Exploring the Benefits and Applications of Video-Span Selection and Search for Real-Time Support in Sign Language Video Comprehension among ASL Learners

SAAD HASSAN, Tulane University, New Orleans, LA, USA
CALUÃ DE LACERDA PATACA, AKHTER AL AMIN, and LALEH NOURIAN, Computing and Information Science, Rochester Institute of Technology, Rochester, NY, USA
DIEGO NAVARRO, National Institute for the Deaf, Rochester Institute of Technology, Rochester, NY, USA
SOOYEON LEE, New Jersey Institute of Technology, Newark, NJ, USA
ALEXIS GORDON, MATTHEW WATKINS, GARRETH W. TIGWELL, and
MATT HUENERFAUTH, School of Information, Rochester Institute of Technology, Rochester, NY, USA

People learning American Sign Language (ASL) and practicing their comprehension skills will often encounter complex ASL videos that may contain unfamiliar signs. Existing dictionary tools require users to isolate a single unknown sign before initiating a search by selecting linguistic properties or performing the sign in front of a webcam. This process presents challenges in extracting and reproducing unfamiliar signs, disrupting the video-watching experience, and requiring learners to rely on external dictionaries. We explore a technology that allows users to select and view dictionary results for one or more unfamiliar signs while watching a video. We interviewed 14 ASL learners to understand their challenges in understanding ASL videos, strategies for dealing with unfamiliar vocabulary, and expectations for an *in situ* dictionary system. We then conducted an in-depth analysis with eight learners to examine their interactions with a Wizard-of-Oz prototype during a video comprehension task. Finally, we conducted a comparative study with six additional ASL learners to evaluate the speed, accuracy, and workload benefits of an embedded dictionary-search feature within a video player. Our tool outperformed a baseline in the form of an existing online dictionary across all three metrics. The integration of a search tool and span selection offered advantages for video comprehension. Our findings have implications for designers, computer vision researchers, and sign language educators.

CCS Concepts: • **Human-centered computing** → *Accessibility systems and tools*; *Graphical user interfaces*; *Empirical studies in interaction design*; *User interface programming*;

Caluã de Lacerda Pataca, Akhter Al Amin, and Laleh Nourian contributed equally to this research.

This article is based upon work supported by the National Science Foundation under Grant No. 1763569, 2212303, and 2125362.

Authors' Contact Information: Saad Hassan (corresponding author), Tulane University, New Orleans, LA, USA; e-mail: saadhassan@tulane.edu; Caluã de Lacerda Pataca, Computing and Information Science, Rochester Institute of Technology, Rochester, NY, USA; e-mail: cd4610@rit.edu; Akhter Al Amin, Computing and Information Science, Rochester Institute of Technology, Rochester, NY, USA; e-mail: aa7510@rit.edu; Laleh Nourian, Computing and Information Science, Rochester Institute of Technology, Rochester, NY, USA; e-mail: ln2293@rit.edu; Diego Navarro, National Institute for the Deaf, Rochester Institute of Technology, Rochester, NY, USA; e-mail: don6763@rit.edu; Sooyeon Lee, New Jersey Institute of Technology, Newark, NJ, USA; e-mail: sooyeon.lee@njit.edu; Alexis Gordon, School of Information, Rochester Institute of Technology, Rochester, NY, USA; e-mail: aag7593@rit.edu; Matthew Watkins, School of Information, Rochester Institute of Technology, Rochester, NY, USA; e-mail: mxw7981@rit.edu; Garreth W. Tigwell, School of Information, Rochester Institute of Technology, Rochester, NY, USA; e-mail: garreth.w.tigwell@rit.edu; Matt Huenerfauth, School of Information, Rochester Institute of Technology, Rochester, NY, USA; e-mail: matt.huenerfauth@rit.edu.

ACM 1936-7236/2024/10-ART14
https://doi.org/10.1145/3690647

Additional Key Words and Phrases: American Sign language, Sign Languages, Continuous Signing, Sign Language Videos, Sign Look-up, Video Selection, Search Interface, Integrated Search, Sign Language Learning, ASL Learning

ACM Reference format:
Saad Hassan, Caluã de Lacerda Pataca, Akhter Al Amin, Laleh Nourian, Diego Navarro, Sooyeon Lee, Alexis Gordon, Matthew Watkins, Garreth W. Tigwell, and Matt Huenerfauth. 2024. Exploring the Benefits and Applications of Video-Span Selection and Search for Real-Time Support in Sign Language Video Comprehension among ASL Learners. *ACM Trans. Access. Comput.* 17, 3, Article 14 (October 2024), 35 pages.
https://doi.org/10.1145/3690647

1 Introduction

Globally, there are over 70 million **Deaf and Hard of Hearing (DHH)** individuals who rely on one of the over 300 sign languages recognized by the World Federation of the Deaf [15, 58]. In the United States, there is also a growing interest among both DHH and hearing individuals to learn **American Sign Language (ASL)**. ASL serves as the primary means of communication for approximately 500,000 people in the country [56]. Since the majority of DHH children are born in hearing families, their parents or teachers are motivated to learn sign languages [67, 76]. Failure to access spoken language or learn sign language during critical developmental years may cause DHH children to experience language deprivation [24].

ASL has one of the fastest-growing enrollments among language classes [21] with nearly 200,000 students in ASL classes [18] at schools or universities. Students trying to understand a challenging video is part of sign language education to develop comprehension skills [22, 36, 46]. Despite the progress in machine translation for converting ASL videos into English text, such technology is still in the developmental stage [62], and the use of this technology would eliminate the educational aspect of students' effort to understand a video themselves. Therefore, there is a need for technology to support learners during a video-comprehension task without completely automating the process.

In the context of learning foreign languages, dictionaries serve as a valuable tool for students when come across an unfamiliar word in a text or audio recording. However, when learning sign languages, students encounter challenges when coming across unfamiliar signs because there is no standard or convenient writing system for students to search for the meaning of a sign based on its visual appearance. The existing search tools are insufficient [4, 29, 30], thus making it difficult for students to use Web sites that ask them to enter linguistic features of a sign and browse through a list of results to find a matching sign based on visual appearance [2, 9, 10, 48, 55, 70, 74]. Research has explored the development of tools that enable students to submit a video of a single sign to conduct a search within an ASL dictionary [4, 6, 14, 16, 32, 45, 74, 78]. However, when attempting to understand an ASL video, students might have challenges with accurately extracting a single sign and replicating the sign themselves into a webcam to initiate a search [4, 28].

To address this issue, we investigate technologies for enabling users to quickly select a span of a video of ASL signing that contains one or more signs that they do not understand. This selection would then trigger a video analysis search, which would provide a set of dictionary results containing potential matches for the signs within the selected span. We focus on the overall task instead of considering the tasks of watching a video and looking up the meaning of an unknown sign separately. We first received feedback from ASL learners on their current issues related to challenging video comprehension, their existing workarounds, and their expectations from a future system. The feedback informed the design of our integrated dictionary system and the choice of video stimuli used in the subsequent studies.

In our second study, eight ASL learners watched challenging ASL videos containing signs that they were unfamiliar with while using a Wizard-of-Oz prototype. The Wizard-of-Oz method typically involves some careful planning to give the impression of a fully functioning system, but the user does not know during use, and it is well established as an approach in **human-computer interaction (HCI)** research [49]. This prototype allowed our participants to play videos, select spans of video, and perform searches for signs but other students used a baseline prototype without the searching functionality. This study provides insights on how ASL learners interact with such a system. Finally, our third lab-based comparative study examined the advantages of including a dictionary-search feature for ASL learners (hearing university students) for video comprehension, compared to their use of an existing dictionary Web site. Our contributions include:

- We present an interview-based study with ASL learners to explore their experiences while watching challenging ASL videos. Our findings reveal their preferences for viewing videos of different genres from various platforms, identified factors that contribute to difficulties in video comprehension, and discovered the strategies they currently use when facing unfamiliar signs. Using design mock-ups, we also received their feedback on different design parameters that informed our design of a dictionary system.
- We present the first observational study of ASL learners as they translated challenging ASL videos using search technology, specifically a Wizard-of-Oz prototype of our proposed system. Our analysis revealed how users selected sub-spans and performed searches, as well as how users benefited from an integrated tool that presented search results alongside the video, allowing users to check results in context. We found that user behavior varied depending on the genre of the signing video, and we discovered an unexpected use of the sub-span selection tool to constrain the video playhead.
- In the context of ASL learners translating challenging ASL videos, we present the first comparative study between our video-player prototype, which includes an integrated dictionary-search feature, and an existing dictionary Web site that offers search-by-feature functionality. The results of our study highlight benefits including the improved quality of translations produced and a reduced workload for the learners. We also present an analysis of differences in time taken to produce translations across the two conditions and how participants use of a span-selector changes when an integrated search system is provided.

1.1 A Continuing Line of Research

This article is an extended version of a paper originally presented at the 2022 ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22) [31].

In our original ASSETS '22 paper, we presented only the challenges with understanding complex ASL videos and existing workarounds of the participants. We also received an audience question during the conference presentation asking about the process behind the prototype design used in the second study. During the interview study, we collected feedback on various interface design variables for our prototype using mock-up designs (e.g., the location of results in reference to the original video). However, due to limited space, we could not include this feedback in our article. In this TACCESS article, we have included a new sub-section based on a qualitative analysis of the feedback on the prototype's interface design to provide additional insights that cannot be captured by our quantitative data. Based on the findings, we also provide some design recommendations for future researchers in the discussion section.

In our original ASSETS paper, we had combined the observational study on users' behavior of interacting with the integrated dictionary system and the comparative study that recruited new participants who interacted with the baseline to uncover any benefits. In this article, we separated the two studies. In study 2 based on observational findings, we incorporate two more behavioral finding categories that were previously omitted due to space constraints and added more comprehensive details about the prototype design and methodology used. For some existing categories, we included additional data and updated graphs and figures.

We ran two new *post hoc* data analyses that are presented in study 3. Our primary objective was to investigate the impact of the search capability on two specific aspects of video watching and transcription writing: the time required to complete video transcriptions and the participants' span-selection behavior. We aimed to determine whether the integrated search system significantly changed the transcription time, as a prolonged duration would likely deter users from utilizing it. To address this, we compared the time taken to transcribe a video, relative to its total duration, under both the integrated search and baseline conditions. Additionally, we sought to examine whether the presence of the integrated search system influenced participants to utilize the span-selector more frequently. We conducted an analysis of the lengths and durations of sub-spans selected by participants across both conditions. The findings are presented at the end of Study 3.

Finally, we have enhanced the discussion based on the new analyses and findings reported. We also added a sub-section on how our findings inform future researchers in different fields.

1.2 Structure of this Article

The article is structured as follows:

Section 2 provides relevant information about the linguistic challenges associated with learning ASL and the existing sign language dictionary systems. This section sets the foundation for understanding the context of complex ASL video comprehension in which the research is conducted.

Section 3 outlines the set of studies conducted and presents the research questions.

Section 4 reports study 1, which explores the experiences of ASL learners when watching challenging ASL videos and their current workarounds for overcoming difficulties. Additionally, this section introduces fresh findings that were not previously shared in the original ASSETS '22 paper. These findings are on the design feedback received from participants who interacted with low-fidelity prototypes of various pages.

Section 5 elaborates on the design of the integrated search system. Afterward, it provides insights into the usage patterns and benefits observed during an observational study. We expanded on our methodology for the observational study, elaborated on some of the categories of usage patterns, and added two new categories.

Section 6 focuses on study 3, which was originally combined with study 2 in the ASSETS '22 paper. We compare the benefits of their integrated search system with a baseline system, expanding on their findings. We also introduce two sub-sections that discuss the differences in time taken to complete translations and the use of playhead to constrain the video timeline across both conditions.

Section 7 provides a comprehensive discussion of the overall results based on our studies and key takeaways for future researchers.

Section 8 outlines the limitations of the work and presents suggestions for future directions.

Section 9 states the conclusions of the article.

2 Background and Related Work

This section starts with a background on sign language linguistics, introducing essential terminology to familiarize the reader. We then summarize prior work on sign language pedagogy to contextualize our work within that domain. Next, Section 2.2 discusses the current state of sign language lookup technologies to highlight key limitations and drawbacks of the existing resources.

2.1 Sign Language Comprehension

The ASL linguistic phenomena might present difficulties for students when trying to understand an ASL video. Although students can search through dictionaries that show videos of an ASL sign's *citation form*, i.e., the standard way in which a sign may appear when produced in isolation, the appearance may vary when signs are produced during sentences continuously. For instance, sign production can differ naturally among individual signers, which may be influenced by demographic or geographic regional variation. Additionally, two or more ASL signs may linguistically combine into a *compound sign* [50]; novice ASL learners may face challenges segmenting them appropriately to look up meanings in an ASL dictionary [11]. *Coarticulation*, broadly, refers to how in continuous signing, the production of one sign may influence the manner in which other nearby signs are produced [23, 68], e.g., the ending location or handshape of one sign may affect the location or handshape of the next. Coarticulation effects may result in the production of a sign in context to differ from its citation form. Coarticulation effects are also possible when ASL signers produce rapid sequences of handshapes during *fingerspelling* [42] (i.e., when alphabetically spelling words), leading to the fingerspelled word not being a straightforward combination of the individual alphabet handshapes. Lastly, ASL signing encompasses the use of *depiction*, where specific linguistic constructions, commonly known as "classifiers," convey spatial information regarding the position, movement, or shape of entities [73].

The majority of previous research on sign language video comprehension has centered on Deaf users, while there has been limited exploration and no observational studies on the behavior of hearing individuals when facing challenging ASL videos or their strategies for dealing with unfamiliar signs, e.g., using sign lookup tools. In a recent review of prior eye-tracking studies with DHH participants [3], one study was discussed which investigated the differences in gaze patterns between Deaf and hearing individuals when observing a live signer [69]. Although understanding sign language in person is different than watching it in a video format (and no sign lookup technologies had been used), that study [69] conducted an analysis on gaze patterns of hearing individuals when attempting to comprehend sign language. Observational studies that involve eye-tracking or other methods can lead to gaining insights into the behaviors of ASL learners during video comprehension, especially given limited prior work. In contrast, substantial literature exists on non-native learners of different spoken languages and their engagement in video comprehension. Additionally, there has been research on translation tasks involving spoken or written languages, with learners using various electronic resources [5, 20, 33, 54, 72, 79].

2.2 State of Sign Language Look-Up Resources

Dictionaries play an important role in helping second language learners in finding the meaning of unfamiliar words. However, when faced with a sign whose meaning is unknown in sign language, it is more challenging to look up the word, considering the absence of a common writing system in sign language, making it hard for users to use a text search or alphabetical listing to search for a sign [6, 35].

