

TYRANNY OF GLOSSING REVISITED: RECONSIDERING REPRESENTATIONAL PRACTICES OF SIGNED LANGUAGES VIA BEST PRACTICES OF DATA CITATION

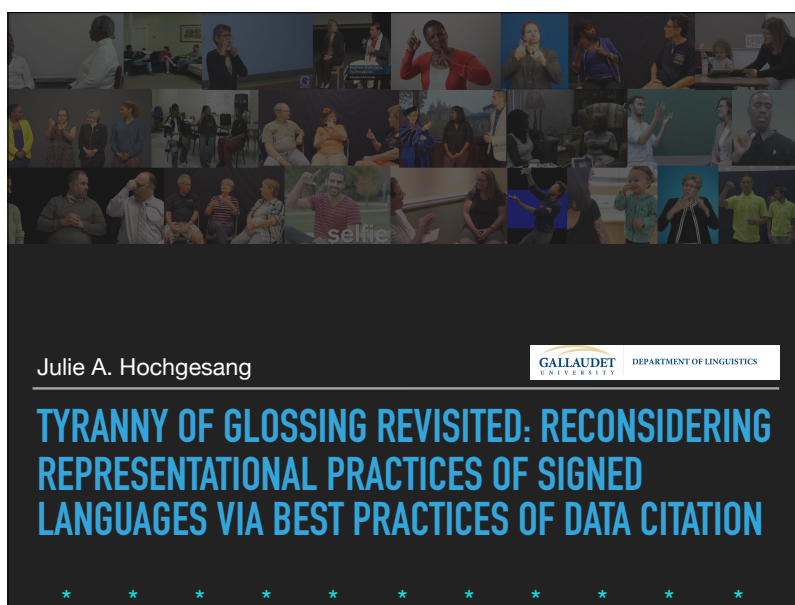
BY JULIE A. HOCHGESANG

SLIDES AND CONTACT INFO AVAILABLE HERE 
(LINK: [HTTPS://SHWCA.SE/GUDA](https://shwca.se/guda))

