

Assisting Corpus Annotation: Automatic BIO-Tagging of Clause-Like Units in PJM. A Pilot Study on Corpus Data

Authors

Piotr Mostowski¹

Anna Kuder²

Joanna Wójcicka¹

Affiliations

¹University of Warsaw

²University College London



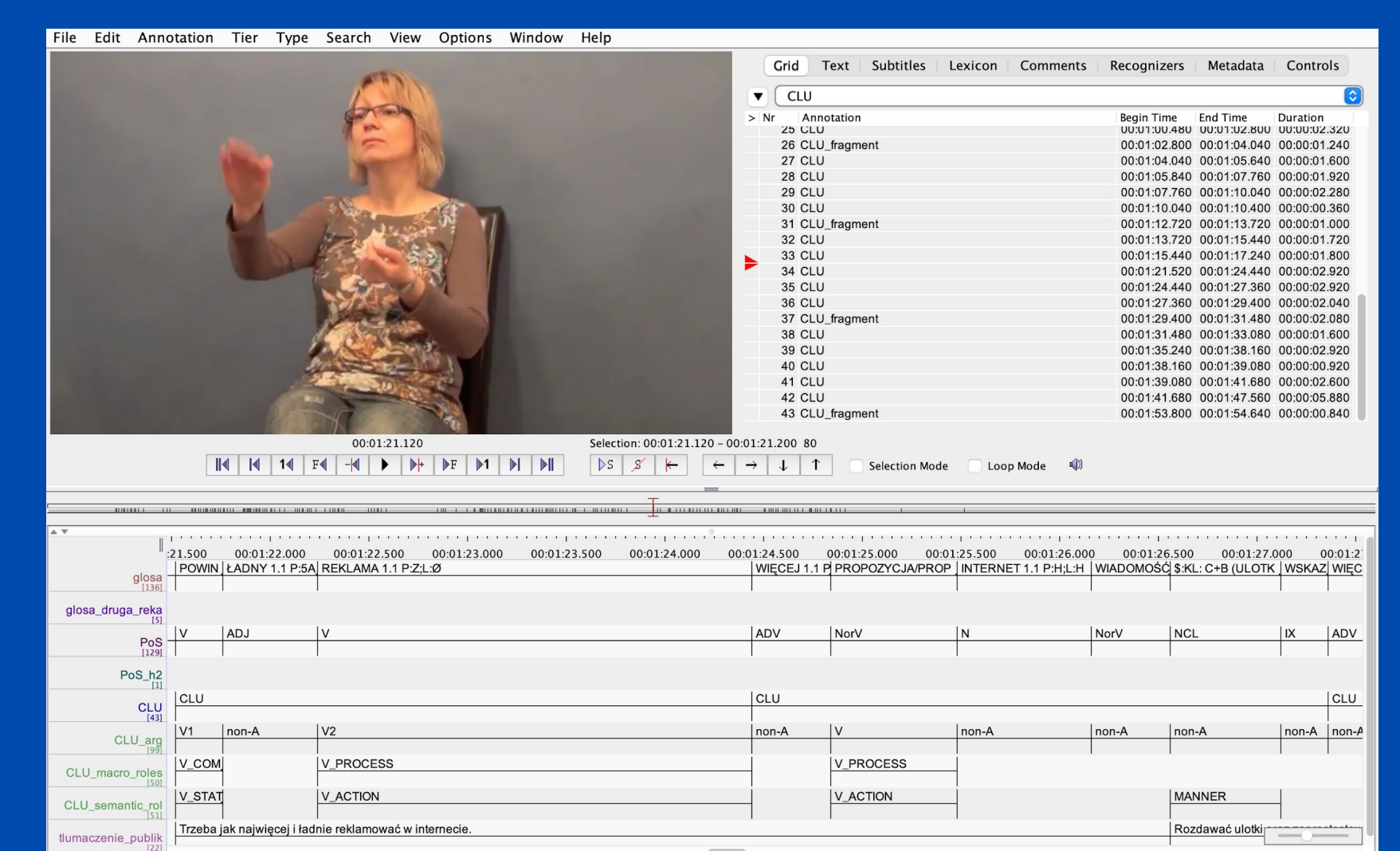
THE CLU ANNOTATION TASK

- **CLU:** a meaningful symbolic unit of utterance characterized by its predicational properties.
- The annotation schema is grounded in the analysis of the predicate-argument structure of the delimited segment (Johnston, 2019).
- Segmenting continuous signing into Clause-Like Units (CLUs) is one of the most resource-intensive tasks in manual annotation process.

SENTENCE BOUNDARIES

- The delimitation and analysis of sentence boundaries in sign language have long been recognized as methodologically challenging (Crasborn, 2007; Fenlon et al., 2008).
- **CLU boundaries:** marked with a mix of syntactic and prosodic cues involving manual (hands) and non-manual markers.
- **The non-manual markers** include: blinks, head movements (tilts and headshakes), facial expressions, eye gaze, and body leans or shifts.

