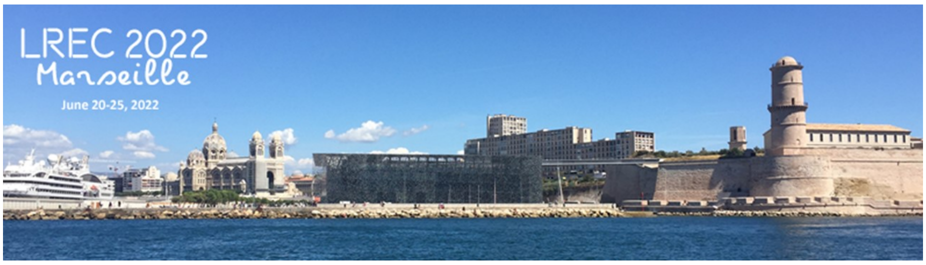



LREC 2022 Workshop
Language Resources and Evaluation Conference
20-25 June 2022

# 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources (sign-lang@LREC 2022)

# PROCEEDINGS

Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke,
Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch,
Marc Schulder (eds.)

# Proceedings of the LREC 2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources (sign-lang@LREC 2022)

Edited by: Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch, and Marc Schulder

**ISBN: 979-10-95546-86-3**
**EAN: 9791095546863**

**For more information:**
European Language Resources Association (ELRA)
9 rue des Cordelières
75013, Paris
France
http://www.elra.info
Email: lrec@elda.org

# Preface

This collection of papers stems from the 10th Workshop on the Representation and Processing of Sign Languages which takes place as a satellite workshop to the Language Resources and Evaluation Conference in Marseille (France).

While there has been occasional attention to sign languages at the main LREC conference, the focus there is on spoken languages in their written and spoken forms. This series of workshops, however, offers a forum for researchers focussing on sign languages, especially on corpus data and corpus technology for sign languages.

This year's hot topic "Multilingual Sign Language Resources" aligns with one of the main conference's hot topics. It stresses the importance of looking across sign languages whenever testing claims about signed modality, but it also addresses the problem that for many sign languages only very few languages resources are available. Combining resources across languages is a promising perspective to draw on richer sets of data.

Please note that this year LREC has two workshops on sign languages: SLTAT7 covers the topics automatic translation and avatar technology. In the corresponding proceedings, you find 19 more sign language-related papers.

The contributions composing this volume are presented in alphabetical order by the first author. For the reader's convenience, an author index is provided as well.

Once again, we would like to thank all members of the program committee who helped us tremendously by reviewing the submissions to the workshop within a very short timeframe!

Finally, we would like to point the reader to the proceedings of the previous workshops that form important resources in a growing field of research. They are all available online from the sign-lang@LREC Anthology at

`https://www.sign-lang.uni-hamburg.de/lrec/`

The site offers an author index across all workshops as well as stable URLs for all workshop papers and posters. If you need bibliographical (BibTeX) data for all workshops, the site now has them per paper, per workshop, per author or all in one. Happy browsing!

The Editors

**Organizers**

Eleni Efthimiou, Institute for Language and Speech Processing, Athens, Greece
Stavroula-Evita Fotinea, Institute for Language and Speech Processing, Athens, Greece
Thomas Hanke, Institute of German Sign Language, University of Hamburg, Germany
Julie A. Hochgesang, Gallaudet University, Washington, USA
Jette Kristoffersen, Centre for Sign Language, Copenhagen, Denmark
Johanna Mesch, Stockholm University, Sweden
Marc Schulder, Institute of German Sign Language, University of Hamburg, Germany

**Program Committee:**

Penny Boyes-Braem, University of Zurich, Zurich, Switzerland
Annelies Braffort, LIMSI/CNRS, Orsay, France
Carl Börstell, University of Bergen, Bergen, Norway
Kearsy Cormier, University College London, London, United Kingdom
Onno Crasborn, Radboud University, Nijmegen, The Netherlands
Eleni Efthimiou, Institute for Language and Speech Processing, Athens, Greece
Michael Filhol, LIMSI/CNRS, Orsay, France
Stavroula-Evita Fotinea, Institute for Language and Speech Processing, Athens, Greece
Thomas Hanke, University of Hamburg, Hamburg, Germany
Julie A. Hochgesang, Gallaudet University, Washington, USA
Amy Isard, University of Hamburg, Hamburg, Germany
Trevor Johnston, Macquarie University, Sydney, Australia
Vadim Kimmelman, University of Bergen, Bergen, Norway
Reiner Konrad, University of Hamburg, Hamburg, Germany
Jette Kristoffersen, Centre for Sign Language, Copenhagen, Denmark
Anna Kuder, University of Warsaw, Warsaw, Poland
Lorraine Leeson, Trinity College Dublin, Dublin, Ireland
John McDonald, DePaul University, Chicago, USA
Johanna Mesch, Stockholm University, Stockholm, Sweden
Hope E. Morgan, Radboud University, Nijmegen, The Netherlands
Carol Neidle, Boston University, Boston, USA
Justin Power, University of Texas, Austin, USA
Paweł Rutkowski, University of Warsaw, Warsaw, Poland
Marc Schulder, University of Hamburg, Hamburg, Germany
Rose Stamp, Bar-Ilan University, Ramat Gan, Israel
Anna Vacalopoulou, Institute for Language and Speech Processing, Athens, Greece
Rosalee Wolfe, Institute for Language and Speech Processing, Athens, Greece

# Table of Contents

*Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages (sign-lang@LREC 2022)*, pages 165–172
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022


# Resources for Computer-Based Sign Recognition from Video, and the Criticality of Consistency of Gloss Labeling across Multiple Large ASL Video Corpora

