

1 About This Element

“Kira: What’s writing? Jen: Words that stay”

(Dark Crystal, 1982).

Humans have known for thousands of years that we need to turn to technology to make our words stay, whether they’re spoken, signed, or felt. There are well-known written orthographies or systems for spoken languages like the one you’re reading now for English. But there aren’t any conventionalized systems for signed languages. I’ve thought a lot about how to make signed words stay and I share one of those ways with you in this Element.

ASL Signbank¹ is a digital tool that stores data such as American Sign Language (ASL) videos, photos, and text, all connected with individual ASL signs for the purpose of annotating ASL videos. The logo, shown in Figure 1, honors both our name sign and iconic teal background used to film signs for ASL Signbank.

Section 1.1 is an introduction and a preface in which I describe my personal motivations for creating and maintaining ASL Signbank. It is also where I acknowledge where I am most at ease – with the practice of data documentation or annotating and organizing bits of data, rather than writing. Section 2 describes what the ASL Signbank is, along with a history of signbanks and what kind of data have been used to build the resource. I also describe our ASL Signbank name sign. Section 3 dives into why signbanks exist – the need for textual representation, using ID glosses for corpora and language documentation, and annotation software used with signbanks. Section 4 is a close look at lemmatization and the principles we used to develop textual labels in ASL Signbank. Section 5 walks the reader through a single entry with detailed descriptions for each entry and possible values. Section 6 briefly describes how to use ASL Signbank with a specific set of annotation conventions with a description of an ELAN template file (.etf) associated with ASL Signbank as an External Controlled Vocabulary (ECV). Finally, Section 7 wraps up with some concluding reflections.

1.1 Introduction

I’m a documenter and a maintainer. That’s what I do – create and maintain records that can be digitally shared. I’ve been documenting signed languages, especially the one I grew up with – ASL – for a while now. As part of this practice, I spend quite a bit of time pondering how to document ASL signs in use, as shown in Figure 2, to be shared in ways the ASL communities appreciate.

¹ <https://aslsignbank.com/>

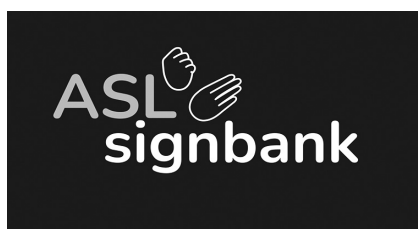


Figure 1 ASL Signbank logo designed by deaf designer Silvia Palmieri (2023).



Figure 2 Images of ASL Signbank actors signing ASL words.²

I invite you to ponder this work with me. How could we document ASL like spoken languages are recorded in text? ASL and other signed languages currently have no conventionalized written systems. Imagine, for example, searching for signs across ASL videos. What would this work entail? Like we see with YouTube videos or social media posts, you might recommend adding captions or a bunch of hashtags, likely using English words that correspond to the

² <https://aslsignbank.com/about/copyright/>

meaning of relevant ASL signs. To search across videos, these captions or hashtags need to be consistent. If you search for “mom,” your results may not show others that you know to be relevant such as “mother,” “mama,” or “mommy”; so to be searchable, you need to use the same choice. To remember those choices, a list of some kind needs to be maintained. It’s going to get long quickly. And you might want to connect them to other signs to help find them easily or point out other kinds of connections like how they look alike or have similar meanings. That’s hard to do with a list. You may be thinking okay, let’s build a database and put them there. Great, good idea. What do we put in it? And what about when we come across new signs? And what about preparing shareable videos of the signs? You might be thinking, this is going to take me some time and resources. Yes, it will. I’ve been thinking about how to represent ASL video data since I started graduate school in 2005 and have been working on ASL Signbank properly since 2015. And I’m still thinking about all of this.

ASL Signbank is an old friend of mine. In its various incarnations, it has kept me company for many years. We’ve grappled with many issues together. It is hard to tell its story because it’s rather complicated and it’s not easy to know where to start. I’ve ended up with an Element that not only tells you a bit about how to do this work but also reflects on it all. ASL Signbank (Hochgesang *et al.* 2025) may appear to be a straightforward product with well-defined edges but is the ongoing result of research processes that I and other researchers undertake daily. It’s also a reflection of how I do research, which is tied to centering deaf signers of all kinds contributing meaningfully to knowledge. These contributions must happen in ways that acknowledge and represent language use as embodied processes and making space for different kinds of signers in a balanced manner. I practice “slow science,” meaning I do science in a way where I value working with collaborators and enjoy the processes themselves. ASL Signbank is also a demonstration of a signing deaf linguist doing work in a mostly hearing space (Hochgesang 2018, 2022a).

