JULIE A. HOCHGESANG AND EMILY SHAW

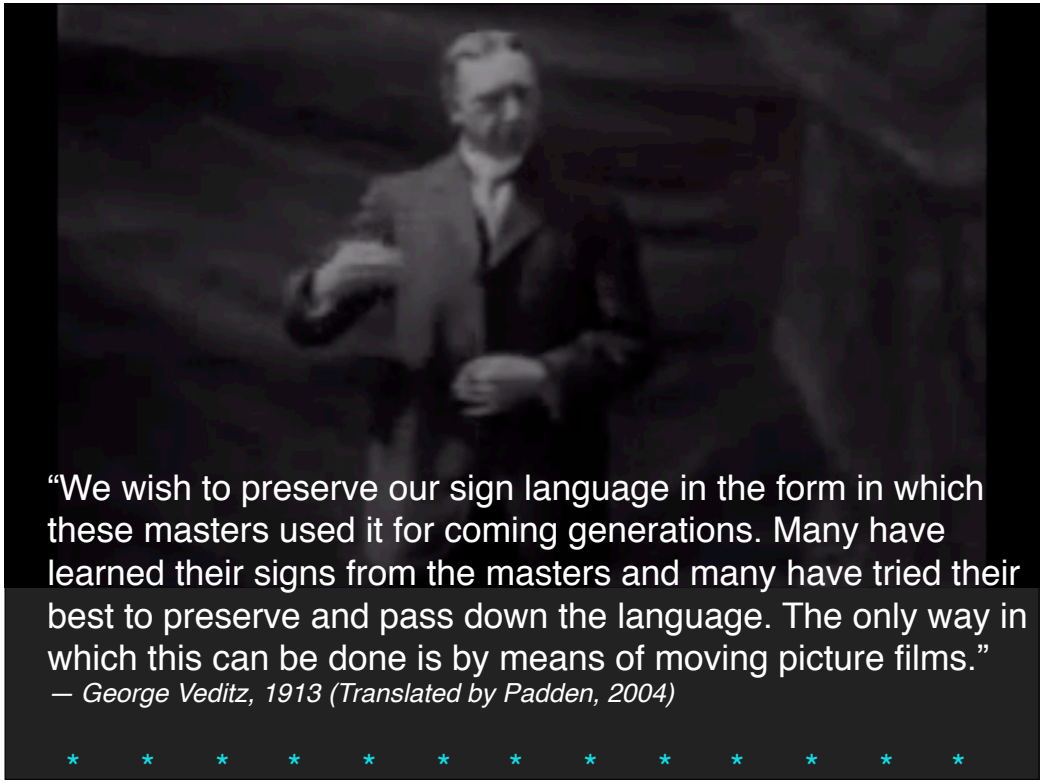
MAINTAINING THE STORIES OF THE DEAF COMMUNITIES AT GALLAUDET

SLIDES AND CONTACT INFO AVAILABLE HERE 👉
(LINK: HTTPS://SHWCA.SE/GUDA)

YOU ARE WELCOME TO LIVE-TWEET OR TAKE PICTURES AND SHARE ON SOCIAL MEDIA.

CONFERENCE HASHTAG #MIII

American Sign Language (ASL), the language of the American Deaf communities, is about two hundred years old. But no widespread written language to represent ASL has ever been developed. This means there is no direct textual representation of ASL - the stories that have been told by deaf people were (and still largely are) necessarily passed down face-to-face. But as soon as video cameras were invented, ASL stories have been filmed. We want to talk about this a bit today.

