CUNY American Sign Language Motion-Capture Corpus: First Release

Pengfei Lu
The City University of New York (CUNY)
Computer Science Program, Graduate Center
365 Fifth Ave, New York, NY 10016 USA
E-mail: pengfei.lu@qc.cuny.edu

Matt Huenerfauth
The City University of New York (CUNY)
Computer Science Department, Queens College
65-30 Kissena Blvd, Flushing, NY 11367 USA
E-mail: matt@cs.qc.cuny.edu

Abstract

We are in the middle of a 5-year study to collect, annotate, and analyze an ASL motion-capture corpus of multi-sentential discourse. Now we are ready to release to the research community the first sub-portion of our corpus that has been checked for quality. This paper describes the recording and annotation procedure of our released corpus to enable researchers to determine if it would benefit their work. A focus of the collection process was the identification and use of prompting strategies for eliciting single-signer multi-sentential ASL discourse that maximizes the use of pronominal spatial reference yet minimizes the use of classifier predicates. The annotation of the corpus includes details about the establishment and use of pronominal spatial reference points in space. Using this data, we are seeking computational models of the referential use of signing space and of spatially inflected verb forms for use in American Sign Language (ASL) animations, which have accessibility applications for deaf users.

Keywords: American Sign Language, motion-capture, corpus, sign language, spatial reference.

1. Introduction

Software to generate American Sign Language (ASL) animations can provide benefits for the significant number of deaf individuals in the United States with relatively low written English literacy. Our research goal is to improve technologies for generating ASL animations through the collection and analysis of a motion-capture corpus of ASL multi-sentence discourse. Our intention is to provide the research community with a sufficient-quality corpus for their future study on ASL linguistic phenomena and to conduct our own analysis of this corpus using statistical modeling and machine learning techniques to create models useful for generating grammatically correct ASL animations.

Signers associate entities under discussion with 3D signing space locations, and signs whose paths or orientations depend on these locations pose a special challenge: They are time-consuming for users of scripting software to produce, and they are not included in the repertoire of most ASL generation/translation software. Our goal is to construct computational models of ASL that could be used to partially automate the work of human authors using scripting software or to underlie generation/translation systems.

Section 2 provides basic information and statistics about the released portion of our corpus. Section 3 describes the linguistics of spatial reference points and inflecting verbs. Section 4 describes the recruitment and prompting strategies we have used to elicit signing performances of the desired form. Section 5 describes the motion-capture equipment, motion-capture data recording, and post-processing. Section 6 describes the annotation process. Section 7 describes a sub-corpus we have collected of ASL inflecting verbs. Section 8 contains conclusions and future research plans.

2. The Released Corpus

After the three years of data collection, we have gathered 246 ASL unscripted multi-sentence single-signer passages from 9 native signers, each signer came to the lab for one recording session on a different day (Lu & Huenerfauth, 2010). While we have recorded and begun annotation on a total of 215 minutes of ASL motion-capture data thus far, we are ready to release to the research community the first sub-portion of our corpus that has been checked for quality.

This paper is the first announcement of this corpus release, which includes 98 passages performed by 3 native signers. The data includes Autodesk Motion Builder files of the motion-capture recording, BVH files (another commonly used file format for motion-capture data), high-resolution video recordings, and annotations for each passage. The annotations are in the form of SignStream™ files (Neidle et al., 2000) and plaintext files. Figure 1 shows two screenshots of videos of a signer in our corpus (a front view and a side view), a screenshot of a visualization in Autodesk MotionBuilder of the motion-capture data recorded (a virtual human body whose movements are driven by the data), and a visualization of the joint locations and orientations as recorded by the sensors (yellow dots on bottom right).