In short, I am interested in the use of ASL because, among a great number of reasons, I’m Deaf and it’s been my primary language since I was one year old along with English (preferably written). I am a linguist, and my interest is in signed languages and how the communities use and talk about them. Video is essential for my work as well as technology such as ELAN³ (Crasborn & Sloetjes 2008) or iLex (Hanke 2002) as annotation software to reduce the labor of finding things again in videos instead of rewatching them in real time. This software allows me to time-link annotations, or textual notes, to videos, which enables finding things in the data. As someone who grew up

³ <https://archive.mpi.nl/tla/elan>

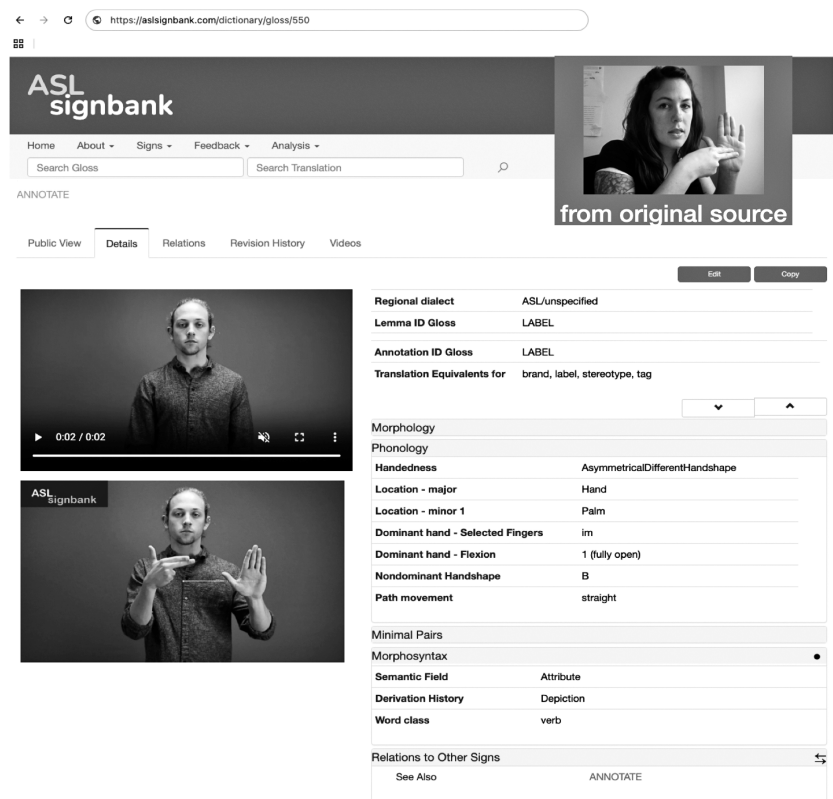


Figure 3 ASL Signbank entry for “label” with source (top right).

constantly browsing dictionaries, this has been a source of immense joy and fascination. As a research assistant for Deborah Chen Pichler and Diane Lillo-Martin for ASL acquisition projects (Hochgesang 2022b), I started organizing and labeling signs in 2006 and have since created thousands of notes, photos, and videos in ways that can be reused for annotation and sharing. As a maintainer of ASL Signbank, I’ve curated or created relevant (meta)data, such as original source videos and screenshots of signs, refilmed sign videos and screenshots, and sign entry information, as demonstrated in Figure 3.

Figure 3 is an example of an ASL Signbank entry I created for the ASL sign “to label,”⁴ where the English word (or “label”) serves as the sign’s identifying “ID” gloss, a consistent and unique textual label (Johnston 2010) to enable

⁴ All glosses in this Element are hyperlinked to ASL Signbank. If you are reading this in print, you can search the glosses at <https://aslsignbank.com/> or browse for them here – Figshare collection: <https://doi.org/10.6084/m9.figshare.30747176>. In fact, most of the online resources I created and linked to in this Element are shared in a single Figshare collection – <https://doi.org/10.6084/m9.figshare.c.7911878>

searching and the like. In my work, I create ID glosses for signs in videos that need to be textually represented in annotation, as indicated by the “draft” version of the sign in front of the top right box in Figure 3.