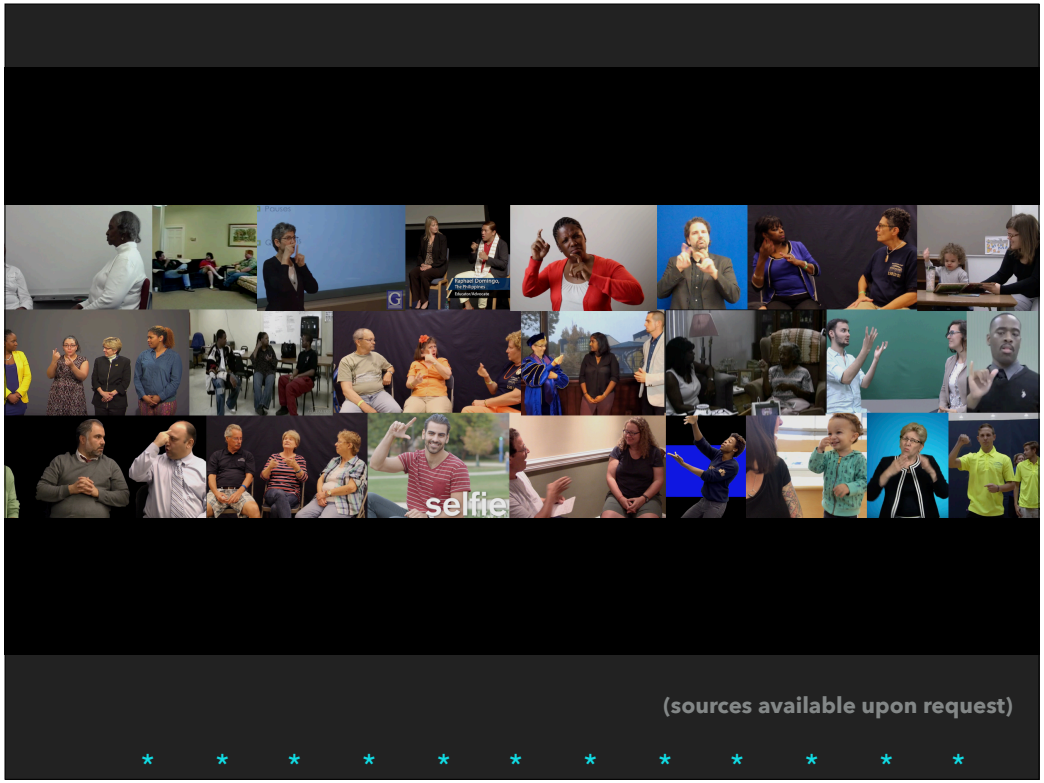


Deaf people all over the world sign to communicate. There's not one signed language but hundreds. Similarly, there's not one American Sign Language (ASL) but hundreds of variations. All of the varieties of ASL exist in the stories of American deaf people that have been passed down across generations. As long as cameras have been around, many of these stories have been documented. We want to preserve them. Many of the films are located here at Gallaudet, one of the oldest institutions serving deaf people in the world.

“We wish to preserve our sign language in the form in which these masters used it for coming generations. Many have learned their signs from the masters and many have tried their best to preserve and pass down the language. The only way in which this can be done is by means of moving picture films.”

— George Veditz, 1913 (Translated by Padden, 2004)

* * * * * * * * * * * * *



For a minute, I want to look at what we're talking about today. Just a minute of video with all kinds of ASL videos edited together.

(Watch video)

After having seen all of these, even if you don't know ASL yourself, it is clear how rich and nuanced language use is. You can see the different representations - different people, different social contexts, different genres. And if you do know ASL, you probably caught some signs and can remember them now - ROCKET, FUNNY, etc. Now imagine trying to go back to an individual video and searching for one specific sign contained within it. Currently, this cannot be easily done. Before being able to query search terms in videos of sign language, we first need to be able to represent the contents of the video into text. We can't study language in its raw form - as sound waves or light waves on the air. We need to freeze that behavior somehow. For signed languages, we are unable to rely on already conventionalized written languages like for English or the International Phonetic Alphabet (IPA) that linguists can use to represent any spoken language.

[redacted]

(sources available upon request)

* * * * * * * * * * * *

As you've just seen, Here at Gallaudet, since cameras have been invented, people have been filming various events and interactions on campus. Examples of such filming include presentations, student campus events, personal narratives, lectures, as well as specific tasks for certain research projects. All of these videos, which are stored on the Gallaudet Video library or Youtube channel, capture some kind of ASL use. Gallaudet was founded in 1864. The first filmed video of ASL was generated in 1910 (in Ohio at the convention of the National Association of the Deaf). It's now 2019 which means we have in our video collection over 100 years of films of ASL. There is so much potential here - especially given that these videos are mostly community-generated (as opposed to scripted or performed events). Some are elicited for research, too. It is worth noting that there are plenty of films in personal collections around North America but because Gallaudet holds a special place among many American deaf communities, I often get requests to help host these collections - home videos, other research projects, and so on.

We have a lot of stories we can use for various ends (both for research and practical purposes). So collecting them isn't so much the issue. We want to take care of the stories we already have.

GALLAUDET
UNIVERSITY
CONNECT. DISCOVER. INFLUENCE.
Gallaudet Video Library
YouTube
Search
Home
My channel
Trending
ASL VIDEOS AT GALLAUDET

Again, there are thousands of videos, starting in the 1900s that have exponentially increased through the years. They're scattered across campus, although some are hosted in the Gallaudet Video library. They're in different formats - VHS, CDs, different digital formats. They're not processed consistently - if at all - meaning they're not named in the same way, organized with the same metadata and so on.