In addition to our primary corpus containing unscripted multi-sentential passages, we are also releasing a small sub-corpus containing several hundred instances of eight ASL inflected verbs (discussed in section 7). As this is our first corpus release, we welcome feedback from other researchers on how to best organize and release this corpus so that it is most useful. Future releases of our corpus may contain revisions of the data formats or annotation for this sub-corpus and additional passages not yet released. Our lab website contains details about accessing the corpus: http://latlab.cs.qc.cuny.edu/

3. Pronominal Spatial Reference Points and Inflected Verbs

Signers frequently associate entities under discussion with locations in the signing space involved in later pronominal reference and other purposes (Liddell, 2003;

Meier, 1990; Neidle et al., 2000). Various ASL constructions can be used to establish a spatial reference point (SRP) for some entity. While sign languages used around the world are not mutually intelligible, they do share certain key linguistic aspects – the use of spatial reference and verb inflection. All of these phenomena involve the use of the 3D space around the signer (often called the "signing space") to represent entities under discussion. During a conversation, signers often associate people, concepts, or other entities under discussion with 3D locations around their bodies. For example, by pointing at a location in the surrounding space at the beginning or at the end of a noun phrase mentioning a new entity, the human signer associates the entity referred to in the noun phrase with that location. Signers remember these spatial associations, and the movements of later signs in the performance may change based on these locations. When referring to one of these entities later in the conversation, a signer may use a pronoun sign (which also looks like a pointing gesture) aimed at the appropriate location in the signing space.

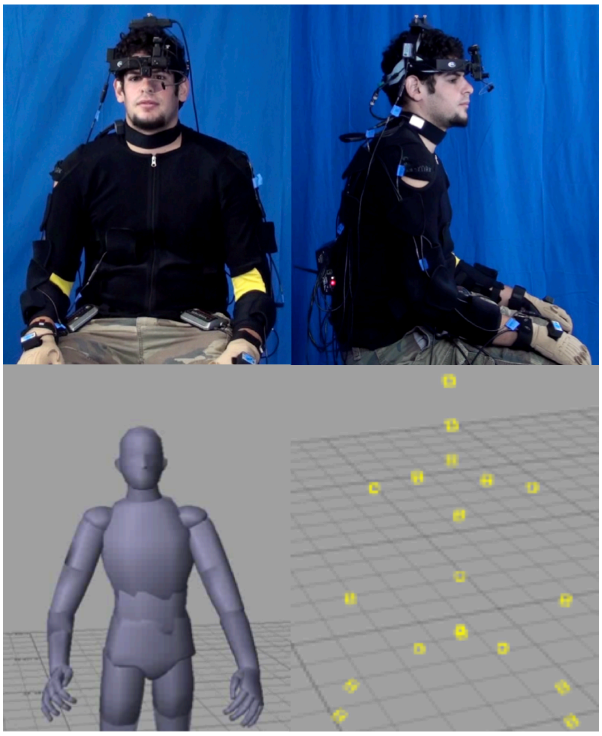

Figure 1: Screenshots of signer in the video recording and visualizations of the motion-capture data.

Some sign language verb signs also change their motion paths or hand orientation to indicate the 3D location where a spatial reference point has been established for their subject, object, or both (Liddell, 2003; Padden, 1988). Generally, the motion paths of these inflecting verbs change so that their direction goes from the subject to the object; however, their paths can be more complex than this. These verbs have been referred to by linguists as "inflecting verbs" (Padden, 1988), "indicating verbs" (Liddell, 2003), or "agreeing verbs" (Cormier, 2002). We call them as "inflecting verbs" in our research. Each inflecting verb has a standard motion path which is affected by the subject's and the object's 3D locations – producing a motion path that is unique to the specific verb, the specific 3D location of the subject, and the specific 3D location of the object.