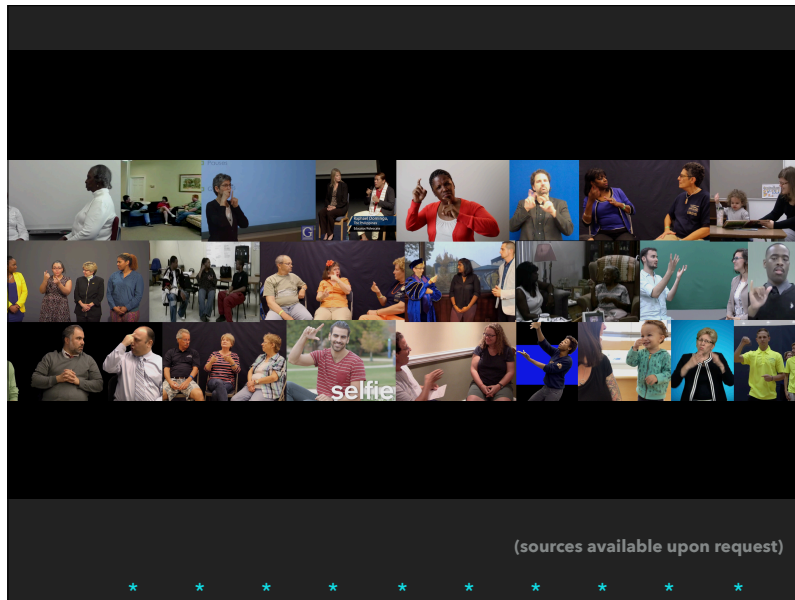


CONFERENCE HASHTAG #DRRS2019

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



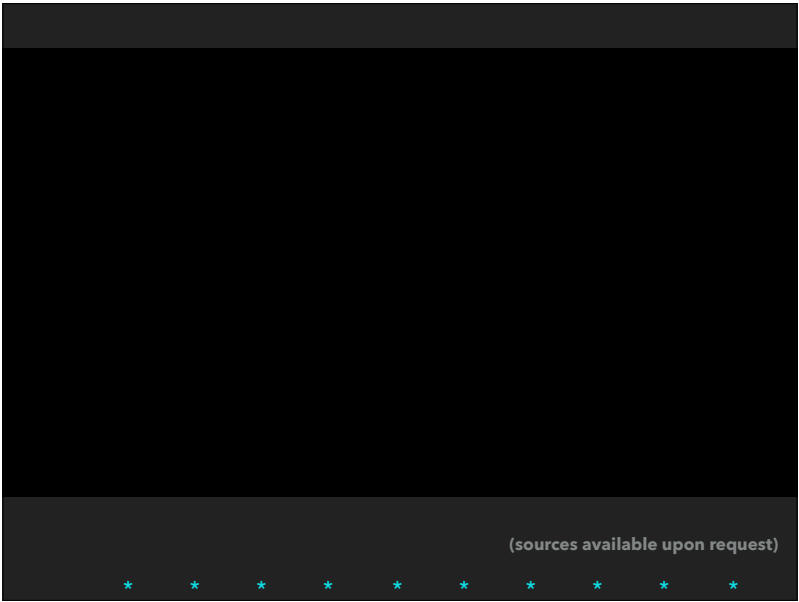
So all us here today work with language and we need to be able to freeze what we're looking at in some kind of textual representation so we can do things with them like count, categorize, sort and all kinds of things language scientists like to do. Most of us represent signed languages through glossing. In a paper published in 2008, Dan Slobin used the phrase "tyranny of glossing" and it stayed with me. I've referred to it often when considering issues of language documentation and corpora. So for today's presentation, I want to talk a bit about that phrase means and how we can consider it when thinking about our own practices for the work we're doing today especially from the perspective of data citation as outlined in the Austin Principles of Data Citation <https://site.uit.no/linguisticsdatacitation/austinprinciples/>.



But first for a minute, I want to look at what we actually study. Just a minute of video with all kinds of ASL videos edited together. (note some of these were taken from YouTube since I used them one of my classes). If you want a reference or link (if available) for any of these, please let me know. They are all used with permission or are available publicly.

(Watch video)

So, after having seen all of these, even if you don't know ASL yourself, you can see 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 and find what you wanted. You need to be able to search for it. We need to be able to represent this into text so we can use that for that purpose. We also use written representation for research. All of us in this room who study language know that we need to rely on some form of frozen representation so we can study it. 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.



NOTHING DS_5 PU THRILL SET-UP FS NOW IX
CANNOT IX WORD WORD SAME-AS BUT SIGN
DIFFERENT SKILL WOW EXPERIENCE ANDasym
CREATE SIGN POSS_1 MOTHERstr STRONG
SUPPORT ORAL DS_3 LOOK IX_1 DS_x ZERO DS_x
SAME-AS WORD SIGN QUOTE NOT WORD IX
WORD THAT APPEAR INSTAGRAM ~SELFIE THIRSTY
SWEET CONTACT DS_cx FS DS_cx DS_b DS_b IX IX
STORYf ABOUTix IX LOOK WORK DS_1 DIFFERENT
UNIVERSITY PU IX_1 DON'T-KNOW WHOflex PU
IX_1 DON'T-KNOW WHOflex FAMOUS BETWEEN
TWO PERSONp PU NOW IX COW DS_s LANGUAGEI
THAT PROCESS FROM MAKEstr FS XXX THAT IX_1
JUST WHY SET-UP FEW GROUPc GROUPc WOW

* * * * *

Now if we want to represent what we just saw in the video, we usually get this kind of thing. All of these represent what was produced on the right hand of every person in the video we just watched. (FYI this is just for the first 20 seconds. I couldn't make all of the words fit on this slide in readable format!) Anyway, right away we see that a lot has been stripped from the original. There's no direct references to the forms of the signs themselves. There's no representation of facial expressions or other phonological modifications of the signs. The original data has been stripped down and these words we see in the slide here are their representations.

This is glossing. Let's talk about that for a minute.

Glossing

(4) Gothic (II Cor. 5:9, Cole et al. (1980, 721))

inuþ-bis usdaudjam... waila Ø galeikan imma
because.of-this we.strive [well Ø_{NOM} please.INF him.DAT]
« Because of this we strive to please him well »

(Haspelmath 2010, 494)

glosses

★ ★ ★ ★ ★ ★ ★ ★

So all of us here today know that glossing is a literal word for word translation of source text to make it accessible to those who don't know the source language. It's used by linguists or people working with languages - spoken, signed.

Here with the example from Haspelmath (2010), we see the source language in the top row, the glosses (a literal translation) in the second row, then in the last row a translation to English. This is how glosses are usually used - along with the source language (albeit in its written form). Glosses are used to give access to source text. For example, here's one from the annotation conventions manual for SLAAASh, a project I'm working on.

