UkrSL: Towards a Ukrainian Continuous Sign Language Dataset

Oleksandr Sobetskyi **Maryna Kosse** **Roman Kyslyi** **Angelina Savchenko**
Ukrainian Catholic University, Kyiv School of Economics, Kyiv Polytechnic Institute
lex.sobieski@gmail.com, mkosse@kse.org.ua,
rkyslyi@kse.org.ua, anhelina.savchenko@edu.kpi.ua

Abstract

We present UkrSL-Annot, an annotated dataset for Ukrainian Sign Language (USL) — one of the most underresourced sign languages in Europe. The dataset comprises 1,456 annotated clips (1,463 with cropped video segments) totalling approximately two hours of signing, sourced from six broadcast videos from Suspilne, Ukraine's public broadcaster. Each clip is annotated with a spoken Ukrainian transcription aligned to the corresponding signing segment. We describe the data collection pipeline, the annotation methodology, and provide a detailed analysis of the dataset's statistics and limitations. The dataset is being actively expanded, and we release this snapshot to support the research community and invite collaboration.

1 Introduction

Sign language recognition (SLR), sign language translation (SLT), and sign language production (SLP) have seen rapid progress in recent years, driven largely by the availability of annotated corpora for established sign languages such as German Sign Language (Koller et al., 2015), British Sign Language (Schembri et al., 2013), and American Sign Language (Duarte et al., 2021). Ukrainian Sign Language (USL), however, remains severely underresourced: no publicly available annotated continuous USL corpus exists to date, leaving Ukrainian Deaf and hard-of-hearing communities without the computational tools increasingly available to speakers of higher-resourced sign languages. Among these tasks, SLP — the automatic generation of sign language video from spoken or written language — is a natural fit for sentence-level aligned corpora such as ours, since it requires parallel data mapping text to signing and can be partially trained without gloss-level annotation, although high-quality SLP typically also benefits from fine-grained motion supervision.

The urgency of this gap has intensified in recent years. Ukrainian public discourse has expanded dramatically online, with sign language interpretation of news broadcasts, political addresses, and public information campaigns becoming widely disseminated on video platforms. These materials represent a valuable, naturally occurring source for dataset construction.

This paper introduces UkrSL-Annot, the first publicly released annotated corpus of continuous Ukrainian Sign Language. We describe the data collection and annotation pipeline, present an exploratory analysis, and discuss the challenges and ongoing efforts to expand coverage. Our contributions are:

1. A release of 1,456 annotated continuous USL clips ($\sim$2 hours of video) sourced from six publicly available broadcast videos;
2. A sentence-level annotation schema pairing signed segments with spoken Ukrainian transcriptions;
3. An open-source annotation tool and pipeline leveraging ASR for temporal alignment;
4. An analysis of dataset statistics, limitations, and a roadmap for community-driven expansion.

2 Related Work

Continuous sign language datasets have been central to the field since the release of the RWTH-PHOENIX-Weather 2014 corpus (Koller et al., 2015), which remains a primary benchmark for German SLR and SLT. Subsequent efforts include the BSL Corpus (Schembri et al., 2013), CSL-Daily for Chinese Sign Language (Zhou et al., 2021), and How2Sign for ASL (Duarte et al., 2021). These corpora range from a few hours to hundreds of hours of annotated video, and their availability has

Proceedings of the Fifth Ukrainian Natural Language Processing Conference (UNLP 2026), pages 53–57

directly enabled state-of-the-art transformer-based models.

For low-resource sign languages, the landscape is sparse. Yin et al. (2021) call on the NLP community to broaden its research agenda to signed languages, noting the strong asymmetry in available resources across sign languages. Existing work on Ukrainian NLP, such as the UberText corpus (Chaplynskyi, 2023) and recent shared tasks for Ukrainian LLMs (Syvokon et al., 2024), has focused on spoken and written Ukrainian, with no dedicated continuous sign language resources. The closest prior efforts for Ukrainian Sign Language are the linguistically oriented USL corpus collected by Bauer (2022) at the University of Cologne (∼3 hours of dialogues with Deaf Ukrainian refugees, intended for linguistic analysis rather than machine learning), and the rule-based machine translation work of Lozynska et al. (2019), which uses small internal sentence-level lexicons that are not released as a continuous video corpus.

Our work follows the philosophy of community-driven, incrementally growing resources used in earlier sign language corpus projects such as the NCSLGR corpus (Neidle et al., 2001) and the BSL/Auslan corpus tradition (Johnston, 2010): we treat the dataset as a living resource that grows with continuing annotation effort, rather than a one-shot release frozen at submission time.

3 Data Collection

Source videos were obtained from Suspilne (*UA:PBC*), Ukraine's public broadcasting company, which granted explicit permission for their use in this research. The videos consist of news broadcasts and public information programs featuring professional sign language interpreters, providing naturally occurring continuous USL in a controlled visual setting. Videos were selected to ensure sufficient temporal coverage and visual clarity (minimal occlusion, adequate framing of the signer's upper body and hands).