ASL VIDEOS AT GALLAUDET

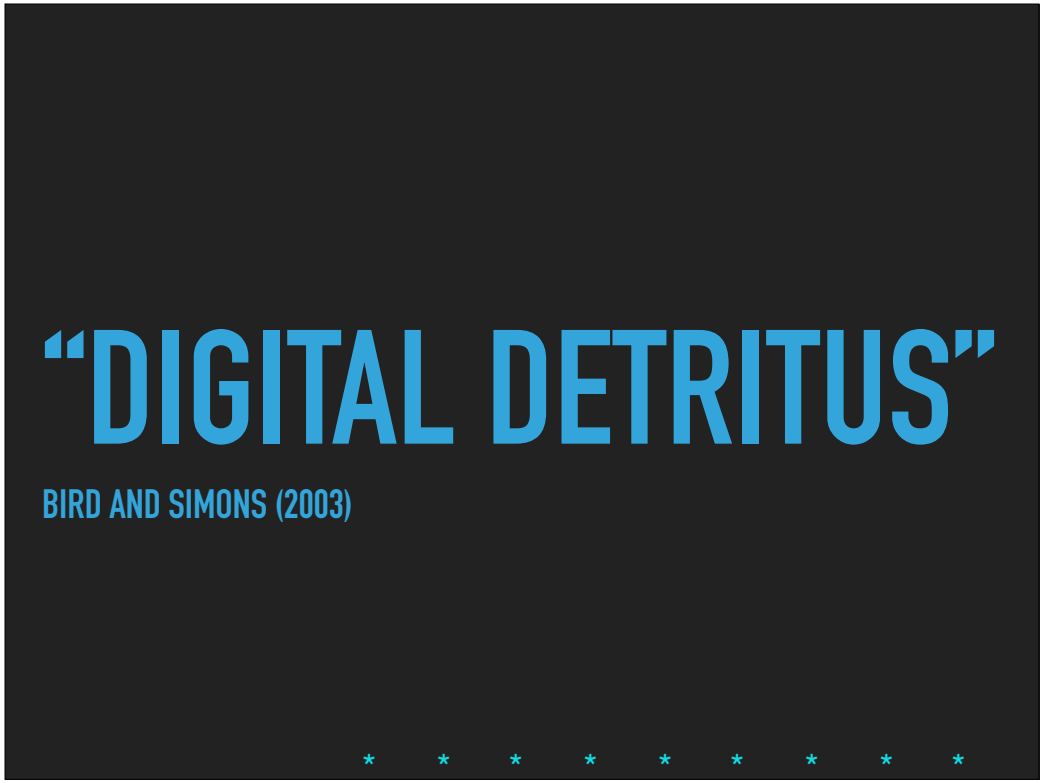
* thousands of videos
* some in Gallaudet Video library or YouTube but most scattered across campus
* different formats
* processed differently (if at all)

* * * * * * * * * *

ASL VIDEOS AT GALLAUDET

* thousands of videos
* some in Gallaudet Video library or YouTube but most scattered across campus
* different formats
* processed differently (if at all)

≠accessible



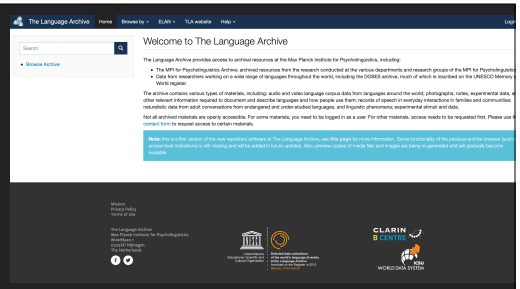
All of this means that they're inaccessible - basically 'digital detritus' as described by Bird and Simons in their 2003 article on portability of data for language documentation. They are inaccessible in the sense that they may (or may not) be public-facing but also they are not processed in precise and consistent ways that would make their contents available for the public, researchers, etc. to search.

Our current need - and this has been echoed by many - is to create a place to store, consistently organize, access (make machine-readable) and share our videos. And it's something that needs to be sustainable - maintained over time.

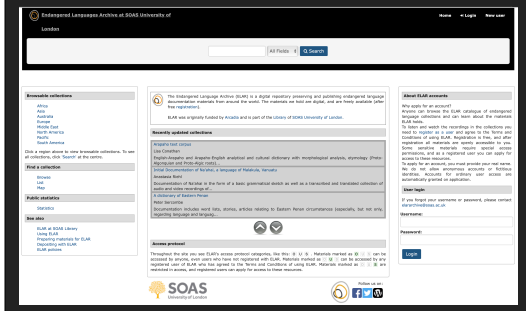
As far as storing and archiving videos at Gallaudet, we've done this management of our videos somewhat individually in departments and in the Gallaudet Video Library which uses the Telescope Digital Assets Management program. The Gallaudet Video Library is currently under the domain of University Communications. But it has no more financial support and needs to be updated or we need to move to other digital resources asset management.