**Carol Neidle[1], Augustine Opoku[1], Carey Ballard[1], Konstantinos M. Dafnis[2], Evgenia Chroni[2], Dimitris Metaxas[2]**
[1]Boston University, [2]Rutgers University
[1]Boston University Linguistics, 621 Commonwealth Ave., Boston, MA 02215
[2]110 Frelinghuysen Road, Piscataway, NJ 08854

carol@bu.edu, augustine.opoku@gmail.com, cmball@bu.edu, kd703@cs.rutgers.edu, etc44@cs.rutgers.edu, dnm@cs.rutgers.edu

## Abstract

The WLASL 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, a critical problem with the accompanying annotations 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. Here we describe a large, linguistically annotated, video corpus of citation-form ASL signs shared by the ASLLRP—with 23,452 sign tokens and an online Sign Bank—in which such correspondences are enforced. We furthermore provide annotations for 19,672 of the WLASL video examples consistent with ASLLRP glossing conventions. For those wishing to use WLASL videos, this provides a set of annotations making it possible: (1) to use those data reliably for computational research; and/or (2) to combine the WLASL and ASLLRP datasets, creating a combined resource that is larger and richer than either of those datasets individually, with consistent gloss labeling for all signs. We also offer a summary of our own sign recognition research to date that exploits these data resources.

**Keywords:** ASL, isolated sign recognition, gloss labels, ASLLRP, WLASL, ASLLVD

## 1. Goals of this Paper

There are several interrelated goals of this paper:

**1)** To disseminate information about resources shared by the American Sign Language Linguistic Research Project (**ASLLRP**), which can be used for linguistic and computational research. These resources have recently been expanded, with new download functionalities.

**2)** To bring to the attention of the many sign recognition researchers who have been using (or who may wish to use) the valuable video data from the **WLASL** (Li et al., 2020) serious issues resulting from inconsistent text-based gloss labeling of signs in that dataset, which adversely affects the use of these data for computer learning.

**3)** To share an alternative set of gloss labels for a large subset of the WLASL data, which follow annotation conventions consistent with those used for ASLLRP data. This provides internally consistent gloss labeling for the WLASL, offering added value to this large set of videos. This also makes it possible to combine WLASL data with any of the ASLLRP datasets, giving rise to a dataset larger and richer than either.

Given space limitations, this paper does not aim to present a comparative survey of datasets available for ASL research, nor an overview of the large literature dealing with *desiderata* for sign language annotation.

## 2. Introduction

Deficiencies in the quality and accuracy of annotated sign language corpora are a key limitation for progress on sign recognition research (Bragg et al., 2019). 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. As an integral part of the research conducted by the American Sign Language Linguistic Research Project (ASLLRP), we have, from the outset of our research, established conventions to ensure a 1-to-1 correspondence between gloss label and ASL sign production, which is essential for use in computational research. See Neidle, Thangali & Sclaroff (2012) for discussion of challenges in establishing glossing conventions, and Neidle & Opoku (2022) for further details about our annotations.

There is widespread recognition of the requirement for unique text-based gloss labels to represent signs. This is enforced in all serious corpus research. We have implemented these principles since the mid-1990s; see, e.g., Neidle (2002). Many others have also written about these and other important issues involved in sign language annotation (e.g., Johnston, 2010; Orfanidou, Woll, and Morgan, 2015; Cormier, Crasborn, and Bank, 2016).

Major 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 are combined. This is the situation for the WLASL (Li et al., 2020), which brings together multiple, publicly shared, ASL video corpora from different sources—thus offering a potentially valuable resource for research. However, internal consistency of labeling is not even enforced within the individual collections that are combined.

## 3. The WLASL Dataset

Li et al. (2020) claim that the WLASL is "by far the largest public ASL dataset to facilitate word-level sign recognition research." They report that it 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 brings together data shared publicly on the Web from different sources; various types of metadata, including a gloss label for each video, are also provided. As they explain: "We select videos whose titles clearly describe the gloss of the sign." However, basing sign identification on filenames is problematic, since there is no standard convention for associating an English-based gloss label with an ASL sign, and no 1-1 relationship between English words and ASL signs; there is also considerable variability in how gloss labels are used. As a result, there are cases where multiple WLASL examples of a single ASL sign are glossed with different English words, as in the sign glossed sometimes as *woman* and sometimes as *lady*, shown in **Figure 1**. Conversely, there are many cases where the same English gloss is used for totally different ASL signs, as shown in **Figure 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 **Figure 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 not correct; 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

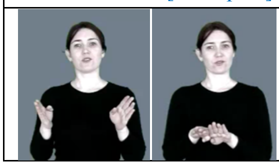
| ID 11257 *close* [ ≠to open ] | ID 37791 *close* [ ~near ] |
|---|---|
|  |  |

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

| ID 67880 *mean*, variant 0 [ ~to signify ] | ID 67881 *mean*, variant 1 [ ~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 (e.g., Hassan, Elgabry, and Hemayed, 2021; Maruyama et al., 2021; Boháček and Hrúz, 2022; Ebrahimi and Ebrahimpour-komleh, 2022); for a partial list of research based on these data, see 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 (Li et al., 2020)).

We have illustrated these problems with the WLASL data in some detail precisely because this dataset has been widely used in recent sign recognition research, and also because, as discussed below, the consistent gloss labels we are providing for use with the WLASL data can greatly increase the value of these data.

## 4. Other Datasets for Sign Recognition

Another dataset used extensively in recent vision-based ASL sign recognition research is our **ASLLVD**; see below. For example, de Amorim & Zanchetti (2021) introduced 2 datasets "derived from one of the most relevant sign language datasets–the American Sign Language Lexicon Video Dataset (ASLLVD)." Several other papers tested new sign recognition methods on datasets including the ASLLVD (Theodorakis et al. (2014): computational phonetic modeling; Elakkiya & Selvamani (2019): "three subunit sign modeling"; Lim et al. (2019): use of CNNs to train hand models; Bilge et al. (2022): new machine learning method; Kumar et al. (2028): sign recognition using computer vision and neural networks; Rastgoo et al. (2022): a combination of neural network methods; among others).