Thing is... in sign language linguistics, it seems that we have gone overboard often showing the glosses without the source text. See these examples from Liddell - one from 1980 where it's not so surprising given the limits of technology. But it persisted even in the 2000s - see the 2003 example from Liddell. My colleague Cecily Whitworth once said that it was like studying French through English translations alone.

Glossing



IX_1 RECENT END-OF POSS_1 FS (phd)

"I just completed my PhD."

glosses



Glossing

- (29) $\overline{[RECENTLY\ DOG\ THATa\ CHASE\ CAT]_{S_1}}_r]_{NP}\ COME$
HOME
'The dog which recently chased the cat came home.'

Liddell 1980, p. 149

* * * * *

Glossing

(no source text)

- (29) $\overline{[[\text{RECENTLY DOG THATa CHASE CAT}]_{S_1}]_{NP} \text{COME HOME}}$
'The dog which recently chased the cat came home.'

Liddell 1980, p. 149

* * * * *

Glossing

(no source text)

PRO-1 SATISFIED
I am satisfied.

MOTHER LOVE PRO-1
Mother loves me.

at A up, right down at A
TOMORROW RAIN YOU GO SWIMMING BEACH YOU
If it rains tomorrow, will you go swimming at the beach?

Liddell 2003, p. 23-24

★ ★ ★ ★ ★ ★ ★ ★

“TYRANNY OF GLOSSING”

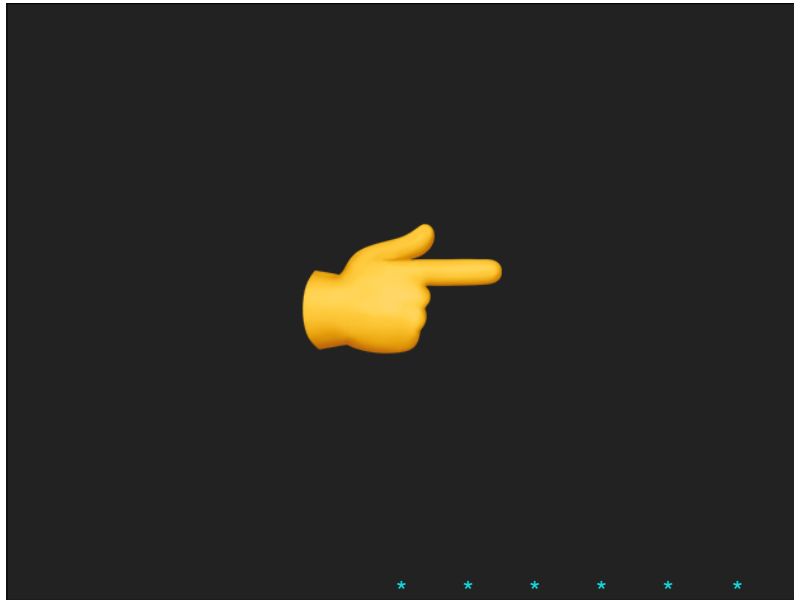
SLOBIN (2008)



The problems with the style of representing signed language data with glosses have been well-documented yet it appears to remain the most common method, a problem that Slobin refers to as the “tyranny of glossing” (2008). That to date is what our documentation looks like. It’s especially problematic for me as a phonologist because there’s nothing about the form of ASL signs in these English glosses and speaking of long-term after many of these signs have changed, how do we know what these signs look like? We don’t. I teach phonology courses at Gallaudet University and inherited Robert Johnson’s notes - you may be familiar with his work on ASL phonology. And his main representation of ASL signs is glossing. There’s a lot of glosses he used that I really couldn’t tell you how to produce.

