

Data Sets available through DAI 2

Addendum to C. Neidle & A. Opoku, *User's Guide to the American Sign Language Linguistic Research Project (ASLLRP) Data Access Interface (DAI) 2 — Version 2. ASLLRP Report No. 18*, Boston University, Boston, MA. <http://www.bu.edu/asllrp/rpt18/asllrp18.pdf>

The DAI 2 website consists of two sections, one providing access to several **ASLLRP continuous signing corpora**, and the other providing access to the **ASLLRP Sign Bank**, as follows:

[1] **Continuous Signing Corpora:**

- (A) The **ASLLRP SignStream® 3 Corpus** - <https://dai.cs.rutgers.edu/dai/s/dai>: an expanding collection of data annotated with SignStream® 3 with a complete list of files listed at the bottom of <https://dai.cs.rutgers.edu/dai/s/dai>; includes:
- Thirty-six SignStream® collections from videos recorded between 2011 and 2013 (as of 3/27/21; this will be expanding). **Includes some new additions** as of March 2021.
 - Annotations of a video article by Ben Bahan: Bahan, B. (2009/2010). Sensory Orientation. *Deaf Studies Digital Journal, Issue 1*, Fall. (Files: 1-Ben-Introduction, 2-Ben-Voice-Identity, 3-Ben-Voice-Life, 4-Ben-Deaf-Speech, 5-Ben-Deaf-Grunts, 6-Ben-Control-of-Sound*, 7-Ben-Social-Msg*, 8-Ben-Control-of-sound-places, 9-Ben-Sensory-Politics*, and 10-Ben-Conclusion*). **New** as of March 2021.
 - Annotations of sentences shared by DawnSignPress (<https://www.dawnsign.com>), which has granted permission for users to view these data on this site, but **not** to download or use or redistribute these data. **New** as of March 2021.
- (B) The **NCSLGR SignStream® 2 Corpus** – <https://dai.cs.rutgers.edu/dai/s/daioriginal> an older corpus released in 2007 (previously available on CD-ROM, <http://www.bu.edu/asllrp/cslgr/pages/> and more recently available from a different website - <http://secrets.rutgers.edu/dai/queryPages/>). See the next page for differences in the annotation conventions for the older data.

[2] **The ASLLRP Sign Bank:** <https://dai.cs.rutgers.edu/dai/s/signbank>

Incorporates the data from (A) above, as well as additional data sets containing isolated, citation-form signs:

- The ASLLVD data set: see <http://www.bu.edu/asllrp/av/dai-asllvd.html>
- Annotations of isolated signs collected by researchers at Rochester Institute of Technology under the supervision of **Matt Huenerfauth** as part of NSF grant IIS-1763569. Thanks to Abraham Glasser, Ben Leyer, Saad Hassan, and Sarah Morgenthal for their roles in helping with data collection and annotation. **New** as of March 2021.
- Annotations of citation-form signs shared by DawnSignPress (<https://www.dawnsign.com>), which has granted permission for users to view these data on this site, but **not** to download or use or redistribute these data. **New** as of March 2021.

New! – note that as of February 2021, it is now possible to view subsets of the data by sign type (lexical signs, loan signs, fingerspelled signs, number signs, classifiers, gestures, compounds).

See also **acknowledgments and credits:**

<http://www.bu.edu/asllrp/dai-credits.html> and <http://www.bu.edu/asllrp/data-credits.html#credits>

Statistics for these data sets are available here:

<https://dai.cs.rutgers.edu/dai/s/runningstats> and <https://dai.cs.rutgers.edu/dai/s/runningstatsdai1>

Please pay careful attention to the **terms of use:** <http://www.bu.edu/asllrp/dai-terms.html>

* *coming soon* (as of March 27, 2021)

Differences in the way data are represented in these online data sets

In addition to our most recent data sets (our expanding **ASLLRP Corpus**) annotated with **SignStream® version 3** and displayed through our improved Data Access Interface – **DAI 2** (click [here](#) for more information), we are also providing access to older data sets (our **NCSLGR Corpus**) annotated with **SignStream® version 2** and displayed through an older Web interface (our **DAI**, updated to work around the expiration of Adobe’s Flash Player as of December 31, 2020).

Representation of gloss information on the two hands has changed significantly between the older and the newer data, because of enhancements to the SignStream® annotation tool. Previously, gloss labels were inserted solely on the upper tier of the 2-hand display, with annotation on the second tier only in case the non-dominant hand was doing something unexpected, or something independent of what was being produced by the dominant hand. So, for a regular two-handed sign, we would only have included a gloss label in the main gloss tier, with nothing annotated for the second gloss tier. In SignStream® version 3, however, labels appear in the non-dominant tier for all signs that involve use of the non-dominant hand; thus all 2-handed signs include gloss information (plus start and end handshapes, also newly added in version 3) on both of the gloss tiers.