Other datasets used in recent computational research include the recently introduced large-scale **How2Sign dataset of American Sign Language** (Duarte et al., 2021; Duarte et al., 2022); and the **MS-ASL Large-Scale Data Set and Benchmark for Understanding ASL** (Joze, Vaezi, and Koller, 2018). This last article also reviews older benchmark datasets, including the **Purdue RVL-SLLL ASL database** (Kak, 2002) and the **RWTH-BOSTON** datasets (Dreuw et al., 2008). It is worth noting that the RWTH-BOSTON data were collected at Boston University through the ASLLRP; those videos are included in our current, much larger, data collection, described next.

## 5. ASLLRP Resources

We describe here ASL data made available through the ASLLRP, including isolated signs (**23,452** sign videos,[1] corresponding to distinct signs, from **33** different signers) and continuous signing corpora (**2,651** utterances, containing a total of **20,560** signs available as video clips segmented from those utterances and in their utterance context, from **19** different signers). It incorporates data collected at Boston University and at the Rochester Institute for Technology (under the supervision of Matt Huenerfauth), as well as videos shared by DawnSignPress. Including the citation-form signs and continuous signing corpora, we have a total of **44,012** sign tokens corresponding to **3,542** distinct signs (not including fingerspelled signs, classifiers, and gestures).

[1] This incorporates our **ASLLVD, American Sign Language Lexicon Video Dataset** (Athitsos et al., 2008; Neidle, Thangali, and Sclaroff, 2012), with >3,300 citation-form signs, produced by 1-6 native ASL signers, for a total of almost 9,800 tokens.

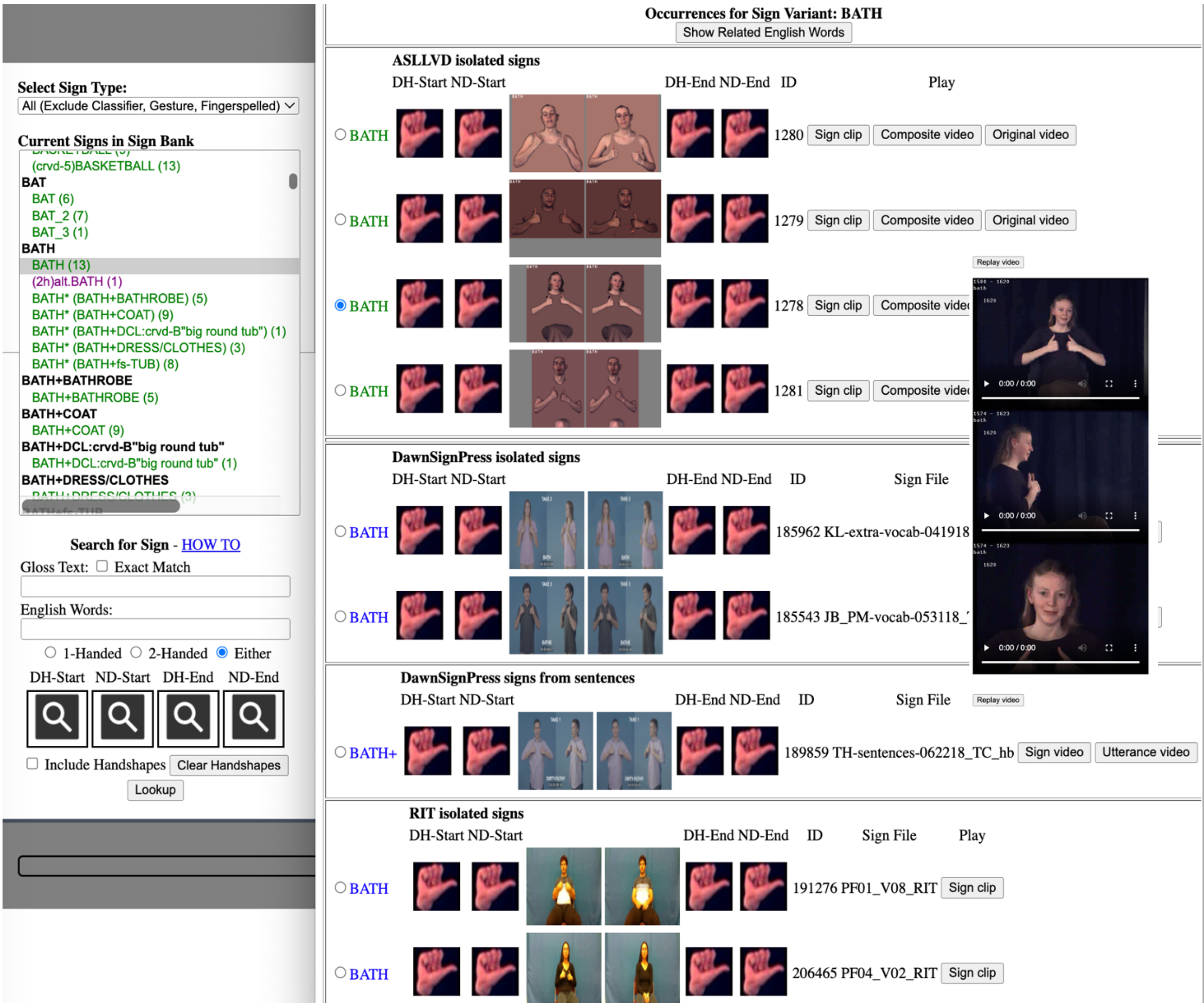



**Figure 4.** Screen shot showing Sign Bank Interface for Searching and Viewing ASL Sign Variants

The data can be searched, browsed, and downloaded. We have enforced, to the best of our ability, consistency in labeling throughout our corpora. Sign-level annotations include gloss labels, annotations of sign type (lexical, loan, fingerspelled, classifier, number, and name signs, as well as gestures and compounds), and phonological properties (e.g., information about hand configurations on the 2 hands). Utterances include sentence-level information about such things as non-manual behaviors and grammatical markings, translations, etc..