“Returning to the practice of capital-letter glossing, this method of representing a signed language—even with the additions of various diacritic markings and comments in a written language—seems to me to be a strange, almost neocolonialist acceptance that the *spoken* language of the surrounding community is somehow relevant to a linguistic analysis of the local *signed* language. No linguist would dream of using English words in capital letters to describe Navajo, for example, simply because the surrounding dominant language is English.” (p. 123, Slobin, 2008)



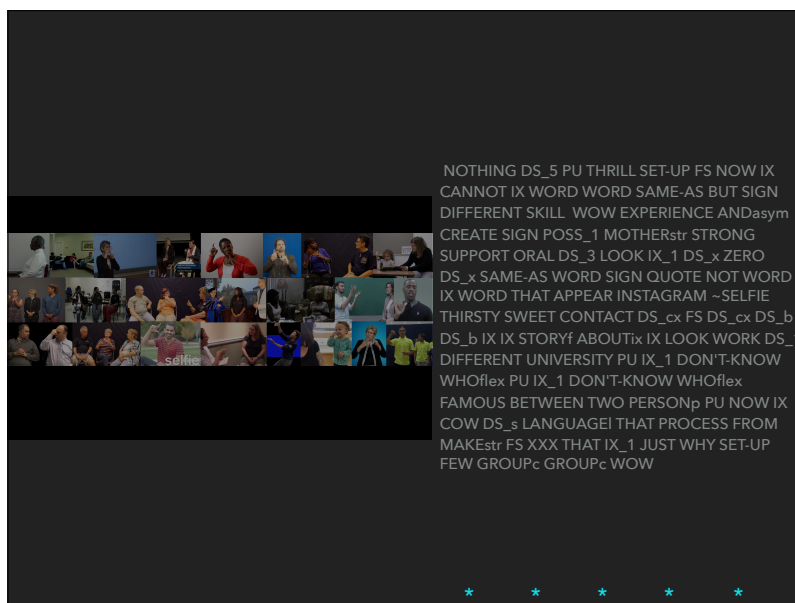


Given the lack of suitable representational replacements for signed languages, we're reliant on the English (or written language) representations here. To do a very simple example of how that's problematic, especially when we're thinking about it from a more mainstream representation, of how it may be construed by the community, I present the pointing sign in the slide.

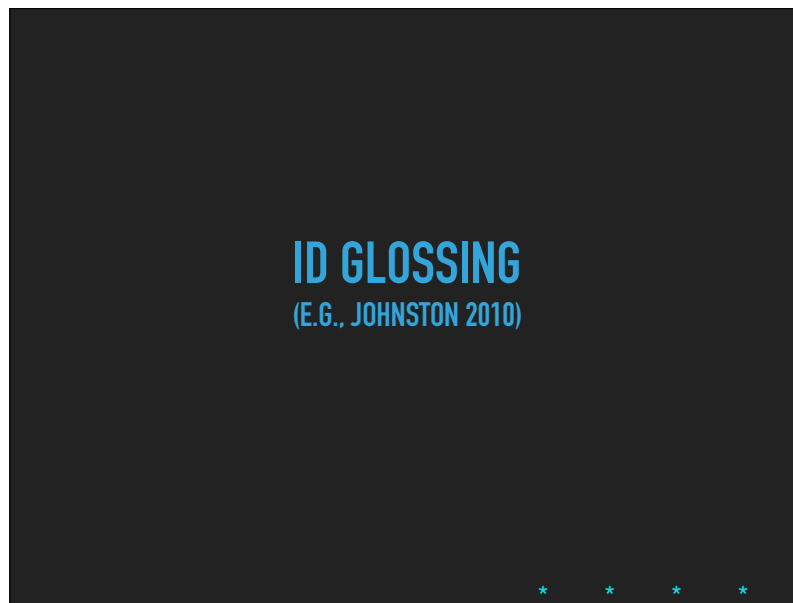
In ASL like many other signed languages, to refer to people or things, we can point. Because this can refer to so many things, it's especially tricky to label. It's so easy because we're using English glosses to represent ASL signs to then think about how it would be represented in English. Once we're there, we start using words like "I, me, you, he, she, it, they". Then it can backfire on us because with those English words, we understand that other things are represented - person (first person like "I" or "me" and second person like "you" and third "he, she, it"), number (singular like "I", "me"...), grammatical case (the nominative "I" versus the accusative "me"). All of these senses in the English words then somehow get linked back to this form in ASL which doesn't do any of that - it points and we use our understanding of the language, the discourse event and the physical surroundings to figure out the reference. But a side effect of glossing, especially as our main choice of representation, is that it obscures the original form as I've demonstrated in this slide. This is just one example of many.

HE IX
(REFERENT) HIM SHE
THEIR IT ME
HER THEY
I
PRO

★ ★ ★ ★ ★ ★



So we get the problems with glossing. But we also need to represent what we're seeing so we can document it and study it. What do we use then? I don't have any answers really. With glossing itself, I haven't found a suitable replacement that works because all of the notation systems we have for signed languages are too technical and take time to learn for most people and those as well as the written systems that have been invented by the signing communities are not yet (or easily made) machine-readable. So I don't know the answer but I do know that I want to minimize the "tyranny of glossing" and bring the primary language itself to the forefront whenever it's represented. So I'll discuss how I've been starting to do that.



