

**American Sign Language  
Linguistic Research  
Project**



**A User's Guide to the American Sign Language  
Linguistic Research Project (ASLLRP)  
Data Access Interface (DAI) 2 — Version 2**

**Carol Neidle & Augustine Opoku**

Boston University, Boston, MA

**Report No. 18  
American Sign Language  
Linguistic Research Project  
<http://www.bu.edu/asllrp/>**

**© May 2020  
Boston University**

For updated information about available data sets (as of March 2022), please see  
<http://www.bu.edu/asllrp/about-datasets.pdf>

For information about new features (as of May 2021) — including ASLLRP Sign  
Bank Search by Related English Words — see:  
<http://www.bu.edu/asllrp/New-features-DAI2.pdf>

## 1 Background Information about the DAI

The Data Access Interfaces (**DAI** and **DAI 2**) are web-based interfaces designed to allow users to browse, search, and download linguistically annotated video data of various kinds collected from native ASL signers as part of the American Sign Language Research Project (ASLLRP - <http://www.bu.edu/asllrp/>). The ASLLRP is a collaborative venture that has included researchers from Boston, Rutgers, and Gallaudet Universities and the Rochester Institute of Technology, in collaboration with DawnSignPress (which publishes American Sign Language (ASL) and Deaf culture-related videos and books). See <http://www.bu.edu/asllrp/people.html> for a list of the many, many people who have made important contributions to the research that made possible the development of the materials described here, including especially Carol Neidle, Stan Sclaroff, Ashwin Thangali, Vassilis Athitsos, Dawn MacLaughlin, Robert G. Lee, Joan Nash, Jessica Scott (BU); Dimitris Metaxas, Gregory Dimitriadis (Rutgers U.); Ben Bahan, Christian Vogler (Gallaudet U.); and Augustine Opoku (consultant). This research has been supported by grants from the National Science Foundation (<http://www.bu.edu/asllrp/nsf.html>). Current work on this project is funded by NSF grants

#1763523, 1763486, and 1763569 “CHS: Medium: Collaborative Research: Scalable Integration of Data-Driven and Model-Based Methods for Large Vocabulary Sign Recognition and Search.”

The video corpora that have been collected at Boston University have been linguistically annotated, in keeping with our annotation conventions (see Section 7), using annotation software developed by our group, SignStream® (described in Section 6). With respect to English-based gloss labels for signs, care has been taken to enforce a 1-to-1 correspondence between **sign** and **gloss label**. The categorization and labeling of signs, as well as establishment of conventions for annotations of non-manual events, has been a major undertaking that posed many challenges (see, e.g., [1], [2], and [7]). The annotations also available in an XML export format (see Section 8.3).

## 2 The original DAI: <http://secrets.rutgers.edu/dai/queryPages/>

The first version of the **DAI** was implemented primarily by Christian Vogler (following up on preliminary work in which many people had been involved). The original **DAI** is still available and provides access to the **National Center for Sign Language and Gesture Resources (NCSLGR) corpus**. It is described in [8].

### 2.1 *National Center for Sign Language and Gesture Resources (NCSLGR) Corpus*

The video data for the **NCSLGR Corpus** of continuous signing were mostly collected at Boston University (but with inclusion of a few videos shared by DawnSignPress). Annotations were carried out using SignStream® version 2; they include start and end frames and labels for all events: English-based gloss labels for manual signs<sup>1</sup> as well as non-manual information, including grammatical markings for things such as question status, negation, topic/focus, conditional/when, relative clauses, etc., and anatomical behaviors, such as head nods and shakes, eyebrow height, eye aperture and gaze.

---

<sup>1</sup>These annotations include a single label per sign and do not overtly display whether the sign is 1- or 2-handed; and they also do not include handshapes. Enhanced representations for manual signs became available in SignStream® 3; thus, there is a richer set of annotations available for the data shared through **DAI 2** (see Sections 3 and 4).

Most of the data in this corpus are from four native ASL signers. The dataset includes 1,866 distinct *signs* (i.e., grouping together very slight variants in production). The total number of sign tokens is 11,854. Restricting consideration to signs other than gestures and classifiers, there are 1,278 distinct signs, and a total of 10,719 tokens. 1,002 of the utterances in this collection are part of 19 short spontaneous narratives. The remaining 885 utterances were elicited to illustrate a variety of constructions and sentence types.

For information about downloading these data, see: <http://www.bu.edu/asllrp/ncslgr-for-download/download-info.html>.

## 2.2 Access to the **NCSLGR Corpus** via the original DAI

Searches can be conducted based on characters contained in the glosses and properties of signs, with the possibility of limiting the results to particular signers or SignStream® collections. The search interface is shown in Figure 1.

The screenshot shows a web-based search interface. At the top, there is a 'Search' button. Below it is a 'DATA SELECTION' section with a 'Sign Options' sub-section. This sub-section contains a search input field labeled 'Search all...'. Below the input field are several radio and checkbox options:
 

- Radio buttons:  All Signs,  All Signs [excl. Gestures],  All Signs [excl. Gestures & CL]
- Checked checkboxes: Lexical Signs, Index Signs, Fingerspelled Signs, Name Signs, Loan Signs.
- Unchecked checkboxes: Gestures.
- Radio buttons for 'Hand':  Dominant,  Non-Dominant,  Don't Care.
- Dropdown menus: 'Classifiers' (with  All Classifiers (# occurrences) selected), and 'Parts of Speech' (with  All Parts of Speech (# occurrences) selected).
- A 'Clear All' button is located at the bottom of the 'Sign Options' section.

 Below the 'DATA SELECTION' section is a 'DATA SOURCE' section with a 'Limit to...' dropdown menu set to 'All Sources'. At the bottom is a 'PARTICIPANT' section with a 'Limit to...' dropdown menu set to 'All Participants'.

**Figure 1. Search interface**

The result of a search through the **NCSLGR Corpus** for the string “FATHER” is illustrated in Figure 2, with the search string highlighted in yellow; and a display of the first several results for “Ben” signing “FATHER” is shown in Figure 3. Clicking on the image in column 3 of Figure 3 will play the video for the sign that was the object of the search query; clicking on the image in column 2 will play the clip of the entire utterance containing that sign.<sup>2</sup>

<sup>2</sup> Flash and pop-ups must be enabled in your browser for the site to display the videos properly. See <https://helpx.adobe.com/flash-player/kb/enabling-flash-player-safari.html> or <https://www.dailydot.com/debug/how-to-enable-flash-chrome/>.