## 5.1 ASLLRP Continuous Signing Corpus

Our continuous signing data can be accessed here: https://dai.cs.rutgers.edu/dai/s/dai. The data can be browsed or searched based on various sign-level and utterance-level properties.

Download options are available for:

- **American Sign Language Linguistic Research Project (ASLLRP) SignStream® 3 Corpus**
  - **47** files with a total of **2,124** utterances; **17,528** sign tokens; and **5** signers

See https://dai.cs.rutgers.edu/dai/s/runningstats for further statistics.

Linguistic annotations for the signs and utterances that can be downloaded are available in XML format. These utterances can also be viewed and further analyzed and annotated within SignStream®, an application we have developed for analysis of visual language data, shared on the Web (http://www.bu.edu/asllrp/SignStream/3/ ; a major new update has just been released).

## 5.2 ASLLRP Sign Bank

An online ASLLRP Sign Bank (Neidle et al., 2018; Neidle, Opoku, and Metaxas, 2022) is also available: https://dai.cs.rutgers.edu/dai/s/signbank. It is possible to search based on various criteria, and to view, for specific signs, both examples from our citation-form sign datasets and segmented signs from our continuous signing corpora (viewable either individually or in their sentential context). **Figure 4** illustrates the interface. It is currently possible to download the citation-form sign datasets and videos from our website for use in sign recognition research, with the ability to download segmented Sign Bank examples from our continuous signing corpora to be provided from the same site in the near future. Datasets currently available for download, with accompanying annotations:

- **Boston University American Sign Language Lexicon Video Dataset (ASLLVD)**
  - **9,748** sign tokens; **6** signers
- **Rochester Institute of Technology (RIT) Dataset**
  - **11,801** sign tokens; **12** signers
- **DawnSignPress (DSP) Dataset**
  - **1,903** sign tokens; **15** signers

Further statistics are available here: https://dai.cs.rutgers.edu/dai/s/runningstats

Linguistic annotations for the videos are available in Excel and csv formats. ASLLRP Sign Bank annotations are explained in http://www.bu.edu/asllrp/rpt20/asllrp20.pdf, Neidle & Opoku (2022), with further description of our general annotation conventions in Neidle (2002, 2007).

## 6. Alignment of Annotations for WLASL

We selected 19,672 sign videos from the WLASL dataset. (Some examples were excluded for one of several reasons, including poor quality of the signing or the video, the presence in the video of a string of signs rather than a single sign, the unavailability of the videos in question, cases where the hands were not within the visible region, etc.) A spreadsheet, at https://dai.cs.rutgers.edu/dai/s/aboutwlasl, provides, for signs already in our Sign Bank, annotations consistent with the rest of the ASLLRP dataset. See **Figure 5**. In cases where the specific signs do not already exist in the ASLLRP dataset, new glosses that follow our existing conventions and that do not conflict with any existing gloss labels were assigned; we will continue to use the same labels for additional examples that may be added to our Sign Bank in the future.

**Figure 6** illustrates how WLASL gloss labeling compares with ASLLVD gloss labels for the sign with ASLLVD class label 'COP'. As is evident, the three different WLASL gloss labels in column 1 (corresponding to possible designations for such a person in English: *cop, police, policeman*) are used indifferently in the WLASL dataset for all occurrences, with no distinction made at all in the gloss labels for the handshape variation that potentially occurs with this sign. In some cases, multiple gloss labels are associated with identical WLASL video examples that bear distinct video IDs. See also "Why Alternative Gloss Labels Will Increase the Value of the WLASL Dataset" (Neidle and Ballard, 2022).

These alternative gloss labels are shared on the Web. So, it would be straightforward to use these labels in conjunction with the WLASL videos and other associated metadata. It is therefore also straightforward to combine the ASLLRP data with the WLASL data for research on sign recognition, to expand the number of examples and distinct signers per sign and to extend the vocabulary beyond what is contained only in one or the other of these datasets.

## 7. Sign Recognition Research using the Modified-Gloss WLASL Data

Recent research by our group has made use of the revised WLASL annotations in conjunction with the WLASL data, combined with the ASLLVD.

### 7.1 Bidirectional Skeleton-Based Isolated Sign Recognition using Graph Convolution Networks (GCNs)

Dafnis et al. (2022b) report on a new skeleton-based learning method for isolated sign recognition involving explicit detection of the start and end frames of signs trained on the ASLLVD dataset. Using linguistically relevant parameters based on skeleton input, this method employs a bidirectional learning approach within a Graph Convolutional Network (GCN) framework. For 18,141 videos of 1,449 lexical signs from the WLASL dataset (with a minimum of 6 examples per sign)—with revised gloss labeling as described earlier in this paper—we achieved a success rate of 77.43% recognition accuracy for top-1 and 94.54% for top-5, outperforming other state-of-the-art approaches. A comparison with the TRN method of Zhou et al. (2018) and the SL-GCN (SAM-SLR-v2) method of Jiang et al. (2021) *on this same WLASL dataset with revised gloss labeling* is shown in **Figure 7**.