ID-glossing (e.g., Johnston, 2010) can help us minimize that tyranny. It basically looks the same as glossing but when used with a database and video annotation, we're able to use glosses as indexical representations. It helps to reduce the impact of the written system. Still plenty of problems there but when used as intended...

For example, take the ASL Signbank which I use as a maintenance of ID glosses for ASL signs (as you know, 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)- remember the glossing we discussed earlier - now it's not the main form of representation like we want but rather a form of labeling data (or ASL signs) so we can find them in the videos we have. This screen recording with the video we showed before but in ELAN with the ASL Signbank linked shows how ASL becomes more prominent in the representation. I think some if not many (or all) of you in this room today have access to this kind of representation.

ASL
signbank

(Hochgesang, Crasborn and Lillo-Martin, 2018)

HomeAboutSignsFeedback

Search glossesSearch translations






Form your query

Search signCSVECVReset

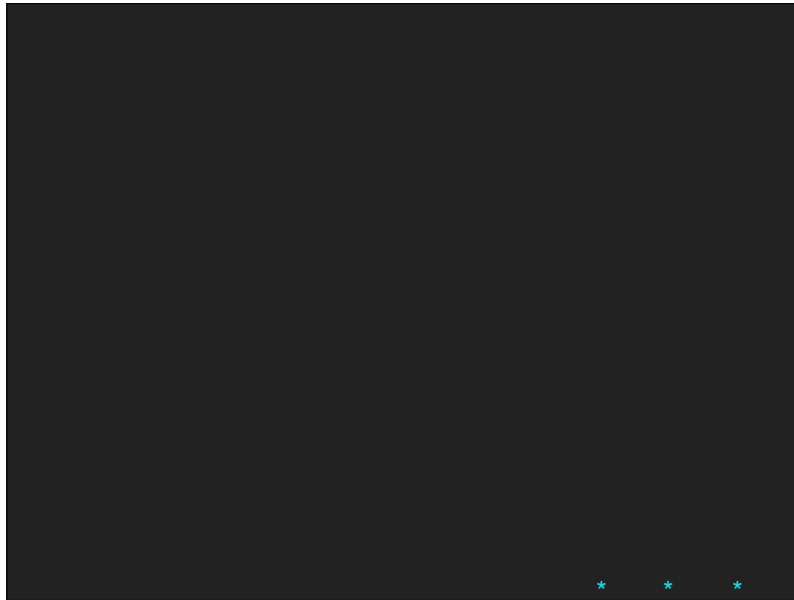
List ViewLemma View

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	-	

[illegible]



These are examples of how I've tried to minimize my reliance on glossing and maintain some kind of connection to the original data. Some are enabled by my use of the ASL Signbank. I also show an example from an article published in the 2018 proceedings of the LREC sign language workshop by the Swedish Sign Language Dictionary team. Two are pulled from my conference posters for TISLR13 - sneak peek! One is the main point shown on the poster and it felt wrong letting English do it solo.

In these examples of my work as well as one by the Swedish Sign Language team, we don't let English glosses serve as the sole representations of the signed forms that are being discussed. There are images as figures or in-line text or even links to online sources with them.

For my own students at Gallaudet, I strongly encourage them to use glosses with signs (either as figures offset in the body or as in-text images or even using specific fonts that represent signs - emojis, hand text and even HamNoSys). And in this last example I share with permission from Gallaudet LIN PhD student Kaj Kraus we see a stunning instance of the ASL sign for "system" by Laurene Simms. In this visualization, we are so much more closer to what we're studying than the English gloss ever allows.



ARTICLE[+]
“different articles”

DIFFERENT[+]

(Hochgesang & Becker, 2019)



that was current then – as Figure 2 shows it was a combination of SIGN and FS(bank). The NAD members were not keen on having fingerspelling as part of the name sign so they offered their own version seen in Figure 3 below. They chose another version of the word “to sign”. The one in Figure 2 is a more neutral and noun-like version referring to the general ability to sign or the modality itself; the one in Figure 3 is more related to the sense of using a word (or sign) itself and is often used to identify people who can sign fluently or are “closer to the Deaf community” than others.



