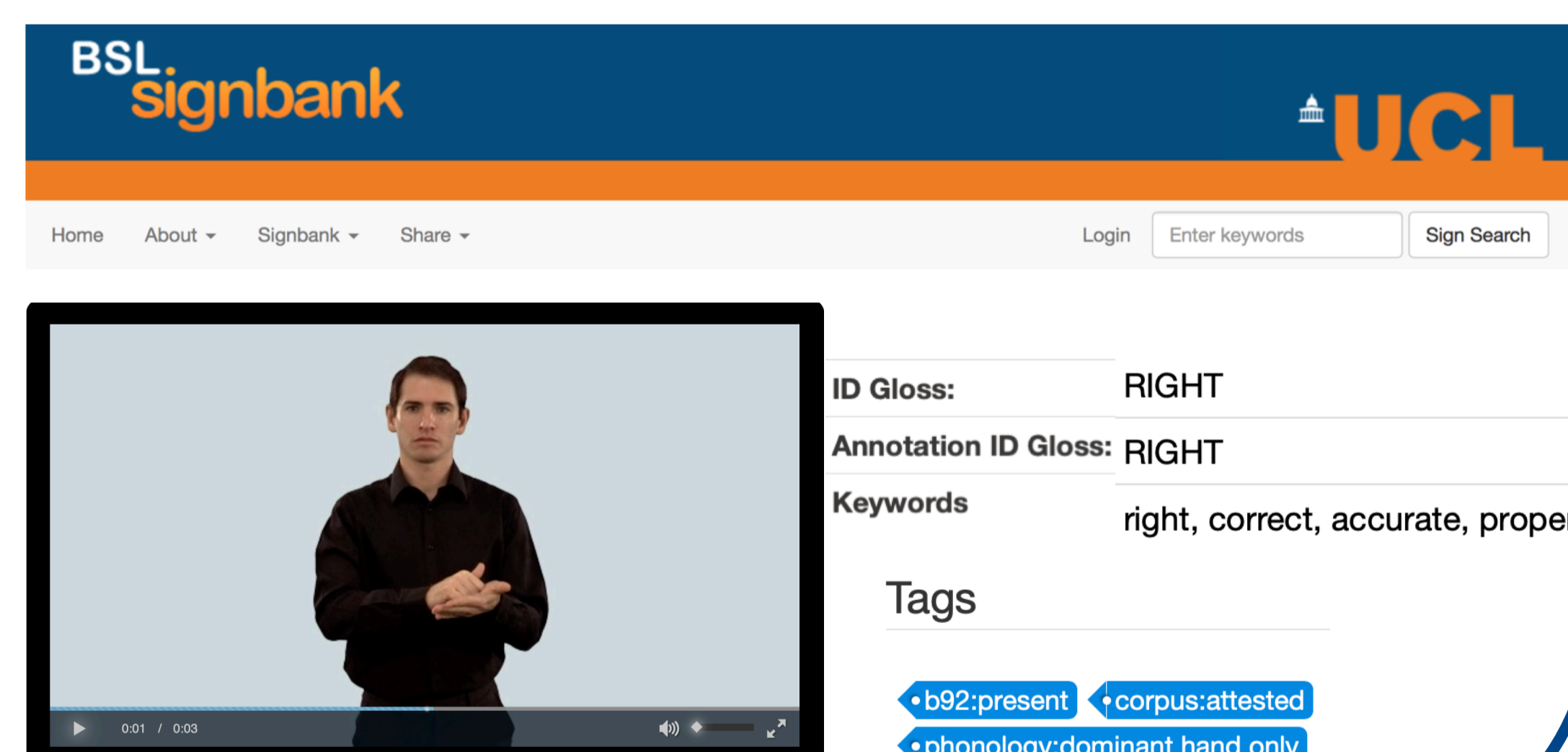
- Large domain of discourse.
 - Too many singletons, not enough data.
- Context, such as placements, is lost in single pairs
- Model longer sequences & long term dependencies.
- Signer independent gloss recognition is still unsolved.
 - Signing style and mannerisms.
- Non-manuals are not utilized.

	Sign Gloss			English		
	Train	Dev	Test	Train	Dev	Test
segments	9207	500	500	← same		
vocab.	5348	915	915	5384	1083	1081
tot. tokens	49786	2678	2729	86208	4697	4585
OOVs	-	177	145	-	116	133
tot. OOVs	-	187	158	-	124	134
singletons	2890	(588)	(580)	2401	(679)	(692)



BSL Signbank

- Usage-based dictionary of BSL based on BSL Corpus
- Lexical database with ID glosses & keywords
- Launched online in 2014 with 3500+ BSL signs & growing
- <http://bsl.signbank.ucl.ac.uk>



Long-term Goals

- **Ultimate aim of project:** World's first British Sign Language to English translation system and first practically functional machine translation system for any sign language
- **Secondary aim:** To create automatic annotation tools to be integrated into the software annotation package ELAN