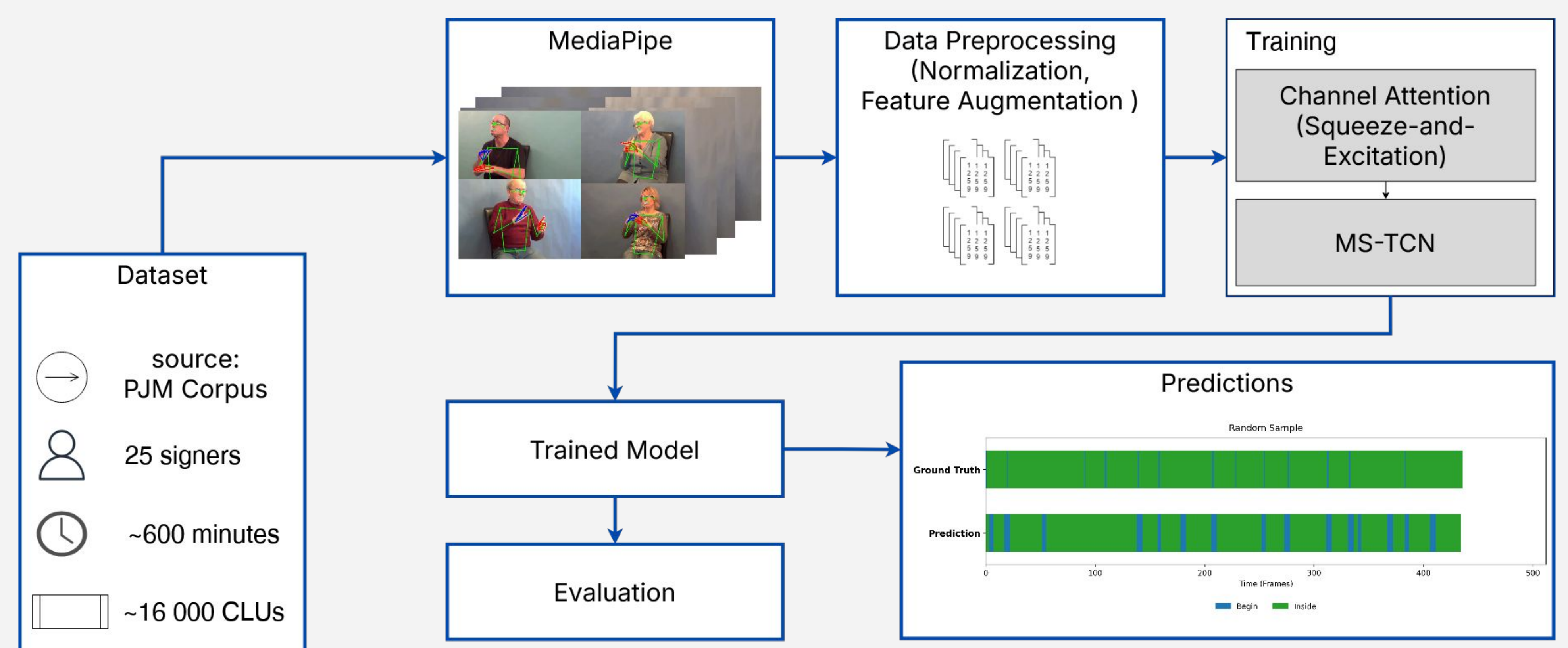
THE ANNOTATION CHALLENGE



- **Objective:** To develop an automated segmentation framework that assists human annotators.
- **Approach:** Trained and evaluated on human-annotated data.

TECHNICAL SETUP

- **Feature Extraction:** 3D coordinates from MediaPipe Holistic + velocity, acceleration derivatives.
- **Model:** A Multi-Stage Temporal Convolutional Network (MS-TCN) (Farha and Gall, 2019) enhanced with a Squeeze-and-Excitation (SE) Channel Attention mechanism (Hu et al., 2018).
- **Logic:** The SE block dynamically weights different modalities, allowing the model to focus on specific articulators depending on the temporal context.



EVALUATION

| Threshold (IoU) | F1-Score | Practical Interpretation |
|-------------------|----------|--|
| @0.10 (Detection) | 75.43% | High reliability in detecting clause presence and temporal onsets. |
| @0.25 (Overlap) | 68.32% | Robust identification of the clause propositional core. |
| @0.50 (Precision) | 57.52% | High precision in boundary alignment with expert labels. |

High detection rates (@0.10) shifts the workflow from manual searching to efficient expert verification.

ABLATION STUDY

Impact of individual modalities on performance (at IoU 0.50):

- Full Multimodal Model: 57.52% (Baseline)
- Hands-only: 47.92% (-9.60 pp)
- Face-only: 43.92% (-13.60 pp)
- The face functions as a prosodic filter, suppressing transitional noise and validating manual signals for coherent boundary detection.

RESULTS & CONCLUSION

- **Reliable Detection:** High recall ensures that nearly all CLUs are identified, with errors limited to minor timing offsets rather than missed segments.
- Facial cues are vital prosodic anchors, both modalities are essential for accuracy.
- High detection rates (@0.10) enable a shift from manual drafting to fast, expert-led verification.
- This scalable framework resolves the annotation bottleneck, supporting larger datasets and robust SL technologies.

REFERENCES

Crasborn, O. (2007). How to recognise a sentence when you see one. *Sign Language & Linguistics*, 10: 103–111. DOI:10.1075/sll.10.2.03cra; Farha, Y. A. and Gall, J. (2019). MS-TCN: Multi-stage temporal convolutional network for action segmentation. *arXiv preprint arXiv:1903.01945*; Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2019). Squeeze-and-excitation networks. *arXiv preprint arXiv:1709.01507*; Johnston, T. (2019). *Auslan Corpus Annotation Guidelines*. Macquarie University, Australia. <https://api.semanticscholar.org/CorpusID:201605087>; Fenlon, J., Campbell, R., and Woll, B. (2008). Seeing sentence boundaries. *Sign Language & Linguistics*, 10(2): 177–200. DOI:10.1075/sll.10.2.06fen