

# ASLLRP DAI

## Database Access Interface

# American Sign Language Lexicon Video Dataset (ASLLVD)

<http://secrets.rutgers.edu/dai/queryPages/search/search.php>

### Important Note:

Flash and pop-ups must be enabled in your browser for the site to display the videos properly. See, e.g., <https://helpx.adobe.com/flash-player/kb/enabling-flash-player-safari.html> or <https://support.digication.com/hc/en-us/articles/115003963468-Enabling-Flash-for-Google-Chrome-Windows-Macintosh>

## Overview

The American Sign Language Lexicon Video Dataset (ASLLVD) consists of videos of >3,300 ASL signs in citation form, each produced by 1-6 native ASL signers, for a total of almost 9,800 tokens. This dataset includes multiple synchronized videos showing the signing from different angles. Linguistic annotations include gloss labels, sign start and end time codes, start and end handshape labels for both hands, morphological and articulatory classifications of sign type. For compound signs, the dataset includes annotations for each morpheme. To facilitate computer vision-based sign language recognition, the dataset also includes numeric ID labels for sign variants, video sequences in uncompressed-raw format, and camera calibration sequences.

## Development of the data set and Web interface / Personnel credits

### *Elicitation of ASL data*

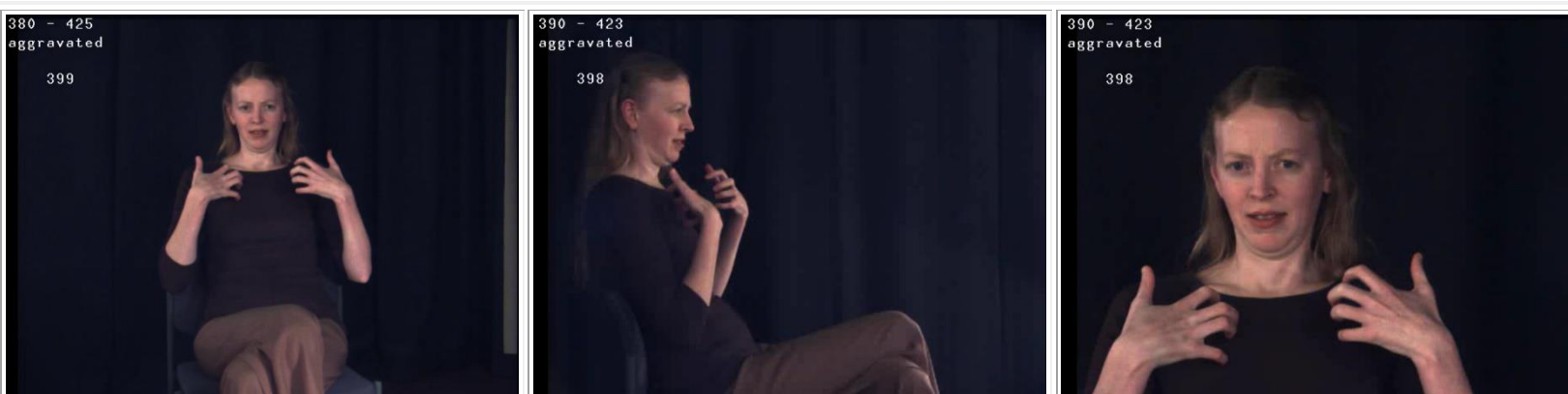
The data were collected at Boston University. The elicitation of linguistic data from ASL native signers was carried out under the supervision of Carol Neidle principally by Joan Poole Nash and Robert G. Lee. We are especially grateful to the native signers who served as subjects for this research: Elizabeth Cassidy, Braden Painter, Tyler Richard, Lana Cook, Dana Schlang, and Naomi Berlove.

Video stimuli (from the *Gallaudet Dictionary of American Sign Language* (Valli, 2002)) were presented to signers [\[see illustrations\]](#), who were asked to produce the sign they saw as they would naturally produce it. In cases where the signer reported that he or she does not normally use that

sign, we did not elicit the sign from this signer. The video stimuli for elicitation were supplemented to include additional signs that were not in the dictionary. It is interesting to note that signers did not always produce the same sign that was shown in the prompt. In cases where a signer recognized and understood that sign but used a different sign or a different version of the same sign, divergences showed up in the data set. So, in reality, a given stimulus resulted in productions that may have varied in any of several different ways: production of a totally different but synonymous sign; production of a lexical variant of the same sign; production of essentially the same sign but differing in subtle ways with respect to the articulation (as a result of regular phonological processes). These variations in production were appropriately distinguished, classified, labeled, and annotated.