Some sign language dictionaries expect users to remember the linguistic properties of ASL—e.g., hand configuration, orientation, location, movement—of the sign that they are searching for, and then to obtain a list of matching signs, the user must enter these properties into a search-query interface [2, 9, 10, 19, 48, 55, 70, 74]. Unfortunately, prior research has indicated that the search-by-feature systems are challenging for ASL students [9]. Other proposed sign language dictionary systems require users to either submit a video of a single sign extracted from a longer video or physically perform a sign into a webcam [6, 10, 14, 16, 45, 74, 78]; the sign-recognition technology conducts a video search within a dictionary to provide potential matches. However, even if a student manages to remember and perform a sign into a webcam, there are technical obstacles in accurately recognizing signs from video due to various factors [63, 77]. Despite recent advancements [57, 62], state-of-the-art continuous sign-recognition software is still imperfect. To mitigate inaccuracies in the video-to-sign matching process, certain proposed dictionaries offer users post-query filtering options. These options allow users to refine and narrow down the set of results that are returned [28]. Overall, existing dictionary systems face several limitations:

(1) These systems require the ASL learner to recall linguistic properties of the desired ASL sign or accurately perform the sign from memory.
(2) These systems assume that users are able to precisely determine the beginning and end of a sign they come across in a video or during a conversation. Fast signing speed or various linguistic factors (Section 4.2.1) make it difficult for ASL learners to precisely select signs in videos.
(3) In these systems, the user is expected to simultaneously initiate a separate search using a dictionary system while engaging in a video-watching and comprehension task. Expecting a user to use a separate tool for the search may result in users losing the contextual information from the sign language video they were originally watching.

In contrast to prior work, we investigate a dictionary-search system that allows the user to select a span (of potentially multiple signs) from a video of continuous sign language. This selected span serves as the basis for a query in the dictionary system, which then presents potential matching signs in an integrated video player and search results interface. This approach may mitigate the need for users to recall specific linguistic properties of the unknown sign, mitigate the need to identify the specific start/end of signs in a continuous video, and enable the user to remain in the context in their video-watching-and-comprehension task.

Certain recent work has sought design feedback using mock-ups and examined ASL learners interacting with Wizard-of-Oz prototype systems for ASL dictionary search to identify factors that affect users' satisfaction with the system [4, 27, 30]. The Wizard-of-Oz method typically involves some careful planning to give the impression of a fully functioning system, but the user does not know during use, and it is well established as an approach in HCI research [49]. Methodologically, our studies also employ a Wizard-of-Oz prototype of an ASL dictionary-search system to understand users' interaction and potential benefits. However, in those prior studies, users had been shown a stimulus video of a native signer performing a single isolated sign (in citation form), and the user was asked to use a dictionary system to identify the sign's meaning. In contrast, our studies examine how ASL learners engage in a search task while in the midst of a video-watching-and-comprehension task.

3 List of Studies and Research Questions

As discussed above, there has been limited research on the experiences of ASL learners when facing challenging sign-language videos and the potential benefits of ASL dictionary lookup technologies in their educational activities. In addition, no prior work has examined how users might benefit from an integrated tool for viewing ASL videos, with users being able to select spans of the video as the basis for dictionary search. To address these knowledge gaps, we investigate the following research questions:

RQ1 What are the challenges that ASL learners currently experience when trying to understand a difficult sign-language video, what workarounds do they employ, and what are their preferences regarding the design of a future tool that could support their video comprehension?

RQ2 How do users interact with a Wizard-of-Oz prototype for viewing an ASL video and conducting dictionary search on selected spans of video, during the video-watching and comprehension task?

RQ3 In a comparison between the experience of users who used our tool and those who used an existing feature-based ASL-English reverse dictionary, is there a difference between:
(a) translation quality?
(b) time needed?
(c) perceived workload?
(d) usage of span-selector feature?

4 Study 1: Understanding the Experiences of ASL Learners While Watching Sign Language Videos

This article comprises of three studies to investigate the research questions above. Study 1 aimed to understand ASL learners' challenges with video comprehension, their current workarounds, and preferences for the design of a tool that would support their video comprehension. The findings from study 1 provided valuable insights for the study design, videos, and prototype development in studies 2 and 3.

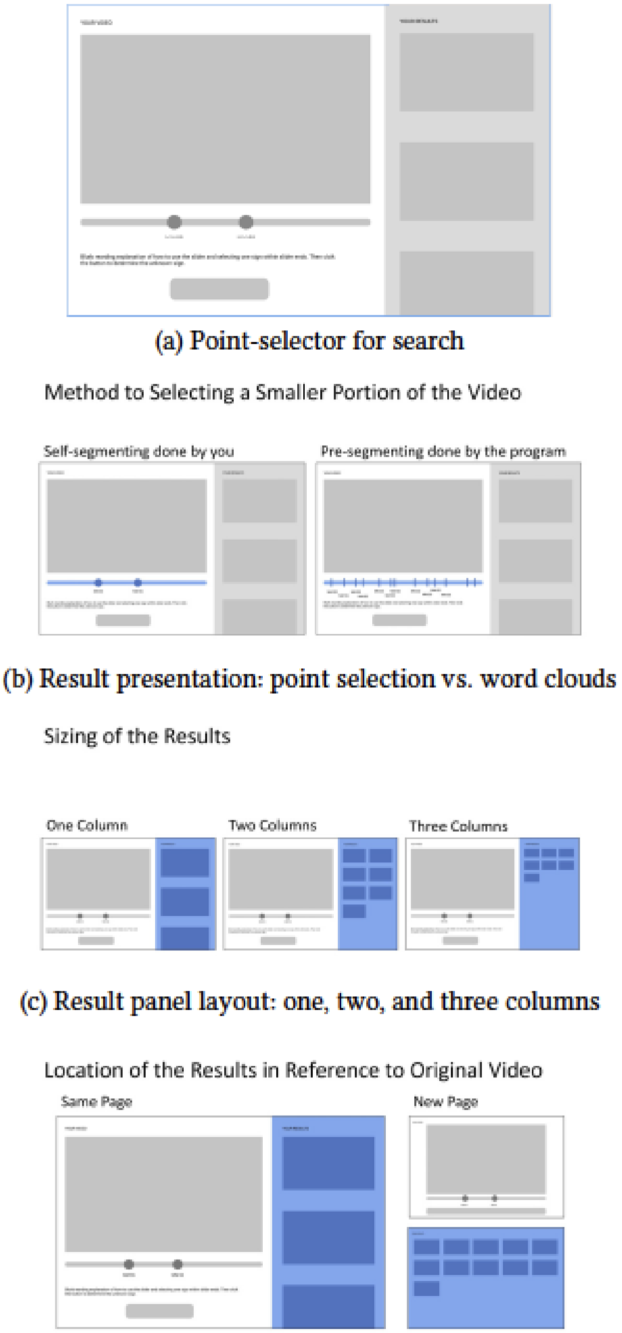
4.1 Study Design

This **Institutional Review Board (IRB)** approved study was conducted either in person or remotely, depending on participants' preferences during the COVID-19 pandemic. We obtained consent from our participants before conducting the study with them. We ran semi-structured interviews, which began with questions about participants' prior experiences watching ASL videos. We also asked the participants about the type of videos they watch, their experiences when they have problems understanding a video, and any workarounds they employ. To provide context for later questions regarding the challenges participants may face in selecting individual signs or spans of multiple signs they do not comprehend, we presented several example videos as a basis for discussion. These videos were taken from advanced ASL or ASL-English interpreting classes, conversational videos between expert signers on YouTube, signing performances at theatres, and interpreted poetry and music. Videos encompassed a range of linguistic phenomena discussed in Section 4.2. Participants were asked to evaluate the difficulty of selecting a sub-span containing one or multiple signs and how they would select a time-range of a video. In the final segment of the interview, we showed the participants some mock-up designs of the system, featuring various variations of the design variables. These mock-up designs were made using Figma and were used to provide context for interview questions. Figure 1 displays four different design variations that were shown to our participants. Notably, we asked about their preferences regarding the design of the span-selector and the presentation of the results.

The average length of each interview was 36.5 minutes.

4.2 Participants and Recruitment

Our participants were recruited through online advertisement on an ASL Reddit channel and by reaching out to professors of introductory ASL courses who could share an advertisement by e-mail to their students. The recruitment contained two screening questions: "Are you currently learning American Sign Language?" or "Have you completed an introductory or intermediate ASL course in the past five years?" We selected participants who responded with yes to at least one of the questions. We recruited a total of 14 participants for our first study (men = 2, women = 11, non-binary = 1). and the median age was 21 years. Participants had studied ASL for a mean of 3.4 years, and all participants confirmed that they had taken fewer than 3 years of formal ASL classes.


Method to Selecting a Smaller Portion of the Video
Self-segmenting done by you
Pre-segmenting done by the program
Sizing of the Results
One Column
Two Columns
Three Columns
Location of the Results in Reference to Original Video
Same Page
New Page

(a) Point-selector for search

(b) Result presentation: point selection vs. word clouds

(c) Result panel layout: one, two, and three columns

(d) Presentation: same page vs. separate page

Fig. 1. Mock-up prototypes of different pages of the integrated search system developed using Figma, demonstrating various design choices in study 1, part 3.

4.3 Analysis and Findings

We employed a mixture of deductive and inductive approaches in our qualitative data analysis. To become familiar with the interview transcripts, 2 authors read all 14, then during a subsequent reading, they individually took notes to produce initial codes, which they collated and collapsed into 2 individual codebooks. Each of the authors then investigated underlying patterns among their codes and formed initial categories, and they consulted the interviewers to get feedback on their initial categories and further improved them. The authors then met to review all of their initial categories, to identify similarities and differences. During two 3-hour meetings, the authors performed an initial thematic grouping, which led to final high-level categories. These high-level categories were then presented to the rest of the team to arrive at a final set of themes and sub-themes presented in this section.

4.3.1 Prior Experiences and Challenges Associated with Watching Signed Video Content. Participants discussed various *motivations* for viewing signing content. Twelve participants mentioned engaging with ASL videos during classroom-related or homework activities. Ten participants also mentioned watching signed content outside of the classroom for their own enrichment or personal exposure to other types of signing, e.g., Deaf theatre or ASL songs. P11 said:

> "It's both in-class, we have different assignments the teacher will give us, and then I also do it on my own time, if I'm looking for a deeper understanding about things, or if I'm looking for specific signs. And I also follow some deaf content creators as well."

Participants also discussed how their lack of familiarity with *regional or dialectical variation* in signing, such as Black ASL [53] used among some African-American signers in the U.S., led to challenges in understanding videos. P5 described their experience in understanding signing among various communities: "I know some white people in the community, [but the] black Deaf community and the interpreter community, I still find hard."

Participants discussed how various *linguistic types of signs* posed comprehension challenges. For instance, P1 described needing to consciously "[...] switch my brain from a sign to actually each letter" when encountering fingerspelling. P11 discussed challenges with "fingerspelling, classifiers, compound words, any of that kind of stuff... fingerspelling is definitely a little tougher for me." Participants discussed how fingerspelled names were challenging to understand in a video, especially when there were multiple individuals with similar names. Participants also described challenges with understanding numbers, e.g., P5 said, "[...] numbers are hard for me, for some reason, I don't know why." Participants also discussed challenges with compound signs, e.g., P13 said "I wasn't sure if that was one or two separate signs. So there were definitely points in the video where they were blending together a little bit, and I wasn't sure."

Overall, participants discussed how *different content sources or genres* pose challenges for comprehension. Participants mentioned viewing signed content on various streaming services, e.g., YouTube and Netflix, as well as on social media, e.g., Instagram and TikTok. P14 said, "I watch ASL videos when I am going through Instagram because I follow some Deaf creators." Participants discussed how the signed content on social media is shorter and more unpredictable in nature, with the topic of the video not always well defined, which poses challenges for comprehension. Participants also discussed how factual signing, e.g., in a documentary, was difficult due to complex vocabulary or increased use of fingerspelling. Other participants mentioned watching videos of ASL poetry and ASL translations of popular songs, contexts in which they described signers as using more depiction and having "their own flow, and they have their own rhythm" (P11). Participants described how videos with multiple signers, e.g., Deaf theatre, pose challenges, as P8 described, "my brain is used to practicing with one signer." Similarly, participants mentioned how natural conversations were difficult to understand, e.g., P6 discussed how signing in such videos tends to be "quicker, and they're a little bit more relaxed."

4.3.2 Workarounds. Our participants also mentioned several workarounds that were useful in understanding challenging signed video content. For instance, several participants discussed using the *context* of a video to understand unknown signs. Participants would consider the description or title of the video, such as on YouTube, or they would consider what was said before or after any unknown signs. For instance, P3 described a situation in which they figured out the sign for a citrus fruit by considering the context of the surrounding signing, which had mentioned lemons. P3 discussed how understanding later signing may clarify a portion of signing that had not been previously understood, explaining how if they become confused, then they "really focus on the next thing they're saying, so I can piece together what they might have said, so I can understand it." Participants discussed various strategies that involved controlling the flow of the video player:

- *Periodically pausing* was a strategy common across several participants. For instance, participants discussed how they paused videos in-between conversational turns in videos with multiple signers; P8 described how they "pause in between each speaker... just enough time to grasp" what had been said.
- *Backtracking and replaying* was another common approach, as P12 explained, "pausing it and replaying it." P3 also discussed how they will "backtrack the video" if needed.
- *Slowing down the video* was also popular, if possible within the video player. For instance, P7 explained how they will "slow down the fingerspelling if...it's on YouTube. If I could alter the speed, I might try to slow it down."

When asked about their current strategies for seeking the meaning of an unknown sign, six participants mentioned using *English-to-ASL dictionaries*, i.e., guessing English meanings of the sign they did not understand to look up that English word in the dictionary to see if the sign displayed visually matched the sign that had not been understood. Participants also mentioned using some ASL-to-English "reverse" dictionaries, i.e., Web sites that allow someone to enter linguistic properties to search for the English translation of an ASL sign. Participants discussed challenges associated with using such tools, e.g., P7 said, "if I think I have an idea of what the sign is, I might use Handspeak, or there's another one I use... [It's] hard to specify handshape in current dictionaries." P6 discussed challenges in entering the various linguistic properties when constructing a query to search for a sign in such systems:

> "I definitely tried using the reverse dictionary stuff online. Usually it doesn't end up being successful, and I have to just end up moving on. Because, the way it's structured, you have the handshape, and the movement, and the location. Sometimes it's a little ambiguous, especially if you don't actually know what that sign is; so, it's hard to end up looking [it] up."

Participants also expressed their frustration with having to launch a web-dictionary in another window while trying to understand a video. P11 said:

> "It's pretty frustrating sometimes when I'm trying to find a specific sign, I have to like go to Google and...then go through all the different pages... If I could just scroll and have the source material right there, I think it would be much more efficient."

Rather than use a specific dictionary Web site, other participants mentioned typing descriptions of what a sign looked like into a *Google search*, e.g., P11 said:

> "I've definitely tried to Google it before, but it's so hard to sometimes describe what it is that you're looking for. I end up being very vague... It's very rare that I go to Google and find what I'm looking for as far as trying to describe a sign."

Finally, several participants mentioned that, if other people are available, they may *ask a teacher or a peer*. As P09 said, "If I'm in class I would ask the teacher. If it's for a class I would either look it up online or if it's in a vocabulary learning unit."

4.3.3 *Expectations from System Design.*

To facilitate discussions on the design of a potential tool to help ASL learners with understanding complex ASL videos, we showed participants a series of images with different design configurations for the video panel, span selector, and results (see Figure 1).