Sign	Occurrences	Benjamin Bahan	Freda Norman	Lana Cook	Marlon Kuntze	Michael Schlang	Norma Bowers Tourangeau	Sam Supalla	Steve McCullough
▼ FATHER (2)	50	18	0	0	0	9	20	0	3
FATHER	35	18	0	0	0	4	11	0	2
FATHERwg	15	0	0	0	0	5	9	0	1
MOTHER+FATHER	9	2	0	0	0	6	1	0	0
▼ GRANDFATHER (3)	4	0	0	0	0	2	0	0	2
(2h)GRANDFATHER	2	0	0	0	0	2	0	0	0
GRANDFATHER	1	0	0	0	0	0	0	0	1
GRANDFATHER:i	1	0	0	0	0	0	0	0	1

Figure 2. DAI search through NCSLGR Corpus for the string “FATHER”

File Name-Utterance	Utterance Video	Sign Video	Full Gloss	Rough Gloss
<input type="checkbox"/> ncslgr10a.xml-21			Show ...	topic/focus: _____foc/top1 main gloss: LEARN+AGENT POSS-3p:i FATHER BUY HOUSE
<input type="checkbox"/> ncslgr10a.xml-24			Show ...	wh question: _____q/wh main gloss: LEARN+AGENT POSS-3p:i FATHER FUTURE BUY WHICH HOUSE part:indef
<input type="checkbox"/> ncslgr10a.xml-31			Show ...	main gloss: FATHER SHOULD+ BUY BOOK+
<input type="checkbox"/> ncslgr10a.xml-32			Show ...	wh question: _____q/wh main gloss: FATHER SHOULD BUY THAT BOOK (Y)WHY
<input type="checkbox"/> ncslgr10a.xml-39			Show ...	main gloss: IX-3p:i fs-JOHN POSS-3p:j FATHER CAR

Figure 3. DAI search through NCSLGR Corpus for the string “FATHER”: First several results for “Ben” signing “FATHER”

### 3 DAI 2: <http://dai.cs.rutgers.edu/dai/s/dai>

Subsequent development has been carried out by **Augustine Opoko**, who has designed and implemented **DAI 2**, with many additional features beyond those available in the original **DAI**. **DAI 2** also provides access to more recently collected and more richly annotated data (as the annotations for the newer datasets were carried out using SignStream® version 3, more powerful than previous versions of the annotation software). The data available through **DAI 2** currently include a large set of utterances that are part of the expanding **ASLLRP SignStream® 3 Corpus**. **DAI 2** provides various ways to browse, search, and download these data.

#### 3.1 The American Sign Language Linguistic Research Project (ASLLRP) SignStream® 3 Corpus

This video corpus of is an expanding collection of continuous signing from ASL native signers, which was collected at Boston University. Synchronized videos display the signing from front and side, as well as providing a close-up view of the face. These videos were annotated with SignStream® 3 (see Sections 6 and 7) with labels and start and end frames for all events: manual English-based gloss labels (with information provided for both hands), sign type, start and end handshapes, as well as non-manual information, such as grammatical markings for things such as

question status, negation, topic/focus, conditional/when, relative clauses, plus anatomical behaviors, such as head nods and shakes, eye aperture and gaze, etc.

As of May 2020, this consists of 38 SignStream® collections, containing a total of 1,810 utterances, with just over 13,000 total signs (representing more than 1950 distinct signs or sign variants) from 4 native ASL signers.

### 3.2 Access to the ASLLRP SignStream® 3 Corpus via DAI 2

#### 3.2.1 Sign-level searches

The search interface is here: <http://dai.cs.rutgers.edu/dai/s/dai>. Searches can be performed and constrained based on the following:

- a specific character string in the gloss on the dominant or non-dominant hand (or either)

**Search** - ASLLRP Linguistic Annotations and Video Files

Search All..

Whole  Partial

**Hand**  Dominant  Non-Dominant  Don't Care

- signs of particular types (lexical, number, loan, fingerspelled, gesture, classifier of various kinds, etc.); compounds, non-compounds, or either

All Signs  All Signs [excl. Gestures]  All Signs [excl. Gestures & CL]

Lexical Signs  Loan Signs  Fingerspelled Signs

Number Signs  Gestures

**+ All Classifiers (# occurrences)**

- Body (e.g., BCL"put arms around friend") (5)
- Body Part (e.g., (2h)BPCL:B"taps foot") (2)
- Descriptive (e.g., DCL"curly hair") (93)
- Instrument (e.g., ICL"playing jacks") (15)
- Locative (e.g., LCL:B"leaf drifting to the ground") (7)
- Plural (e.g., PCL:4"long line of people") (6)
- Semantic (e.g., SCL:1"person walking stiffly") (102)

Only Compounds  Only Non-Compounds  Either

- handshapes on the dominant and/or non-dominant hand, at the start or end; presence of reduplication or passive base arm

**Restrict Search:** Restrict your search to **non-compounds** with the following options:

1-handed    2-handed    Name Signs

---

**Restrict Search:** Restrict your search to **lexical signs only**, also meeting these criteria:

Marked number of hands

**Dominant Hand**

Same start/end hs  
 Different start/end hs

**Non-Dominant Hand (if present)**

Same start/end hs  
 Different start/end hs

---

Start HS   End HS

Start HS   End HS

Reduplication    Passive base arm

- specific SignStream® collection

Available collections are listed, with the number of signs contained in each in parentheses, e.g.

- [+DataSource \(32603\)](#)
  - Cory\_2013-6-27\_sc112.xml (1870)
  - Cory\_2013-6-27\_sc113.xml (1771)

- specific signer

Available singers are listed, with the number of signs from each contained in parentheses, e.g.

- [+Participant \(32603\)](#)
  - Ben (338)
  - Cory (12947)
  - Jonathan (7441)
  - Rachel (11877)

For example, searching for whole word “FATHER” in the glosses gives the following results:

Sign	Occurrences	Cory	Jonathan	Rachel
▼ FATHER (2)	<a href="#">51</a>	<a href="#">28</a>	<a href="#">12</a>	<a href="#">11</a>
FATHER	<a href="#">44</a>	<a href="#">22</a>	<a href="#">12</a>	<a href="#">10</a>
FATHER+	<a href="#">7</a>	<a href="#">6</a>	<a href="#">0</a>	<a href="#">1</a>
FATHERwg	<a href="#">7</a>	<a href="#">0</a>	<a href="#">7</a>	<a href="#">0</a>
MOTHER+FATHER	<a href="#">44</a>	<a href="#">14</a>	<a href="#">2</a>	<a href="#">28</a>

**Figure 4. Search for the partial string "FATHER"**

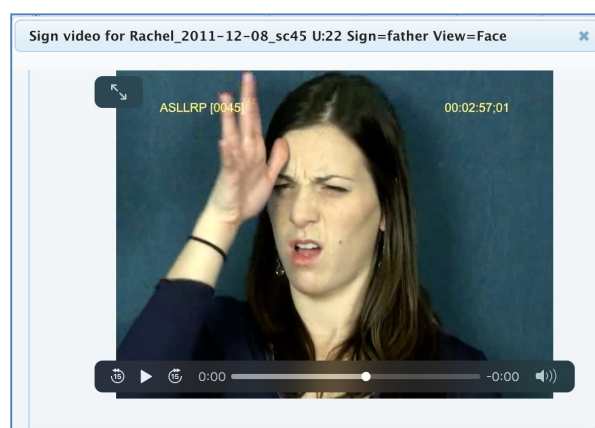
The + sign indicates that movement of the end of the sign is repeated once. FATHERwg is a variant of the sign FATHER that involves finger wiggling. And MOTHER+FATHER is a compound meaning ‘parent(s)’. The table shows total number of occurrences, as well as productions by specific signers. Clicking on the number in any cell will display those results.

Selecting the set of examples from “Rachel” for “FATHER” will display ten utterances, with the search term highlighted; the first two results appear in Figure 5—with the start and end handshapes for each gloss left- and right-aligned to that gloss. The principal non-manual grammatical markers are also displayed. So, the first example below includes a wh-question followed by a negative statement. The second example has an initial topic, as marked.