### ***Video recording and processing***

Videos were captured using four synchronized cameras, providing: a side view of the signer, a close-up of the head region, a half-speed high resolution front view, and a full resolution front view. ([See details below](#))



The Computer Science personnel responsible for the recording and processing of the video data included Stan Sclaroff, Ashwin Thangali, and Vassilis Athitsos, as well as Alexandra Stefan, Eric Cornelius, Gary Wong, Martin Tianxiong Jiang, and Quan Yuan. Much of the video processing was done by Ashwin Thangali.

Video sequences collected during data capture were processed to format the video for viewing on a website. This processing aims to produce video with high fidelity in the hand and face regions as these are widely regarded as conveying the most salient information in signs. Automatic skin region segmentation is applied to each video frame. The frames are cropped to the skin region and then normalized to ensure uniform brightness within the video sequence. The processing reduces variance introduced by differences in the capture setup among different data capture sessions. The processed videos includes synchronized front and side views. The side camera was positioned to the signer's right. This is because the ASL consultants participating in the ASLLVD data collection were right hand dominant. For interested users, unprocessed and uncompressed raw video files are available for download as described further [below](#). Since these files are significantly larger (on the order of 1-2Gb each) than the web formatted video sequences, we request users to exercise caution when downloading these files.

Additional information regarding the video format chosen for website display is described [here](#).

### ***Linguistic classification, annotation, and verifications***

Lexical variants of a given sign were grouped together, and each distinct lexical variant was assigned a unique gloss label. Variants that differed only in variations attributable to regular phonological processes were not assigned distinct gloss labels. The gloss labels are consistent with those in use for our other data sets, cf. <http://secrets.rutgers.edu/dai/queryPages/>.

Linguistic annotations include unique gloss labels, start/end time codes for each sign, labels for start and end handshapes of both hands, morphological classifications of sign type (lexical, number, fingerspelled, loan, classifier, compound), and articulatory classifications (1- vs. 2-handed, same/different handshapes on the 2 hands, same/different handshapes for sign start and end on each hand, etc.). For compound signs, the dataset includes annotations as above for each morpheme.\*

See [below](#) for documentation of glossing conventions.

Linguistic classifications of signs, as well as annotations and verifications of the annotations — involving an extraordinary time investment — were carried out at Boston University under the supervision of Carol Neidle. The major contributors and consultants on this project included: Jessica Scott, Joan Poole Nash, Tory Sampson, Donna Riggle, Braden Painter, Amelia Wisniewski-Barker, Indya Oliver, Corbin Kuntze, Rachel Benedict, Ben Bahan, Jonathan McMillan, Alix Kraminitz, Daniel Ferro, Chrisann Papera, Caelen Pacelli, Jaimee DiMarco, Jon Suen, and Isabel Zehner.

The annotations were initially carried out using SignStream® software developed to facilitate the linguistic annotation of video data. The original version of SignStream® (through version 2.2.2) was implemented as a Mac Classic application, under the direction of Carol Neidle and others at Boston University (including Dawn MacLaughlin, Ben Bahan, Robert G. Lee, and Judy Kegl) by David Greenfield, under the direction of Otmar Foelsche, at Dartmouth College. A Java reimplementaion (version 3) introduced new features to enable annotation of phonological and morphological information (especially in relation to handshapes). SignStream® 3 has been implemented to date by Iryna Zhuravlova, and it still under development.

Verifications and corrections of the annotations of the lexical data, as well as morphological groupings of related signs, were greatly facilitated by a very powerful software tool developed by Ashwin Thangali. Ashwin, in conjunction with his dissertation research, developed a remarkable interface to facilitate this work: the Lexicon Viewer and Verification Tool (LVVT). See:

\* Ashwin Thangali [2013] [\*Exploiting Phonological Constraints for Handshape Recognition in Sign Language Video\*](#). Doctoral Dissertation, Boston University.

\* Ashwin Thangali, Joan P. Nash, Stan Sclaroff, and Carol Neidle [2011] [\*Exploiting Phonological Constraints for Handshape Inference in ASL Video\*](#), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2011.

### ***Search interface***

In order for these data to be shared publicly, the materials from this dataset were then incorporated into our Data Access Interface, for which Christian Vogler (Gallaudet University) has been the principal developer, cf.

\* Carol Neidle and Christian Vogler [2012] [\*"A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface,"\*](#) 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey, May 27, 2012).