In prior experimental studies, we determined that the use of spatially inflected verbs in an ASL animation significantly increased viewers' comprehension of the animations (Huenerfauth & Lu, 2012). However, most sign language animation generation software lacks sophisticated models of this phenomenon. A current focus of our research has been to develop computational models of how the motion-paths of inflecting ASL verbs change based on the 3D location in the signing space associated with the subject and/or object of the verb. During the construction of our ASL motion-capture corpus, in addition to the unscripted ASL passages in our released corpus, we also collected recordings of signers performing instances of spatially inflected verbs, and we also release some of those spatially inflected verb samples as a sub-corpus in this paper (discussed in section 7).

4. Recruitment and Elicitation

For the data recording sessions for our corpus, all instructions and interactions were conducted in ASL. Advertisements posted on Deaf community websites in New York City asked whether potential participants had grown up using ASL at home or whether they attended an ASL-based school as a young child. Of the 3 participants in the current corpus release: 3 grew up with parents who used ASL at home, 1 was married to someone deaf/Deaf, 3 used ASL as the primary language in their home, 3 used ASL at work, and 3 had attended a college where instruction was primarily in ASL. The signers were 3 men of ages 22-33 (mean age 25.7).

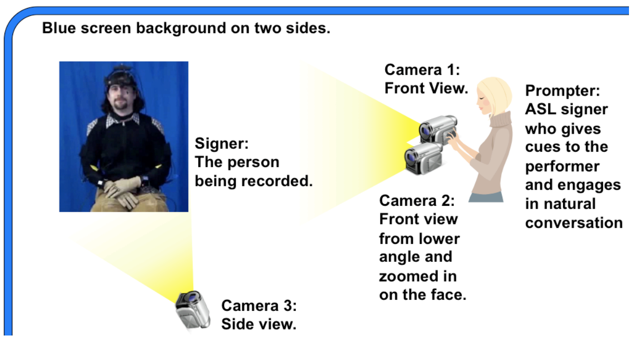

Figure 2: Diagram of an overhead view of our motion-capture studio setup.

Prior to data collection, a prompter who is an ASL signer engaged in natural ASL conversation sitting facing the signer being recorded (Figure 2); during data collection, the prompter gave a prompt to the recorded signer. All of our motion-capture recording sessions are videotaped to facilitate later linguistic analysis and annotation (details in section 5). Three digital high-speed video cameras film front view, facial close-up, and side views of the signer (Figure 2); a similar camera placement has been used in video-based ASL-corpora-building projects (Neidle et al., 2000). The front view is similar to the top left image in Figure 1, but it is wider. The facial close-up view is useful when later identifying specific non-manual facial expressions during ASL performances, which are essential to correctly understanding and annotating the collected data. Videotaping the session may facilitate "clean up" of the motion-capture data in post-processing, during which algorithms can be applied to adjust synchronization of different sensors or remove "jitter" or other noise artifacts from the recording.

To record natural ASL signing (with spontaneous and fluent use of spatial reference points in a multi-sentential single-signer discourse), we are collecting non-scripted passages; so, it has been essential for us to identify appropriate ways to prompt for the type of passages we wish to collect, to support our research. It is important for our research that we collect sign language passages in which signers establish different numbers of points in space to refer to people, places, or things under discussion (SRPs). Further, it is important that the passages do not contain too many classifier predicate (CP) expressions, which are a linguistic construction in ASL that also uses the space around the signer's body. CPs are not our current research focus, and because they lead signers to use space around their bodies in a different way than SRPs, we don't want to record stories that contain a lot of CPs, relative to the number of SRPs.

During our multi-year project, we have experimented with different forms of prompting strategies to elicit ASL signing in which signers establish different numbers of pronominal reference points in space, continuing signing for more time, and in which they do not make frequent use of CPs. Thus, the analysis of the different prompting strategies in one year of our project guided our data collection procedure for the following year. We identified the most effective prompts, and we stopped using some prompts with high CP/SRP ratios. We continued to use those prompts that led to long passage lengths, high number of SRP points established, and low CP/SRP ratios (Lu & Huenerfauth, 2011a; Huenerfauth & Lu, 2010a). The topics of the passages include signers discussing their personal histories, their recollection of news stories/movies, their explanation of encyclopedia articles, their opinion about a hypothetical scenario, their comparison of two persons or things, their description of a page of photos, and their recounting short narratives (Lu & Huenerfauth, 2011a; Huenerfauth & Lu, 2010a). Table 1 lists the prompts we used in the collection of this released corpus data, and brief description of each prompting strategy. Some of our prompting approaches involved showing pictures to a signer. Figure 3 shows an example of what a page of photos looked like for the "photo page" prompts.