| gloss | video_id | CLASS | MAIN ENTRY | [entry/variant] |
|---|---|---|---|---|
| cop | 13244 | COP | COP | COP |
| cop | 13246 | COP | COP | COP |
| cop | 13247 | COP | COP | COP |
| cop | 13249 | COP | COP | COP |
| cop | 13252 | COP | COP | COP |
| cop | 13253 | COP | COP | COP |
| police | 43519 | COP | COP | COP |
| police | 43525 | COP | COP | COP |
| police | 43527 | COP | COP | COP |
| police | 43528 | COP | COP | COP |
| police | 43531 | COP | COP | COP |
| police | 43534 | COP | COP | COP |
| police | 43535 | COP | COP | COP |
| policeman | 43536 | COP | COP | COP |
| policeman | 43538 | COP | COP | COP |
| policeman | 43539 | COP | COP | COP |
| cop | 13245 | COP | COP | COP_2 |
| cop | 13248 | COP | COP | COP_2 |
| cop | 13250 | COP | COP | COP_2 |
| police | 43522 | COP | COP | COP_2 |
| police | 43523 | COP | COP | COP_2 |
| police | 43526 | COP | COP | COP_2 |
| police | 43529 | COP | COP | COP_2 |
| police | 43532 | COP | COP | COP_2 |
| police | 43533 | COP | COP | COP_2 |
| police | 66306 | COP | COP | COP_2 |
| policeman | 43537 | COP | COP | COP_2 |
| policeman | 43540 | COP | COP | COP_2 |
| policeman | 43542 | COP | COP | COP_2 |
| policeman | 67087 | COP | COP | COP_2 |
| police | 43524 | COP | COP | COP_3 |

**Figure 5.** Excerpt from spreadsheet establishing correlations between WLASL signs (glossed as in Column 1) and ASLLRP-based gloss labels (with class labels used as the basis for our sign recognition research)

### 7.2 Combining Data from the WLASL and ASLLVD Datasets

In more recent work, Dafnis et al. (2022a) have been combining the WLASL data used in Dafnis et al. (2022b) with lexical signs from the ASLLVD dataset, again selecting those signs for which we had a minimum of 6 examples per sign—this time from those *combined* datasets; we ended up with **1,480** total signs (and **22,853** total video examples). There is an additional challenge involved in combining these datasets, because signers in the WLASL are standing, whereas the ASLLVD signers are seated; see **Figure 8**. It should be noted that this makes the combined dataset especially valuable, since in the real world, signers may be either sitting or standing.

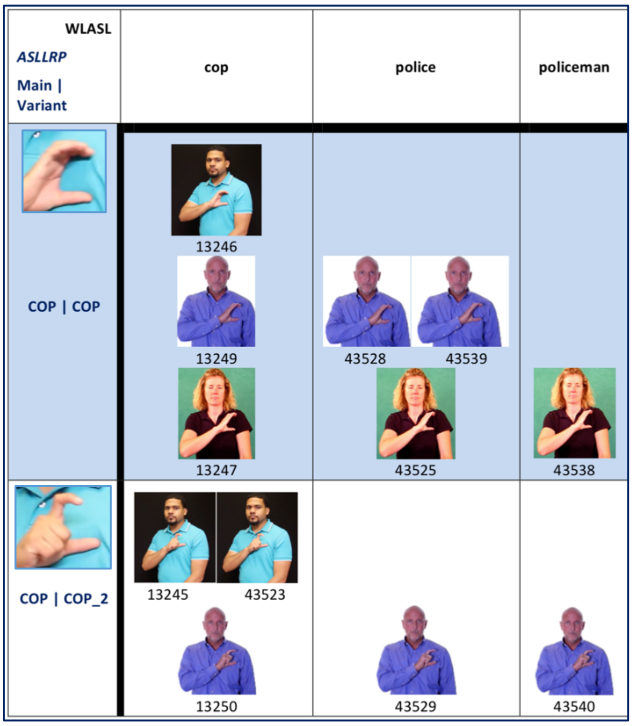



**Figure 6.** Comparison of WLASL and ASLLRP labeling of signs and sign variants. The WLASL labels ***cop***, ***police***, and ***policeman*** are used indifferently for these examples; the ASLLRP class label COP is used for all of them, with variant labels COP *vs.* COP_2 distinguishing the handshapes.

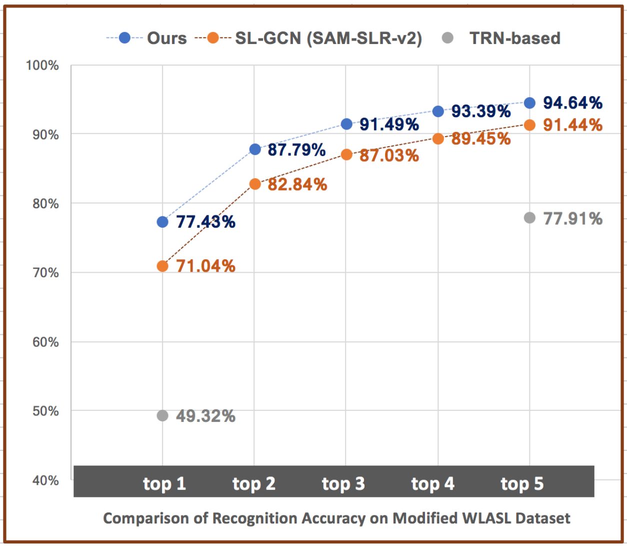



**Figure 7.** Comparison of Recognition Accuracy for 18,141 videos of 1,449 Lexical Signs in the WLASL Dataset with Modified Gloss Labeling