Five participants expressed a preference for undertaking the video segmentation process, particularly in segments they found challenging or unclear. Most of them expressed a preference for on-demand searching when encountering unfamiliar signs within those segments. One participant (P6) highlighted this preference, stating: "I think I would still probably prefer doing [segmenting] myself. Just kind of filter out the parts that I did understand versus I didn't understand. I guess it'd be a little bit easier to have that kind of control."

Only two participants expressed interest in an AI-powered system that automatically segments the video and provides the words contained in each segment. P2 expressed their interest in automatic segmentation and word search for difficult videos: "For harder videos, maybe the pre-segmented approach would work better. And for easier videos, I feel like doing your own segmenting would be better."

Two participants also mentioned that the preference for pre-segmented videos versus self-segmentation would depend on the learner's skills. P11 stated:

> "I think it will be more tedious maybe to be able to drag and select myself... I think I would prefer that [doing the segmentation on their own] because it gives me control. It might just be on one side, and I'm not just missing an entire phrase. So, I think that would be my preference—to be able to select it myself and focus on specific points because my level of ASL is going to be totally different than somebody else's. Being able to pick out specifically what applies to me would be really helpful."

We inquired with the participants who expressed interest in searching selected spans regarding their preferred method of specifying an input, such as choosing a span or utilizing a point selector (see Figure 1(a) and (b)). The majority of participants indicated a clear preference for utilizing a span-selector to focus on a segment of the video while watching and searching, rather than opting for a point selector. P8 shared their perspective, stating, "Probably the chunk, and I feel like I would have a hard time being able to pause the video in a spot that would be helpful to me."

We asked participants whether they would prefer the results to be presented on the same page as the video or on a separate page (see Figure 1(c) and (d)). The majority of participants expressed a strong interest in having the results accessible to them while watching the video, alluding to the benefits of a unified viewing and searching experience. Participant P14 articulated this sentiment, stating:

> "Yeah, it's kind of like having everything on the same page because with a new page, if it's a sign you're not familiar with, looking at all these other signs can be confusing. I feel like I'm getting mixed up, trying to remember how it was signed. I'd probably go back and forth, comparing, when I could just look at it simultaneously."

One participant who expressed a preference for viewing the results on a separate window also conveyed a desire to have both windows open simultaneously. This emphasizes the convenience and efficiency of accessing related resources without constantly switching between different pages or sources:

> "I like having another page open because that way I can still have the original video up while simultaneously viewing a page with the results. Yeah, I prefer having both open. I don't really like having to remember things and then navigate back to the video, losing the source material."

We also inquired about participants preferences' regarding the structure of the results bar or page. A significant majority of participants expressed a preference for a two-column layout of the results, highlighting its advantages in terms of faster visual browsing and facilitating cross-comparison. P14 stated, "The 2 columns, because you can kind of see it all at once, which is useful." Participants also expressed an interest in having linguistic details accompanying each result. One participant mentioned, "If there were parameters, I like the details below..." This indicates a desire for additional linguistic information to provide context and enhance the understanding of the results. A minority of participants expressed a preference for displaying the results on-demand using a toggle button. P11 explained, "In the second video where I'm missing, you know, some like the numbers or if I'm missing, you know, like where are we in this story, what's the context, it might be helpful to be able to have that toggle."

Finally, at least six participants mentioned YouTube's flexible design options as a reference for their preferences, for example:

> "Oh, kind of like how on YouTube you could choose to have an autoplay or not, like, you know, play the next video, that kind of thing where it's like you choose what setting, I guess, to keep... I think if you had the choice to have that on or not based on what they need, I think that's a really good idea." - P3

In general, participants appreciated the ability to customize settings, such as autoplay, and expressed a desire for similar flexibility in the snippet and result presentation design.

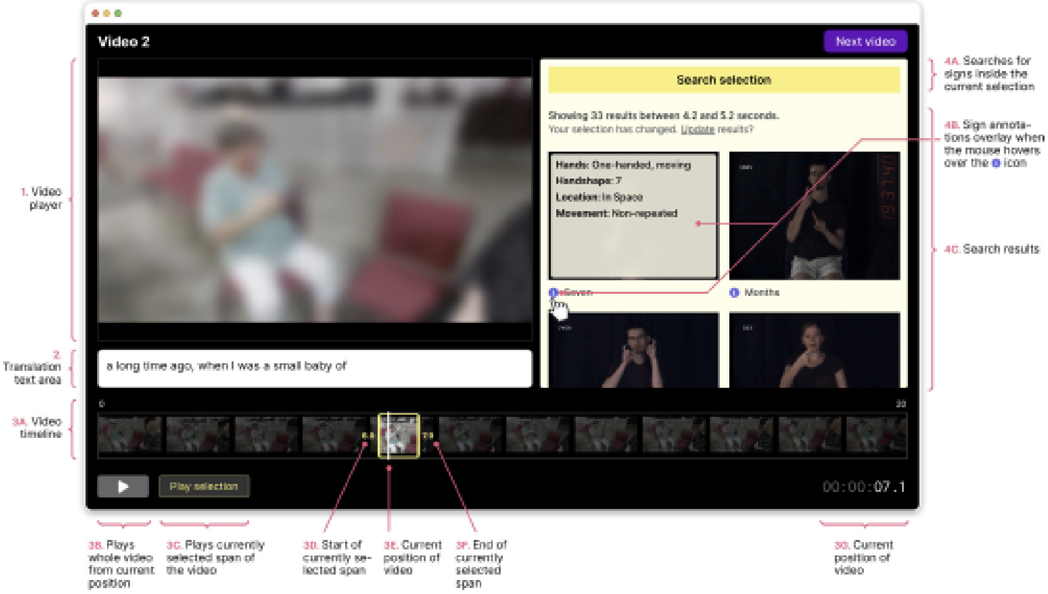
5 Study 2: Observational Study of ASL Learner Interactions with an Integrated-Search Dictionary

Study 1 findings highlighted the need for a tool that integrates video-playing and sign-searching capabilities, while also revealing participants' preferences for such a tool. To explore how users would interact with this tool, we conducted a second study using a Wizard-of-Oz prototype. The Wizard-of-Oz method typically involves some careful planning to give the impression of a fully functioning system, but the user does not know during use, and it is well established as an approach in HCI research [49]. This prototype ran on a desktop computer in a controlled laboratory environment. The insights gained from study 1 not only influenced the prototype's design but also guided the selection of videos for study 2, as described in the following section.

5.1 Prototype Design

Given that this study focused on users' interaction and behavior, the prototype was designed (Figure 2) so that its underlying sign-recognition was simulated through a Wizard-of-Oz approach, without recurring to actual automatic video analysis.

The steps of the experiment are shown in Figure 3. First, participants entered an identification number. Next, they were provided with an interface to ensure that the size and aspect ratio of their browser window remained consistent. For the third step, as depicted in Figure 2, participants


Video 2
Next video
Search selection
Play selection
00:00:07.1
0
33
Seven
Months
1. Video player
2. Translation text area
3A. Video timeline
3B. Plays whole video from current position
3C. Plays currently selected span of the video
3D. Start of currently selected span
3E. Current position of video
3F. End of currently selected span
3G. Current position of video
4A. Searches for signs inside the current selection
4B. Sign annotations overlay when the mouse hovers over the ● icon
4C. Search results

Showing 33 results between 4.2 and 5.2 seconds.
Your selection has changed. Update results?

Hands: One-handed, moving
Handshape: 7
Location: In Space
Movement: Non-repeated

a long time ago, when I was a small baby of

Fig. 2. Screen shot of the prototype with labels identifying the different interface elements, including (1) the video player at the top left, (2) the text box for inputting the video's translation below, (3) a video timeline with a span-selection interface along the bottom, and (4) a dictionary-search panel on the right.

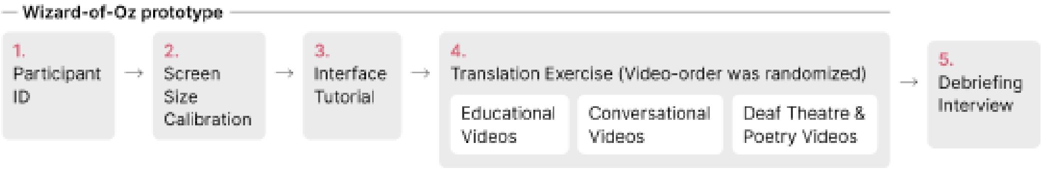


Fig. 3. Diagram illustrating the sequence of steps taken by participants in both studies 2 and 3.

accessed the experiment's main interface, both for an initial tutorial and for the translation exercise, which accounted for most of the experiment. Lastly, there was a debriefing interview.

Within the prototype's interface, a large ASL video could be controlled using a play/pause button located at the bottom-left corner of the screen. The current position in the video was indicated by a white line, known as the playhead, which users could manipulate by clicking on the timeline.

The timeline also included a selection span, represented by a rectangle with yellow edges, that highlighted the desired portion of the video. Initially, when users first entered the page, the video span encompassed the entire video, but users had the flexibility to increase or decrease it at will by dragging its edges. Additionally, they could change the position of the selection but not its duration by dragging the span. A "Play selection" button was available to play the portion of the video that fell within the selected span.

On the top-right corner, a "Search selection" button allowed users to search for signs within the currently selected span. The search results were displayed in a scrollable window located just below the button. Each result included a video featuring a sign from the **American Sign Language Lexicon Video Dataset** (ASLLVD) [60], along with a label indicating its closest English gloss. Users could access more information about each result by clicking on a "more information" icon, which displayed linguistic properties.

An important finding from the first study was that participants experienced frustration when switching between watching a video and accessing electronic dictionaries in a separate context. Additionally, during the third phase of interviews, the majority of participants expressed a strong preference for viewing the results on the same page. They specifically emphasized their interest in a two-column bar design, with the inclusion of linguistic metadata beneath each result. This feedback directly informed our design decisions regarding the presentation of the results bar. Moreover, participants' insights into the workaround strategies they employed when faced with challenging videos, such as replaying and backtracking, inspired us to implement the "Play selection" button. This feature allows users to replay a specific segment of the video by constraining the playhead to their selected span.

Study 1 also informed which videos were used in study 2. We took into account participants' comments about how video genre and linguistic phenomena are related to its difficulty and thus chose videos from three genres: educational videos, conversational videos involving multiple signers with turn-taking, and videos showcasing Deaf theatre and poetry. Considering participants' specific challenges with fingerspelling and compound signs, we made it a point to include instances of these signing types in the chosen videos. Moreover, we addressed participants' comments on conversational signing and regional dialects by selecting videos that covered these aspects.

The nine videos selected—three for each genre plus an educational video used for the initial tutorial—were sourced from online platforms and fourth-year ASL interpreting courses. Videos had an average duration of 23.7 seconds. Details about each can be found in an external electronic resource at https://saadh.info/a11y/publication/hassan-taccess-24/.

5.1.1 Selection of Signs Appearing in the Results List. We adopted a Wizard-of-Oz methodology to simulate an automated search-recognition system rather than spend time on technical development so that we could focus on understanding questions related to participants' behavior and interaction with the interface. The choice of signs included in the results was determined through prior video pre-processing. The protocol outlining this procedure is described below:

(1) One of the authors, who is Deaf and has native ASL fluency, watched all of the 10 videos, identified each video's sequence of signs, along with their corresponding starting and ending timestamps.
(2) For each sign, a set of dictionary-search results was manually prepared to simulate the experience of using an actual automatic dictionary-search system. Specifically, one of the authors was responsible for selecting 11 signs that closely matched the target sign in terms of appearance using the ASLLVD collection [60]. Their aim was to find signs that shared as many properties as possible with the target sign, including handshape, number of hands, movement, and location.
(3) During the experiment, when a span of video was selected, there was a chance that its start and end boundaries would not perfectly overlap the timestamps of any given sign. For the purposes of simulating the search, we established that the system would consider a sign as if within the span if the selection span covered at least half of its total duration.
(4) Additionally, the displayed list of dictionary-search results was created by combining the match lists from all the signs within that span,[1] as long as they adhered to the rule above, and as follows: the top item from a randomly selected match-list was taken, without replacement, and added to the final-results list. This first sign was correct, but as the process is repeated and the match-lists empties, then the chances of selection of an incorrect sign increases. The

[1] Note that a selection span may include more than one sign.

process was repeated until the final-results list had 50 items or until all individual match-lists were emptied.

5.2 Study Design and Analysis Plan

Participants followed an IRB-approved study protocol. After providing their informed consent to participate, they were asked to give an ASL-to-English translation for a video while using a video-playing and sign-searching prototype. At the start of the study, a researcher used a sample video to demonstrate how the prototype worked. Once they felt they understood the prototype, participants could start working on translating the nine videos that followed. Their interactions were recorded as follows:

(1) The prototype's backend software automatically recorded the major user interactions with its interface, including changes to the start and end points of every span the participant selected on the video's timeline, the number of signs withing each of these spans, the moments the search function was triggered, and the final text of the English translation typed by the participant.
(2) A Tobii Nano [59] 60 Hz screen-based eye-tracking device was attached to a 19-inch monitor. Participants' faces were at a distance of approximately 65 cm, and their gaze direction was recorded using the iMotions (v9.1) [37] software.
(3) One of the authors, who was a fourth-year English-ASL interpreting student at the university at the time of the experiment, sat 2 meters away from the participant, taking observational notes during the course of the experiment. The participants' gaze movements across the user interface of the prototype were displayed in real time on a secondary monitor, visible only to the researcher.

After the participant completed the translation of the nine videos, a debriefing interview was conducted to gather their impressions of the system, their interactions with it, and any other recommendations they had. The data from these interviews were transcribed and coded using the same methodology as in study 1.

Two of the authors conducted a qualitative analysis of the aforementioned data. Their objective was to identify typical sequences of interaction behavior for each video session. To achieve this, they reviewed screen and eye-gaze recordings, plotted eye-gaze patterns, analyzed data captured by the software prototype, and examined the observer's notes. After coding the data, the two researchers discussed their findings and reached a consensus on a set of categories for the observed behaviors, as presented in Section 5.4.[2]

5.3 Participants and Recruitment

For our second study, we enlisted the participation of eight ASL students. The recruitment criteria and methods employed were consistent with those used in the first study. The participants had a median age of 20, comprising of 7 women and 1 non-binary individual. On average, the participants had studied ASL for 3.5 years, and they all affirmed that they had not taken more than 3 years of formal ASL classes.[3] None of the participants from the first study were included in the second study.

5.4 Findings

5.4.1 Using the Span Selection to Constrain the Playhead. To support incremental progress through the videos, six participants used the span selection tool, which allowed them to restrict

[2] After viewing and translating each video, participants' English translation texts were saved, and participants completed a NASA TLX [25, 26]. NASA-TLX is a task load index created by NASA to capture a subjective score for participants' perception of experienced workload. These two sets of data were retained for later analysis as part of a "Study 3" described in Section 6.
[3] Years of studying ASL included independent learning as well.

the portion of the video played at any given time. As they typed the English translation text, participants gradually selected video spans. On average, the selected spans had a duration of 11.43 seconds. The trend of span selection over time is depicted in Figure 4(a), where the positions of the selected spans are plotted. The initial span selected, referred to as Span 1, is shown at the bottom of the image.