Figure 3: Example of what a page of photos looked like for the "photo page" prompts.

Type of Prompt	**Description of This Prompting Strategy**
News Story	Please read this brief news article (about a funny or memorable occurrence) and recount the article.
Compare (people)	Compare two people you know: your parents, some friends, family members, etc.
Compare (not people)	Compare two things: e.g. Mac vs. PC, Democrats vs. Republicans, high school vs. college, Gallaudet University vs. NTID, travelling by plane vs. travelling by car, etc.
Photo Page	Look at this page of photos (of people who are in the news recently) and then explain what is going on with them.
Personal Narrative	Please tell a story about an experience that you had personally.
Personal Intro/Info	Introduce yourself, describe some of your background, hobbies, family, education, etc.
Recount Movie Book	Recall a book you've read recently or a movie you saw, and then explain the story as you remember it.
Opinion / Explain Topic	Please explain your opinion on this topic (given) or explain the concept as if you were teaching it to someone.
Wikipedia Article	Read a brief Wikipedia article on some topic and then explain/recount the information from the article.

Table 1: Types of prompts used.

Figure 4 lists how many passages we have collected using each of the different prompting strategies in this released corpus. The total number of passages of each prompt-type collected from each signer varies because the recording session was intentionally kept relaxed and conversational to promote more natural signing. Sometimes performers were verbose in their response to a prompt, but other times, they could think of little or nothing to say for a particular prompt. Further, since performers were recorded for only one hour (after the motion-capture equipment was set-up and calibrated), we rarely had sufficient time to try all of the different prompt-types during each performer's recording session.

This release of our corpus contains 9717 glosses in total (signer #1: 3962 glosses, signer #2: 3121 glosses, signer #3: 2634 glosses). The total length of video is 87.7 minutes (signer #1: 2048 seconds, signer #2: 1786 seconds, signer #3: 1426 seconds). The average number of glosses per passage is 54 (signer #1: 82 glosses per passage, signer #2: 53 glosses per passage, signer #3: 37 glosses per passage). The average video length of the passages collected is 99 seconds (signer #1: 158 seconds per passage, signer #2: 92 seconds per passage, signer #3: 68 seconds per passage). Figures 5 and 6 show histograms of passage length for each signer (measured in the number of signs performed or the number of seconds of the video recording). Figure 7 shows a sample of a transcript of a passage, in which the signer was elicited using the "Photo Page" style of prompt (Figure 3). Table 2 explains the notations we used for annotation in the transcript.

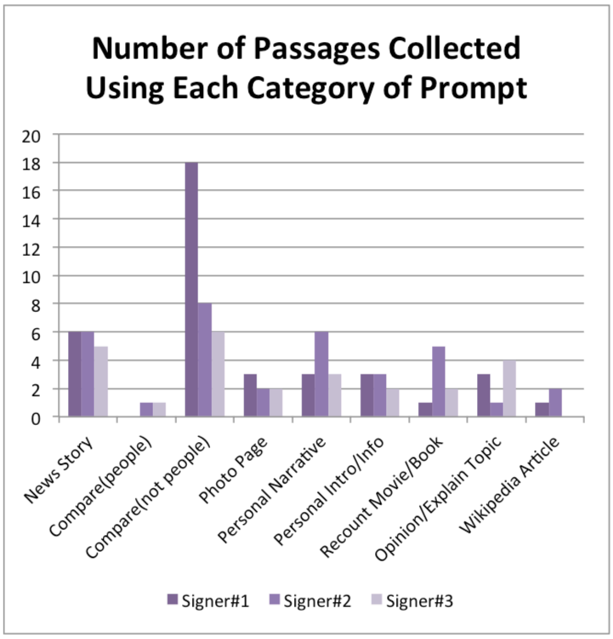

Figure 4: The number of passages in our released corpus that were collected using each category of prompt.