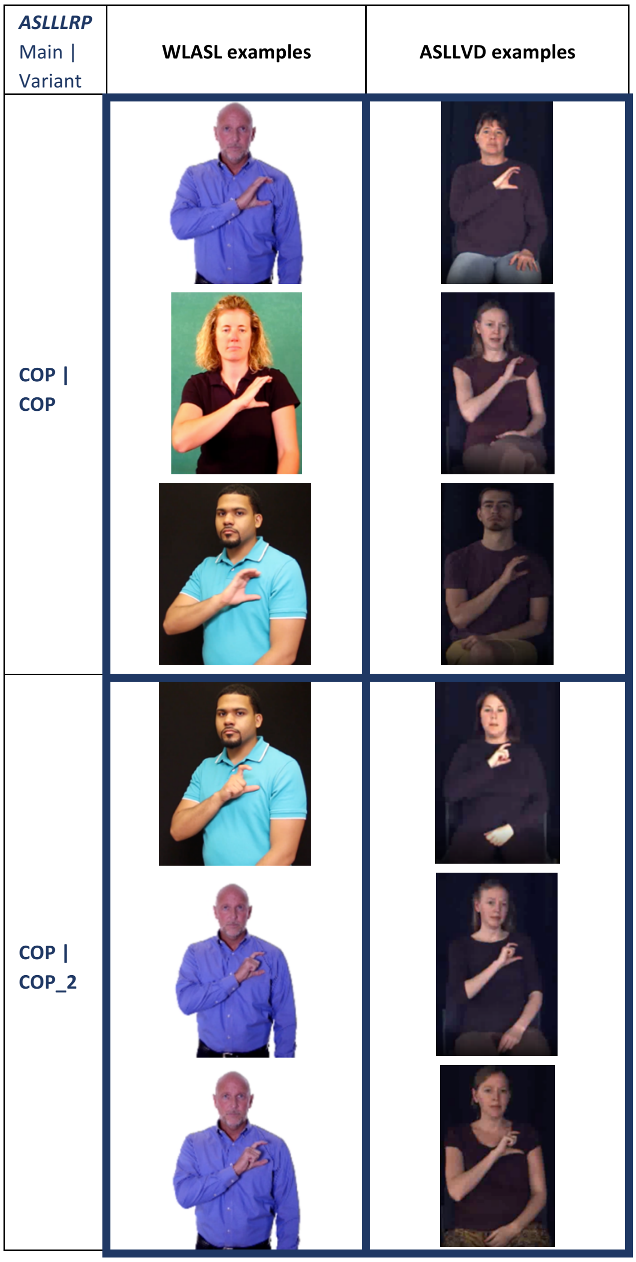



**Figure 8**. Pooling examples from ASLLVD and WLASL

| Sign Types | Min. # samples per sign | Total # distinct signs (class labels) | Total # samples | Top-1 | Top-5 |
|---|---|---|---|---|---|
| **Lexical** | **6** | 1,480 | 22,853 | **78.54%** | **94.72%** |
| **Lexical** | **12** | 983 | 18,362 | **84.23%** | **96.69%** |
| **All *** | **6** | 1,502 | 23,016 | **78.70%** | **94.79%** |
| **All *** | **12** | 990 | 18,482 | **84.70%** | **96.56%** |

*** Includes lexical signs, loan signs, and compounds**

**Figure 9.** Sign Recognition Accuracy for Different Sets of Signs (all with WLASL & ASLLVD combined)

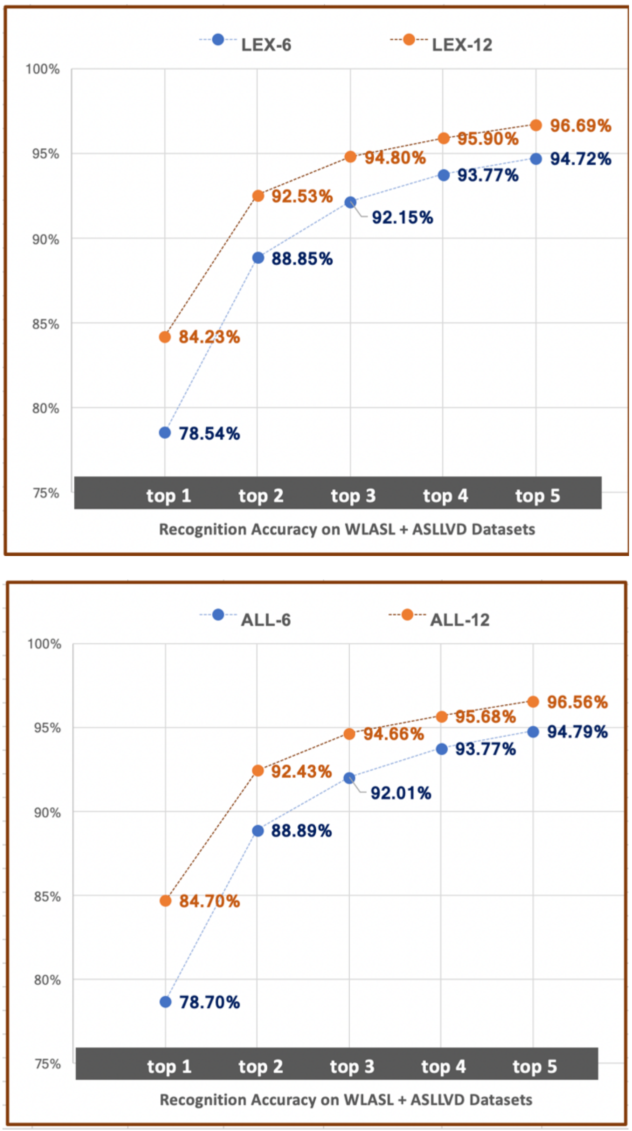



**Figure 10.** Sign Recognition Accuracy for Different Datasets (WLASL & ASLLVD combined): **for lexical *vs.* all signs (incl. compounds & loan signs) with minimum of 6 or 12 samples**

This research is based on a spatial-temporal GCN architecture for modeling skeleton keypoints, with use of both the forward and backward data streams for joints and bones for isolated sign recognition, following Dafnis et al. (2022b).

In preliminary results—with further improvements anticipated as our research proceeds—we achieved a success rate of **78.54**% for top-1, and **94.72**% for top-5; see the top graph in **Figure 10**.

We also explored how increasing the minimum number of examples per sign from **6** to **12**—thereby also decreasing the total number of signs from **18,362** total examples to **983**, the number of distinct signs for which we have at least that many examples, resulting in a more balanced dataset overall—improved recognition accuracy.