The accessibility and sustainability issues will be addressed next.

http://www.language-archives.org



https://tla.mpi.nl

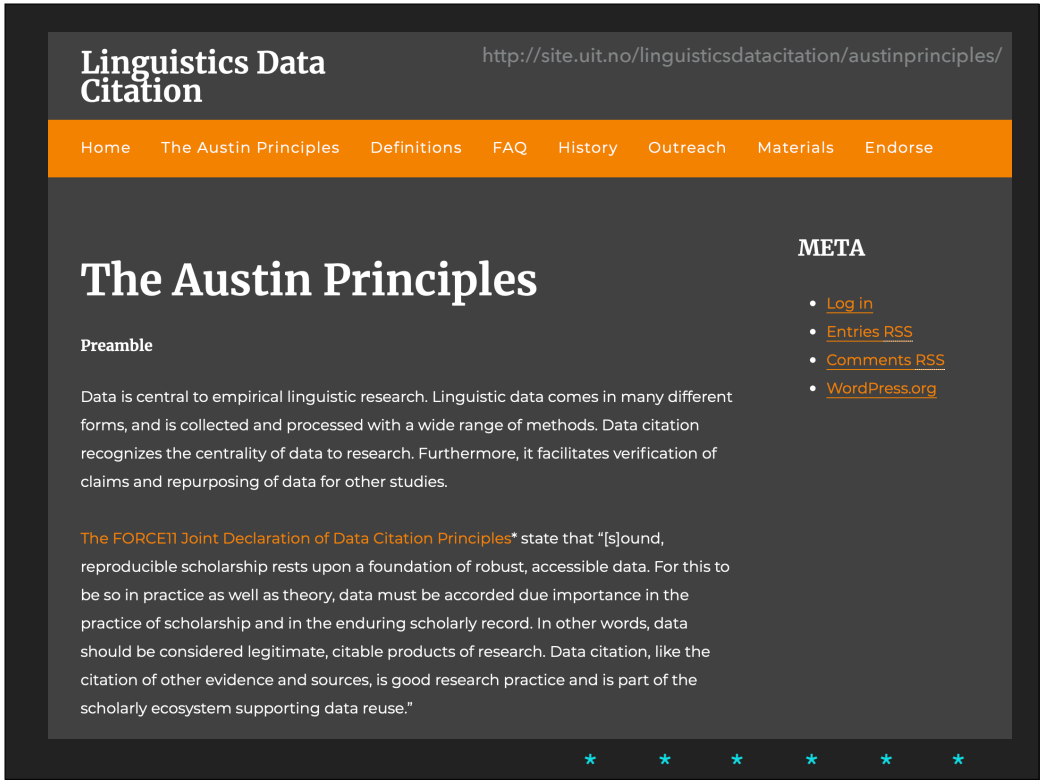


https://elar.soas.ac.uk

http://www.endangeredlanguages.com

* * * * * * *

Basically we need language archives, something like these, which would be a stable repository with consistent metadata as well as dedicated technicians who would update the materials as technology evolves. We would require regular updates to terms of use, graded permission levels if needed, and so on.



Linguistics Data Citation

http://site.uit.no/linguisticsdatacitation/austinprinciples/

Home | The Austin Principles | Definitions | FAQ | History | Outreach | Materials | Endorse