File Name-Utterance	Utterance Video	Sign Video	View	Rough Gloss
<input type="checkbox"/> Rachel_2011-12-08_sc45-U-22			Gloss.. Sign: Utterance:	negative: _____ neg wh question: _____ q/wh dh: (25)WHY FATHER (P)UPSET (1h)part:indef KNOW +(1h)NEG dh shape: engl trans: Why is dad upset? I have no idea...
<input type="checkbox"/> Rachel_2012-02-14_sc63-U-22			Gloss.. Sign: Utterance:	topic/focus: _____ topic dh: POSS-3p:i FATHER GO-OUT 5"perplexed" dh shape: ndh: _____ 5"perplexed" ndh shape: engl trans: His/Her father left.

**Figure 5. First two examples of Rachel signing FATHER**

The name of the SignStream® collection plus the utterance number within that collection is included in the first column of Figure 5. The checkboxes make it possible to add specific examples to the Download Cart; see section 8. Clicking on the icons in column 4 will display the sign video or the entire utterance video – front, face, or side view; an example is shown in Figure 6 of the facial close-up for the sign video. The arrow keys make it possible to step through the video frame by frame. The filename and time within the whole video are also shown.



**Figure 6. Playing the close-up of the face for the selected video**

Clicking on the blue Gloss button visible in Figure 5 reveals a more complete display of the non-manual events (a rough approximation of the precise time-alignment that would be visible in the SignStream® display). For the first sentence above, this is shown in Figure 7.

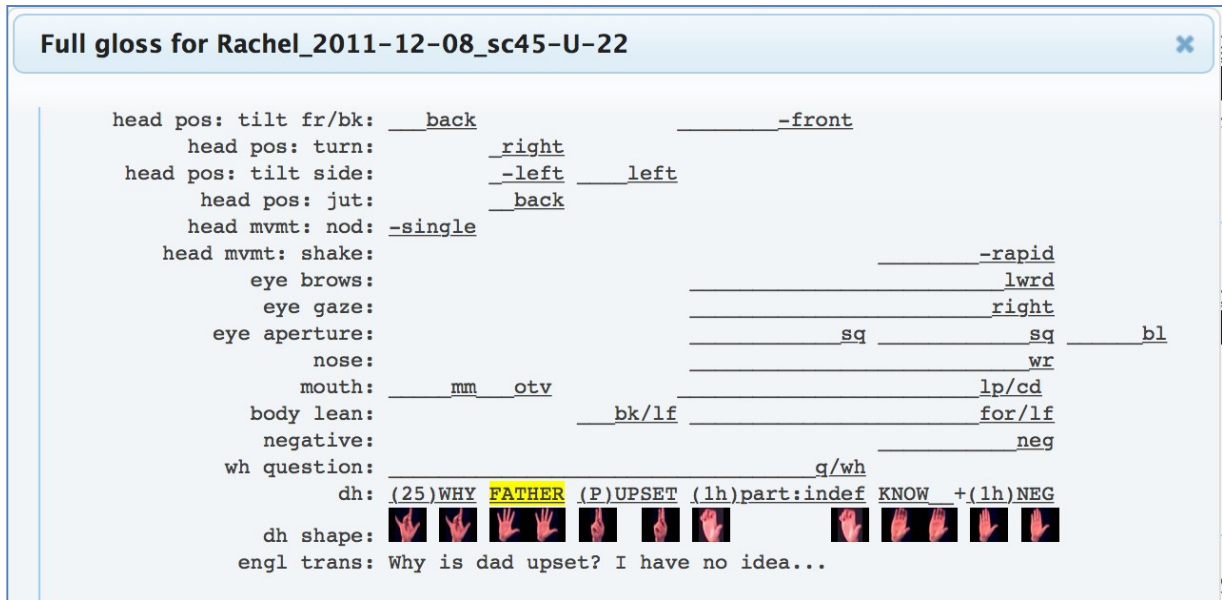


Figure 7. More detailed view of non-manuals (with approximate time-alignments)

### 3.2.2 Utterance-level searches

Utterance-level searches make it possible to view all utterances in the collections, or to limit the selection based on any or all of the following:

- the presence of a specific grammatical marking and/or specific anatomical non-manual actions, some of which are shown below

Utterance Search  SUBMIT

Limit to (Limit the search results to the following options)

- [+ Grammatical Construction \(2594\)](#)
  - negative (532)
  - wh question (277)
  - yes-no question (236)
  - [+ rhetorical question \(232\)](#)
    - rhq (200)  rhq2 (32)
  - [+ topic/focus \(877\)](#)
    - sentence-initial adverb (91)
    - focus (149)  topic (637)
  - [+ conditional/when \(363\)](#)
    - conditional (250)  when (113)
  - [+ role shift \(38\)](#)
    - rs:other (17)  role shift other person (21)
  - relative clause (39)



And (Combine grammatical construction or non-manual markings)

- + Non-manual Markings (30003)
- + Head Position (9499)
- + Head Movement (2618)
- + head mvmt: nod (852)
- slight slow head nod (30)  slight single head nod (112)
- rapid (127)  slow (82)  single (293)  slight rapid head nod (208)
- + head mvmt: shake (1119)
- rapid (259)  slow (189)  single (185)  slight rapid head shake (367)  slight slow head shake (55)
- slight single head shake (64)
- + head mvmt: side to side (37)
- rapid (12)  slow (13)  single (12)
- + head mvmt: jut (610)
- further forward (4)  slightly forward (49)
- slightly forward (14)  slightly forward (2)  slightly forward (2)
- forward (316)  back (60)  slightly forward (24)  slightly back (111)
- left (1)
- slightly forward (19)  further back (3)  further forward (2)
- slightly forward (3)
- + face (17302)
- + eye brows (3731)
- left raised/right lowered (3)  raised (1221)  lowered (694)
- slightly lowered (609)  left raised/right furrowed (2)
- slightly raised (564)  further lowered (200)
- left raised/right furrowed (2)  further raised (220)
- raised-furrowed (2)  raised-furrowed (187)
- left raised/right furrowed (1)
- right raised/left furrowed (9)
- further raised-furrowed (1)
- + eye gaze (1665)
- right (303)  up/left (17)  to addressee (77)  into space (228)  up (23)
- up/right (25)  down/left (82)  down (107)  left (611)  watch hands (40)
- down/right (152)
- + eye aperture (8932)
- slightly lowered (632)  further lowered (551)
- slightly squinted (896)  further squinted (416)  wider (96)
- slightly wide (799)  blink (2427)  squint (1070)  wide (849)  lowered lid (977)
- closed (219)
- + nose (280)
- slightly wrinkled (1)  wrinkle (112)  slightly wrinkle (1)
- tensed (109)  slightly tensed (24)  further tensed (3)
- slightly wrinkled (3)  wrinkle left (22)  wrinkle right (5)
- + mouth (2587)
- + cheeks (107)
- puffed (48)  tensed (19)  puff right (4)  puff left (3)  less tensed (3)
- tensed right (18)  tensed left (11)  more tensed (1)

- the presence of a particular character string in the English translation

Search English Translation (Limit the search results to the following...)

- specific SignStream® collection
- specific signer

For example, a search for the presence of negative marking would display the 532 utterances with that marking, the first two of which are shown in Figure 8. The utterance video—front, face close-up, or side view—can be played.