The current release draws from six source videos. The number of clips per video ranges from 18 to 508 (Table 1), with an imbalance that reflects the varying length of available source material rather than a balanced design. Addressing source diversity is a primary goal of the ongoing collection phase.

Video clips were cropped to isolate individual signing segments. All clips are encoded as H.264 MP4 files at 510×510 pixels and 30 fps, a resolution and frame rate sufficient for capturing hand and finger motion.

4 Annotation Methodology

Each source video contains both a signed and a spoken Ukrainian track. We use the Whisper large-v3 automatic speech recognition model (Radford et al., 2023) to obtain word-level transcriptions and their corresponding timestamps. The raw ASR output is then passed through a lightweight preprocessing step that merges consecutive words into sentence-level segments based on punctuation and pause heuristics. These sentence-level boundaries serve as candidate segmentation points for initial clip extraction.

Annotation tool. To support efficient and consistent annotation we developed a custom web-based annotation tool (Figure 1a), deployed on Hugging Face Spaces with Firebase Realtime Database as the backend. The interface embeds the source YouTube video alongside an editable caption table, providing playback controls (speed adjustment from $0.25\times$ to $2\times$, frame-level seeking with ± 100 ms and ± 1 s buttons) and per-caption editing fields for start time, end time, and subtitle text. Annotators can mark individual captions as *aligned* and flag entire videos as complete. The tool stores annotations in a hierarchical structure keyed by YouTube video ID, where each caption record contains the spoken text, start and end timestamps (in seconds), and an alignment flag.

Annotation workflow. For each video, an annotator: (1) reviews the Whisper-generated caption table alongside the video; (2) corrects any ASR transcription errors in the spoken Ukrainian text; (3) adjusts start and end timestamps so that each segment boundary aligns precisely with the onset and offset of the corresponding signing; and (4) marks each caption as aligned once satisfied. Annotation is carried out by fluent USL speakers from a broader pool of volunteers; two annotators contributed to the current release, compensated at 400 UAH per hour. Because our annotators are fluent signers with practical expertise in USL, we treat their annotations as reference-quality and do not measure inter-annotator agreement — a choice we revisit in the Limitations section.

Each resulting clip is paired with the spoken Ukrainian sentence as transcribed by Whisper and

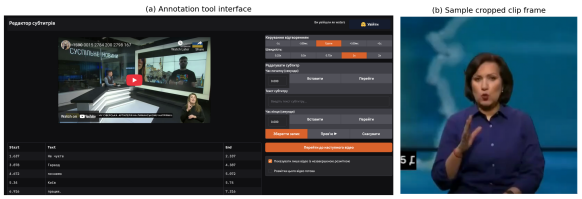

Figure 1: (a) The web-based annotation tool showing the embedded video player, playback controls, and editable caption table. (b) A sample cropped clip frame (510 × 510 px) showing a USL interpreter during a Suspilne broadcast.

Property	Value
Total annotated clips	1,456
Clips with video	1,463
Source videos	6
Annotators	2
Total video duration	∼2 h (119.8 min)
Mean clip duration	4.91 s ($\sigma = 3.65$)
Median clip duration	4.00 s
Min / Max duration	0.23 s / 36.30 s
Resolution	510 × 510 px
Frame rate	30 fps
Video codec	H.264
Total text tokens	14,514
Unique text tokens	5,876
Mean words per clip	10.0
Total file size	347.6 MB

Table 1: UkrSL-Annot dataset statistics (current release).

corrected by the annotator, representing the spoken-language transcript that accompanies the signing. The original audio track is preserved in the source videos but is not redistributed in the released clips, which contain only the cropped signer view; the spoken transcription is therefore stored as a separate textual layer aligned to the video by start and end timestamps. We note that broadcast signing may diverge from a verbatim rendering of the spoken source: in this release we adopt the spoken transcript as the reference text without explicitly flagging interpreter paraphrasing or summarisation, and we revisit this design choice in the Limitations section. Final annotations are exported to a pipe-delimited CSV with fields for clip identifier, spoken text, and annotator identity.

5 Dataset Analysis

Table 1 summarizes the key statistics of the current release. The dataset contains 1,456 annotated clips with a total video duration of approximately two hours. Clip durations are right-skewed (mean

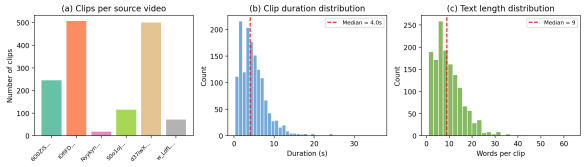

Figure 2: Dataset statistics. (a) Clip distribution across six source videos. (b) Clip duration histogram; red dashed line indicates the median (4.0 s). (c) Text length distribution (words per clip).