As an academic, I could have started this introduction with the seemingly objective statement: “ASL Signbank is a lexical database that stores ID glosses, unique textual labels, that use English text linked to ASL signs, in order to annotate ASL videos. It has been built upon the template of other signbanks developed by those working with signed language corpora.” That statement may be accurate but is devoid of oft messy and ongoing questions and processes that come with developing and maintaining any scientific product. We created ASL Signbank to represent the signed data we had and wanted to use for our research. We also wanted to make it useful to the signing communities. Along the way, I used what I know about ASL as a Deaf signer and as a signed language linguist with an interest in phonetics and phonology to develop an annotation tool that could be used with ELAN, which is well-used among signed language researchers.

1.2 Writing about This Resource and Reflection

This Element was hard to write. Not just the typical writer’s block and the anxious procrastination that comes with it. Not just the apocarevolutiondemic days (Hochgesang 2022a) and imposter syndrome (Chua *et al.* 2022). But because resources like dictionaries and databases can be a source of violence (Craft *et al.* 2020), including gatekeeping or discrimination. Researchers working with sociolinguistic (Bucholtz 2000), raciolinguistic (Rosa & Flores 2017), educational (Charity Hudley & Mallinson 2018), and crip linguistic (Henner & Robinson 2023) perspectives have noted that linguistic resources, and the idea of “Linguistics,” are often used to push a specific linguistic, teaching, or cultural agenda. One only has to look at the long, extractive, and colonial history of museums or even the practice of collecting to see the harms that have been done by imposing a narrow set of experiences on what is actually a wide-ranging set of bodies of knowledge.⁵ And this has persisted in how we process and store knowledge digitally as digital archivists⁶ or librarians know all too well (Reidsma 2019).

Linguistic disciplines, including language documentation, share the same issues (Leonard 2018). While we have FAIR⁷ guidelines (Wilkinson *et al.* 2016) to help us consider how to responsibly process and share data openly,

⁵ www.rolecollective.org/

⁶ <https://taeyoonchoi.com/soft-care/distributed-web-of-care/> (last accessed August 2023)

⁷ Findable, Accessible, Interoperable, and Reuseable.

we also have the CARE⁸ guidelines (Carroll *et al.* 2020) to remind us that these data come from people and should be treated with utmost respect and care. Management of that data, both primary and meta, is essential for citation and subsequent accountability (Berez-Kroeker *et al.* 2022).

And we have considered what it means to work with the signed communities when doing research (Harris *et al.* 2009; McCaskill *et al.*, 2011; Kusters 2012; Lucas 2013; Singleton *et al.* 2015; Moriarty 2020; Hochgesang & Palfreyman 2022; Hochgesang 2023; Hou & Ali 2024). There are many claims that have been made in early signed language research – often based on data of a small number of signing consultants and not openly shared – that have shaped much of our working methodological and theoretical assumptions today. They need to be revisited and evaluated. This work cannot be done without revisiting accessible data and has not been done much for signed languages (Bennett 2024). Even if ASL is one of the most studied signed languages out there, there remains much we do not know or have not asked.

I have been thinking about the lessons shared by all of who I just mentioned and have reflected on how to do the work I do. And because I know my work cannot, and should not, represent all the ASL communities (Henner & Robinson 2023; Hou & Namboodiripad 2025), I work to create resources that do not center or privilege a specific group as language resources often do, for example, “White Deaf native signerism” (Lim and Hou (Lim 2025)). I have tried to ensure that the signers seen in ASL Signbank videos represent a wide variety of signing experiences tied to different kinds of embodied socio-identities. Even with its limitations, I propose that continually maintained signbanks, with some flexibility and constant reflection in design and care, are useful strategies for archiving contemporary language use – eventually becoming historical sources.

Historically, the work of developing a language resource has usually meant that one person or a small group of people make decisions about a language that belongs to millions. Language description requires entering a conversation that is decades old and rife with theoretical assumptions, such as considering native speakers as ideal sources of data although they aren’t (Cheng *et al.* 2021), that were made by hearing nonsigners who barely consulted deaf ones, who, if they were lucky, were the assistants or lab managers.

And it irks me to no end that we rely on glosses from ambient spoken languages to represent signed languages, like English glosses for ASL (“tyranny of glossing” (Slobin 2008; Hochgesang 2019)). But to draw upon an experience that the modern person can perhaps relate to, glosses are like passwords. People hardly want to use them. They’re a pain to remember. It takes convoluted

⁸ Collective benefit, Authority to control, Responsibility, and Ethics.

methods to use them, but they're necessary for access to tools and information. And to serve that purpose, they – both the passwords and the glosses – need to be consistent and shared. And with glosses, that work should be done with reflection while operationalizing CARE FAIRly.