File Name-Utterance	Utterance Video	View	Rough Gloss
Cory_2013-6-27_sc109-U-1		 <a href="#">Gloss..</a> Utterance: 	negative: dh: #NO++ POSS-1p ROOMMATE NOT (1h)GRAB RECENT-PAST HAVE EAT +MORNING WIT dh shape:  ndh: ROOMMATE MORNING WIT ndh shape:  engl trans: No, my roommate wasn't arrested; I just had breakfast with her.
Jonathan_2012-11-27_sc98-U-1		 <a href="#">Gloss..</a> Utterance: 	negative: neg dh: IX-1p (2h)NOT-YET (P)PAY fs-RENT dh shape:  ndh: (2h)NOT-YET (P)PAY ndh shape:  engl trans: I haven't yet paid rent.

Figure 8. First two results of a search for utterances with negative marking

#### 4 Access to the *ASLLRP Sign Bank* via DAI 2: <http://dai.cs.rutgers.edu/dai/s/signbank>

The *ASLLRP Sign Bank* was initially constructed from the *ASLLVD* dataset, a collection of isolated, citation-form signs.

##### 4.1 *The American Sign Language Lexicon Video Dataset (ASLLVD)*

The *ASLLVD*—collected at BU through a collaborative effort of Stan Sclaroff, Vassilis Athitsos, Ashwin Thangali, and Carol Neidle, as well as many ASL consultants and BU students (<http://www.bu.edu/asllrp/people.html>)—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, and morphological and articulatory classifications of sign type. For compound signs, the dataset also includes annotations for each morpheme. Further information about this data set is available in [7], and from <http://www.bu.edu/asllrp/av/dai-asllvd.html>. Those wishing to download the *ASLLVD* data should go to: <http://www.bu.edu/asllrp/av/dai-asllvd.html#download>.

##### 4.2 *Expansion of the ASLLRP Sign Bank*

DAI 2 provides access to the *ASLLRP Sign Bank*, and the *Sign Bank* has now expanded beyond the original *ASLLVD* to incorporate sign examples extracted from our expanding *ASLLRP SignStream® 3 Corpus* of continuous signing (described in Section 3.1). The *ASLLRP Sign Bank* will also soon be augmented to provide access to two other large sets of linguistically annotated ASL data, including both citation-form signs and examples from continuous signing.

### 4.3 Accessing the ASLLRP Sign Bank via DAI 2

#### 4.3.1 Selecting signs in the *Sign Bank* from the main sign list, for viewing

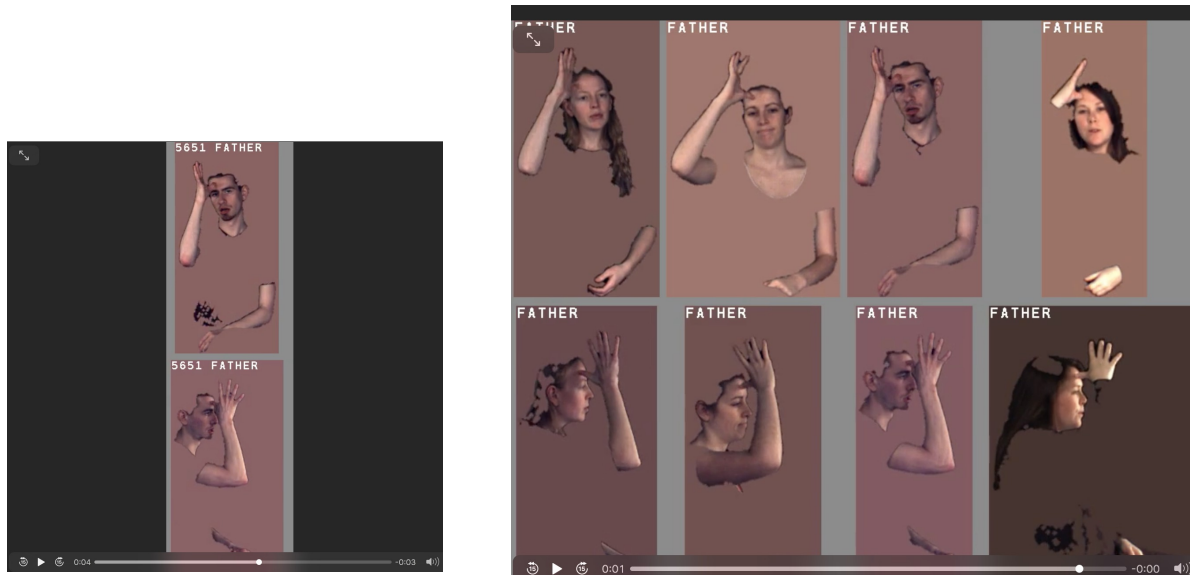
The *Sign Bank* has a list at the left, arranged alphabetically by the glosses for the primary entries. Multiple “entry/variants” may be grouped under a single primary entry in cases where there are lexical variants or closely related forms. (We consider signs that differ only with respect to final reduplications marked with +’s to be instances of the same “entry/variant” although they are distinguished by the +’s in their gloss labels.) So, for example, under FATHER, in Figure 9, we see the entry variant FATHER, as well as the entry/variant FATHERwg discussed above. It is also the case that the sign FATHER occurs as part of compounds. In such cases, \*FATHER indicates that there is a compound with something that occurs *before* FATHER (e.g., MOTHER+FATHER), and FATHER\* indicates that FATHER is occurring as part of a compound *followed by* another sign (as in FATHER+LAW). The number in parentheses after each entry/variant corresponds to the total number of examples of that sign in the *Sign Bank*.<sup>3</sup>

One canonical example of each entry/variant is displayed when the user clicks on that gloss. The right and left handshapes of the start of the sign are shown to the left of the thumbnails of the signer. In this case, since it is a one-handed sign, there is no handshape for the non-dominant hand. The handshapes of the two hands found at the end of the sign are shown, in the same order, to the right of the thumbnails. One can play that sign video by clicking on the button “Play Sign Video”), as shown in Figure 10. Clicking on “Play Composite Video” will simultaneously play all of the examples of citation-form signs available from the ASLLVD, as is also shown in Figure 10.

The screenshot displays the ASLLRP Sign Bank interface. On the left, a scrollable list titled "Current Signs in Sign Bank" shows various sign entries with their counts in parentheses. The entries include: FATHER (57), \*FATHER (MOTHER+FATHER) (27), \*FATHERwg (TURN+FATHERwg) (2), FATHER\* (FATHER+IN+LAW) (1), FATHER\* (FATHER+LAW) (2), FATHERwg (9), FATHERwg\* (FATHERwg+LAW) (1), FATHER+IN+LAW, FATHER+IN+LAW (1), FATHER+LAW, FATHER+LAW (2), FATHERwg+LAW, FATHERwg+LAW (1), FAULT, FAULT (4), FAVORITE/PREFER, FAVORITE/PREFER (23), FAX, FAX (2), (Vulcan)FAX (1), and FEATHER. Below this list is a "Search for Sign" section with a "Key Word:" input field, radio buttons for "Exact Word", "1-Handed", "2-Handed", and "Either" (selected), and checkboxes for "DH-Start", "ND-Start", "DH-End", and "ND-End". There are four magnifying glass icons below these checkboxes, and a "Lookup" button at the bottom. On the right, the "Sign Search Results" section shows the selected entry "FATHER (57)" with buttons for "View Occurrences", "Play Sign Video", and "Play Composite Video". Below these buttons are two sets of handshape thumbnails (left and right hands) and two video thumbnails showing a signer performing the sign. At the bottom of the interface, there are navigation links: "DAI | Sign Bank | Help | Reset Password | Log in".

