

## Overview of resources

Norwegian Sign Language (NTS, *Norsk Tegnspråk*) is officially recognized in Norway, but remains under-resourced for sign language technology.

We survey existing and potential NTS resources and describe their accessibility, format, licensing and computational suitability.

### Curated datasets

Most existing resources are valuable for documentation and linguistic analysis, but cannot be easily used for data-demanding recognition and translation pipelines.

Dataset	Format	Size / Duration	Icons
Halvorsen	Video	118 MB (18 min)	🎥
Pilot Corpus	Video	8,74 GB (5,5 h)	🔒
Depicting Perspective	Video	9,8 GB (7,5 h)	🔒
Language Ecology	Video	13,8 GB + 68 GB	🎥 / 🔒
Alphabet	Images	418 MB (24.300)	🔒
Numbers	Video		🚫

🔒: openly available; 🎥: non-commercial; 🔒: available upon approval; 🚫: private.

**Gap:** no publicly available continuous-signing dataset with rich annotations ready for computational NTS research.

### Alternative data sources

Alternative sources include single-sign videos from a sign language dictionary, SignWriting lexicons, public news broadcasts and community video platforms such as tegn.tv.

Dataset	Format	Size / Duration	Icons
Tegnbanken	Video + Glosses	2,72 GB (5,29 h)	🎥
SignPuddle	Gloss + SignWriting	1,8 MB	🔒
NRK News	Video + Subtitles	38,92 GB (17,35 h)	©

🔒: open-source; 🎥: non-commercial; ©: under copyright, but available for research.

## Conclusions

### Data reuse and representation bottlenecks:

- Copyright and access restrictions limit reuse of many useful video sources.
- Signing videos contain personally identifiable information. Re-identification risk is high in small signing communities.
- Single-sign dictionary videos are accessible and clear, but slower and less representative of natural signing.
- Community-created resources vary in completeness and consistency.
- News broadcasts are interpreted; subtitles are imperfectly aligned; copyright restrictions still apply.

### SignWriting transcription:

- SignWriting is a promising representation system, but current language-independent segmentation and transcription tools are not yet reliable for NTS.
- Evaluation is limited by lack of expert-validated references.

**Takeaway:** NTS resources exist, but licensing, privacy and annotation constraints limit their usefulness for computational applications. Reliable NTS technology needs expert-validated data and robust tools (e.g. for segmentation).

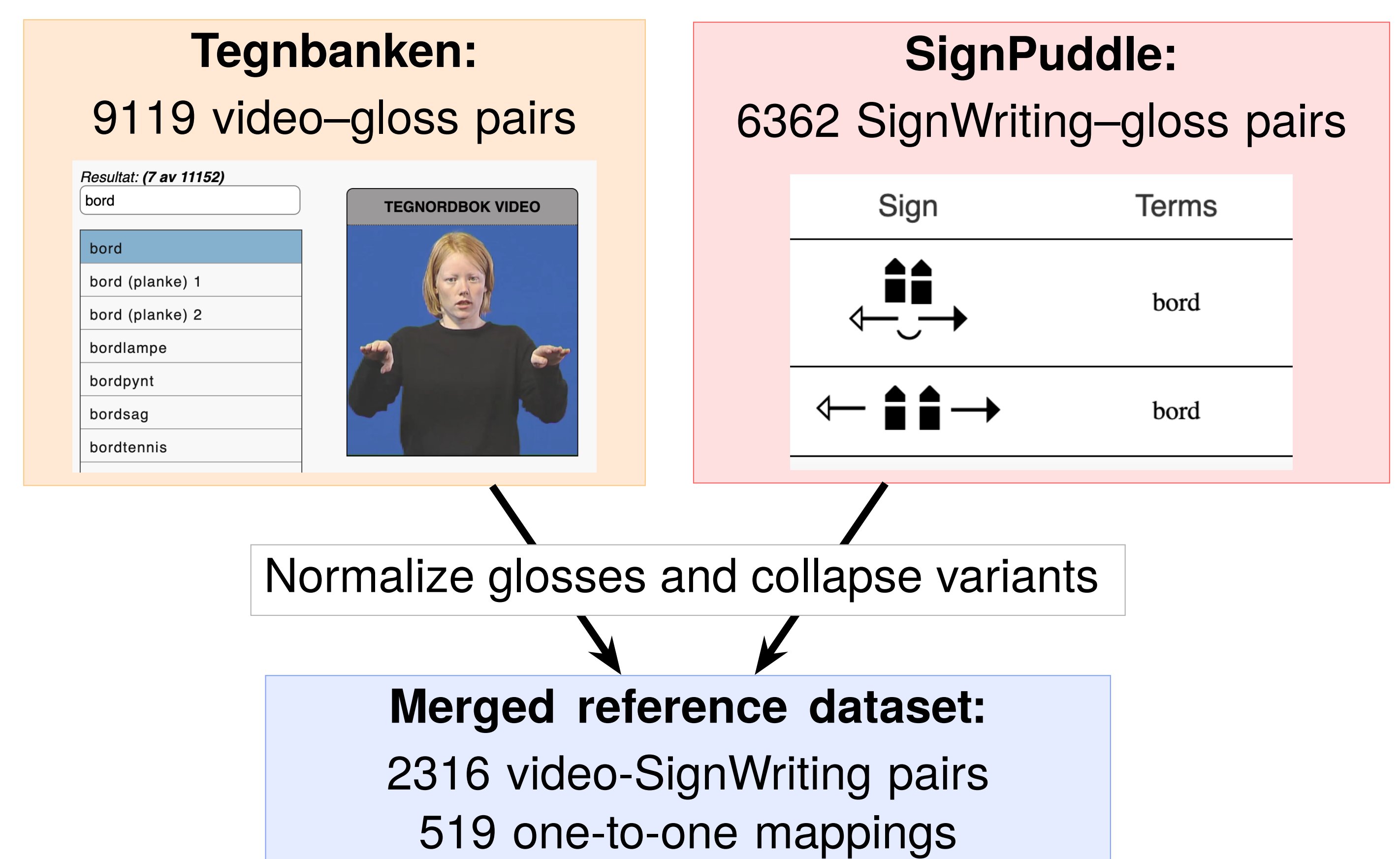
## SignWriting transcription experiments