Hochgesang, 2018



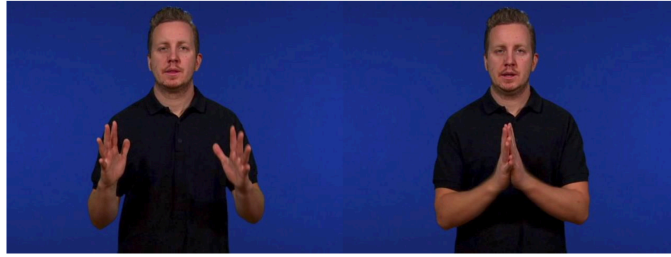





Figure 4: The two-handed sign TIO(Y) (SSLID ID 11951) ('ten')


²The ID numbers in the figure captions are linked to the sign entry in the SSLD.




Kankkonen et al 2018

I demonstrate the use of the diacritic in the following signs where the thumb has

behaved unpredictably. In BLUE  [B•], HORSE  (P1) [H•], and STUCK

 [L or G•], the thumb is unopposed and abducted when it is usually flexed (as

compared to other child and target productions. In HELP  [A•], the thumb is abducted (or spread away from the palm) when it is usually adducted (held close to the side of the fingers). In NOT

 [A•], OTHER  [A•], and RED  (P1) [L

(Hochgesang, 2013)



PREPARE SHARE ACQUIRE CORPUS

“Preparing to share acquisition corpus”

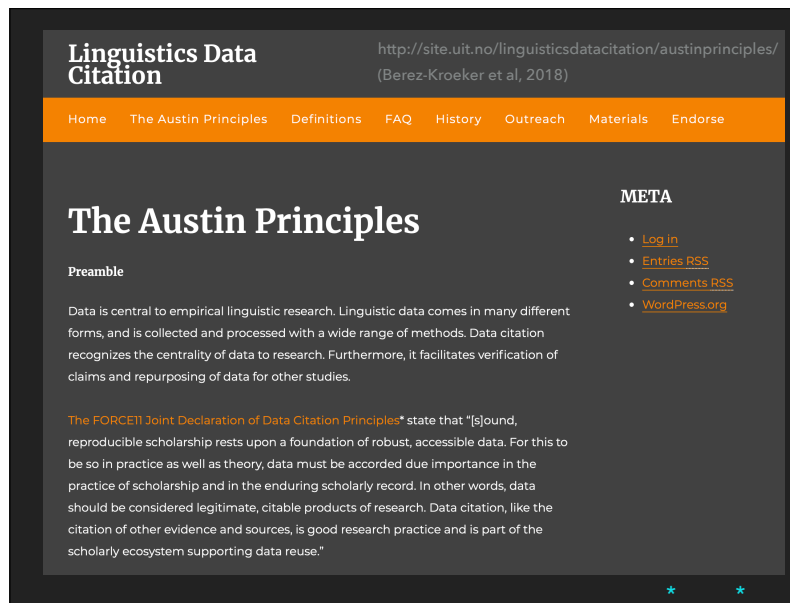
(Hochgesang, et al 2019)

The ID gloss of the first sign I will discuss is SYSTEMs, as shown in figure 4.



Figure 4. Simms signing "SYSTEMs"

(Used with permission, class paper by Gallaudet LIN PhD student Kaj Kraus)



So I've made two main points so far - first, glossing is our field's main choice of representation, especially in our information dissemination (published articles, conference posters and presentations, etc). Second, I've demonstrated how that's problematic and how I've tried to minimize the use of glosses although still rely on it for making primary data (the videos I work with) machine-readable. The third and last point I want to make today - why should we care? Because we as signed language researchers should make access to data a priority. To discuss that, I turn to the Austin Principles of Data Citation - which is a general push by linguists to make primary data (spoken, written or signed) accessible. It's based on the FORCE11 Joint Declaration of Data Citation Principles. If you haven't seen this site or read their article yet, highly recommended.

For today, I'd like to point out two relevant aspects of this. First is reproducibility - one of the reasons why we're all here today! If we want our work to resonate in our field and our communities, they need to be reproducible and that cannot be done without making access to our data a priority and that data is signed forms, not written glosses of the ambient spoken languages.

The second is data citation - which is simple. Data is what we're looking at and we need maintain as close as a link as we can to it in everything we do. What I shared today (in how I try to minimize the reliance on glosses) is a start but needs much more work.