The Austin Principles

Preamble

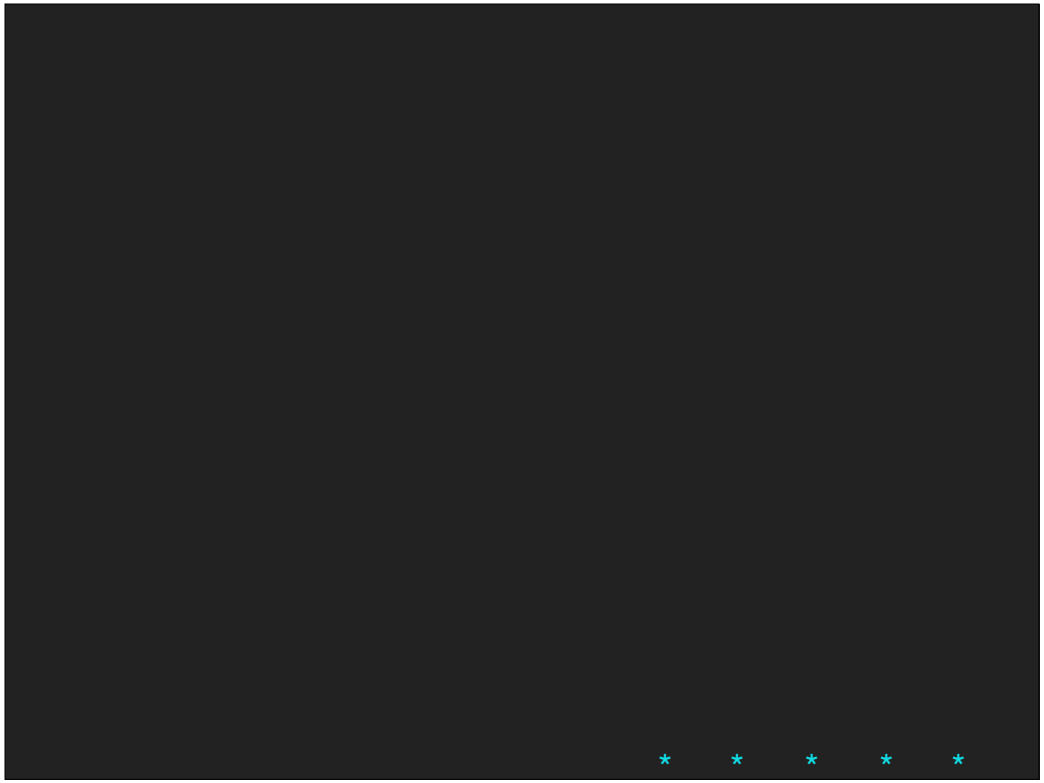
Data is central to empirical linguistic research. Linguistic data comes in many different forms, and is collected and processed with a wide range of methods. Data citation recognizes the centrality of data to research. Furthermore, it facilitates verification of claims and repurposing of data for other studies.

The FORCE11 Joint Declaration of Data Citation Principles* state that "[s]ound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse."

META

- Log in
- Entries RSS
- Comments RSS
- WordPress.org

The Austin Principles of Data Citation - is a general push by linguists to make primary data (spoken, written or signed) accessible. We call on these principles in the work that we do to help guide us as we develop best practices in making our ASL video collections more accessible to everyone. Personally for me, this is essential as a linguist and as a Deaf woman because it ensures access to the primary data - something that's very hard to do when looking at publications - usually text on paper - which creates challenges for representing signed languages which do not have standardized written forms.



Speaking of not having written systems for signed languages… while we don't have standardized writing systems, we have something called *glossing* which means we use the written system of the majority spoken language to represent each sign. For the U.S., that means using written English in upper case form to represent ASL. We don't have time to get into the process of glossing but I do want to mention that it is fundamentally problematic to use written English to represent ASL signs. This practice reinforces the secondary representation (English) as tied to the primary representation (ASL) which is an entirely different system. One way we have gotten around this in our field is to use something called ID-glossing (e.g., Johnston, 2010). It basically looks the same as glossing but when used with a database and video annotation (pictured here in this slide), we are able to use glosses as indexical representations. It helps to reduce the impact of the written system on our analysis of ASL. Still plenty of problems there (because the English text is still functioning as a label of some sort) but when used as intended, the person who is assigning ID-glosses would have the representation of the sign visibly accessible along with the written text (thereby minimizing the prominence of the written word in relation to the sign).

For example, take the ASL Signbank which I use to maintain ID glosses for ASL signs (there are other sign banks for other sign languages - we modeled ours after theirs). ID-glosses are "identifying glosses" (unique labels used to represent the same signs). ID-glosses are not the main form of representation we would want but they are a form of labeling data (in this case ASL signs) so we can find them in the videos that are annotated. This screen recording shows one of the videos we showed before but in ELAN, a particular software annotation program. The software has been linked to the ASL Signbank so when a person assigns glosses to the signs they see in the videos, the ASL Signbank record is made prominent next to the gloss.

ID GLOSSING

(JOHNSTON 2010)

* * * * *

(Hochgesang, Crasborn, and Lillo-Martin, 2018)

ASL signbank (Hochgesang, Crasborn and Lillo-Martin, 2018)