Figure 9. Alphabetical listing of *primary entries*, along with the corresponding *entry/variants*

<sup>3</sup> For statistics concerning the current total number of entries and signed examples in the (expanding) *Sign Bank*, see <http://dai.cs.rutgers.edu/dai/s/runningstats>.



**Figure 10. Playing the sign video (left), or the composite video (right)**

Clicking the “View Occurrences” button in Figure 9 will display examples in two parts of the window. On top, as seen in Figure 11, are the examples of citation-form signs from the **ASLLVD**, and below are additional examples from the continuous signing **ASLLRP SignStream® 3 Corpus**. For the **ASLLVD** citation forms, you can play the sign video or the composite video, as shown in Figure 10. The composite video shows all of the signers producing the sign in parallel, at the same time.<sup>4</sup> Note that these video images, as shown in Figure 10, have undergone some post- processing.<sup>5</sup> There is also an option to “View Original Video”.<sup>6</sup>

Figure 11 only shows the first couple of examples from the continuous signing data (**ASLLRP SignStream® 3 Corpus**); there are 54 examples in all external to the **ASLLVD**. Note that for the continuous signing data, it is possible to view just that sign clip, or the entire utterance that contains the sign, as is also shown in Figure 11.

<sup>4</sup> In a few cases, the signs in the composite video don’t exactly correspond to the occurrences listed. Some corrections to the classification of occurrences of signs occurred after those composite videos were produced.

<sup>5</sup> RAW frame captures → Bayer interpolation (un-calibrated color profile) → Skin color region detection and segmentation using skin color FG/BG model trained on signers in the data set → Background fill-in with average skin color → Crop the video frame to skin color segments → Color and contrast enhancement heuristic → Convert image sequence to .mov format video with gloss and frame number overlay.

This was done was to facilitate presentation of video collated from several signers together for hand shape annotation purposes, to create videos that work within available screen real-estate and to provide consistent color, handshape contrast, and brightness across different signers.

<sup>6</sup> In some cases, the display of camera 2 (the side view) is not yet available for the original videos as of May 2020; those missing views will be added soon. Where both views are available, the two will start to play, in synchronized fashion, when the window is opened. After that, the controllers for the two views operate independently.

Occurrences for Sign Variant: FATHER							
Citation-form examples							
	DH-Start	ND-Start		DH-End	ND-End	ASLLVD ID	
<input type="radio"/> FATHER+						3615 <a href="#">Play Sign Video</a> <a href="#">Play Composite Video</a> <a href="#">Play Original Video</a>	
<input type="radio"/> FATHER+						3618 <a href="#">Play Sign Video</a> <a href="#">Play Composite Video</a> <a href="#">Play Original Video</a>	
<input checked="" type="radio"/> FATHER+						3617 <a href="#">Play Sign Video</a> <a href="#">Play Composite Video</a> <a href="#">Play Original Video</a>	
<input type="radio"/> FATHER+						3616 <a href="#">Play Sign Video</a> <a href="#">Play Composite Video</a> <a href="#">Play Original Video</a>	
Examples from continuous signing							
	DH-Start	ND-Start		DH-End	ND-End	Occurrence ID	Utterance
<input type="radio"/> FATHER						43093	Cory_2013-6-27_sc112-U-70 <a href="#">Play Sign Video</a> <a href="#">Play Utterance Video</a>
<input type="radio"/> FATHER						46028	Jonathan_2012-11-27_sc93-U-5 <a href="#">Play Sign Video</a> <a href="#">Play Utterance Video</a>

Figure 11. Viewing the occurrences of FATHER

### 5.1.1 Navigation and searching within the Sign Bank

If you click anywhere in the sign list and then type a sequence of characters quickly, the selection will advance to the next sign that starts with those characters. So, if you are at the top of the list and start typing “FAT” the sign FAT will be selected and will become visible.

Alternatively, you can scroll to the entry you are interested in. Or you can use the search tools at the bottom left in Figure 9. If you click on the magnifying class corresponding to the hand shape you wish to select, a palette will appear, and you can click on the start and/or end handshape for the search. In Figure 12, we search for 1-handed signs that start with the 5 handshape.

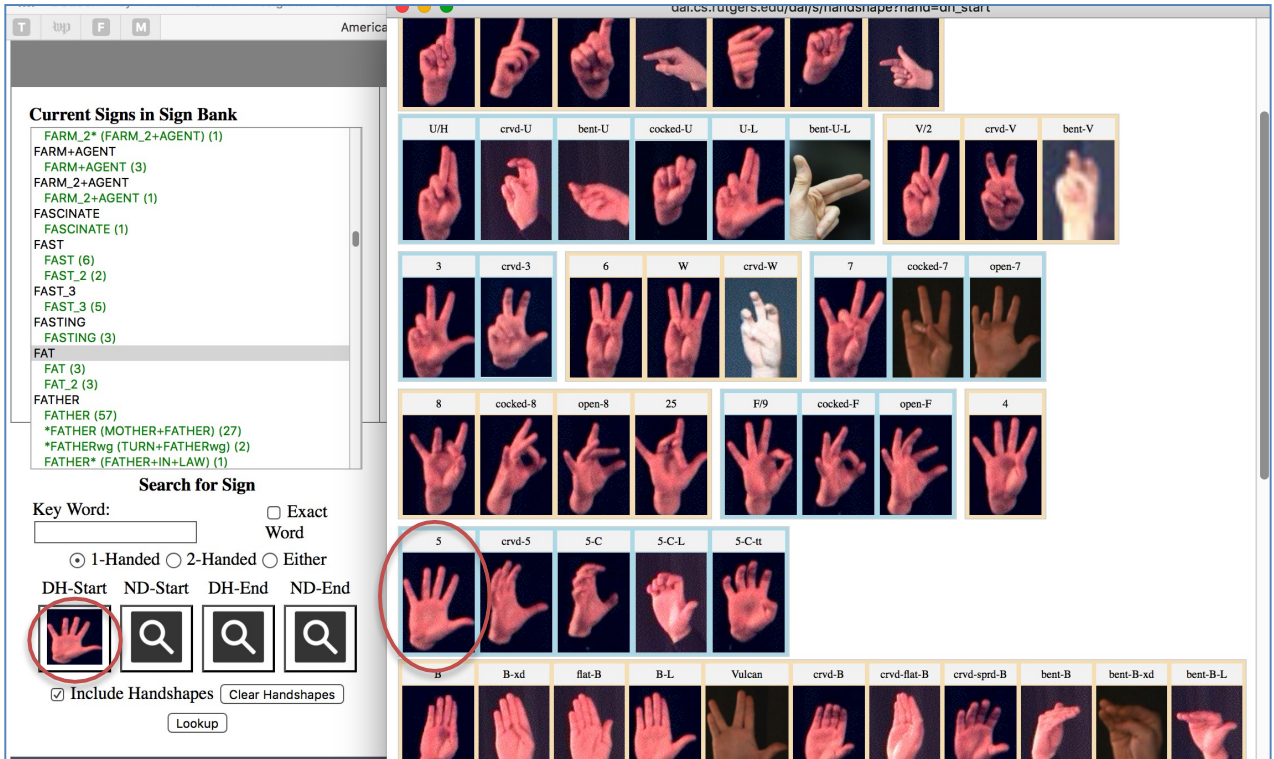


Figure 12. Searching for 1-handed signs with dominant start handshape of 5  
 There are a lot of signs meeting this description, the first few of which are displayed in Figure 13.