And finally, about how data citation should give us access to the “data themselves”, especially because we owe it to the communities themselves to honor the languages they shared with us.

How can we think about these principles for data citation in terms of representing signs? There are many questions that remain. Especially how to make sure these sources remain accessible. How do we archive them? How do we make sure the citations persist? And so on. But that’s why we’re here today - to discuss this issue. I look forward to working on addressing these issues with you all.

The Austin Principles

META

- [Log in](#)
- [Entries RSS](#)

“[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 ...

The FORCER 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.”



The Austin Principles

META

- [Log in](#)
- [Enter RSS](#)

Data citation recognizes the centrality of data to research.

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 FORCET 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."



5. Access

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.

Linguistic data should be as open as possible, in order to facilitate reproducibility; and as closed as necessary**, to honor relevant ethical, legal and speaker community constraints.

citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse."

"Most scientists rely on written and graphic representations, at the same time that they maintain that they do not study maps, photographs, spectra, charts or graphs but study frogs, star clusters, T-cells, and chromosomes."

- Michael Erard, 2000, p. 21

In closing, I'll just leave you with this one last quote by Michael Erard from his dissertation, "Inscribing language: Writing and scientific representation in American linguistics."

(You may know Michael Erard from his popular science articles/books - quite a few related to signed languages, e.g., Atlantic Monthly article on signing gloves - <https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/>)

REFERENCES

- Berez-Kroeker, A., Gawni, 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 Zwitterlood 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://elam.mpi.nl/tools/elan/elan/>.
- Erard, M. (2000). Inscribing language: Writing and scientific representation in American linguistics. (Doctoral dissertation, University of Texas at Austin, 2000). *Dissertation Abstracts International*, 62(02), 549.
- Haspelmath, Martin. (2010). https://www.researchgate.net/publication/240290554_The_change_of_behavioral_and_coding_properties_in_constructional_grammaticalization
- Hochgesang, J.A. (2018). SLAAASH and the ASL Deaf Communities (or "so many gifs!") In *Involving the Language Community: Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages*. Ed by E. Efthimiou, E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, & J. Mesch (Eds). 63-68. 11th International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan. Paris. ELRA.
- Hochgesang, Julie A. and Amelia Becker. (2019). Absence of reduplication in American Sign Language nominal plural morphology. Conference Poster presented at TISLR13, the 13th Conference of Theoretical Issues in Sign Language Research, Hamburg, Germany (September 26-28, 2019).
- Hochgesang, J.A., O. Crasborn, & D. Lillo-Martin. (2018). ASL Signbank. New Haven, CT: Haskins Lab, Yale University. <https://aslsignbank.haskins.yale.edu/>.
- Hochgesang, Julie A., O. Crasborn, D. Lillo-Martin. 2018b. Building the ASL Signbank: Lemmatization Principles for ASL. In *Proceedings of the LREC 2018 Workshop 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* 11th International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, edited by by Eleni Efthimiou, Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen, & Johanna Mesch. 69-74. ELRA. http://lrec-conf.org/workshops/lrec2018/W17/pdf/book_of_proceedings.pdf.
- Hochgesang, Julie A., Dorcasan Catt, Deborah Chen Pfeiffer, Corina Goodwin, Carmelina Kennedy, Lea Prunier, Doreen Simons, and Diane Lillo-Martin. (2019). Sign Language Acquisition, Annotation, Archiving and Sharing: The SLAAASH Project Status Report. Conference Poster presented at TISLR13, the 13th Conference of Theoretical Issues in Sign Language Research, Hamburg, Germany (September 26-28, 2019). <https://doi.org/10.6084/m9.figshare.9742529>
- Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15:1,106-131.
- Kankkonen, Nikolas R., Thomas Björkstén, Johanna Mesch & Carl Böttzell. 2018. Crowdsourcing for the Swedish Sign Language Dictionary. In Mayumi Bono, Eleni Efthimiou, Stavroula Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen, Johanna Mesch & Yutaka Osugi (eds.), *Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* (Language Resources and Evaluation Conference (LREC)), 171-176. Miyazaki: European Language Resources Association (ELRA).
- Liddell, Scott. (1980). *American Sign Language Syntax*. Mouton.
- Liddell, S.K. (2003). *Grammar, Gesture and Meaning in American Sign Language*. Cambridge.
- Slobin, D.I. (2008). Breaking the Molds: Signed Languages and the Nature of Human Language. *Sign Language Studies*, 8 (2), 114-130.
- Stokoe, W., D. Casterline, and C. Croneberg. 1965. *A Dictionary of American Sign Language on Linguistic Principles*. Silver Spring, MD: Linstok Press

TYRANNY OF GLOSSING REVISITED...