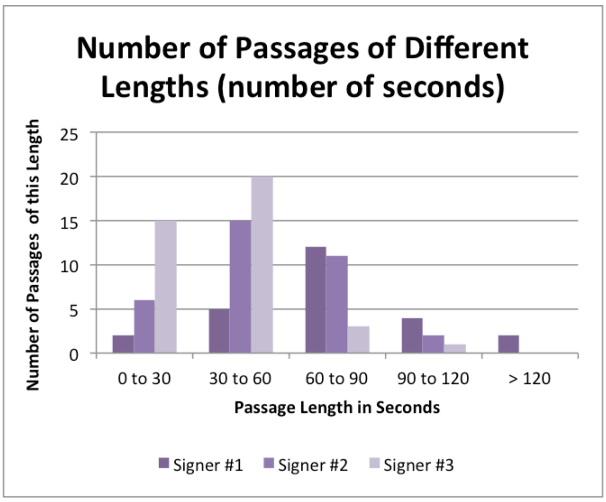

Figure 5: Length of ASL passages collected for each signer.

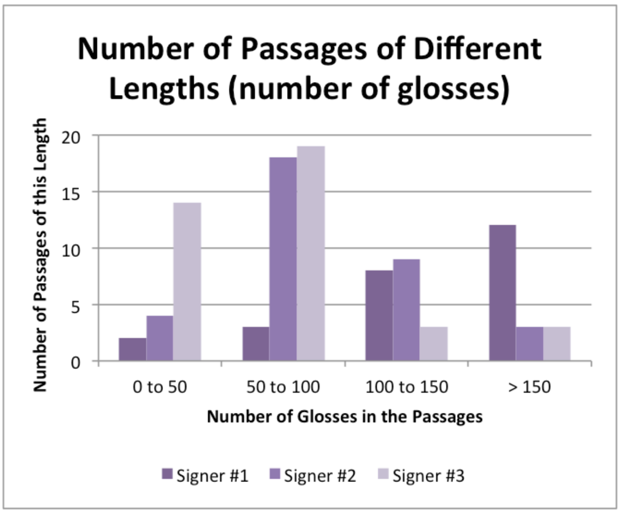

Figure 6: Number of glosses in the collected ASL passages.

PRESIDENT fs-OBAMA NOW IX-1-s:1
OUR PRESIDENT PRESIDENT fs-BUSH
RECENT FINISH THREE YEAR AGO
IX-1-s:2 FOURTY-THREE #TH
fs-OBAMA FOURTY-FOUR BOTH
DIFFERENT #WHAT BLACK IX-1-s:1
FIRST BLACK PRESIDENT WOW
IX-1-s:S LIKE fs-OBAMA BECAUSE
IX-1-s:1 DEMOCRAT …

Figure 7: A sample excerpt of a transcript of a passage.

Type of notation	Explanation of this notation
fs-X	Fingerspelled word
IX-1-s:1	Index sign (pointing), handshape-#1, singular, spatial reference point #1.
IX-1-s:2	Index sign (pointing), handshape-#1, singular, spatial reference point #2.
#X	Lexicalized fingerspelled word.
IX-1-s:S	"I" or "ME": Index sign (pointing), handshape-#1, singular, signer/self.

Table 2: The notations in the transcript in Figure 7.