I hope readers of this Element exercise the same kind of care. In these days of so-called Artificial Intelligence (AI) and oft-extractive natural language processing (NLP) (Bender & Friedman 2018; Bender & Hanna 2025), it is even more essential. While on the topic of “AI,” please note that signers should always accompany ASL Signbank use whether it's browsing, annotating, or teaching. It was entirely made with them in mind (Section 1.4). There are plenty of resources out there for ASL learners, very few for those who already use them as their daily primary language. As for what I write about in this Element, I can only be explicit and reflective. I may offend people who say that science is supposed to be neutral and objective. That's just untrue. Science has been led by those with power who have decisions about what is “neutral” and “objective” and want you to believe it is so (e.g., Clemons 2024).

As will be clear here, I don't see myself as a traditional academic. As I thought about how to write this, I found myself thinking what do readers want me to say? What is the “right” way to do this, but then I realized that's not what I can write about. I'm a Deaf ASL community member trained in linguistics. I see the beauty and importance of practices and feel the shapes of things better as I do them, as I interact with them and reflect on the process. That's what I write about. This Element will not be a “traditional academic” product, a decision I have made precisely because I question the label “traditional academic.” It will be what I would have liked to see when I was a young scholar. It's for deaf people. It's for people who don't see themselves as traditional academics but have the knowledge although they haven't seen themselves as having authority for constructing and sharing that knowledge.

1.3 Whose Story and for What Purpose?

As Roberts (1997) wrote: “every decision about how to transcribe tells a story. The question is, whose story and for what purpose?” (p. 169). I started working on my dissertation in 2008. When I first started out, I decided that its title would be “whose story and for what purpose?” When I finished in 2013, the title was completely different (Hochgesang 2013) although the “why” driving my work was still the same. I wanted to do a dissertation focused on representation of data. Dear-well-meaning-but-naive-2008-Julie imagined she'd do each chapter focusing on different levels of representation in signed language research – from phonetic to morphological, syntactic to discourse level – and dive into the

histories of representational choices and practical demonstrations using actual signed language data from different signed languages. From this work, surely, I'd be able to examine the different stories of signed language researchers and the purposes they had.

Luckily, Bob Johnson, my dissertation chair, counseled me to make it smaller, much smaller. I ended up focusing on phonological representation of the handshake in ASL using child acquisition data consisting of one child's set of signs over two years and examining how four notation systems did the job. While the focus was narrow, the work was intensive and consuming. I was able to focus on "the small bits" of language use, and in the small things, I could still see the big. And the big with that work was realizing how nuanced the transcription process is, and how difficult it is to identify the one system that will solve all our problems.⁹ The other members of my dissertation committee also helped me realize this, especially Onno Crasborn. In describing his work for a phonological database for signed languages, Crasborn discussed the "database paradox" (Crasborn, van der Hulst & van der Kooij 2001) where one needs to make decisions about what to represent while also doing the representation which is quite the metalinguistic (read: mind-boggling) task!

I'll admit. . . . I started my dissertation thinking I'd find the one system to rule all, the one system that would be best for anything. I quickly realized that was foolish and impossible. This would be like saying there should be only one kind of kitchen utensil. Tools for language representation vary because they must – people have different preferences, goals, and skill sets. Their access to resources like funding, colleagues, students, research assistants, and tools will always vary based on where they work and with whom. I've given up on finding the one answer. But I haven't given up on wondering about the work that goes into it, why, and by whom. "Whose story and for what purpose?" are important questions. What is considered data? (Finnegan 2008). How do people perceive, collect, represent, share, use, analyze, and cite the data? What motivates them? This matters because these ways have the potential of becoming other people's ways. While shaped by knowledge, our practices, and our choices, our representation in return shapes knowledge (Ochs 1979; Edwards 1993; Murphy 2021).