The comments from the debriefing interviews further reinforce this observation. P7 gave insight into their approach, explaining that they selected a specific span width and then dragged it across the video, which allowed them to progressively watch different portions. They expressed satisfaction with this feature, stating, "I like how you can just maintain the length and you just drag it over so you're getting the same length of a chunk of the video; that was easy to use."

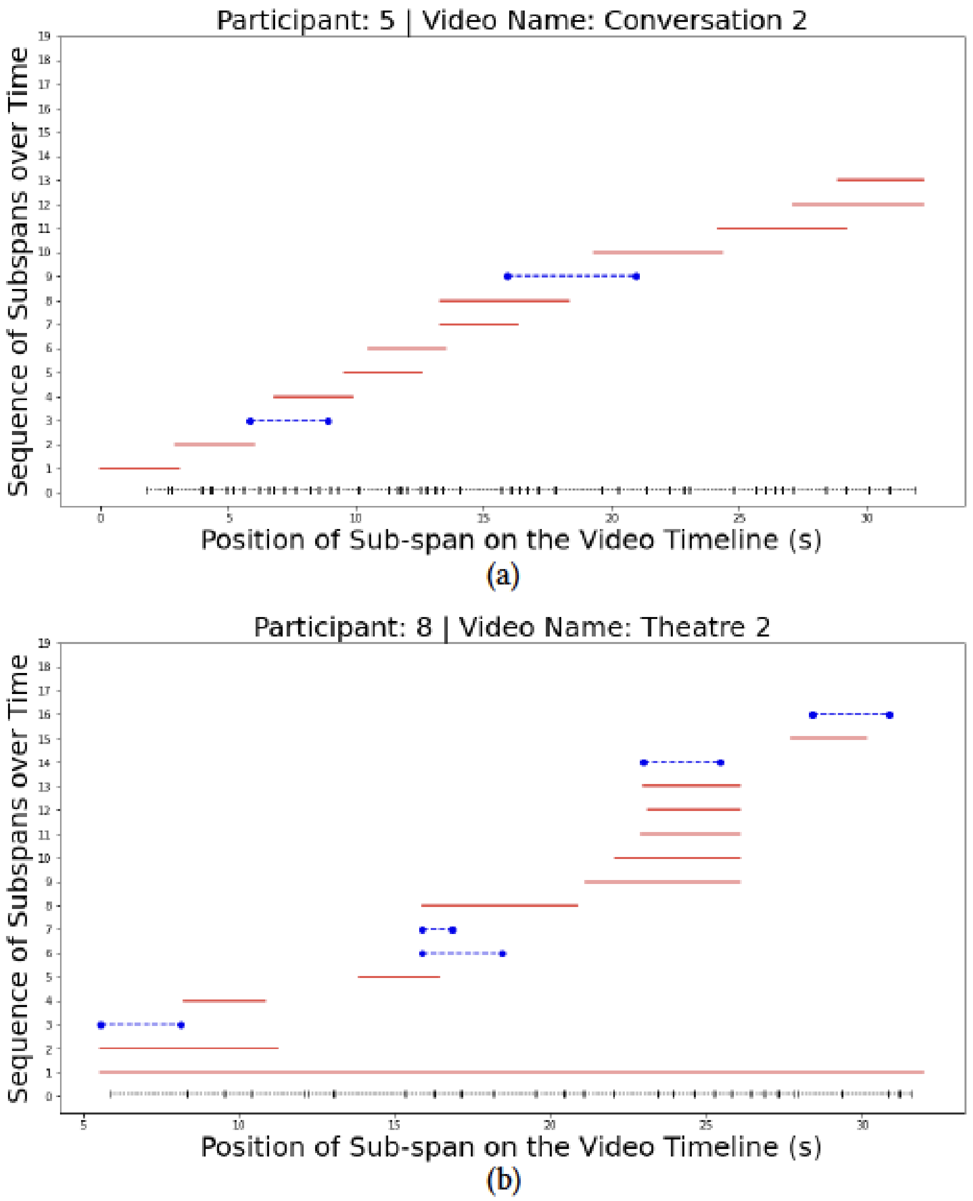
5.4.2 Dragging One of the Span Selection Boundaries Back and Forth. While the majority of participants predominantly used span-selections to browse through the video, at least three participants also used a different approach. These participants frequently used a single boundary of the span-selector, treating it as a point selector to navigate the video. For instance, participant P8 repeatedly adjusted the starting point of the span-selector to closely examine a sign. This behavior is demonstrated in Figure 4(b) between the span sequence 9 and 13. Similarly, participant P2, who exhibited this behavior, remarked, "I've just been moving the cursors [span boundaries] and then just replaying—that was very, very easy."

5.4.3 Using the Span Selection to Constrain a Sentence or a Context Window. Five participants used the span selection to focus on specific context windows that they found more challenging. Figure 5 shows P6 focusing on two different context windows. During the interview, they said: "I chose based on context clues, and like hope for the best. The interface was helpful in understanding the context." Similarly, P7 also employed the span selector to capture contextual information. They said, "I had selected a broader one because... I was more just looking for the meaning of what was being signed rather than like a certain sign or a certain word that I didn't know."

5.4.4 Approaching Task Linearly, Sometimes after Initial Overview. For most videos, participants followed a linear approach while viewing the videos, generating transcripts as they watched concise video segments, typically consisting of single sentences. In some cases, participants first watched the complete video before returning to the beginning and sequentially viewing shorter video segments. This linear progression can be visualized in Figure 4(b), where a span covering the entire video is watched before the subsequent selection of shorter progressive spans.

5.4.5 Using Dictionary Search to Inform Translation. Out of the 72 video sessions, participants made use of the dictionary-search feature in 62 sessions to look up the meanings of unknown signs in the videos. The search tool's output, illustrated in Figure 6, informed participants' translation decisions as they progressed through the video in a linear manner. During the debriefing interview, P6 expressed how the tool helped them, stating, "I knew what he was saying in general, I just couldn't think of the exact English words and that one came up right away." P7 highlighted the tool's advantages in cases of rapid signing, mentioning, "It was definitely useful, especially when the signers were going really fast because then I could double check to make sure that what I thought I saw was actually what I saw."

5.4.6 Gradually Making the Span Shorter Prior to Search. When faced with challenging video sections, participants frequently reduced the size of their selected span before running a dictionary search. In cases where participants chose a span solely for viewing purposes (without conducting a search), the average span width was 8.17 seconds (equivalent to 10.83 signs). However, spans immediately preceding the search had an average duration of 2.33 seconds (equivalent to 3.25 signs).



(a)

(b)

Fig. 4. In the charts above, horizontal bars indicate spans selected with respect to the total duration of the video. Spans in dotted blue lines are those for which the participant used the dictionary-search function. Black lines at the bottom delineate the start and end moments of the actual signs on the video timeline. In (a), P5 watched the video linearly, selecting spans to progressively view short video segments. In (b), P8 initially selected and watched the entire video and then proceeded to reduce the span to a shorter duration, progressively moving it forward along the video.

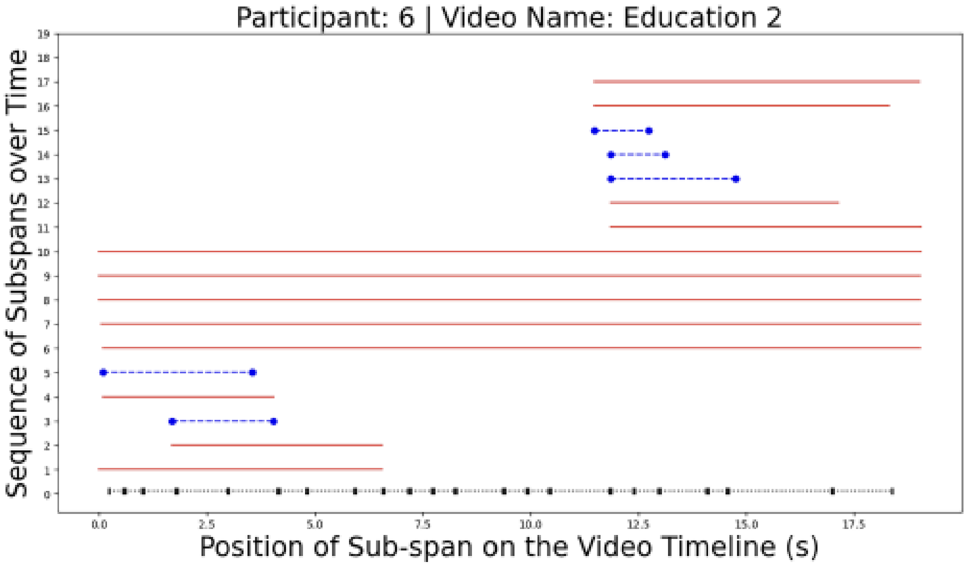


Fig. 5. P6 utilizing the constrained playhead to focus on a specific segment corresponding to a sentence. They proceeded to watch the entire video multiple times, writing the translation, and eventually shifted their attention to a different sentence.

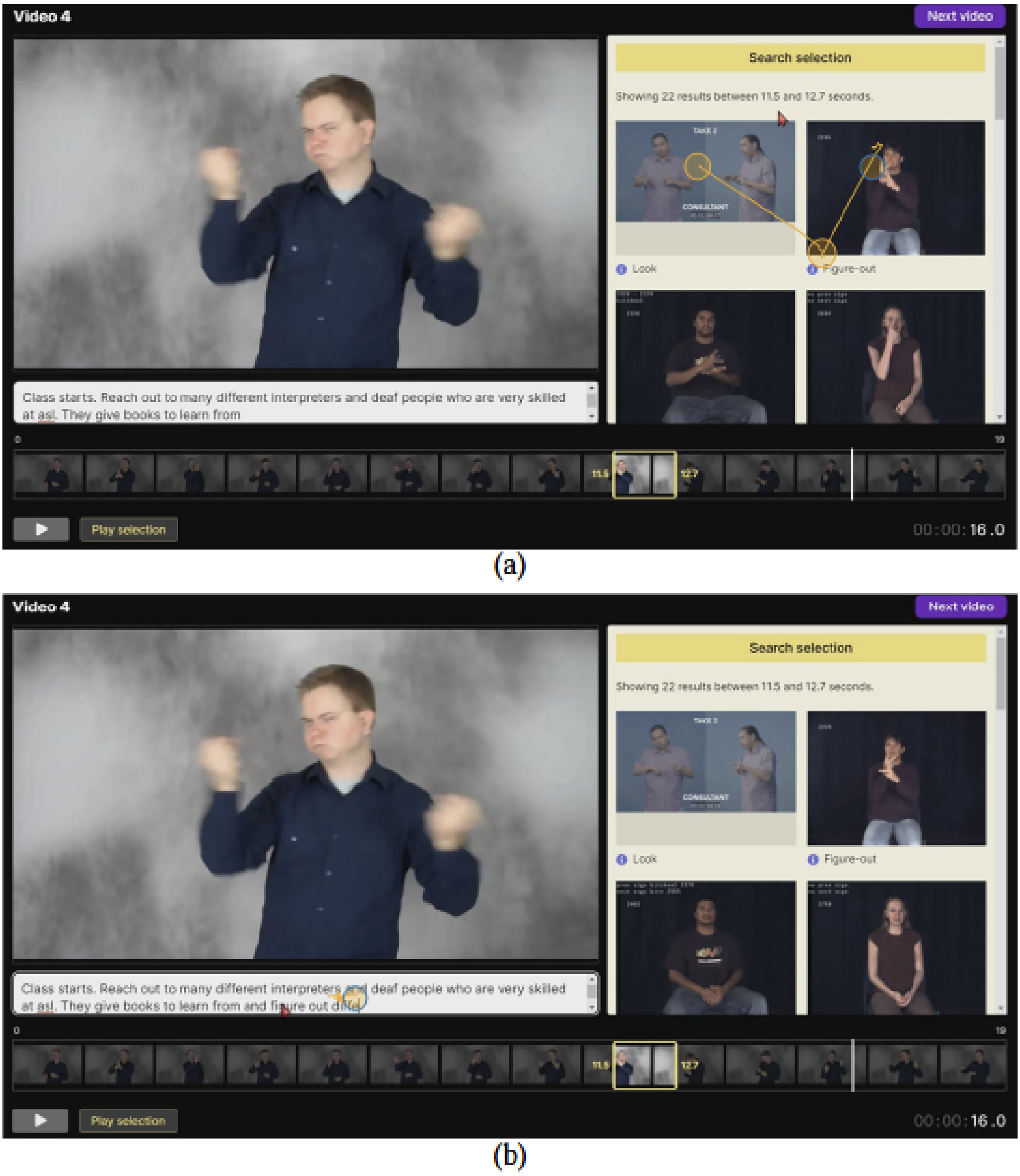
If the search results did not provide sufficient clarity regarding the meaning of signs within a specific span, some participants progressively narrowed the width of their spans to pinpoint the precise video portion that confused them. This iterative process involved running additional dictionary searches, as illustrated in Figure 7(a). While adjusting spans, participants' gaze alternated between the main video region and the span-selection control, as depicted in Figure 7(b).

During the debriefing interviews, participants shared their strategies of initially watching longer video segments to grasp the overall context before focusing on narrower spans that posed comprehension challenges. P5 described their approach as "narrowing it down and then pressing search." On the other hand, some participants discussed the advantages of starting with a search encompassing a wider span. For instance, P7 elaborated on this approach, stating:

> "If I was just trying to get the general idea of a section, it was helpful that sometimes there were more signs in the up results besides like the specific signs that I had selected because it gave more context and it was easier to understand."

5.4.7 Using Dictionary Search to Confirm Results after Initial Translation. Out of the 62 video sessions where participants utilized the dictionary-search tool, we noted that in 40 cases, participants employed the tool after completing a full translation of the entire video. As depicted in Figure 8, participants engaged in a process where, after writing a translation for the entire video, they reviewed specific earlier segments and utilized the search tool to confirm their comprehension of particular signs. This allowed them to ensure the accuracy of their understanding of those specific sections.

During the debriefing interviews, participants shared various ways in which the search tool proved helpful in achieving a more accurate translation. For example, P7 mentioned how the search results served as motivation to refine their wording in the translation, stating, "I would go back and use the tool to make my translation more precise, I guess, so I could fix the sentences and the



Showing 22 results between 11.5 and 12.7 seconds.