5. Motion-Capture Equipment, Recording, and Post-Processing

As shown in Figure 2, three high-definition digital video cameras recorded front view, side view, and facial close-up views of the signer. The video in this corpus release has been separated into individual video clips for each passage; each clip has been trimmed from the full-length video recording of the entire data collection session. The start and end keyframes (marking the beginning and end of each passage) were identified by ASL native signers in our research group who watched the videos at the end of the recording session. To facilitate synchronizing the three video files (front, side, face close-up) during our post-processing, a strobe light was flashed once at the start of the recording session. Thus, as soon as the start- and end-times for each passage were identified in one of the three videos, it was straightforward to calculate the appropriate start- and end-times for the other two videos. All the videos are released as QuickTime MOV format files, of size 640x480, no audio, with a frame rate of 29.97 fps. If there is an interest from researchers in obtaining the original high definition video of the full recording session, this may be available in a future corpus release.

Since our goal in creating this corpus was to learn how to control the movements of an animated signing character, we knew that we would need to identify hand locations and joint angles of the human signer's body throughout the performance. Asking human annotators to write down 3D angles and coordinates from a video recording is time-consuming and inexact. Using computer vision techniques to automatically track the movements of a human's body in a video is also challenging due to the complex shape of the hands/face, rapid speed, and frequent occlusion of parts of the body during sign language. Thus, we chose to employ motion capture technology during the collection of our corpus, as a more reliable and accurate way to record a precise level of movement detail from a human sign language performance.

Full details about our equipment configuration have been previously described in (Huenerfauth & Lu, 2010a), but this information is briefly summarized here. For our corpus, we record handshape; hand location; palm orientation; eye-gaze vector; and joint angles for the wrists, elbows, shoulders, clavicle, neck, and waist. We use a novel combination of commercially available motion-capture equipment for this project, which includes: two Immersion CyberGloves®, an Applied Science Labs H6 head-mounted eye-tracker, an Intersense IS-900 inertial/acoustic tracker (for tracking the location and orientation of the signer's head, which is necessary for calculating an eye-gaze vector in a room), and an Animazoo IGS-190 bodysuit which uses a set of magnetic/inertial sensors.

To facilitate synchronization of the three videos and the data stream from the Animazoo IGS-190 body suit, we asked the signer in the motion-capture equipment to perform a very quick head movement (turning the head to one side) immediately after the strobe light was flashed at the start of the recording; this action was easily identifiable in the videos and the motion-capture data.

Our motion-capture data was recorded using Autodesk MotionBuilder, which is a general-purpose 3D animation program that enables the input of motion-capture data streams during a live recording session. For the convenience of future researchers viewing the motion-capture data recording, we have also inserted a virtual human figure into the MotionBuilder data file with body proportions that are based on the human signer being recorded. The body segments of this virtual human are linked to the data streams of the motion-sensors on the body suit. If we had not inserted such a virtual human figure into the MotionBuilder file, then the data stream recorded would merely consist of the individual location and orientation values of the sensors on the body suit. There would be no easy way for future researchers to quickly visualize of the sensor data. Of course, the raw sensor data is also accessible in the MotionBuilder file if researchers require it.

Since we do not use motion capture techniques to record facial expressions of the signers being recorded, the virtual human figure inserted into the MotionBuilder files does not have any facial details. In addition, we only record the upper body movement (from the hip joint upwards) of the human signers while they sit on a stool, so the position of the legs of the virtual human character in the MotionBuilder file is not meaningful. The eye-tracking data recorded from the signers will require additional post-processing by our research team, and it has not been included in this initial corpus release.

To minimize errors in the motion-capture data we recorded, we carefully calibrated the cybergloves worn by signers; details of the cyberglove calibration protocol we have designed for use in sign language recording projects appears in (Lu & Huenerfauth, 2009). Evaluations of the resulting hand motion-capture accuracy we achieve with the cybergloves is also included in (Lu & Huenerfauth, 2009).