At Rutgers University, Jessy Sheng and Gang Yang successfully developed a searchable online database to allow access to the annotated ASLLVD video examples (digital assets). They developed a relational database schema to persist the indices of the various keywords and terms as well as matching digital assets, and adapted the DAI web interface to allow for easy search and retrieval of the lexical data. As part of this effort, they created an optimized search algorithm that resulted in improved search times for the end users. This work resulted in a user friendly website for searching the sign language lexicon for matching keywords and terms. Augustine Opoku provided assistance in designing the database and search algorithm for this application and provided supervision for this project.

Further integration, to allow for connections between our continuous signing data and the citation-forms contained in the ASLLVD, is planned for the near future for the DAI.

## Information about the annotations and the search interface

### ***Documentation***

Each distinct sign (and each distinct lexical variant of a given sign) has a unique gloss label. Information about the annotation conventions in use is available from:

\* Carol Neidle [2002] [SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project](#). American Sign Language Linguistic Research Project Report No. 11, Boston University.

\* Carol Neidle [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.

The plus sign ("+") at the end of a gloss indicates a repetition/reduplication of the end portion of the sign (beyond any repetition that is part of the base form of the sign). A "+" is also used to connect two parts of a compound.

The handshape labels used for these annotations are displayed here: <http://www.bu.edu/asllrp/cslgr/pages/ncslgr-handshapes.html>

For further details about the dataset, see:

\* Carol Neidle, Ashwin Thangali and Stan 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 27, 2012.

### ***About the search interface***

You can search for a partial or complete text string in the glosses. If you leave the search box empty and choose a partial search, you'll get a list of all the items in the database.

You can expand the triangles to see the items that are lexical variants (which have distinct glosses) and/or that have different numbers of +'s.

You can click on the cell that indicates the number of items, and you'll get a screen that shows the glosses, images of start and end frames, plus start and end handshapes, etc. You can choose "play movie" to view any individual movie, or "play composite" to see all tokens of a given sign variant (potentially with differing numbers of +'s) together.

Additional information regarding the video format chosen for website display is described [here](#).

## Statistics

As displayed in Figure 1 below, taken from [Neidle, Thangali, and Sclaroff \[2012\]](#), we collected a total of 3,314 distinct signs, including variants (for a total of 9,794 tokens). Among those were 2,793 monomorphemic lexical signs (8,585 tokens) and 749 tokens of compounds. Column 4 shows the total number of sign variants we have as produced by 1 signer, 2 signers, etc. Since in some cases we had more than one example per signer, the total number of tokens per sign was, in some cases, greater than 6.

Class of signs	Number of signs	Number of sign variants	# sign variants with { 1, 2, 3, 4... } consultants		# tokens (examples) per sign { 1,2,...,6, >6 }		Number of sign tokens	
Monomorphemic lexical signs	2,284	2,793	x1	621	587	x1	8,585	Two-handed
			x2	989	858	x2		5,713
			x3	394	386	x3		67%
			x4	563	491	x4		
			x5	85	142	x5		One-handed
			x6	141	154	x6		2,873
					175	>6		33%
Compound signs	289	329	x1	129	117	x1	749	
			x2	106	107	x2		
			x3	48	46	x3		
			x4	33	33	x4		
			x5	4	11	x5		
			x6	9	13	x6		
					2	>6		
Number signs	76	88					260	
Loan signs	46	52					136	
Classifier constructions	27	31					38	
Fingerspelled signs	21	21					25	
ALL	2,742	3,314	--	--	--	--	9,794	

Figure 1. Overview of statistics from the dataset.

To make it clear how this chart should be read, a total of 2,284 monomorphemic lexical signs were collected. For some signs, there is more than one variant, resulting in a total number of distinct sign variants that is greater: 2,793. For 621 of those sign variants, we have examples from a single signer; for 989 of them, we have examples from 2 signers, etc., and for 141 of those sign variants, we have examples from all 6 of our native signers. Since we have more than one example from a given signer in some cases, the total number of tokens per sign may be greater than the total number of signers whose productions of that sign are included in our data set. In fact, for 175 of the signs, we have more than 6 tokens. (For 2 of the signs, we have as many as 19 tokens.)

## Download video data and annotations

Linguistic and video information for signs in the lexicon dataset are available as an Excel file with the caveat that all this is work in progress:

[http://www.bu.edu/asllrp/dai-asllvd-BU\\_glossing\\_with\\_variations\\_HS\\_information-extended-urls-RU.xlsx](http://www.bu.edu/asllrp/dai-asllvd-BU_glossing_with_variations_HS_information-extended-urls-RU.xlsx)

Explanation of what is in the spreadsheet:

L5														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Main New Gloss	Consultant	Main New Gloss	Gloss Variant	D Start HS	N-D Start HS	D End HS	N-D End HS	Passive Arm	Combined	Separate	Session	Scene	Start
1	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
2	TWENTY	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
3		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
4		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
5		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
6		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
7		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
8	ALONE	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
9		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
10		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
11		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
12		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
13		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
14		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
15		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
16		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
17	LONELY	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
18		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
19		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
20		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
21		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
22	BACHELOR	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
23		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
24		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
25		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
26		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
27		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
28		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
29		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
30		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
31		*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****

Glosses have been assigned only so as to assure a unique gloss for each sign variant. Lexical variants of a sign have been grouped together, i.e., with the same gloss in Column D but a distinct glosses for each variant in Column E.