Class starts. Reach out to many different interpreters and deaf people who are very skilled at asl. They give books to learn from

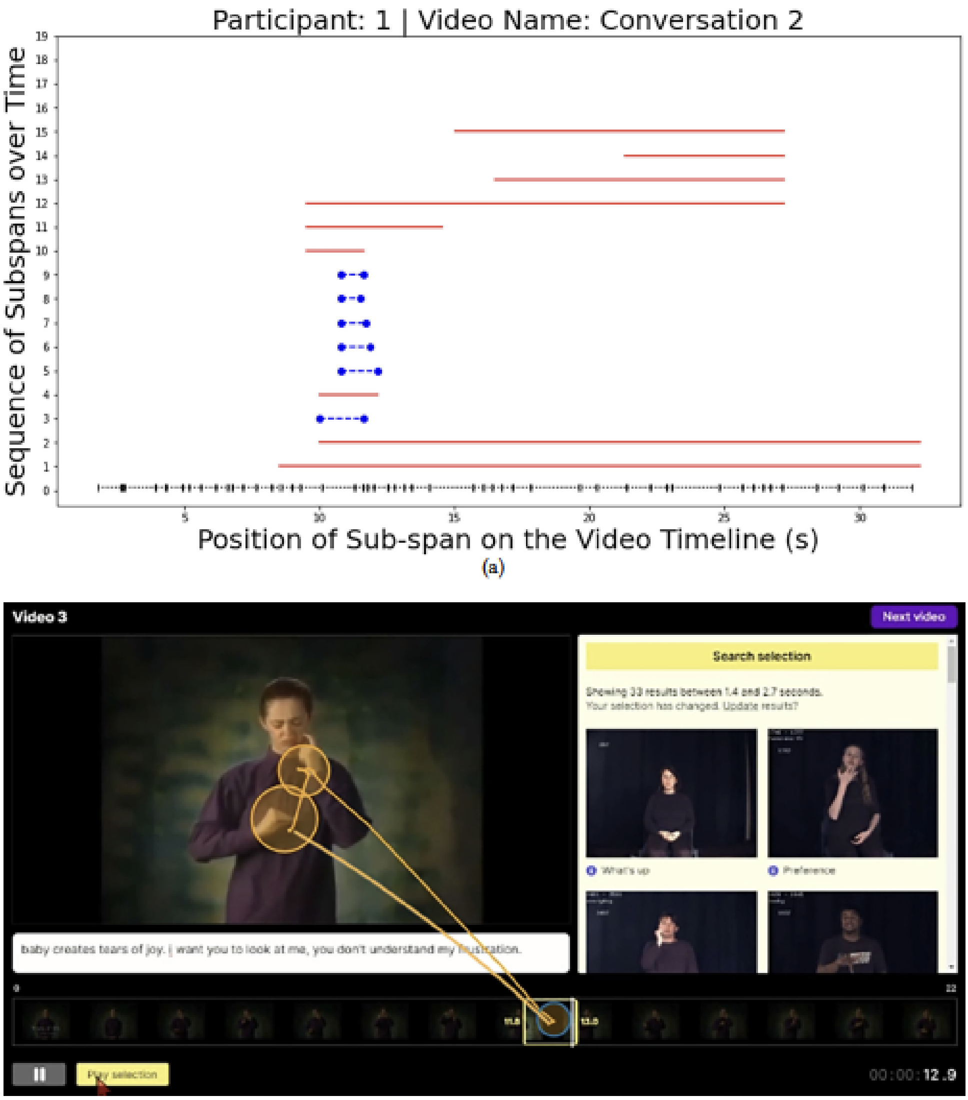
Showing 22 results between 11.5 and 12.7 seconds.

Class starts. Reach out to many different interpreters and deaf people who are very skilled at asl. They give books to learn from and figure out diffe

Fig. 6. Eye-gaze patterns from P5's use of the tool to create a translation, first (a) browsing the list of search results given for a selection span and then, after identifying the meaning of the sign FIGURE-OUT, (b) typing into the translation text box to continue the sentence: "They give books to learn from and figure out..."

wording." At times, the search results simply boosted their confidence in their understanding of the video, as P7 expressed, "sometimes that helped to confirm what I thought I saw."

5.4.8 *Participants Struggled to Find a Sign If a Different Version Was Being Signed in the Results.* On several occasions, participants encountered confusion when the citation form of a sign displayed in the search results did not align with the variation of the sign observed in the main video,



Showing 33 results between 1.4 and 2.7 seconds.
Your selection has changed. Update results?

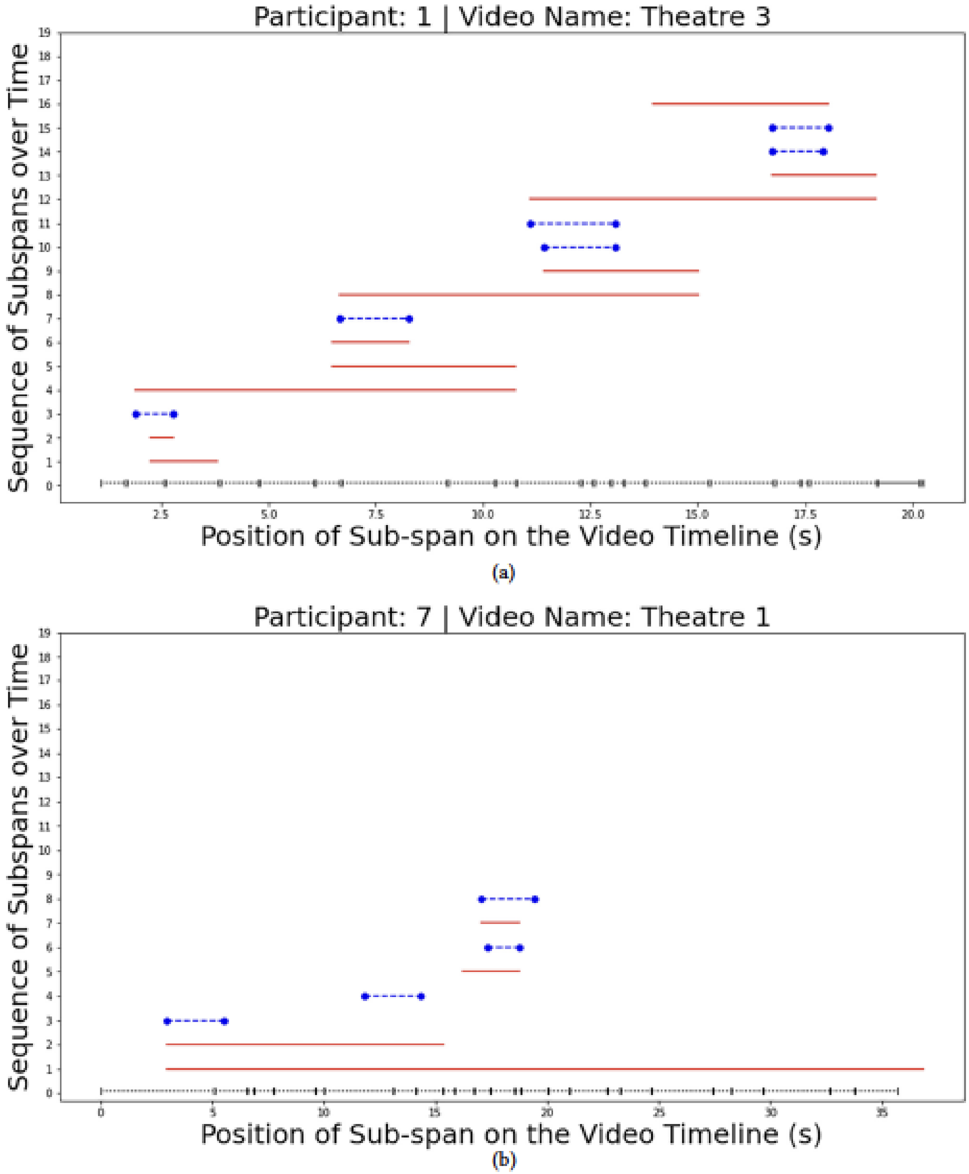
baby creates tears of joy. i want you to look at me, you don't understand my frustration.

(a)

(b)

Fig. 7. P1 repeatedly adjusted the width of a span over time, with search requests between each adjustment. (a) Horizontal blue dotted lines show spans for which a dictionary search was made, while red lines show spans that had no associated search request. Notably, spans 7 to 10 gradually decreased in width as time progressed. In (b), and as they fine-tune a span's width, P8's eye-gaze moves between the video and timeline regions, as can be seen by the yellow lines connecting both regions.

particularly in cases involving compound signs and depictions. In Figure 9, P5 conducted a search; however, the specific appearance of the sign in the video differed from the citation form presented in the dictionary-search results. As a result, the participant engaged in back-and-forth glancing between the video and the search results to make comparisons. Ultimately, this discrepancy led to an incorrect translation for that particular segment of the video, suggesting that the participant did not realize there was a match between the observed sign and its citation form.

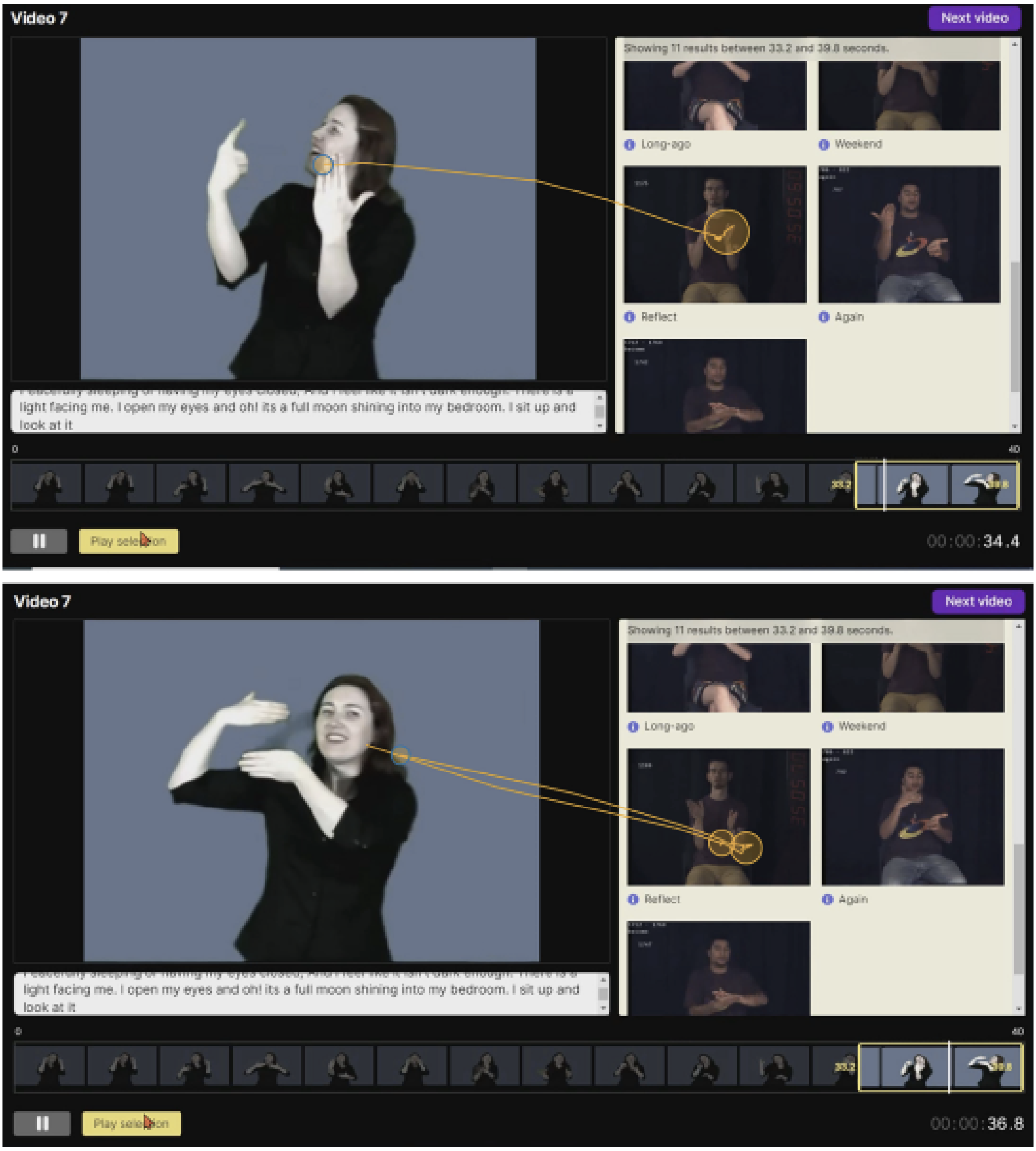


(a)

(b)

Fig. 8. (a) P1 watches video segments sequentially, rewatching some sub-segments, and using constrained playhead for targeted searches. (b) After completing an English translation for the entire video, P7 revisited earlier sections and ran dictionary searches to verify the accuracy of their translation for specific segments.

During the debriefing interviews, participants highlighted the challenges they faced when attempting to match dictionary-search results to signs in the video, particularly in certain video genres. For example, P4 discussed the difficulty of aligning a sign produced with strong emotions, such as in a theater video, with a dictionary-search result that had a more neutral effect. They described their struggle when encountering a video where the signer "was showing emotion and



light facing me. I open my eyes and oh! its a full moon shining into my bedroom. I sit up and look at it

Fig. 9. P5's eye-gaze pattern, illustrated by the yellow lines, reveals their alternating focus between the dictionary result for the sign REFLECT and the corresponding sign in the video, which had been executed differently. This caused P5 to mistakenly perceive these as distinct signs, resulting in an erroneous translation.

then you would go in the searches and they wouldn't. So it's like, I guess you can get mixed up about the emotion."

This led to some participants suggesting potential improvements to the dictionary-search system. They recommended incorporating dialectical variations of each sign result and providing examples of each sign's usage within sentences. P3 expressed the desire for the system to showcase the sign in a context that included more facial expressions or within a sentence, similar to what is available in other dictionaries.

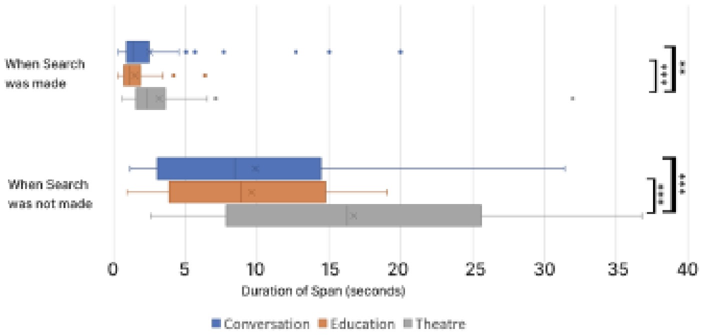


Fig. 10. Box and whisker plots showing the span widths for each video genre in two scenarios: spans selected immediately before a search (top of graph) and spans for which no search was run (bottom of graph). The plots demonstrate that participants consistently opted for wider spans when dealing with videos from the theater genre. The vertical bars within each box indicate the median, while the "x" represents the mean, and outliers are represented by dots. Significant pairwise differences are denoted by asterisks: ** if $p < 0.01$ and $***$ if $p < 0.001$.

5.4.9 Differences in Span Selections across Genres of Videos. As previously mentioned, the study included a total of nine videos from three different genres: natural conversations, educational videos, and theatre/poetry performances. Analysis of the span-selection data captured by the prototype showed that participants tended to choose wider spans when interacting with theatre videos, as depicted in Figure 10.

The average duration of span selected across three video genres when participants did not use the search feature were: natural conversation (M = 9.86s, σ = 8.00s), educational videos (M = 9.61s, σ = 5.88s), and theatre/poetry performances (M = 16.74s, σ = 9.60s). To further analyze this observation, a Kruskal-Wallis test was conducted, revealing a significant effect [H(2) = 24.28, $p < 0.00001$, η^2 = 0.043 (small effect)]. Subsequent Mann-Whitney *post hoc* tests, with Bonferroni corrections applied, confirmed that participants selected wider spans for theatre videos when choosing a span solely to constrain the playhead for video playback.

