Why Alternative Gloss Labels

Will Increase the Value of the

WLASL Dataset

Carol Neidle and Carey Ballard

Boston University, Boston, MA

1. The WLASL dataset

The WLASL [2] purports to be "the largest video dataset for Word-Level American Sign Language (ASL) recognition." It brings together various publicly shared video collections that could be quite valuable for sign recognition research, and it has been used extensively for such research. However, there is a critical problem with the accompanying annotations that has heretofore not been recognized by the authors, nor by those who have exploited these data: There is no 1-1 correspondence between sign productions and gloss labels.

Deficiencies in the quality and accuracy of annotated sign language corpora are a key limitation for progress on sign recognition research [1]. Sign recognition research based on gloss labels for signs faces a serious challenge, given that: (1) there is no 1-1 correspondence between English words and ASL signs; and (2) there are also no established glossing conventions shared by the ASL/research community. For our ASLLRP projects, our research group has established conventions to ensure a 1-to-1 correspondence between gloss label and ASL sign production [3-5], which is critically important for use in computational research. See [5] for discussion of the challenges posed in establishing glossing conventions. These conventions also govern the gloss labels used for our ASLLRP Sign Bank, with 23,452 citation-form sign tokens. These can be viewed online from **https://dai.cs.rutgers.edu/dai/s/signbank**. The ASLLRP Sign Bank also incorporates segmented examples of signs from our continuous signing corpora, making for a total of 44,012 sign tokens corresponding to 3,542 distinct signs (not including fingerspelled signs, classifiers, or gestures).

Serious problems arise, however, when researchers use datasets where 1-1 gloss label to sign correspondences have not been enforced; or when multiple datasets using inconsistent glossing conventions are combined. This is the situation for the WLASL dataset, which brings together multiple, publicly shared, ASL video corpora from different sources; and internal consistency of labeling is not even enforced within the individual collections that are combined.

The authors [2] report that the WLASL dataset contains "2,000 common different words in ASL" (although for reasons discussed below, the count of distinct gloss labels does not necessarily correlate with the number of distinct signs). The WLASL dataset brings together data shared publicly on the Web from different sources, providing various types of metadata, including a gloss label for each video. As they explain: "We select videos whose titles clearly describe the gloss of

the sign." It seems that they do not realize that there is no standard convention for associating an English-based gloss label with an ASL sign, and, since there is no 1-1 relationship between English words and ASL signs, that there is considerable variability in how gloss labels are used (and, thus, how video files may be named).[1]

As a result, there are many cases where WLASL has examples of a single ASL sign glossed with more than 1 English word, as in the sign glossed sometimes as *woman* and sometimes as *lady*, shown in Fig. 1. Conversely, there are many cases where the same English gloss is used for totally different WLASL ASL signs, as shown in Fig. 2 for the gloss label *close*: the sign on the left is a verb, the opposite of 'open,' whereas the sign on the right is an adjective, meaning 'near'. Another example is shown in Fig. 3, for *mean*. The sign on the left is a verb in ASL meaning 'to signify,' whereas the sign on the right is an adjective meaning 'unkind'. They classify these as 'dialectal variants,' but that is simply wrong, and the designation of dialectal variants throughout the WLASL dataset is highly problematic.
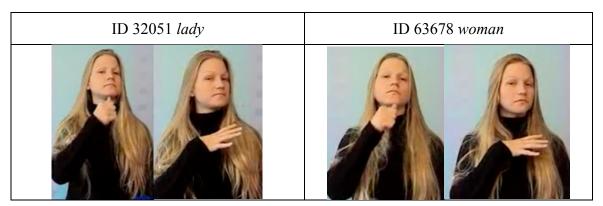
ID 32051 *lady*	ID 63678 *woman*

Figure 1. WLASL: same ASL sign, different English glosses

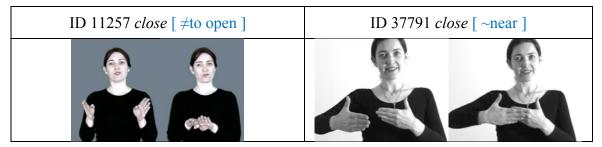| ID 11257 *close* [≠to open]	ID 37791 *close* [~near]

Figure 2. WLASL: same English gloss, different ASL signs

---

[1] They also state: "if the gloss annotations are composed of more than two words in English, we… remove those videos to ensure that the dataset contains words only." This also reflects a lack of understanding of the relationship between the 2 languages.

ID 67880 *mean*, variant 0	ID 67881 *mean*, variant 1
[~to signify]	[~unkind, cruel]

Figure 3. Supposed Dialectal Variants in WLASL

The issues exemplified above are pervasive in the WLASL data, posing critical obstacles to using this dataset reliably for computational research, despite the fact that it has been widely used for such research; a partial list of research based on these data can be found at: **https://paperswithcode.com/dataset/wlasl**. This surely explains, at least in part, the low recognition rates that have been reported (e.g., less than 63% for top-10 accuracy on 2,000 words/glosses [2]).

2. The WLASL dataset: Value added by consistent gloss labeling

This serious limitation of the WLASL dataset could be overcome if consistent text-based gloss labeling were to be provided for the rich set of ASL videos that are being shared. This is precisely what we have done by sharing on the Web a spreadsheet that reorganizes and relabels the data based on the glossing conventions used for the ASLLRP Sign Bank (and for other ASLLRP data). This spreadsheet, available from <**https://dai.cs.rutgers.edu/dai/s/signbank**>, can be used in conjunction with the data distributed on the WLASL website: **https://dxli94.github.io/WLASL/** .

3. But wait! There's more

Another huge advantage of having consistent gloss labeling across the ASLLRP and WLASL datasets is that these datasets can also be combined, to create a valuable resource that is richer and more extensive than either of the datasets on its own. The citation-form signs from the ASLLRP Sign Bank are also available for download from < **https://dai.cs.rutgers.edu/dai/s/signbank**>.

References

[1]   Bragg, D., O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef, a. C. Vogler, and M. Ringel Morris. (2019) Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19). Association for Computing Machinery, New York, NY, USA, pp 16–31. DOI:**https://doi.org/10.1145/3308561.3353; https://dl.acm.org/doi/pdf/10.1145/3308561.3357743**.

[2]   Li, D., C. R. Opazo, X. Yu, and H. Li. (2020) Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison, *arXiv:1910.11006*. **https://arxiv.org/abs/1910.11006**

[3]   Neidle, C., and A. Opoku. (2021) Update on Linguistically Annotated ASL Video Data Available through the American Sign Language Linguistic Research Project (ASLLRP), Boston University, ASLLRP Project Report No. 19, Boston, MA. **http://www.bu.edu/asllrp/rpt19/asllrp19.pdf**

[4]   Neidle, C., A. Opoku, G. Dimitriadis, and D. Metaxas. (2018) NEW Shared & Interconnected ASL Resources: SignStream® 3 Software; DAI 2 for Web Access to Linguistically Annotated Video Corpora; and a Sign Bank. 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. LREC 2018, May 2018; Miyagawa, Japan. **https://open.bu.edu/handle/2144/30047**

[5]   Neidle, C., A. Thangali, and S. Sclaroff. (2012) Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. LREC 2012, Istanbul, Turkey. May 2012. **https://open.bu.edu/handle/2144/31899**