The use of the + sign indicates one repetition/reduplication beyond what would be the base form of a sign. Glosses containing different numbers of +'s are considered equivalent for purposes of grouping (i.e., are not considered as distinct sign variants).

The dominant start handshape, non-dominant start handshape (if any), dominant end handshape, and non-dominant end handshape (if any) are listed in columns H and I. The handshape palette showing the handshapes associated with these labels is available from: <http://www.bu.edu/asllrp/cslgr/pages/ncslgr-handshapes.html>.

Example:

B	C	D	E	F	G	H	I	J
ACCIDENT	*****	*****	*****	*****	*****	*****	*****	*****
	Liz	ACCIDENT	ACCIDENT	S	S	S	S	N
	Naomi	ACCIDENT	ACCIDENT	S	S	S	S	N
	*****	*****	*****	*****	*****	*****	*****	*****
	Liz	ACCIDENT	(5)ACCIDENT	5	5	A	A	N
	Naomi	ACCIDENT	(5)ACCIDENT	5	5	10	10	N
	Lana	ACCIDENT	(5)ACCIDENT	5	5	S	S	N
	Dana	ACCIDENT	(5)ACCIDENT	5	5	S	S	N
	*****	*****	*****	*****	*****	*****	*****	*****
	Tyler	ACCIDENT	(3)ACCIDENT	3	3	S	A	N

In this case, the sign for 'accident' has three lexical variants, which are distinguished by handshape but which have otherwise the same basic movement. These are considered to be lexical variants, and they have distinct glosses, in this case with the distinguishing handshape noted as part of the gloss label (although that is not necessarily the case for lexical variant glosses). See illustration of start and end handshapes for these three variants. In some cases, the alternation in handshape, e.g., between the A and S hand shapes shown for the end hand shapes

of (5)ACCIDENT, is quite productive under appropriate phonological conditions and is not a property associated specifically with this lexical item.

Columns K and L contain hyperlinks to the combined and individual movie files, respectively. Columns Q and R contain alternative links to those in K and L, enabling download from the Rutgers mirror site rather than the BU site.

Columns M, N, O, and P are useful to obtain the unprocessed video sequences: session, scene, start frame, end frame.

The URL to download the unprocessed video sequences in VID format is the following (the associated software for reading VID files is described further below):  
<http://csr.bu.edu/ftp/asl/asllvd/asl-data2/<session>/scene<scene#>-camera<camera#>.vid>

The corresponding MPEG-4 movie files to easily view the data can be obtained using:  
<http://csr.bu.edu/ftp/asl/asllvd/asl-data2/quicktime/<session>/scene<scene#>-camera<camera#>.mov>

Note about the availability of QuickTime files: QuickTime MOV files for camera1 (standard definition, front view) are available for all the sessions. However, only a small number of sessions contain Quicktime format files for camera views 2 through 4. Users who require the other camera views would need to download the VID format files

You would need to use either Matlab mex files or the C++ API in the 'vid\_reader' library to read frames from a VID format video file,  
[http://csr.bu.edu/ftp/asl/asllvd/asl-data2/vid\\_reader/vid\\_reader.tar.bz2](http://csr.bu.edu/ftp/asl/asllvd/asl-data2/vid_reader/vid_reader.tar.bz2)  
(vid\_reader is pre-compiled for Windows and Linux, it is straightforward to build on a new platform using mex\_cmd.m)

A quick example for how this library works in Matlab ( see main.cpp for standalone C++ ):

```
>> [ vid_handle, vid_info ] = vidOpenMex( 'ASL_2008_01_11/scene10-camera1.vid' );  
>> vid_info
```

```
vid_info =
```

```
640          %          width          in          pixels  
480          %          height         in          pixels  
2020 % number of frames in video
```

```
>> video_frame = vidReadMex( vid_handle, 100 ); % frame number input is in range [0,  
#frames          -          1          ]  
>>          imshow(          uint8(          video_frame          )          );  
>> vidCloseMex( vid_handle );
```

### **\*Important notes about compounds and morpho-phonological classifications**

1. The ASLLVD contains a rich collection of compound signs. Although these signs have been coded with linguistic attributes for the constituent morphemes, such attributes have not yet been exported into the above Excel spreadsheet. We do not unfortunately have an estimate for when these attributes will become available in an easily accessible format. The start/end frame numbers



and start/end handshapes displayed in the spreadsheet should therefore be taken to refer to the entire compound sign, rather than the constituent parts.