(LINK TO SLIDES [HTTPS://SHWCA.SE/GUDA](https://shwca.se/guda))

- ▶ julie.hochgesang@gallaudet.edu
- ▶ @jahochcam (Twitter)



Just in case

textual representation

All images - ASL Signbank, 2018

FEMALE (*glossing*) **US**^vx* (*Stokoe*) **F**^v (*ASLWrite*)

transcription *notation* *writing*

Although there are plenty of overlaps with how information can be represented textually, transcription here means representing primary language (spoken or signed) in a way that can be used for analysis (so here we see glossing). Then notation means representation of the form for phonetic or phonological information (here we see Stokoe notation). Then we have writing - cultural invention for disseminating information - not widespread in ASL or any signed language. I show one attempt here (ASLWrite).

Further "reading" **emojis as representation** 🙌 = WONDERFUL (in ASL)



shoshannah7

Liked by erinkfisher and 383 others

shoshannah7 🙌🙌🙌 (ASL for yes for the non-fluent) we get FOUR more episodes of #ThisClose this week on @sundancetv. Two on Thursday and the final two of our season on Friday! 🙌🙌 by @unitstills

<https://www.instagram.com/p/B2q4B12lpwb/>

Julie Ann Hochgesang @jahochcam

Replying to @gab_hodge @AdamCSchembri and @c_borstell

The sign LESBIAN in ASL (co-opting this emoji for it! 🙌) - depending on where the contact with the cheek occurs on the finger is telling of your experience with that community (there's an article on it but I can't remember anything about it!)

2:10 PM · Apr 26, 2019 · Twitter for iPhone

<https://twitter.com/jahochcam/status/1121839024245813250?s=20>

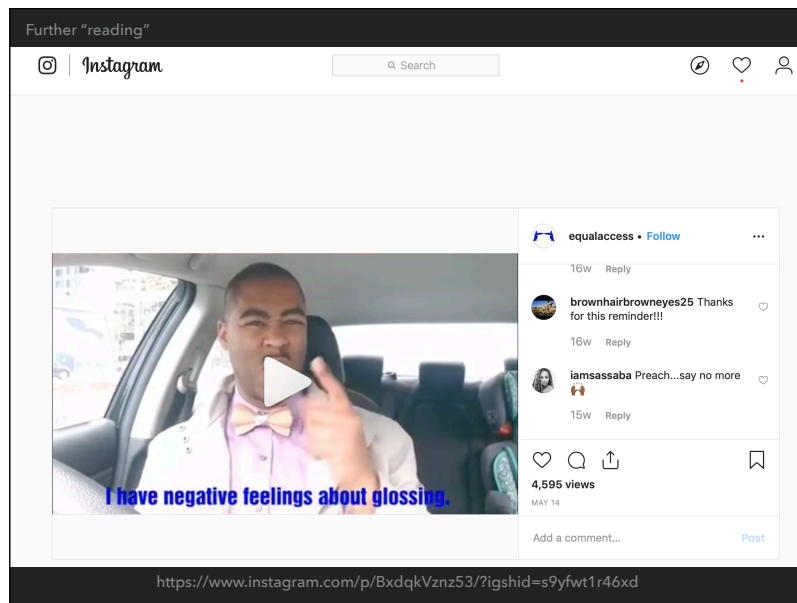
ASL SLAASH @AslSlaash

On our insta, we used 🙌 to label the handshape in the ASL sign for "airplane" and someone asked what that emoji then meant. She was confused as to whether it meant "I love you" or was a heavy metal gesture 1/n

8:47 AM · Jan 28, 2019 · Twitter for iPhone

<https://twitter.com/AslSlaash/status/1089882576708427776?s=20>

I wanted to make this fit my presentation but I couldn't so much in that this isn't the best response to the issues I've raised today BUT it is indicative of the signing community's preference to use something that resembles their forms and in a way that's machine-readable. Emojis to represent signs. Obviously they still cannot represent everything (or even much) but that the practice, however limited it is, seems common and indicates something about what the signing community prefers.



An opinion on glossing from someone in the signing communities (specifically ASL)