Home | About | Signs | Feedback | Search gloss | Search translation

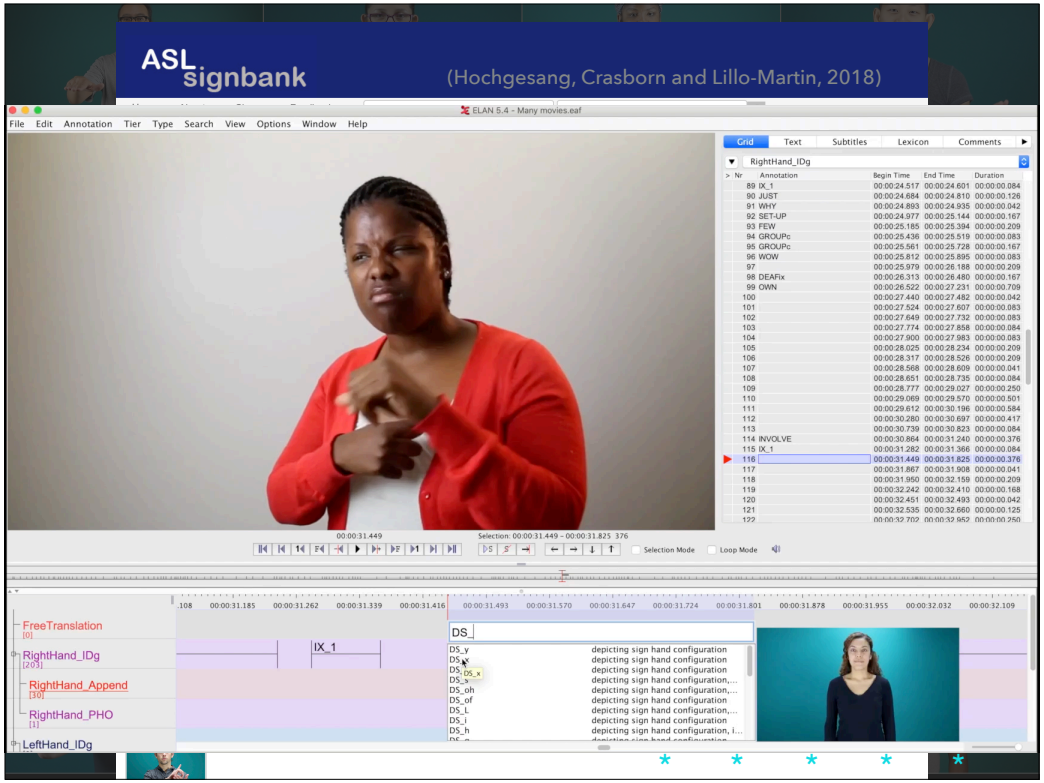
Form your query

Search sign | CSV | ECV | Reset

List View | Lemma View | Number of matches: 11 out of 3130.

Add New Sign

	Annotation ID Gloss (American English)	Lemma ID gloss (American English)	Translation equivalents	Handedness	Dominant Handshape	Nondominant Handshape	Location - major	Tags
	ARTICULATE-SIGN	ARTICULATE-SIGN	articulate-sign, chat, conversation, discuss, fluent, fluency, sign, socialize, to perform in sign, to pronounce, to sign something	SymmetricalOrAlternating	-	-	-	
	SIGN	SIGN	sign, sign language, signing	SymmetricalOrAlternating	-	N/A	-	
	SIGNAGE	SIGNAGE	billboard, board, board-sign, notice, road sign, sign, signage, square, traffic sign, SIGNboard	SymmetricalOrAlternating	-	N/A	-	
	SIGNATURE	SIGNATURE	autograph, contract, sign, sign a paper, signature, SIGNpaper	AsymmetricalDifferentHandshape	-	B	-	
	DESIGN	DESIGN	design, draw	AsymmetricalDifferentHandshape	-	B	-	


ASL signbank
(Hochgesang, Crasborn and Lillo-Martin, 2018)
ELAN 5.4 - Many movies.eaf
File Edit Annotation Tier Type Search View Options Window Help
Grid Text Subtitles Lexicon Comments
RightHand_IDg
FreeTranslation
RightHand_IDg
RightHand_Append
RightHand_PHO
LeftHand_IDg
IX_1
DS_
DS_y depicting sign hand configuration
DS_x depicting sign hand configuration
DS_s depicting sign hand configuration,...
DS_oh depicting sign hand configuration,...
DS_of depicting sign hand configuration
DS_L depicting sign hand configuration,...
DS_i depicting sign hand configuration
DS_h depicting sign hand configuration, i...

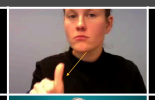
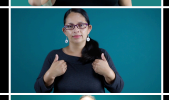
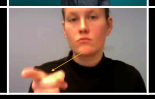
Frequency

examining how often words
occur in a corpus (usage-based)