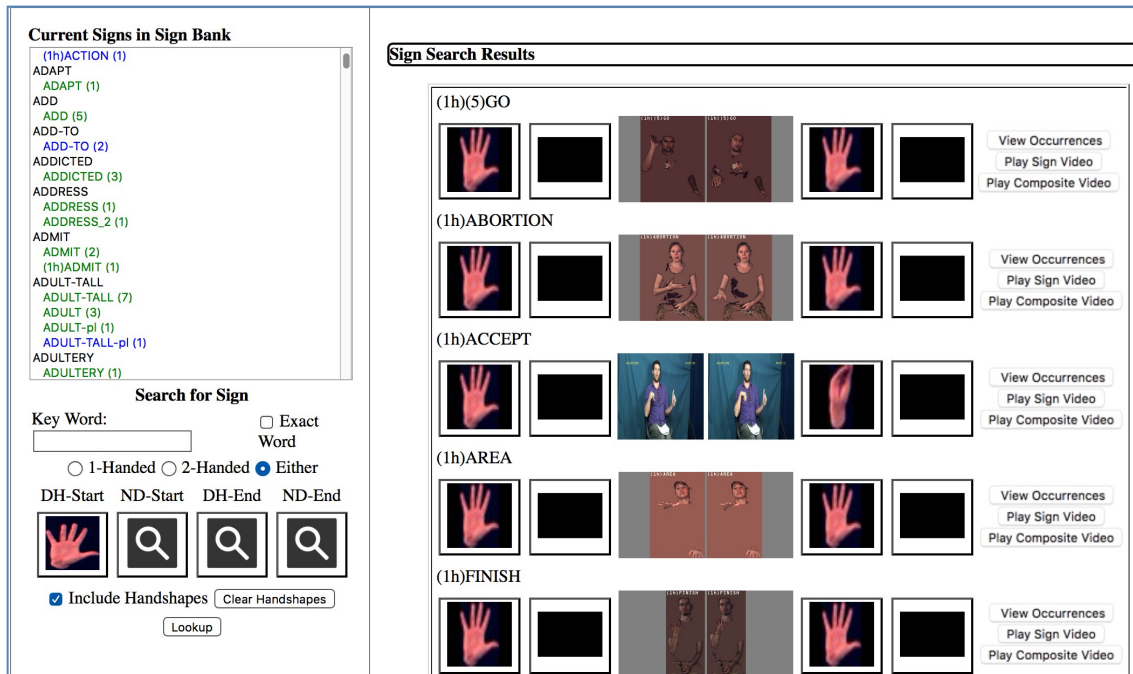


Figure 13. First results for search for 1-handed signs with dominant start handshape 5

Note that the “(1h)” prefix indicates that the examples in question are normally signed with 2 hands, but are exceptionally signed as 1-handed in these cases.

For additional information about **DAI 2** and the datasets to which it provides access, see [6].

## 5 Information about the Data Sets: Statistics

Statistics about the contents of the various datasets are provided through **DAI 2** at <http://dai.cs.rutgers.edu/dai/s/runningstats>.

## 6 SignStream® – the Software used for the Linguistic Annotations

All of our video corpora have been linguistically annotated using SignStream®, an application for linguistic annotation of visual language data developed by our group and freely shared. It provides an intuitive interface for labeling and time-aligning manual and non-manual aspects of the signing while viewing up to 4 synchronized videos showing different views. The original Mac Classic version of the application (through version 2.2.2) was a collaborative effort, with David Greenfield at Dartmouth University as the principal developer, working with Otmar Føelsche.

SignStream® version 3 is a Java reimplementaion, with many additional features; see <http://www.bu.edu/asllrp/SignStream/3/>. The principal programmer is Gregory Dimitriadis, working with Douglas Motto at Rutgers University. A screen shot is shown in Figure 14. The annotations from SignStream® are uploaded to the **DAI** or **DAI 2** to enable shared access, but are also available in an XML export format (see XML Export Format)

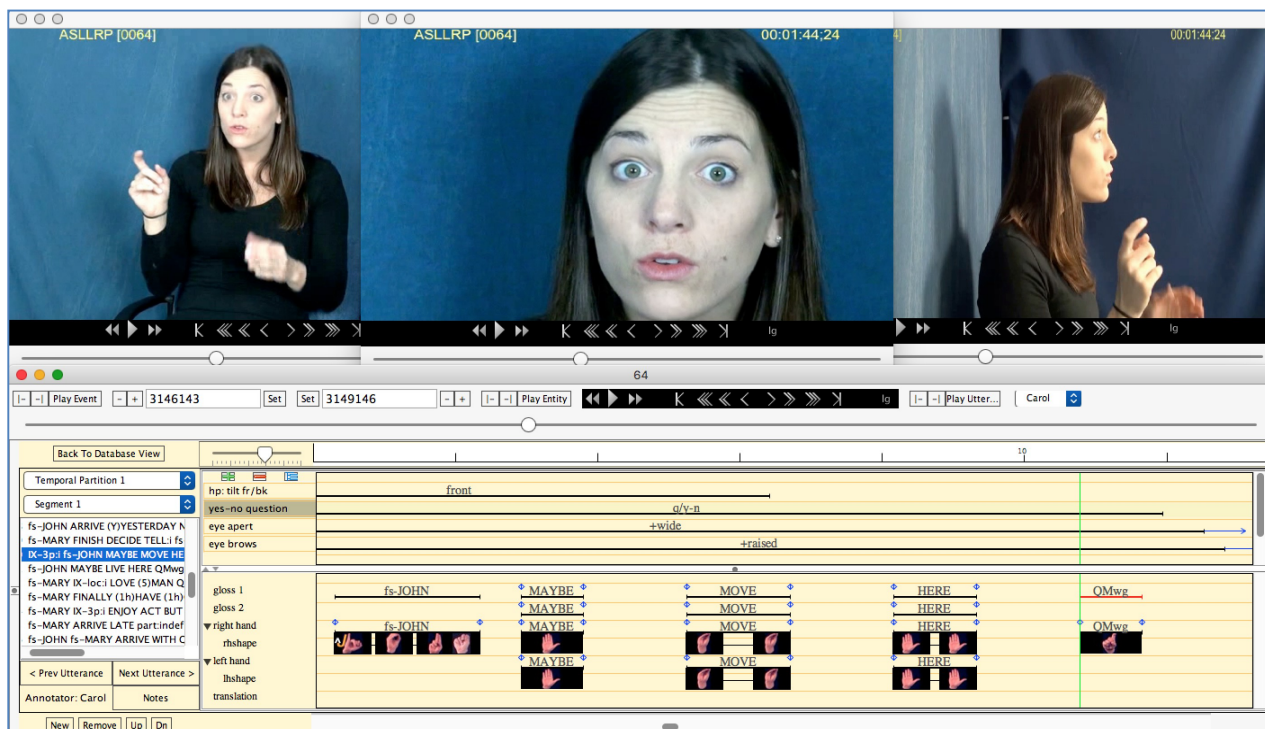


Figure 14. Screen shot of SignStream® Version 3

See [3,4,5] for SignStream® 3 User’s Guides.

## 7 Annotation Conventions

The conventions used for annotations in all of our corpora are documented in [1, 2]. A list of the field and value names and labels for non-manual events is also contained in the Appendix of [5], as is the structure of the XML export format.