2. The data included in the spreadsheets that are currently available also do not include the articulatory classifications (1- vs. 2-handed, same/different handshapes on the 2 hands, same/different handshapes for sign start and end on each hand, etc.).

Video	capture	setup
Camera1 (the front view), camera2 (the side view -- signer's right) and camera3 (face closeup) are 60fps 640 x 480, while camera4 (a high-resolution front view) is 30fps 1600 x 1200. All four cameras are time-synchronized. Geometric calibration sequences are available for most sessions. Color calibration sequences using a Munsell color chart are available for more recent video capture sessions. The last one or two scenes in each session are typically calibration sequences.		

### Video format information

The start/end numbers in the Excel spreadsheet, as well as those displayed within the video in the upper left corner, are "frame numbers". These two numbers should correspond exactly.

The processed videos clipped to individual signs (available through the DAI as well as through the spreadsheet) have a 50 frame buffer at the start as well as at the end. The ones that have a smaller buffer are those signs that were very close to the start of a video capture or towards the end of the video capture. The first case (which is the useful case) is easy to identify because those signs have a start frame number less than 50.

The composite videos do not have a buffer; the sign starts right away.

The videos were all captured at 60 frames a second in the studio. The processed videos are displayed at 1/4th that rate (i.e., 15fps), with no missing frames. If these are played at 4x the speed (for example in VLC player), the user sees the captured frame rate.

All videos in the dataset (both MOV and VID) have been encoded to ensure that each video frame is a key frame (B frames in MPEG). There are no interpolated frames (I frames in MPEG). Users can step frame-by-frame with video players that support this feature (e.g., [QuickTime](#)).

## Caveat about older data set from 2008

The data collection made available here supersedes a very early set of data that had formed the basis for V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Q. Yuan, & A. Thangali, "The ASL Lexicon Video Dataset." First IEEE Workshop on CVPR for Human Communicative Behavior Analysis. Anchorage, Alaska, Monday June 28, 2008. That early research made use of some data that had *not* undergone the (critically important) linguistic analysis, grouping, and annotation that were done for the current collection. (The signs were grouped there based on the stimulus that had been used to elicit them, not based on what the signers actually produced -- which often diverged in important and interesting ways from the stimulus.) Please **do not use** the data that was shared in conjunction with that 2008 paper.

# Acknowledgment of grant support

We are very grateful to funding from the National Science Foundation, which made this research possible:

"HCC: Large Lexicon Gesture Representation, Recognition, and Retrieval." Award #0705749 - Stan Sclaroff and Carol Neidle.

"III: Medium: Collaborative Research: Linguistically Based ASL Sign Recognition as a Structured Multivariate Learning Problem." Award #0964597 - Dimitris Metaxas; 0964385 - Carol Neidle.

"II-EN: Infrastructure for Gesture Interface Research Outside the Lab." Award #0855065 - Stan Sclaroff, Carol Neidle, and Margrit Betke.

"Collaborative Research: II-EN: Development of Publicly Available, Easily Searchable, Linguistically Analyzed, Video Corpora for Sign Language and Gesture Research." Award #0958442 - Carol Neidle and Stan Sclaroff; #0958247 - Dimitris Metaxas; #0958286 - Vassilis Athitsos.

"Collaborative Research: CI-ADDO-EN: Development of Publicly Available, Easily Searchable, Linguistically Analyzed, Video Corpora for Sign Language and Gesture Research." Award #10592180 - Carol Neidle and Stan Sclaroff; #1059281 - Dimitris Metaxas; #1059221 - Benjamin Bahan and Christian Vogler; #1059235 - Vassilis Athitsos.

## Terms of use

The data available from these pages can be used for research and education purposes, but cannot be redistributed without permission.

Commercial use, without explicit permission, is not allowed, nor are any patents and copyrights based on this material.

Those making use of these data must, in resulting publications or presentations, cite: The American Sign Language Lexicon Video Dataset (ASLLVD) Corpus and these publications:

\* Carol Neidle, Ashwin Thangali and Stan 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 27, 2012.

\* Carol Neidle and Christian Vogler [2012] "A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface," Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey.

and also include the following URL: <http://secrets.rutgers.edu/dai/queryPages/search/search.php>.