* * * *

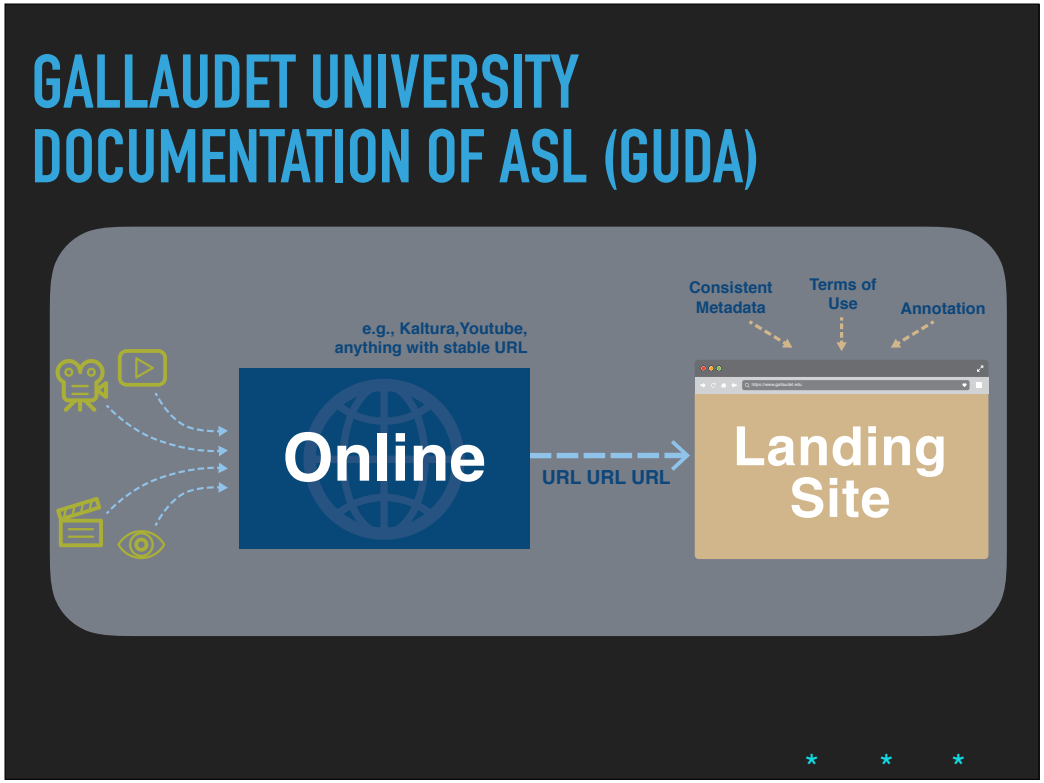
Once we annotate the videos in their entirety - meaning we capture the signs that occur on camera and assign them ID-glosses in the ELAN software annotation system - then that information becomes machine-readable. We can search the contents of the video just like we search for text on Google. Immediately, with annotated videos, we can do things like find out the most frequent words in ASL. This kind of information is incredibly useful to linguists, ASL teachers and students, and interpreters especially; until now, most of our work has relied on native signer intuitions about language use rather than representative samples of the language.

lexical frequency

THAT		1162	NOT		606
WHAT-PU		995	DEAF		603
LOOK		787	BUT		545
HAVE		754	KNOW		458
LANGUAGE		628	REAL		321

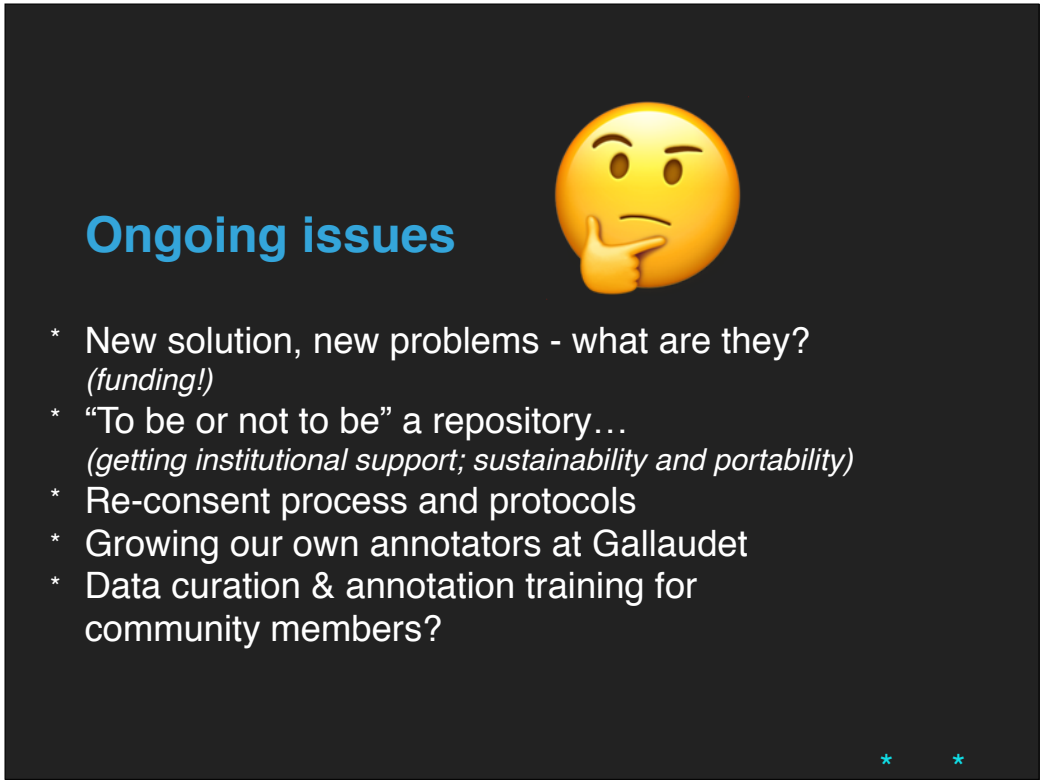
(Out of 85,674 items from 942 .eafs)