### Key research questions:

1. Can we leverage existing language-independent transcription tools to overcome the scarcity of annotated NTS data?
2. Is SignWriting a useful language-independent representation system?
3. Long-term goal: What is the most efficient way to identify individual signs in continuous-signing videos?

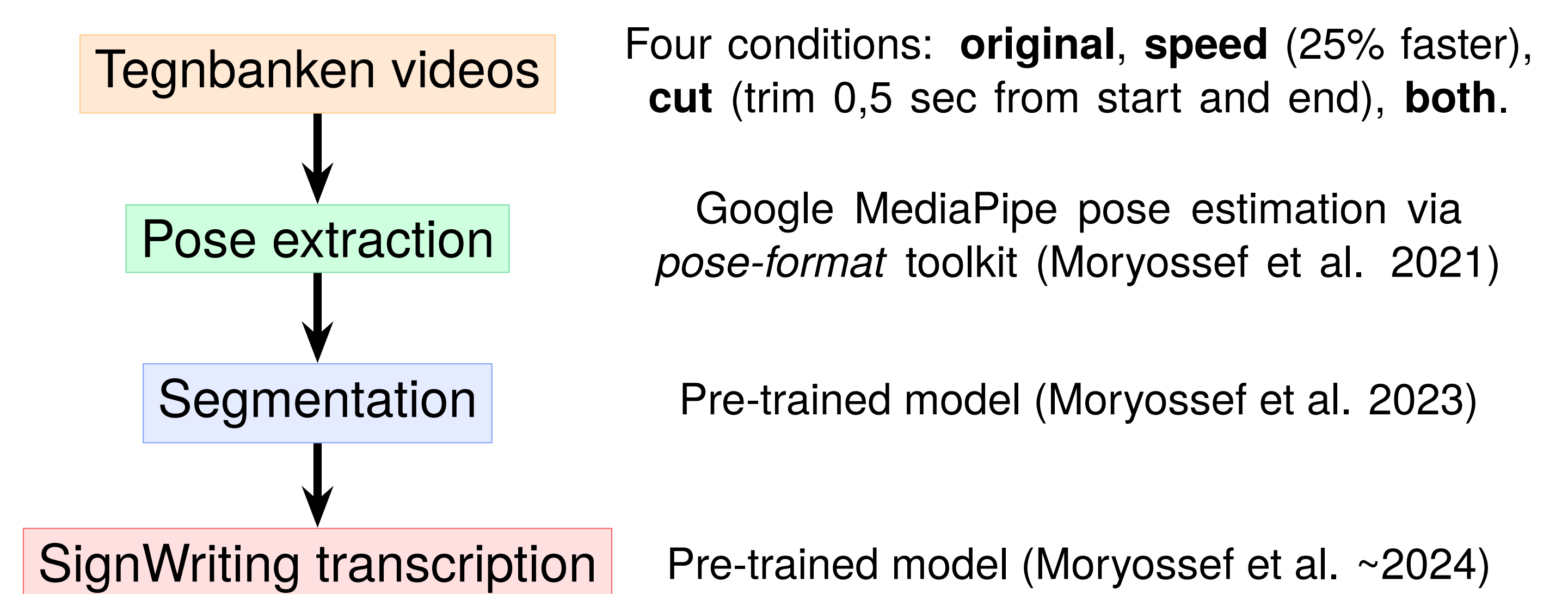
### Building a reference dataset

Merge Tegnbanken and SignPuddle by pivoting on the glosses.



**Limitations:** Many-to-many mappings and unverified alignments affect dataset quality.

### Automatic SignWriting transcription pipeline



## Results

The transcriptions were evaluated against the reference with four metrics: chrF, CLIPScore, SymbolDistances and TokenizedBLEU.

	ChrF	CLIPScore	SymbolDistances	TokenizedBLEU
Original	0.2428	0.8708	0.3269	0.0845
Speed	0.2427	<b>0.8712</b>	0.3265	<b>0.0857</b>
Cut	<b>0.2434</b>	0.8694	<b>0.3188</b>	0.0841
Both	0.2431	0.8701	0.3197	0.0843

- Low scores overall, automatic transcriptions rarely correspond to ground truth.
- The model is somewhat sensitive to speed and length, but there is no clear winning setup.
- The expected output is one sign per video, but the average annotation length is about two signs per video.