Similarly, when selecting a span for the purpose of performing a search, the average duration of span selected across three video genres were: natural conversation (M = 2.50s, σ = 3.29s), educational videos (M = 1.50s, σ = 1.09s), and theatre/poetry performances (M = 3.18s, σ = 4.15s). Kruskal-Wallis test [H(2) = 24.3031, $p < 0.00001$, $\eta^2 = 0.12$ (medium effect)] indicated that participants also opted for wider spans when dealing with theatre videos.

During the debriefing interviews, participants provided insights into their reasons for selecting spans of different widths for different genres. P6 expressed the challenge of understanding theatre/poetry videos, noting that "This one had more of a poetic meaning and display; so, it did take more focus to understand it for the translation." Furthermore, P5 discussed their approach to selecting spans while watching conversational videos compared to theatre/poetry performances:

> "The conversational ones, when I chose these subspans were shorter, because they go back and forth a lot [between signers]. But the poetry ones I feel are more conceptual... so you can watch longer pieces. You don't need to cut it down."

P7 explained that the width of the selected span varied depending on the overall signing pace of the video. However, they noted that theatre/poetry videos generally required longer spans due to their distinctive style of signing:

> "Some of the ones that were very visual, like the mushroom one and the moon one and all those ones that were ASL storytelling type of very figurative language... It's typically slower paced, and sometimes there's a lot of repeated signs. Or there's a lot of just a depiction that's very visual and doesn't have a lot of strictly vocabulary to go with it, but it's more classifiers. I found that I would sometimes need a longer chunk in order to use the tool and actually get relevant results of what was being signed."

Participants had the freedom to select spans that did not precisely align with the boundaries of individual signs, although more accurate span selection would result in more precise dictionary-search results. An analysis of the mean error in terms of seconds between each span selection boundary and the nearest actual sign boundary revealed that participants showed less accuracy when selecting spans during theatre/poetry videos. The mean error for natural conversation videos was 0.19 seconds, for educational videos was 0.23 seconds, and for theatre/poetry videos was 0.55 seconds.

A Kruskal-Wallis test [H(2) = 5.2174, p = 0.02236, $\eta^2 = 0.041$ (small effect)] with Mann-Whitney *post hoc* testing, employing Bonferroni correction, indicated that the error in the case of theatre/poetry videos was significantly higher compared to the other two genres. During the debriefing interviews, P2, P6, and P7 discussed the challenges they faced in aligning span selection with actual sign boundaries, particularly for theatre/poetry videos. P7 pointed out that the type of signs used in these videos, specifically the use of depiction and classifiers, made it more difficult to identify clear boundaries in the signs.

6 Study 3: Comparing Integrated-Search and State-of-the-Art Non-Integrated Search Approach

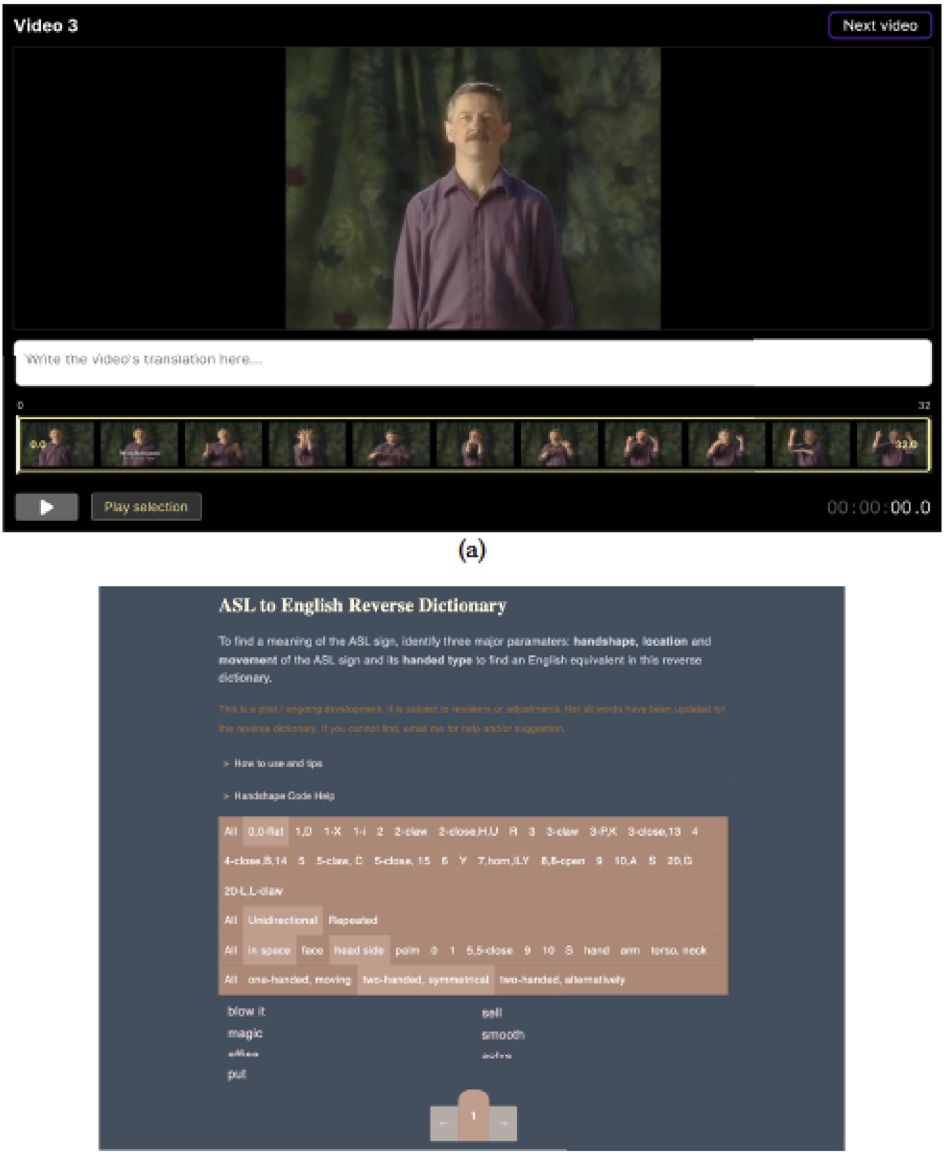
The previous section reports on a study that examined participants' experience with a prototype of an integrated tool for watching ASL videos, which included a span-based dictionary-search feature. However, this study did not demonstrate whether the use of such a tool had measurable effects on a video comprehension task. Therefore, we now present "Study 3," which extends "Study 2" to allow for a between-subjects comparison of perceived workload and task performance. This comparison is conducted between the original complete prototype and a stripped-down version that includes only a video player and a separate Web site featuring a state-of-the-art search-by-feature ASL dictionary.

6.1 Study Design

We developed a modified version of the prototype used in study 2 to allow for a "baseline" reference point. While similar to the previous prototype, this version lacked the integrated Wizard-of-Oz dictionary-search feature, as depicted in Figure 11(a). Instead, participants were instructed to open a separate web browser window and use the HandSpeak reverse dictionary,[4] as illustrated in Figure 11(b). The HandSpeak Web site allows users to search for individual ASL signs by selecting specific text labels corresponding to different linguistic aspects of a sign, such as handshape, hand location, hand movement, and hand orientation. The search results display a list of English gloss labels for signs that match the selected query options. Participants were not allowed to access any other Web sites or external resources.

Both participants of studies 2 and 3 were instructed to complete a NASA TLX after viewing each video. NASA-TLX is a task load index created by NASA to capture a subjective score for participants'

[4] https://www.HandSpeak.com/word/asl-eng/



(a)

ASL to English Reverse Dictionary

To find a meaning of the ASL sign, identify three major paramaters: **handshape, location** and **movement** of the ASL sign and its **handed type** to find an English equivalent in this reverse dictionary.

This is a pilot / ongoing development. It is subject to revisions or adjustments. Not all words have been updated for this reverse dictionary. If you cannot find, email me for help and/or suggestion.

How to use and tips
Handshape Code Help
All 0,0-flat 1,D 1-X 1-I 2 2-claw 2-close,H,U R 3 3-claw 3-P,K 3-close,13 4
4-close,B,14 5 5-claw, C 5-close, 15 6 Y 7,horn,ILY 8,8-open 9 10,A S 20,G
20-L,L-claw
All Unidirectional Repeated
All in space face head side palm 0 1 5,5-close 9 10 S hand arm torso, neck
All one-handed, moving two-handed, symmetrical two-handed, alternatively
blow it
sell
magic
smooth
put
1

(b)

Fig. 11. (a) The prototype for the baseline condition in study 3, similar to the one from study 2 but missing the simulated dictionary-search ability. (b) The HandSpeak ASL-English reverse dictionary Web site, which participants used in the study 3 baseline condition. As users click on linguistic properties, the list of English gloss labels at the bottom of the window updates to list matching signs.

perception of experienced workload [25]. Additionally, each of their final English translation texts was saved—it is worth noting that the same set of videos was used for both studies.[5]

[5] Previous research has already investigated how students interact with existing dictionary-search Web sites, e.g., [9, 28]. As such, we omit a detailed observational analysis of study 3's participant's interaction with the baseline prototype given its limited novelty vis-à-vis prior literature.

Table 1. NASA TLX Sub-Scale Scores from Participants in the Dictionary-Search (n = 8) and Baseline Prototype (n = 7) Conditions, with Scores Scaled to a 0-to-100 Range

TLX Sub-Scale	Dictionary-Search Prototype	Baseline Prototype	Significance Testing
Mental Demand	41.667	56.508	p = 0.042, U = 10 *
Physical Demand	14.583	31.492	p = 0.223, U = 8
Temporal Demand	27.014	45.682	p = 0.007, U = 4 **
Effort	38.473	43.508	p = 0.327, U = 19
Performance	33.611	45.651	p = 0.223, U = 17
Frustration	18.730	44.634	p = 0.013, U = 6 *

In all sub-scales, a lower score is better and indicates less perceived demand, effort need, frustration, or a better sense of performance success. The significance testing column displays results from twotailed Mann-Whitney U tests (** indicates $p < 0.01$, * indicates $p < 0.05$).

6.2 Participants and Recruitment

Recruitment for both studies 2 and 3 was done in tandem, with participants randomly assigned to either the prototype-with-dictionary-search or baseline-prototype condition. For the latter, there were 4 women and 3 men, with a median age of 21 years. All were ASL students, having studied it for a mean of 3.7 years, with all participants confirming they had taken fewer than 3 years of formal ASL classes.

6.3 Comparative Analyses and Findings

6.3.1 Transcription Quality. We adopted the approach by Castilho et al. [12] for the assessment of translation quality. In it, a human judge searches for translation errors, such as wrong or omitted words, and assigns an overall translation-accuracy score (out of 10). For our studies, we recruited a fourth-year ASL interpreting student who had completed a university course on ASL linguistics. The student, unaware of whether the translations were generated using the dictionary-search prototype or the baseline prototype, identified errors and assigned translation accuracy scores for each document. On average, translations from the dictionary-search prototype received an 8.03 score, while translations from the baseline prototype received a 6.67 score. The distributions in scores between the two conditions differed significantly (Mann-Whitney U = 10, $n_{search} = 8$, $n_{baseline} = 7$, $p = 0.0424 < 0.05$ two-tailed, small effect).

6.3.2 Workload. Table 1 shows mean scaled NASA-TLX sub-score values (physical demand, temporal demand, performance, effort, and frustration) and results of two-tailed Mann-Whitney U tests comparing sub-scores across conditions. These same results are depicted in Figure 12. Participants of the dictionary-search prototype scored significantly lower—i.e., better—on measures of mental demand (how much mental and perceptual activity was required), temporal demand (how much time pressure was felt), and frustration (how insecure, discouraged, irritated, or stressed they felt). A copy of the NASA TLX instrument and details of these scales appear in [25, 26].

6.3.3 Time Taken. Since videos were of different durations, the time required by a participant to view and translate each video was normalized by expressing it as a percentage of the total video time. The average time taken by participants in the dictionary-search condition was 547.491%, and for participants in the baseline condition, 1244.4%. The distributions in time between the two conditions differed significantly (Mann-Whitney U = 21, $n_{search} = 8$, $n_{baseline} = 6$, $p = 0.042$, two-tailed).

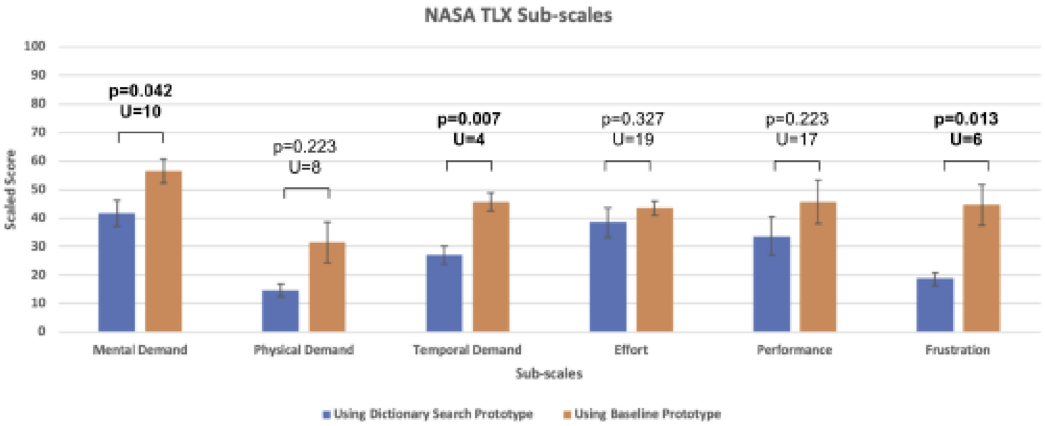


Fig. 12. NASA TLX sub-scale scores from participants in the dictionary-search (n = 8) and baseline prototype (n = 7) conditions, with scores scaled to a 0-to-100 range. For all sub-scales, lower scores are better, i.e., indicating less perceived demand, less effort needed, less frustration, or better sense of performance success. Significance testing results from two-tailed Mann-Whitney U tests are also presented on top of the bars.

6.3.4 *Interactions with Span-Selector.* We conducted a comparative analysis of span durations between participants who interacted with the dictionary-search prototype and those who used the baseline prototype. As previously mentioned in the observational findings of study 2, participants who engaged with the dictionary-search prototype had an average span duration of 8.17 seconds (10.83 signs). For searches that were not part of a search session, the span duration was 11.43 seconds (14.85 signs). Conversely, when participants narrowed down the span as part of a search session without performing a search, the average duration was 3.51 seconds (4.13 signs).

In the baseline condition, the average duration was found to be 12.44 seconds (16.66 signs). The distribution of average span durations was not significantly different between the baseline condition and the dictionary-search condition for spans that were not part of the search (Mann-Whitney $U = 20$, $n_{search} = 8$, $n_{baseline} = 6$, $p = 0.944$, two-tailed).