## 8 Downloading Data from the ASLLRP SignStream® 3 Corpus via DAI 2

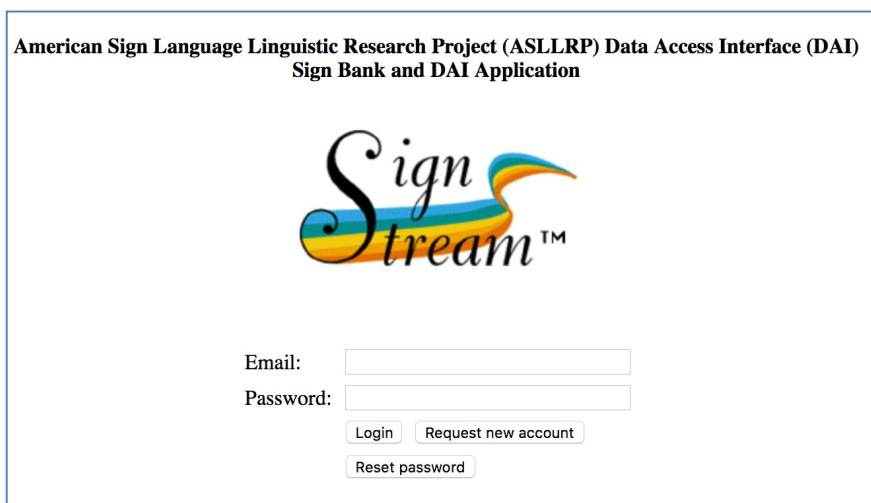
### 8.1 Account – for access to download

If you are interested in downloading data from **DAI 2**, you should request a (free) account by clicking on the “login” link. See Figure 15.

The purpose of the account is to help keep track of prior downloads and downloads in progress. Once your account is set up and you have logged in, you will be able to place items in the Download Cart, and then, from the Download Page, select the items and the formats of those items that you wish to download.

No account is needed to browse and search all of the data.

You can also submit a request to reset your password from that same page, as shown in Figure 15. You will receive an email message with a temporary password; you should log in promptly and change your password.



American Sign Language Linguistic Research Project (ASLLRP) Data Access Interface (DAI)  
Sign Bank and DAI Application

**SignStream™**

Email:

Password:

Figure 15. Request an account or reset password



## 8.2 Download cart

### 8.2.1 Selecting and downloading annotation and video files

There are two ways to download files, once you have logged in.

1) You can go directly to the Download Cart, where you will be offered a selection of items from which to select those that you wish to download, as shown in Figure 16. You can select the entire set of available collections, or any combination of individual collections. You can choose to download the SignStream® file, the annotations in XML export format (see Section 8.3), or both. Once you have selected the format of the annotations that you wish to download, you can click on the “Collection” button to download them. You can also download the corresponding video file by clicking on the “Video” button. You can speed up the selection by selecting all items in any row (by clicking the box in column 1) or column (by choosing “Select All” from the top).

Column 2 displays the name of the SignStream® collection file. There are occasionally small corrections made to the annotations, in which case the file version number, shown in Column 3, is incremented. If you cite a collection, you should include the version number.

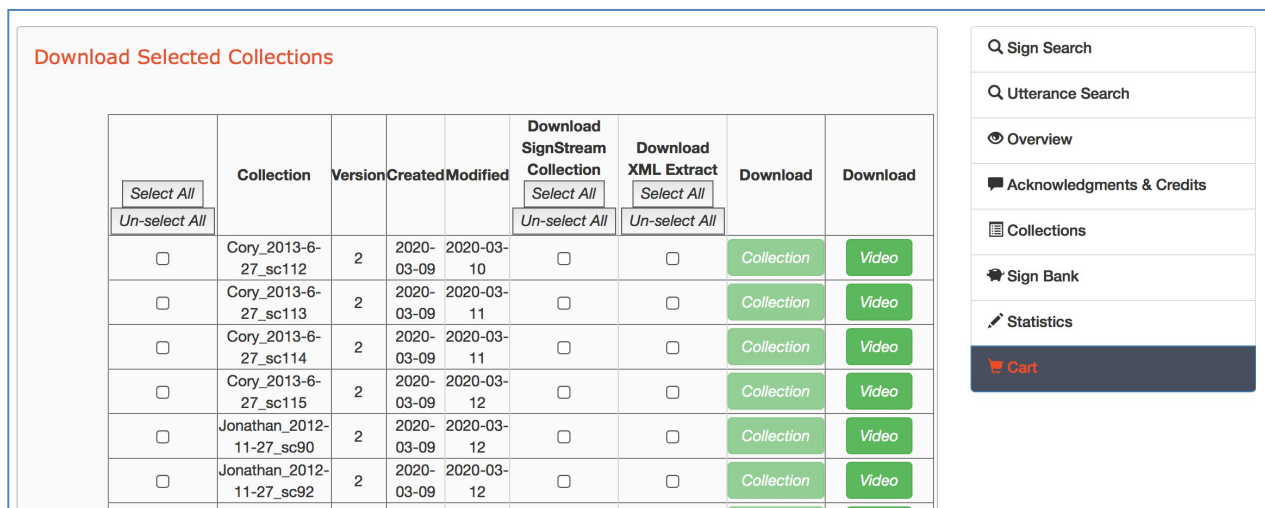


Figure 16. Download cart - download options

2) From the Search Results window, you can select specific collections to be added to the Download Cart, as shown in Figure 17. You can then click on “go to cart” to finalize your selections, as described above, and download the flagged collections.

### 8.2.2 Download History

The history of your downloads is displayed at the bottom of the page, as shown in Figure 18. You can re-download anything that you might wish to re-download. In case there is a more recent version of the annotations than the one that you last downloaded, that will be displayed there, enabling you to update to the most recent version of the annotations, if you wish to do so.

**American Sign Language Linguistic Research Project (ASLLRP) Data Access Interface (DAI) 2**

File Name-Utterance	Utterance Video	Sign Video	View	Rough Gloss
<input checked="" type="checkbox"/> Rachel_2011-12-08_sc45-U-22			Gloss.. Sign: Utterance:	negative: wh question: dh: (25)WHY FATHER (P)UPSET (h)part:indef KNOW dh shape: encl trans: Why is dad upset? I have no idea...
<input checked="" type="checkbox"/> Rachel_2012-02-14_sc63-U-32			Gloss.. Sign: Utterance:	topic/focus: dh: POSS-3p:1 FATHER GO-OUT 5"perplexed" dh shape: ndh: ndh shape: encl trans: His/Her father left.
<input checked="" type="checkbox"/> Rachel_2012-02-14_sc63-U-34			Gloss.. Sign: Utterance:	topic/focus: dh: POSS-1p FATHER LONG-AGO+ ADULT-TALL NOW dh shape: ndh: ndh shape: encl trans: My father used to be tall but now he's short

**Figure 17. Saving results to Download Cart**

**Download History** (You can re-download or update to the most recent version)

Collection	Download Date	Previously Downloaded Version	Latest Version	SignStream Collection	XML Extract	(Re-)download Annotations	(Re-)download
Cory_2013-6-27_sc112	2020-05-10 12:09:58.0	Version: 2 <input checked="" type="radio"/>	Version: 2 <input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="button" value="Download"/>	<input type="button" value="Video"/>