Furthermore, we expanded the set of signs considered from the combined datasets to include loan signs and compounds, in addition to lexical signs,[2] thereby increasing the number, of total examples for which we have at least 6or 12 examples per sign to **23,016**, representing **1,502** distinct signs, or **18,482** representing **990** signs, respectively. The sign recognition accuracy achieved by fusion of the forward and backward video streams is shown in **Figure 9** and **Figure 10**. This research is reported in Dafnis et al. (2022a), but for present purposes, we offer these examples of the usefulness of the consistent gloss labeling across the ASLLVD and WLASL datasets in enabling sign recognition research on the larger and richer combined dataset.

## 8. Benchmark Datasets

Details about the datasets used for our published research on sign recognition, including identification of videos used for training, validation, and testing, are available on our website: http://www.bu.edu/asllrp/signrec.html.

## 9. Conclusions

Thus, our belief is that the spreadsheet we provide with internally consistent gloss labeling for the WLASL greatly increases the value of that dataset for use in research. The fact that these gloss labels are also consistent with those used for the ASLLRP Sign Bank (i.e., the ASLLVD and other available ASLLRP data) makes it possible to use these datasets in combination, resulting in a resource that is substantially larger and richer than those datasets individually. The preliminary research on sign recognition reported in Section 7 gives an indication of the promise offered by this approach.

Furthermore, the high accuracy with which a sign can now be recognized from video within the top-5 makes this technology potentially useable in applications, such as search by video example (from the signer's webcam or a video clip identified by the user) in an all-ASL dictionary, where a user could be presented with 5 choices and asked to confirm the selection. Our research group is, in fact, currently working to develop such functionality.

## 10. Acknowledgments

For invaluable assistance with this project, we thank Indyaloreal Oliver, Gregory Dimitriadis, and Douglas Motto. We are grateful to Matt Huenerfauth for carrying out the ASLLRP Sign Bank data collection at RIT, with the help of Abraham Glasser, Ben Leyer, Saad Hassan, and Sarah Morgenthal; and to DawnSignPress for sharing video data for the ASLLRP Sign Bank. See http://www.bu.edu/asllrp/people.html for a list of the

[2] Lexical signs still represented a very large proportion of this expanded 'All' dataset; the total number of signs did not increase by a large amount. As shown here, this expansion made only a negligible difference in the recognition accuracy. However, it should be noted that the current methodology did not take advantage of the linguistic constraints on the internal structure of lexical signs, something we plan to explore further in the future, as this has proven to increase recognition accuracy in our prior research (Neidle et al., 2013; Dilsizian et al., 2014).

many, many, many people who have made important contributions to the research that gave rise to the materials described here. We also gratefully acknowledge Stan Sclaroff, Vassilis Athitsos, Ashwin Thangali, Joan Nash, Ben Bahan, Rachel Benedict, Naomi Caselli, Elizabeth Cassidy, Lana Cook, Braden Painter, Tyler Richard, Tory Sampson, and Dana Schlang for collaboration on the development of the ASLLVD dataset. The work reported here has been supported in part by grants from NSF: no. 2040638, 1763486, 1763523, and 1763569. Any opinions, findings, or conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## 11. Bibliographical References

Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., and Thangali, A. (2008). The American Sign Language Lexicon Video Dataset. *Proceedings of the IEEE Workshop on Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis*. http://vlm1.uta.edu/~athitsos/publications/athitsos_cvpr4hb2008.pdf.

Bilge, Y.C., Cinbis, R.G., and Ikizler-Cinbis, N. (2022). Towards Zero-shot Sign Language Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, July 2009*. doi: 10.1109/TPAMI.2022.3143074.

Boháček, M. and Hrúz, M. (2022). Sign Pose-based Transformer for Word-level Sign Language Recognition. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. pp. 182-191. https://ieeexplore.ieee.org/document/9707552.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., and Ringel Morris, M. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19). Association for Computing Machinery*, New York, NY. doi: 10.1145/3308561.3353774.

Cormier, K., Crasborn, O., and Bank, R. (2016). Digging into Signs: Emerging Annotation Standards for Sign Language Corpora. *Proceedings of the Workshop on Representation and Processing of Sign Languages: Corpus Mining*. 10th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA). Portorož, Slovenia, May 2016.

Dafnis, K.M., Chroni, E., Neidle, C., and Metaxas, D. (2022a). Isolated Sign Recognition using ASL Datasets with Consistent Text-based Gloss Labeling and Curriculum Learning. *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual (SLTAT 2022).* 13th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA). Marseille, France, June 2022.

Dafnis, K.M., Chroni, E., Neidle, C., and Metaxas, D.N. (2022b). Bidirectional Skeleton-Based Isolated Sign Recognition using Graph Convolution Networks. *Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA). Marseille, France, June 2022.

de Amorim, C.C. and Zanchetti, C. (2021). ASL-Skeleton3D and ASL-Phono: Two Novel Datasets for the American Sign Language. *arXiv:2201.02065*. doi: 10.48550/arXiv.2201.02065.

Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. (2014). A New Framework for Sign Recognition based on 3D Handshape Identification and Linguistic Modeling. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA). Reykjavik, Iceland, May 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/1138_Paper.pdf.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., and Ney, H. (2008). Benchmark Databases for Video-Based Automatic Sign Language Recognition. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA). Marrakech, Morocco, May 2008. http://www.lrec-conf.org/proceedings/lrec2008/pdf/287_paper.pdf.

Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J., and Giró-i-Nieto, X. (2021). How2sign: a large-scale multi- modal dataset for continuous American Sign Language. *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR),* 2021. Dataset website: https://how2sign.github.io.