Participants made an average of 3.46 changes to the selected span per video while interacting with the baseline prototype. When using the integrated search prototype, they made an average of 8.29 adjustments to the selected span. These adjustments included, on average, 3.36 changes per video solely to constrain the playhead for video playback, 2.26 changes for fine-tuning the span, and 2.57 changes made to initiate a search.

7 Discussion

Previous studies in second-language pedagogy have explored the challenges that arise when students encounter complex vocabulary in texts [20]. However, there is a gap in the literature regarding studies that, like ours, utilize direct interviews and observational methods to investigate the specific experiences and challenges faced by ASL students. In our research, participants described how comprehension difficulties can arise from dialectical variations, different linguistic sign types, and diverse genres of ASL content. We discovered that, in order to overcome the comprehension challenges posed by ASL videos featuring unfamiliar signs, students typically rely on online platforms for video streaming, sharing, and social media, as these platforms provide a wide range of natural signing examples. It is worth noting that most previous studies on ASL video comprehension (e.g., [17]) were conducted prior to the widespread availability of video social media and streaming services ([75]). Our findings motivate the need for a tool to (a) support students viewing videos they seek from *diverse sources* and (b) with *challenging content* to support developing comprehension skills.

Study 1 also examined the strategies employed by students when encountering videos with difficult signing, revealing their dissatisfaction with the currently available ASL dictionary resources. This aligns with prior research (e.g., [9]), emphasizing the need for improved tools to identify the meaning of unfamiliar signs. Building upon this, our study specifically focused on the challenges associated with searching for signs in a dictionary while trying to comprehend a challenging video. We discovered that students expressed frustration with the need to switch between the video playback context and the separate Web site for sign searches. Consequently, they resorted to workarounds such as repeatedly pausing and rewinding the videos. We also conducted an initial exploration of the design space of integrated search dictionaries that informed the design of our prototype. Our findings also serve as a motivation for HCI research in developing tools that enable (a) the *viewing and repetition* of short video segments and (b) the *integration* of dictionary-search functionality within the video playback experience.

Study 2 involved the development of a Wizard-of-Oz prototype to observe the actual user interaction with an integrated tool such as was suggested by the findings of study 1. Previous research has explored how students interacted with ASL dictionary-search interfaces, such as [4, 9, 28, 30]. However, their focus was primarily on looking up specific signs, either from memory or from videos of isolated signs. Our findings are novel in that they demonstrate how users engage with a dictionary-search tool in the context of their overall video comprehension task. This includes actions like replaying sections of the original video and comparing them to potential sign matches. The observational findings from study 2 align with those obtained from the interviews conducted in study 1, where students expressed frustration regarding the loss of context when utilizing a separate dictionary tool.

Our observations also revealed that students utilized the span-selection interface of the video player in a *dual-function* manner: (1) to restrict the playhead to specific sections of the video and (2) to enter input for dictionary searches. Notably, the first usage provided valuable insights into the "window size" of the video that students considered while navigating through challenging content. While prior studies on video analysis or annotation tools for linguists studying ASL videos had incorporated a span-selection interface for labeling video segments (e.g., [47, 60]), none had explored the application of span selection in the context of ASL learners watching videos. Therefore, our findings serve as motivation for further research in the field of HCI regarding span-selection interfaces within this unique context.

In Study 3, *we identified the potential advantages associated with the dual utilization of the span-selection interface.* By comparing study 2's full prototype with a baseline that included span selection but lacked an integrated dictionary-search tool, we found that the integrated search tool resulted in higher transcription accuracy, reduced participant workload ratings, and less time taken to produce translations. Additionally, we also observed that participants made greater use of the span-search feature in the integrated search condition, while their usage of the feature solely to constrain the video was not significantly different across the two conditions. Although previous research (e.g., [44]) has explored the use of *in situ* dictionaries to enhance comprehension for non-native speakers, our study was the first to investigate this within a sign-language context.

7.1 How Can Others Make Use of This Work?

7.1.1 Accessibility and HCI Researchers. Accessibility and HCI researchers may find interest in the current experiences and workarounds of ASL students regarding video comprehension tasks, as well as their calls for integrated tools. Our exploration of the design space provides design guidance for integrated sign-language dictionary systems and reveals some remaining open questions, which may be a basis for future research studies.

- The majority of participants expressed a preference for manually segmenting the boundaries of unknown video spans themselves, indicating that ASL learners prefer an on-demand search system over one that provides pre-segmented video spans.
- Most participants preferred viewing the results based on the selected span on the same page, while some also suggested having a toggle option.
- Participants favored a two-column layout for displaying results, as it allows for faster navigation across the outcomes. In cases where a selected span contains multiple signs, a two-column result section increases the likelihood of more relevant signs appearing on the same page.
- Participants highlighted the importance of having more linguistic information accompanying each result item. This preference aligns with the design guidelines outlined in previous research on ASL dictionaries [28].

7.1.2 ASL Linguistics and Education Researchers. ASL linguistics and education researchers may find value in students' perspectives on the challenging factors of ASL video comprehension. Our findings highlight and confirm the challenges related to regional and dialectal variations [8], linguistic types of signs [66, 71], and different genres [65, 80].

ASL educators can integrate this tool into their curricula by selectively incorporating it into comprehension practice assignments. Students could also be encouraged to use the tool for independent learning, uploading diverse videos, and practicing comprehension through transcript writing. A fully functional tool like this would grant students greater autonomy, allowing them to engage with a wider range of ASL content at their own pace. These tools could also ease the burden on ASL teachers, as students can independently look up unfamiliar signs and confirm transcripts with minimal feedback. Moreover, the tool's ability to work with any recorded video reduces the need for teachers to create and annotate additional videos for exercises.

ASL education researchers may also find the potential use of span-selection video players as research tools. Our data in study 2 also showed that users' interactions with the span-selection interface is a *useful probe* to understand participants' comprehension strategies while viewing the videos. We saw, for example, that different video genres elicited differences in span duration, with participants commenting that their selections were related to linguistic properties, with wider spans used for the more challenging theater and poetry videos. It was also possible to compare how well the boundaries of the span-selections intersected with the actual sign boundaries, with a marked increase in error for the more complex theater and poetry videos. This post-study analysis of spans selected and viewed may reveal insights for researchers of ASL linguistics or education. Additionally, a real-time analysis of spans could lead to adaptive educational programs that can identify when students are currently facing difficulties while watching a video.

7.1.3 Computer Vision Researchers. Computer vision researchers may benefit from our findings on the advantages of using extracted spans of continuous-signing video for ASL dictionary search. Regarding this last point, our findings uncover a new sign-recognition task: using a sub-segment of a continuous video as input for ASL recognition. Traditionally, ASL recognition has focused on fully automated machine translation of ASL videos to English text. However, our findings revealed that students could benefit from a use case where, in order to understand a challenging ASL video, they are presented with an integrated sign-searching supportive tool. This presents a novel challenge because the input videos used for sign-searching may be fragments of longer videos with boundaries that are not necessarily perfectly aligned with sign boundaries. The reason for this is that signs at the boundaries may have been affected by co-articulation effects with signs beyond the selected extract, resulting in a loss of contextual information available for the recognition system.

7.2 Broader Generalizability

In a broader context, our research findings contribute to the existing literature on user *interaction with videos*, particularly when closely examining video content, such as when utilizing specialized video-editing software. Recently, YouTube introduced a new feature enabling users to share a continuous loop of a 5–60 second clip from a video directly on the original video's watch page [34]. The span-selector design employed in this feature bears a resemblance to the one utilized in our own studies. While such span-selection functionalities are less prevalent in other video-player systems, various commercial video-editing systems (e.g., [38, 40]) offer users the ability to select specific video segments. Previous observational research has explored users selecting spans while engaged in video-editing tasks [41], and our study expands this span-selection interaction to the domains of ASL education and video search. Other earlier works have investigated educational software tools that allow students to select portions of spoken-language lecture videos while simultaneously taking notes [13, 52], as well as integrated approaches to editing, sharing, and controlling spoken-language educational lecture videos [13, 72]. Although differences exist between the task of understanding ASL videos and comprehending educational lecture videos in spoken language, our findings regarding the advantages of span-selection interfaces in controlling playback and facilitating integrated search tools may have relevance to that particular domain.

8 Limitations and Future Work

The ASL video stimuli employed in our second and third studies encompassed a range of characteristics, including size, frame rate, bit-depth, compression level, and frame scan method (e.g., interlaced or progressive). These attributes have been recognized as influential factors impacting video comprehension [36]. Given that the videos were consistent between prototype conditions, the aforementioned factors did not exert an influence on our results. However, in subsequent investigations, it would be worthwhile to explore the experiences of ASL students when exposed to videos that exhibit variations in these dimensions. While our research primarily centers on ASL learners and videos, future endeavors could extend to encompass learners of other sign languages.

In our prototype, the Wizard-of-Oz dictionary-search output emulated a constant level of accuracy in sign recognition. However, future investigations can explore the impact of varying output quality on users' experiences, thereby providing valuable insights to computer vision researchers regarding the necessary level of accuracy. Similar research has already been conducted on systems employing isolated-sign search-by-video [4, 29, 30].

The transcripts were evaluated by a senior ASL interpreting student using a method adapted from prior research. However, we recognize that the labeling and final score computation are subjective, and there may be differences in assessment among other ASL experts or interpreters.

Although our research has shed light on the advantages of integrated, span-based dictionary search in ASL video comprehension, there is still room for future exploration in the design aspects of span selection and search result presentation. Furthermore, while study 3 demonstrated the positive impact on translation quality and reduced workload scores, it did not delve into the potential learning effects that may arise when students engage in the task of watching challenging videos. Hence, future research could focus on examining both short-term and long-term benefits that may emerge from such activities.

Our research has primarily focused on ASL due to practical considerations, as we are based in the United States and affiliated with an institution known for its prominent ASL and ASL-English interpreting program. However, there are several different sign languages used worldwide that are unique to their culture of origin in terms of syntax, hand movements, facial expression, non-manual markers, vocabulary, and grammar [1]. Furthermore, it is worth noting that HCI and accessibility research have primarily focused on Western and Global North countries [7, 51]. As a result, there is an opportunity for future studies to explore the design of integrated search systems for sign languages in other regions of the world. This can be achieved by investigating the specific learning needs of local sign language learners, as well as designing user interfaces that are tailored to their cultural preferences since cultural background has been shown to be an important consideration in design [43, 64]. By doing so, we can enhance the inclusion of cultural sensitivity in sign language search system design, while future knowledge gained on those possible design modifications could be incorporated as design examples for students learning about accessible design because current courses do not include enough discussions around culture [61].

While our studies have primarily focused on ASL learners, it is worth noting that the video-playing and span-searching tool we developed could have broader applications for various user groups. For instance, linguists working on annotating ASL videos or experienced ASL interpreters tasked with translating complex technical videos could benefit from utilizing our tool. Its potential extends beyond ASL transcription, as it can be employed in other contexts where human annotators aim to identify and label specific components of human movements in videos. This could open up possibilities for applications such as dance script writing or sports analysis, where annotators may be interested in labeling specific movements, positions, or actions performed by subjects in the videos. Future research could delve into the design of tools tailored for these related tasks.

9 Conclusion

Understanding challenging videos plays a crucial role in the development of comprehension skills for students learning ASL. While sign-lookup tools are valuable for this process, they do have their limitations. In our interview study, we learned about ASL learners challenges faced when trying to understand such videos, the strategies they use to overcome the shortcomings of existing sign-lookup tools, and their expectations from a future integrated dictionary system.

These findings inspired a second study, in which we utilized a Wizard-of-Oz tool featuring a video player with an integrated dictionary search based on span selection. An analysis of user interaction patterns shed light on how participants used the tool, highlighting the influence of video genre and linguistic complexity. The integrated design offered advantages, and participants found value in utilizing span selection both as an input for the sign-lookup dictionary and as a means of controlling the video player's playhead.

Finally, in our third study, we compared integrated search to a baseline version of the prototype, which involved an existing ASL dictionary Web site. We found that the integrated tool resulted in improved accuracy in translating ASL to English, a reduced sense of workload, and time taken to produce translations. There was no significant difference in the use of span-selector to constrain the video player's playhead.

Our work serves as motivation for further research and the development of tools that cater to the specific needs of ASL learners who are working on comprehending challenging videos. The findings from our study provide valuable insights for future designers of ASL systems, computer vision researchers working on sign-matching technologies, as well as sign-language educators and linguists. Moreover, these findings can also offer guidance for the design of video comprehension tools in other contexts.

References

[1] Rahib H. Abiyev, Murat Arslan, and John Bush Idoko. 2020. Sign language translation using deep convolutional neural networks. *KSII Transactions on Internet & Information Systems* 14, 2 (2020), 613–653.

[2] Alikhan Abutalipov, Aigerim Janaliyeva, Medet Mukushev, Antonio Cerone, and Anara Sandygulova. 2021. Handshape classification in a reverse dictionary of sign languages for the deaf. In *From Data to Models and Back*. Juliana Bowles, Giovanna Broccia, and Mirco Nanni (Eds.), Springer International Publishing, Cham, 217–226.

[3] Chanchal Agrawal and Roshan L. Peiris. 2021. I see what you're saying: A literature review of eye tracking research in communication of deaf or hard of hearing users. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)*. ACM, New York, NY, Article 41, 13 pages. DOI: https://doi.org/10.1145/3441852.3471209

[4] Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. 2019. Effect of automatic sign recognition performance on the usability of video-based search interfaces for sign language dictionaries. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. ACM, New York, NY, 56–67. DOI: https://doi.org/10.1145/3308561.3353791

[5] Stavroula Sokoli Athens and Stavroula Sokoli. 2007. Stavroula Sokoli (Athens) learning via subtitling (LvS): A tool for the creation of foreign language learning activities based on film subtitling. In *Proceedings of the Audiovisual Translation Scenarios: Conference Proceedings (MuTra '06)*. MuTra, Copenhagen, Denmark, 8 pages.

[6] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Ashwin Thangali, Haijing Wang, and Quan Yuan. 2010. Large lexicon project: American sign language video corpus and sign language indexing/retrieval algorithms. In *Proceedings of the Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT '10)*, Vol. 2, European Language Resources Association (ELRA), Valletta, Malta, 11–14.