**Figure 18. Download History – you can re-download or update to the most recent version**

### 8.3 XML Export Format

Below is an illustration of the structure of the XML export format.

```

</MANUALS>
<SIGN ID='17408'>
  <LABEL>'OPPORTUNITY'</LABEL>
  <SIGN_TYPE>'Lexical Signs'</SIGN_TYPE>
  <TWO_HANDED>'T'</TWO_HANDED>
  <PASSIVE_BASE_ARM>'F'</PASSIVE_BASE_ARM>
  <MARKED_HANDS>'F'</MARKED_HANDS>
  <START_END_HS>'DIFFERENT'</START_END_HS>
  <START_END_LEFT_HS>' '</START_END_LEFT_HS>
  <START_END_RIGHT_HS>' '</START_END_RIGHT_HS>
  <D_START_HS>'0'</D_START_HS>
  <ND_START_HS>'0'</ND_START_HS>
  <D_END_HS>'P/K'</D_END_HS>
  <ND_END_HS>'P/K'</ND_END_HS>
  <TWOHANDED_HANDSHAPES>'SAME'</TWOHANDED_HANDSHAPES>
  <DOMINANT_HAND START_FRAME='6579573' END_FRAME='6593587' />
  <MEDIA-FILES>
</MEDIA-FILES>
</SIGN>
</MANUALS>

```

```
<NON_MANUALS>
  <NON_MANUAL ID='28963' START_FRAME='6504498' END_FRAME='6533527'>
    <LABEL>'head pos: tilt fr/bk'</LABEL>
    <VALUE>'slightly back'</VALUE>
  </NON_MANUAL>
</NON_MANUALS>
```

Fields and value labels are included in the SignStream® field specifications.

## 9 About the Website

The **DAI 2** is a secure site, and session activity is monitored for abuse. Technical support for the site is provided by Augustine Opoku (augustine.opoku@gmail.com) and Carol Neidle (carol@bu.edu) and. The site is hosted by LCSR at Rutgers University.

## 10 References

- [1] Neidle, Carol [2005] SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research. *ASLLRP Project Report No. 11*, Boston University. <http://www.bu.edu/asllrp/asllrpr11.pdf>
- [2] Neidle, Carol [2007] SignStream™ Annotation: Addendum to Conventions used for the American Sign Language Linguistic Research Project. *ASLLRP Project Report No. 13*, Boston University. <http://www.bu.edu/asllrp/asllrpr13.pdf>
- [3] Neidle, Carol [2017] A User's Guide to SignStream® 3. Boston, MA: *ASLLRP Project Report No. 16*, Boston University. [http://www.bu.edu/asllrp/SignStream/3/SS\\_User-guide.pdf](http://www.bu.edu/asllrp/SignStream/3/SS_User-guide.pdf)
- [4] Neidle, Carol [2018] What's New in SignStream® 3.1.0. Boston, MA: *ASLLRP Project Report No. 17*, Boston University. [http://www.bu.edu/asllrp/SignStream/3/SS\\_update.pdf](http://www.bu.edu/asllrp/SignStream/3/SS_update.pdf)
- [5] Neidle, Carol [2020] What's New in SignStream® 3.3.0. Boston, MA: American Sign Language Linguistic Research Project Report No. 18, Boston University. [http://www.bu.edu/asllrp/SignStream/3/SS\\_update-3\\_3.pdf](http://www.bu.edu/asllrp/SignStream/3/SS_update-3_3.pdf)
- [6] Neidle, Carol, Opoku, Augustine, Dimitriadis, Gregory, and Metaxas, Dimitris [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. LREC 2018, Miyagawa, Japan. May 2018*. <https://open.bu.edu/handle/2144/30047>
- [7] Neidle, Carol, Thangali, Ashwin, and Sclaroff, Stan [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. LREC 2012, Istanbul, Turkey. May 2012*. <http://www.bu.edu/linguistics/UG/LREC2012/LREC-asllvd-final.pdf>
- [8] Neidle, Carol and Vogler, Christian [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*. <http://www.bu.edu/linguistics/UG/LREC2012/LREC-dai-final.pdf>

## **Table of Figures**

Figure 1. Search interface.....	2
Figure 2. <b>DAI</b> search through <b>NCSLGR Corpus</b> for the string “FATHER”.....	3
Figure 3. <b>DAI</b> search through <b>NCSLGR Corpus</b> for the string “FATHER”: First several results for “Ben” signing “FATHER”.....	3
Figure 4. Search for the partial string "FATHER" .....	5
Figure 5. First two examples of Rachel signing FATHER .....	6
Figure 6. Playing the close-up of the face for the selected video .....	6
Figure 7. More detailed view of non-manuals (with approximate time-alignments) .....	7
Figure 8. First two results of a search for utterances with negative marking.....	9
Figure 9. Alphabetical listing of <i>primary entries</i> , along with the corresponding <i>entry/variants</i> . .....	10
Figure 10. Playing the sign video (left), or the composite video (right).....	11
Figure 11. Viewing the occurrences of FATHER .....	12
Figure 12. Searching for 1-handed signs with dominant start handshape of 5.....	13
Figure 13. First results for search for 1-handed signs with dominant start handshape 5.....	13
Figure 14. Screen shot of SignStream® Version 3.....	14
Figure 15. Request an account or reset password.....	15
Figure 16. Download cart -download options .....	16
Figure 17. Saving results to Download Cart .....	17
Figure 18. Download History – you can re-download or update to the most recent version .....	17

## **Table of Contents**

1	Background Information about the <b>DAI</b> .....	1
2	The original <b>DAI</b> : <a href="http://secrets.rutgers.edu/dai/queryPages/">http://secrets.rutgers.edu/dai/queryPages/</a> .....	1
2.1	<b>National Center for Sign Language and Gesture Resources (NCSLGR) Corpus</b> .....	1
2.2	Access to the <b>NCSLGR Corpus</b> via the original <b>DAI</b> .....	2
3	<b>DAI 2</b> : <a href="http://dai.cs.rutgers.edu/dai/s/dai">http://dai.cs.rutgers.edu/dai/s/dai</a> .....	3
3.1	The <b>American Sign Language Linguistic Research Project (ASLLRP) SignStream® 3 Corpus</b> .....	3
3.2	Access to the <b>ASLLRP SignStream® 3 Corpus</b> via <b>DAI 2</b> .....	4
3.2.1	Sign-level searches .....	4
3.2.2	Utterance-level searches .....	7
4	Access to the <b>ASLLRP Sign Bank</b> via <b>DAI 2</b> : <a href="http://dai.cs.rutgers.edu/dai/s/signbank">http://dai.cs.rutgers.edu/dai/s/signbank</a> .....	9
4.1	<b>The American Sign Language Lexicon Video Dataset (ASLLVD)</b> .....	9
4.2	Expansion of the <b>ASLLRP Sign Bank</b> .....	9
4.3	Accessing the <b>ASLLRP Sign Bank</b> via <b>DAI 2</b> .....	10
4.3.1	Selecting signs in the <b>Sign Bank</b> from the main sign list, for viewing .....	10
4.3.2	Navigation and searching within the <b>Sign Bank</b> .....	12
5	Information about the Data Sets: Statistics .....	14
6	SignStream® – the Software used for the Linguistic Annotations .....	14
7	Annotation Conventions .....	15
8	Downloading Data from the <b>ASLLRP SignStream® 3 Corpus</b> via <b>DAI 2</b> .....	15
8.1	Account – for access to download .....	15
8.2	Download cart .....	16
8.2.1	Selecting and downloading annotation and video files .....	16
8.2.2	Download History .....	16
8.3	XML Export Format .....	17
9	About the Website .....	18
10	References .....	19
	Table of Figures .....	20