4.91 s, median 4.00 s), with a long tail of clips exceeding 15 seconds (Figure 2b).

The text vocabulary comprises 5,876 unique word types over 14,514 tokens. The average clip contains 10 words of spoken Ukrainian text, reflecting the sentence-level granularity of the segmentation. Figure 2c shows the distribution of text lengths across clips.

Source video coverage is uneven: the two largest source videos contribute over 1,000 clips combined, while the smallest contributes only 18 (Figure 2a). For downstream model training, we recommend treating source video identity as a stratification variable in any train/validation/test split to prevent models from exploiting video-specific visual cues rather than learning the sign language itself.

A cross-check between annotations and video files reveals that 1,463 cropped video segments are available, with 1,456 of these having corresponding annotations (99.5% coverage).

6 Ongoing Work

Active work to expand UkrSL-Annot includes: (1) expansion to additional source videos and a target of 5,000+ annotated clips; (2) addition of gloss-level annotation to enable sign language recognition tasks; (3) addition of signer identity metadata per clip, enabling signer-independent evaluation splits; (4) integration with pose estimation (MediaPipe Holistic) to extract skeletal keypoint features alongside raw video; (5) a planned train/dev/test split stratified by source video and signer.

We invite researchers and members of the Ukrainian Deaf community to contribute annotations, report errors, and suggest additional source material. The dataset and annotation guidelines are available at the repository.

7 Conclusion

We have presented UkrSL-Annot, the first publicly released annotated corpus of continuous Ukrainian

Sign Language, comprising approximately two hours of annotated video across 1,456 clips. The dataset provides sentence-level alignment between signed video segments and spoken Ukrainian transcriptions, supported by a reproducible collection and annotation pipeline. We hope this release will lower the barrier for the NLP community to engage with Ukrainian Sign Language and will serve as the foundation for a substantially larger resource in future releases.

Limitations

Although UkrSL-Annot has grown substantially, two hours of annotated video remains small compared to established benchmarks (e.g., PHOENIX-2014T contains $\sim$11 hours). The current annotation provides only sentence-level spoken text; gloss-level and sub-lexical annotations (handshape, movement, non-manual features) are not yet available, which limits applicability for phonological research and standard SLR evaluation. Source diversity is limited to six videos from a single broadcaster, which may introduce topical and stylistic biases. While we rely on fluent USL speakers for annotation quality, inter-annotator agreement has not been formally measured; future releases will include a doubly-annotated subset to quantify consistency. Because the reference text is the corrected spoken transcription rather than a gloss of what was actually signed, cases where the interpreter paraphrases, summarises or reorders the spoken source are not explicitly marked; downstream users training translation models should be aware that the alignment is text-to-signing and not gloss-to-signing. Finally, we have not yet established baseline model performance on this dataset.

Ethics Statement

All source videos were obtained with explicit permission from Suspilne (UA:PBC), Ukraine's public broadcasting company. Annotators are fluent USL speakers who are compensated for their work at a fair hourly rate. We are mindful of the ethical considerations surrounding sign language data, including the representation and agency of Deaf communities, and we commit to consulting with members of the Ukrainian Deaf community as the project expands. No personally identifiable information beyond publicly visible signer appearances in broadcast footage is included.

References

Anastasia Bauer. 2022. The Ukrainian Sign Language corpus. University of Cologne. https://ifl.phil-fak.uni-koeln.de/. Project page accessed April 2026.

Dmytro Chaplynskyi. 2023. Introducing UberText 2.0: A corpus of modern Ukrainian at scale. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 1–10.

Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. 2021. How2Sign: A large-scale multimodal dataset for continuous American Sign Language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2735–2744.

Trevor Johnston. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1):106–131.

Oscar Koller, Jens Forster, and Hermann Ney. 2015. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125.

Olga Lozynska, Maksym Davydov, and Volodymyr Pasichnyk. 2019. Rule-based machine translation into Ukrainian Sign Language using concept dictionary. In *Proceedings of the International Workshop on Modern Machine Learning Technologies and Data Science (MoMLeT+DS)*.

Carol Neidle, Stan Sclaroff, and Vassilis Athitsos. 2001. SignStream: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior Research Methods, Instruments, & Computers*, 33(3):311–320.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 28492–28518.

Adam Schembri, Jordan Fenlon, Ramas Rentelis, Sally Reynolds, and Kearsy Cormier. 2013. Building the British Sign Language corpus. *Language Documentation & Conservation*, 7:136–154.

Oleksiy Syvokon, Mariana Romanyshyn, and Roman Kyslyi. 2024. The UNLP 2024 shared task on fine-tuning large language models for Ukrainian. In *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING*.

Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. Including signed languages in natural language processing. In

Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP), pages 7347–7360.

Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, and Houqiang Li. 2021. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1316–1325.