[7] Giulia Barbareschi, Manohar Swaminathan, Andre Pimenta Freire, and Catherine Holloway. 2021. Challenges and strategies for accessibility research in the global south: A panel discussion. In *Proceedings of the 10th Latin American Conference on Human Computer Interaction (CLIHC '21)*. ACM, New York, NY, Article 20, 5 pages. DOI: https://doi.org/10.1145/3488392.3488412

[8] Patrick Boudreault. 1999. Grammatical processing in American sign language: Effects of age of acquisition and syntactic complexity. Retrieved from https://escholarship.mcgill.ca/concern/theses/5x21th32w

[9] Danielle Bragg, Kyle Rector, and Richard E. Ladner. 2015. A user-powered American sign language dictionary. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '15)*. ACM, New York, NY, 1837–1848. DOI: https://doi.org/10.1145/2675133.2675226

[10] Fabio Buttussi, Luca Chittaro, and Marco Coppo. 2007. Using web3D technologies for visualization and search of signs in an international sign language dictionary. In *Proceedings of the 12th International Conference on 3D Web Technology (Web3D '07)*. ACM, New York, NY, 61–70. DOI: https://doi.org/10.1145/1229390.1229401

[11] Naomi K. Caselli, Zed Sevcikova Sehyr, Ariel M. Cohen-Goldberg, and Karen Emmorey. 2017. ASL-LEX: A lexical database of American sign language. *Behavior Research Methods* 49, 2 (01 Apr. 2017), 784–801. DOI: https://doi.org/10.3758/s13428-016-0742-0

[12] Sheila Castilho, Stephen Doherty, Federico Gaspari, and Joss Moorkens. 2018. *Approaches to Human and Machine Translation Quality Assessment*. Springer International Publishing, Cham, 9–38. DOI: https://doi.org/10.1007/978-3-319-91241-7_2

[13] Konstantinos Chorianopoulos and Michail N. Giannakos. 2013. Usability design for video lectures. In *Proceedings of the 11th European Conference on Interactive TV and Video (EuroITV '13)*. ACM, New York, NY, 163–164. DOI: https://doi.org/10.1145/2465958.2465982

[14] Christopher Conly, Zhong Zhang, and Vassilis Athitsos. 2015. An integrated RGB-D system for looking up the meaning of signs. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '15)*. ACM, New York, NY, Article 24, 8 pages. DOI: https://doi.org/10.1145/2769493.2769534

[15] David M. Eberhard, Gary F. Simons, and Charles D. Fennig (Eds.). 2024. *Ethnologue: Languages of the World*. (27th ed.). SIL International, Dallas, TX. Retrieved from https://www.ethnologue.com/subgroup/4395/

[16] Ralph Elliott, Helen Cooper, John Glauert, Richard Bowden, and François Lefebvre-Albaret. 2011. Search-by-example in multilingual sign language databases. In *Proceedings of the 2nd International Workshop on Sign Language Translation and Avatar Technology (SLTAT '11)*. SLTAT, Dundee, Scotland, 8 pages.

[17] Karen Emmorey, Robin Thompson, and Rachael Colvin. 2009. Eye gaze during comprehension of American sign language by native and beginning signers. *Journal of Deaf Studies and Deaf Education* 14, 2 (2009), 237–243.

[18] National Center for Education Statistics (NCES). 2018. Digest of education statistics number and percentage distribution of course enrollments in languages other than English at degree-granting postsecondary institutions, by language and enrollment level: Selected years, 2002 through 2016. Retrieved from https://nces.ed.gov/programs/digest/d18/tables/dt18_311.80.asp

[19] José L. Fuertes, Ángel L. González, Gonzalo Mariscal, and Carlos Ruiz. 2006. Bilingual sign language dictionary. In *Computers Helping People with Special Needs*. Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer (Eds.), Springer, Berlin, 599–606.

[20] Susan M. Gass, Jennifer Behney, and Luke Plonsky. 2020. *Second Language Acquisition: An Introductory Course* (5th. ed.). Routledge, New York, NY. 774 pages.

[21] Nelly Furman, David Goldberg, and Natalia Lusin. Enrollments in languages other than English in United States institutions of higher education, Fall 2009. Retrieved from https://eric.ed.gov/?id=ED513861

[22] Debbie B. Golos and Annie M. Moses. 2011. How teacher mediation during video viewing facilitates literacy behaviors. *Sign Language Studies* 12, 1 (2011), 98–118.

[23] Michael Andrew Grosvald. 2009. *Long-Distance Coarticulation: A Production and Perception Study of English and American Sign Language*. University of California.

[24] Wyatte C. Hall, Leonard L. Levin, and Melissa L. Anderson. 2017. Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins. *Social Psychiatry and Psychiatric Epidemiology* 52, 6 (2017), 761–776.

[25] Sandra G. Hart. 2006. NASA-task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, Sage Publications, Los Angeles, CA, 904–908.

[26] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*, Vol. 52, Elsevier, Amsterdam, Netherlands, 139–183.

[27] Saad Hassan. 2022. Designing and experimentally evaluating a video-based American sign language look-up system. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*. ACM, New York, NY, 383–386. DOI: https://doi.org/10.1145/3498366.3505804

[28] Saad Hassan, Akhter Al Amin, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022a. Design and evaluation of hybrid search for American sign language to English dictionaries: Making the most of imperfect sign recognition. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, 13 pages. Retrieved from http://latlab.ist.rit.edu/pubs/Hassan-et-al-2022-CHI.pdf

[29] Saad Hassan, Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. 2020. Effect of ranking and precision of results on users' satisfaction with search-by-video sign-language dictionaries. In *Sign Language Recognition, Translation and Production (SLRTP) Workshop-Extended Abstracts*, Vol. 4, Computer Vision – ECCV 2020 Workshops, 6 pages.

[30] Saad Hassan, Oliver Alonzo, Abraham Glasser, and Matt Huenerfauth. 2021. Effect of sign-recognition performance on the usability of sign-language dictionary search. *ACM Transactions. on Accessible Computing* 14, 4, Article 18 (Oct. 2021), 33 pages. DOI: https://doi.org/10.1145/3470650

[31] Saad Hassan, Akhter Al Amin, Caluã de Lacerda Pataca, Diego Navarro, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022. Support in the moment: Benefits and use of video-span selection and search for sign-language video comprehension among ASL learners. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22)*. ACM, New York, NY, Article 29, 14 pages. DOI: https://doi.org/10.1145/3517428.3544883

[32] Saad Hassan, Akhter Al Amin, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022. Design and evaluation of hybrid search for American sign language to English dictionaries: Making the most of imperfect sign recognition. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, Article 195, 13 pages. DOI: https://doi.org/10.1145/3491102.3501986

[33] Saad Hassan, Aiza Hasib, Suleman Shahid, Sana Asif, and Arsalan Khan. 2019. Kahaniyan - Designing for acquisition of Urdu as a second language. In *Human-Computer Interaction – INTERACT 2019*. David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.), Springer International Publishing, Cham, 207–216.

[34] Google Help. 2023. Share clips - youtube help. Retrieved from https://support.google.com/youtube/answer/10332730?hl=en

[35] Robert J. Hoffmeister. 2000. *A Piece of the Puzzle: ASL and Reading Comprehension in Deaf Children*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 143–163.

[36] Simon Hooper, Charles Miller, Susan Rose, and George Veletsianos. 2007. The effects of digital video quality on learner comprehension in an American sign language assessment environment. *Sign Language Studies* 8, 1 (2007), 42–58.

[37] iMotions A/S. 2019. iMotions Biometric Research Platform. imotions. Retrieved from https://imotions.com/academy/

[38] Adobe Inc. 2008. Adobe Premiere Pro. Retrieved 3 March 2022 from https://www.adobe.com/products/premiere.html

[39] Apple Inc. 2008. Apple Finalcut. Retrieved 3 March 2022 from http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control-software/

[40] Apple Inc. 2008. Apple iMovie. Retrieved 3 March 2022 from https://www.apple.com/imovie/

[41] Tero Jokela, Minna Karukka, and Kaj Mäkelä. 2007. Mobile video editor: Design and evaluation. In *Proceedings of the 12th International Conference on Human-Computer Interaction: Interaction Platforms and Techniques (HCI '07)*. Springer-Verlag, Berlin, 344–353.

[42] Jonathan Keane, Diane Brentari, and Jason Riggle. 2012. Coarticulation in ASL fingerspelling. Retrieved from https://pubs.jonkeane.com/pdfs/Keane2012aa.pdf

[43] Ji Hye Kim and Kun Pyo Lee. 2005. Cultural difference and mobile phone interface design: Icon recognition according to level of abstraction. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (Mobile HCI '05)*. ACM, New York, NY, 307–310. DOI: https://doi.org/10.1145/1085777.1085841

[44] Annette Klosa-Kückelhaus and Frank Michaelis. 2022. The design of internet dictionaries. In *The Bloomsbury Handbook of Lexicography*, Howard Jackson (Ed.), Bloomsbury Publishing, 405.

[45] Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. 2018. A position and rotation invariant framework for sign language recognition (SLR) using Kinect. *Multimedia Tools and Applications* 77, 7 (2018), 8823–8846.

[46] Marlon Kuntze, Debbie Golos, and Charlotte Enns. 2014. Rethinking literacy: Broadening opportunities for visual learners. *Sign Language Studies* 14, 2 (2014), 203–224.

[47] The Language Archive. 2018. ELAN - The Max Planck Institute for Psycholinguistics. Retrieved from https://archive.mpi.nl/tla/elan

[48] J. Lapiak. 2021. Handspeak. Retrieved from https://www.handspeak.com/

[49] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann.

[50] Scott K. Liddell and Robert E. Johnson. 1986. American sign language compound formation processes, lexicalization, and phonological remnants. *Natural Language & Linguistic Theory* 4, 4 (1986), 445–513.

[51] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI? In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY. DOI: https://doi.org/10.1145/3411764.3445488

[52] Ching (Jean) Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. NoteStruct: Scaffolding note-taking while learning from online videos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, 1–6. DOI: https://doi.org/10.1145/3290607.3312878

[53] Carolyn McCaskill, Ceil Lucas, Robert Bayley, and Joseph Christopher Hill. 2011. *The Hidden Treasure of black ASL: Its History and Structure*. Gallaudet University Press, Washington, DC.

[54] John Milton and Vivying S. Y. Cheng. 2010. A toolkit to assist L2 learners become independent writers. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids (CL & W '10)*. Association for Computational Linguistics, 33–41.

[55] Daniel Mitchell. 2021. British sign language BSL dictionary. Retrieved from https://www.signbsl.com/

[56] Ross Mitchell, Travas Young, Bellamie Bachleda, and Michael Karchmer. 2006. How many people use ASL in the United States? Why estimates need updating. *Sign Language Studies* 6 (Mar. 2006). DOI: https://doi.org/10.1353/sls.2006.0019

[57] Anshul Mittal, Pradeep Kumar, Partha Pratim Roy, Raman Balasubramanian, and Bidyut B. Chaudhuri. 2019. A modified LSTM model for continuous sign language recognition using leap motion. *IEEE Sensors Journal* 19, 16 (2019), 7056–7063.

[58] J. Murray. 2020. World Federation of the deaf. Retrieved from http://wfdeaf.org/our-work/

[59] Tobii Pro Nano. 2014. *Tobii Pro Lab*. Tobii Technology. Retrieved from https://www.tobiipro.com/

[60] Carol Neidle and Christian Vogler. 2012. A new web interface to facilitate access to corpora: Development of the ASLLRP data access interface (DAI). In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC*. Citeseer, OpenBU, Istanbul, Turkey, 8 pages. DOI: https://open.bu.edu/handle/2144/31886

[61] Laleh Nourian, Kristen Shinohara, and Garreth W. Tigwell. 2023. Understanding discussions around culture within courses covering topics on accessibility and disability at U.S. universities (CHI '23). In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, Article 221, 14 pages. https://doi.org/10.1145/3544548.3581533

[62] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2021. Sign language recognition: A deep survey. *Expert Systems with Applications* 164 (2021), 113794. DOI: https://doi.org/10.1016/j.eswa.2020.113794

[63] Kishore K. Reddy and Mubarak Shah. 2013. Recognizing 50 human action categories of web videos. *Machine Vision and Applications* 24, 5 (2013), 971–981.

[64] Katharina Reinecke and Abraham Bernstein. 2013. Knowing what a user likes: A design science approach to interfaces that automatically adapt to culture. *MIS Quarterly* 37, 2 (2013), 427–453. DOI: http://www.jstor.org/stable/43825917

[65] Michael Richardson. 2018. The sign language interpreted performance: A failure of access provision for Deaf spectators. *Theatre Topics* 28, 1 (2018), 63–74.

[66] Wendy Sandler. 2017. The challenge of sign language phonology. *Annual Review of Linguistics* 3 (2017), 43–63.

[67] Jerry Schnepp, Rosalee Wolfe, Gilbert Brionez, Souad Baowidan, Ronan Johnson, and John McDonald. 2020. Human-centered design for a sign language learning application. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '20)*. ACM, New York, NY, Article 60, 5 pages. DOI: https://doi.org/10.1145/3389189.3398007

[68] Jérémie Segouat. 2009. A study of sign language coarticulation. *SIGACCESS Accessible Computing* 1, 93 (Jan. 2009), 31–38. DOI: https://doi.org/10.1145/1531930.1531935

[69] Zatorre R. J., Shiell M. M., Champoux F. 2014. Enhancement of visual motion detection thresholds in early deaf people. *PloS One* 9, 2 (2014), e90498. DOI: https://doi.org/10.1371/journal.pone.0090498

[70] ShuR. 2021. SLintoDictionary. Retrieved from http://slinto.com/us

[71] Kristin Snoddon. 2018. Whose ASL counts? Linguistic prescriptivism and challenges in the context of parent sign language curriculum development. *International Journal of Bilingual Education and Bilingualism* 21, 8 (2018), 1004–1015.

[72] Namrata Srivastava, Sadia Nawaz, Joshua Newn, Jason Lodge, Eduardo Velloso, Sarah M. Erfani, Dragan Gasevic, and James Bailey. 2021. Are you with me? Measurement of learners' video-watching attention with eye tracking. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference (LAK '21)*. ACM, New York, NY, 88–98. DOI: https://doi.org/10.1145/3448139.3448148

[73] Ted Supalla. 1982. *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. Ph. D. Dissertation. University of California, San Diego.

[74] Nazif Can Tamer and Murat Saraçlar. 2020. Improving keyword search performance in sign language with hand shape features. In *Computer Vision – ECCV 2020 Workshops*. Adrien Bartoli and Andrea Fusiello (Eds.), Springer International Publishing, Cham, 322–333.

[75] Carolina Tannenbaum-Baruchi and Paula Feder-Bubis. 2018. New sign language new (S): The globalization of sign language in the smartphone era. *Disability & Society* 33, 2 (2018), 309–312.

[76] Kimberly A. Weaver and Thad Starner. 2011. We need to communicate! Helping hearing parents of deaf children learn American sign language. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '11)*. ACM, New York, NY, 91–98. DOI: https://doi.org/10.1145/2049536.2049554

[77] Polina Yanovich, Carol Neidle, and Dimitris Metaxas. 2016. Detection of major ASL sign types in continuous signing for ASL Recognition. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC '16)*. European Language Resources Association (ELRA), Portorož, Slovenia, 3067–3073. DOI: https://www.aclweb.org/anthology/L16-1490

[78] Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. 2011. American sign language recognition with the kinect. In *Proceedings of the 13th International Conference on Multimodal Interfaces (ICMI '11)*. ACM, New York, NY, 279–286. DOI: https://doi.org/10.1145/2070481.2070532

[79] Mikhail A. Zagot and Vladimir V. Vozdvizhensky. 2014. Translating video: Obstacles and challenges. *Procedia - Social and Behavioral Sciences* 154 (2014), 268–271. DOI: https://doi.org/10.1016/j.sbspro.2014.10.149

[80] Ulrike Zeshan. 2015. "Making meaning": Communication between sign language users without a shared language. *Cognitive Linguistics* 26, 2 (2015), 211–260.

Received 28 June 2023; revised 28 June 2023; accepted 15 August 2024