Since I've been at Gallaudet, both as a graduate student and a professor in the linguistics department, I've worked with data representation. I've taught phonology and field methods. I've trained students in the use of notation systems for signed languages and annotation software linked to videos. As described in Hochgesang (2022b), I've been considering how to represent signs – usually produced by one or two hands/arms in front of the signer, along with considerable

⁹ xkcd.com/927/

information on the face, torso, and even the rest of the body. I've thought endlessly about how to capture what I observe about signers languaging – from the small to the big, keeping in mind how design choices shape the representation and subsequent use of. As Roberts (1997) writes:

If talk is a social act, then so is transcription. As transcribers fix the fleeting moment of words as marks on the page, they call up the social roles and relations constituted in language . . . transcribers bring their own language ideology to the task. In other words, all transcription is representation, and there is no natural or objective way in which talk can be written. (1997, pp. 167–168)

As a signing deaf person from a midwestern middle-class hearing family who grew up in mainstreamed educational environments with frequent deaf socialization through deaf friends and deaf events with signing deaf adults, I have seen firsthand the intense and complicated language attitudes regarding English and ASL, not to mention other spoken and signed languages. I've battled with them myself. Like most work for spoken language linguistics, English is the dominant form of textual representation for signed languages, especially ASL, for both transcription and dissemination. I've explored this in previous work, where I advocate for prioritizing visual representation for visual signed data (#Glossgesang)¹⁰ (Hochgesang 2019, 2022b). What's especially important to note here is that this practice of using English as our main form of representation, often leaving out the original data (the signed data), has deeply shaped our current knowledge of signed language research. For example, representing  with English “he,” “she,” or “it” means there's also the senses associated with those English pronouns – gender, third person, and singular – that could be carried over.

Some linguists are now pushing for language data to be shared publicly or as open access (e.g., Seyfeddinipur *et al.* (2019) citing Christen (2012)). Under these research models, research should then be freely accessible and available. Ironically, this free access is not cheap in terms of resources. Resources include immense people power, stable technological infrastructure, and ongoing institutional support (Hochgesang *et al.* 2023). It also includes the power of language representation, such as who can “decide” what kind of knowledge to include. I know firsthand that many signed language researchers either don't know how to sign or cannot sign much. I also know that some researchers can be extractive and exploitative. This makes it difficult to trust their “read” of the data, a problem that is compounded when the data is not available. I share work

¹⁰ Coined by Carl Börstell (2019), www.github.com/borstell/signglossR, #Glossgesang is the principle: “always present sign language data in a visual format (videos/images) without relying solely on glossing.”

because I want to see the data firsthand and make sure others in the signing communities can access this tool (Berez-Kroeker *et al.* 2022). To make sure we are accountable by ensuring knowledge is accurately represented while acknowledging that “accurate” is subjective.

While I will be writing about ASL Signbank in this Element to explain the tool so others may use it for their own purposes, I will also be reflecting on choices that I made. The care I took while building and maintaining this resource. Language use is not to be distilled into simple labels devoid of interaction and studied on their own. But rather the labels along with recorded language use helps us make sense of it all.

Lest I make maintaining the ASL Signbank seem like a chore, let me be perfectly clear – working on ASL Signbank is a joy for quite a few reasons. Among them, celebrating the variation and actual language use of signers in North America. I loved dictionaries as a child and still do. While ASL Signbank is not a dictionary, there are processes that are like making a dictionary. There is joy in cataloging our language use for the use of others. When I annotate videos, I see many things that make me go VEE: The sign choices people make. The certain signs of places, identities, or situations. The ways bodies move to articulate signs. For example, people love fingerspelling on the way to the next sign. No wonder learners or algorithms have a hard time with fingerspelling! But I love it – they’re sneaky little words born on the go and disappear upon arrival but still manage to contribute to the message overall. All of this is fascinating to me as a linguist observing and documenting people doing language. That’s the whole point of observing language – to see how people move their bodies to communicate, perform, express identity, and so on. It puzzles me that we’d outsource this “labor” to algorithms. I enjoy sharing these observations with my communities in nontraditional ways that are more accessible to people such as VEEing emboxed discourse, that is, signing while using video-mediated communication like Zoom or Facetime, during the early days of the COVID-19 pandemic in 2020 (Hochgesang 2025b) and doing language documentation projects (Hochgesang 2023).¹¹ I often share them on social media or my own website, which allows me to take advantage of digital tools to express my ideas in a multimodal manner – with written English, emojis, signs in photos, gifs or videos, and links to ASL Signbank entries.

1.4 ASL Signbank Is Not for “AI” Extraction

ASL Signbank is not intended for “AI” extraction. My objections to this practice are numerous and include skepticism about how “AI” researchers view data, acknowledgement for attribution and authority of knowledge,

¹¹ Each of those citations point to a source with an ASL video with captions.