Duarte, A., Albanie, S., Giró-i-Nieto, X., and Varol, G. (2022). Sign Language Video Retrieval with Free-Form Textual Queries. *Arxiv 2201.02495v1*. doi: 10.48550/arXiv.2201.02495.

Ebrahimi, M. and Ebrahimpour-komleh, H. (2022). Rough Set with Deep Neural Network for Sign Language Recognition. *TechRxiv. Preprint.* doi: 10.36227/techrxiv.19255742.v1.

Elakkiya, R. and Selvamani, K. (2019). Subunit sign modeling framework for continuous sign language recognition. *Computers and Electrical Engineering, 74*, pp. 379–390. doi: 10.1016/j.compeleceng.2019.02.012.

Hassan, A., Elgabry, A., and Hemayed, E. (2021). Enhanced Dynamic Sign Language Recognition using SlowFast Networks. *Proceedings of the 17th International Computer Engineering Conference (ICENCO)*. pp. 124-128. doi: 10.1109/ICENCO49852.2021.9698904.

Jiang, S., Sun, B., Wang, L., Bai, Y., Li, K., and Fu, Y. (2021). Skeleton aware multi-modal sign language recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* Workshops, 2021, pp. 3413-3423.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics, 15*(1), pp. 106-131. doi: 10.1075/ijcl.15.1.05joh.

Joze, H., Vaezi, R., and Koller, O. (2018). MS-ASL: A large-scale data set and benchmark for understanding American Sign Language. *arXiv preprint arXiv:1812.01053*. doi: 10.48550/arXiv.1812.01053.

Kak, A.C. (2002). Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language. *Proceedings of the IEEE Int. Conf. on Multimodal Interfaces (ICMI)*. https://dl.acm.org/doi/abs/10.1109/ICMI.2002.1166987.

Kumar, S.S., Wangyal, T., Saboo, V., and Srinath, R. (2018). Time Series Neural Networks for Real Time Sign Language Translation. *Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018,* pp. 243-248. doi: 10.1109/ICMLA.2018.00043.

Li, D., Rodriguez, C., Yu, X., and Li, H. (2020). Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020. pp. 1459-1469.

Lim, K.M., Tan, A.W.C., Lee, C.P., and Tan, S.C. (2019). Isolated sign language recognition using Convolutional Neural Network hand modelling and Hand Energy Image. *Multimedia Tools and Applications (2019), 78*, pp. 1-28. doi: 10.1007/s11042-019-7263-7.

Maruyama, M., Ghose, S., Inoue, K., Prathim Roy, P., Iwamura, M., and Yoshioka, M. (2021). Word-level Sign Language Recognition with Multi-stream Neural Networks Focusing on Local Regions. *Arxiv abs/ 2106.15989*. doi: 10.48550/arXiv.2106.15989.

Neidle, C. (2002). SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project. American Sign Language Linguistic Research Project Report No. 11: Boston University, Boston, MA. http://www.bu.edu/asllrp/asllrpr11.pdf.

Neidle, C. (2007). SignStream™ Annotation: Addendum to Conventions used for the American Sign Language Linguistic Research Project. American Sign Language Linguistic Research Project Report No. 13: Boston University, Boston, MA. http://www.bu.edu/asllrp/asllrpr13.pdf.

Neidle, C., Thangali, A., and Sclaroff, S. (2012). Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon,* 9th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA). Istanbul, Turkey, May 2012. https://open.bu.edu/handle/2144/31899.

Neidle, C., Thangali, A., Nash, J.P., and Sclaroff, S. (2011). Exploiting phonological constraints for handshape inference in ASL video. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2011*, pp. 521-528. https://www.semanticscholar.org/paper/Exploiting-phonological-constraints-for-handshape-Thangali-Nash/5e404d87b84cb5db11fb6a0c01291ff59528dceb.

Neidle, C., Opoku, A., Dimitriadis, G., and Metaxas, D. (2018). NEW Shared & Interconnected ASL Resources: SignStream® 3 Software; DAI 2 for Web Access to Linguistically Annotated Video Corpora; and a Sign Bank. *Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community.* 11th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA). Miyagawa, Japan, May 2018. https://open.bu.edu/handle/2144/30047.

Neidle, C. and Ballard, C. (2022). Why Alternative Gloss Labels Will Increase the Value of the WLASL Dataset Boston, MA: Boston University, ASLLRP Project Report No. 21. http://www.bu.edu/asllrp/rpt21/asllrp21.pdf.

Neidle, C. and Opoku, A. (2022). Documentation for Download of ASLLRP Sign Bank Citation-Form Sign Datasets. Boston, MA: Boston University, ASLLRP Project Report No. 20. http://www.bu.edu/asllrp/rpt20/asllrp20.pdf.

Neidle, C., Opoku, A., and Metaxas, D. (2022). ASL Video Corpora & Sign Bank: Resources Available through the American Sign Language Linguistic Research Project (ASLLRP). *arXiv:2201.07899*. doi: 10.48550/arXiv.2201.07899.

Orfanidou, E., Woll, B., and Morgan, G., eds. (2015). *Research Methods in Sign Language Studies: A Practical Guide,* West Sussex, UK: John Wiley & Sons.

Rastgoo, R., Kiania, K., and Escalerab, S. (2022). Word separation in continuous sign language using isolated signs and post-processing. *Arxiv 2204.00923v3*. doi: 10.48550/arXiv.2204.00923.

Theodorakis, S., Pitsikalis, V., and Maragos, P. (2014). Dynamic–static Unsupervised Sequentiality, Statistical Subunits and Lexicon for Sign Language Recognition. *Image and Vision Computing, 32*(8), pp. 533-49. doi: 10.1016/j.imavis.2014.04.012.

Zhou, B., Andonian, A., Oliva, A., Torralba, A. (2018). Temporal Relational Reasoning in Videos. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11205.* Springer, Cham. doi: 10.1007/978-3-030-01246-5_49.