* * * *



Based on the need to store and organize videos digitally, armed with ideas of language archives and the guiding vision of the Austin Principles of Data citation and ID glosses stored in the ASL Signbank - we have set out to build the necessary infrastructure and protocols to make the vision a reality. We call this project the Gallaudet University Documentation of ASL (GUDA)
The graphic here in the slide shows our overall vision of the project.

Essentially, we are interested in finding all of those existing video collections that are scattered across the university (and also the U.S.) and making sure these collections have a stable URL. Once we have identified which collections can be connected to GUDA, our landing site will then point to these stable URLs. The data contained on the stable URLs will be have consistent metadata (done in similar format), terms of use, and annotation tied to them to make the videos searchable.



These ongoing issues (“whither a corpus for ASL”) come from our discussion about the ongoing “devaluing” of our project by critical stakeholders (funders, in particular, and administrators).

Ongoing issues - funding (we need some); sustainability (this landing site is not an archives - it’s not a long-term solution. It’s an initial step in something that will take our entire careers to do); portability (we need to consider how to make sure this is accessible across platforms and over time. e.g., Bird and Simons 2003; re-consent (seeking permission from those on video to continue to share their data). We also have the overwhelming task of annotating the videos. The skills required to do so are held by a very small population of signing people. We are currently working on projects that will “grow our own” annotators from the signing student body at Gallaudet. We hope to include these activities as a part of certain courses and even to expand training opportunities to members of the deaf community. The effort required to generate machine-readable videos of ASL is one of the biggest barriers we face right now in terms of making a corpus of ASL.

GUDA is then the maintenance of ASL stories if we can actually manage to combine the collection of ASL videos in a stable and accessible digital place with data archiving protocols. If this project could ever become a reality, it makes sense that it would happen here at Gallaudet University, the only university in the world for deaf and hard of hearing people.

Berez-Kroeker, A., Gawne, L., Kung, S., Kelly, B., Heston, T., Holton, G., Pulsifer, P., Beaver, D., Chelliah, S., Dubinsky, S., Meier, R., Thieberger, N., Rice, K., & Woodbury, A., (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. Linguistics, 56(1). 1-18.

Bird, Steven and Gary Simons. 2003. Seven dimensions of portability for language documentation and description. Language, 79(3). 557-582.

Crasborn, Onno and Han Sloetjes. 2008. Enhanced ELAN functionality for sign language corpora. In Construction and Exploitation of Sign Language Corpora. [Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation, LREC 2008, Marrakech.], edited by Onno Crasborn, Thomas Hanke, Eleni Efthimiou, Inge Zwitserlood and Ernst Thoutenhoofd. 39-43. ELRA, Paris. http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf

ELAN (Version 5.4) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

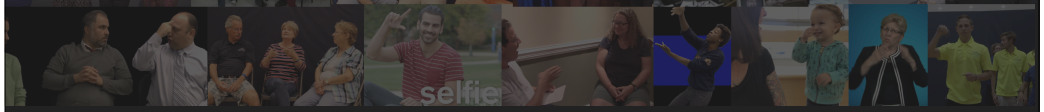
Hochgesang, J.A., O. Crasborn, & D. Lillo-Martin. (2018). ASL Signbank. New Haven, CT: Haskins Lab, Yale University. https://aslsignbank.haskins.yale.edu/

Padden, C. (2004). “Translating Veditz” Sign Language Studies, 4(3), 244-260.

Sources of video available upon request (Translation of Veditz from (Padden, 2004)

JULIE A. HOCHGESANG AND EMILY SHAW

MAINTAINING THE STORIES OF THE DEAF COMMUNITIES AT GALLAUDET



SLIDES AND CONTACT INFO AVAILABLE HERE 👉
(LINK: HTTPS://SHWCA.SE/GUDA)

CONFERENCE HASHTAG #MIII