The Animazoo bodysuit system requires information about the lengths of the body segments (the bone lengths) of the human being recorded; this data is needed so that the system can determine how the human is posed during a recording session based on the data from the sensors placed on each segment of the body. Prior to the recording session, we measured the body proportions of signer by photographing each of them while standing in a cube-shaped rig of known size. In this way, we obtained bone-length information for each signer (which can be determined from the resulting photographs using software which accompanies the Animazoo system).

While great care was taken in calibrating the various motion-capture equipment, there are still some errors in the body position that are visually apparent in the motion-capture data. For instance, sometimes when one of the human's hands touches the other, it is apparent that the hands of the virtual human character do not touch precisely. So, there are some retargeting errors in the motion-capture data stream, which future researchers using this data may need to further process, depending on their research goals. We may seek additional methods of cleaning-up and post-processing the collected motion-capture data for our future corpus releases.

Sign Performed	fs-OSAMABINLADEN	IX-1-s:1	AMERICA	NUMBER	ONE	MOST	WANT	FIND	MAN	FINALLY	fs-US	CAPTURE	IX-1-s:1
SRP#1 Establishment		OSAMABINLADEN											
SRP#1 References		e											r

Figure 7: Example of a timeline from a passage from our corpus that contains an SRP.

The motion-capture data has been post-processed to adjust the timing synchronization of the motion-capture equipment. We found that it was challenging to perfectly synchronize the body movement data from the body suit and the hand movement data from the cybergloves, due to inconsistencies in the data transfer rate of the equipment and its small drift over time. To fix this timing issue, we asked researchers at our lab who are native ASL signers to watch the virtual human figures in our MotionBuilder files and to carefully edit (delay or advance) the timing of the glove data relative to the body suit data. In this way, we were able to verify that we have an accurate synchronization of the glove and body suit data streams; each of the recorded passages in our corpus was checked in this manner.

We are releasing our motion capture data in two files formats: FBX files and BVH files. FBX format files are the original file format owned by Autodesk and used by Autodesk MotionBuilder; this is the original recording file with the virtual human character (based on the human signer's body proportions) inserted into it. Next, we converted the FBX files into BVH files, which is a popular file format for 3D animation analysis and processing. BVH files are ASCII format files that contain two types of information: (1) a hierarchy of body segments sizes and joints for a figure and (2) rows of numerical data that correspond to information for all of the joints on a frame-by-frame basis. This corpus release also includes time mapping information between the motion-capture files and the videos for each signer.

6. Corpora Annotation

A team of native ASL signers (including students from deaf high schools in New York) annotated the data using the SignStream™ annotation tool (Neidle et al., 2000). The linguistic annotations for each passage have been cross-checked by at least two other native ASL signers on our research team. The long-term goal of our project is to annotate: sign glosses (with time alignment to the recorded video); part-of-speech of each gloss; syntactic bracketing (NP, VP, clause, sentence); and non-manual signals (role shift, negation, WH-word questions, yes-no questions, topicalization, conditionals, and rhetorical questions).

In addition, we annotate spatial reference points (SRPs) when they are established during a passage, which discourse entity is associated with each SRP, when referring expressions later refer to an SRP, and when any verbs are spatially inflected to indicate an SRP. These SRP establishments and references are recorded on parallel timeline tracks to the glosses and other linguistic annotations.

Figure 7 shows an example of a timeline from a passage from our corpus that contains an SRP; it is a timeline of an ASL passage discussing when Osama bin Laden was captured. In the example, the first time that the signer points to a location in 3D space around his body (glossed as "IX-1-s:1"), he establishes an SRP at that location to represent "Osama bin Laden." This SRP is referred to again later in the passage when the signer performs another "IX-1-s:1" sign. A loose translation of the passage in Figure 7 would be: "Osama bin Laden was America's No. 1 most wanted man; finally, the US captured him…"

Figure 7 shows the following rows of information:

- Row 1: Sign Performed: This row shows the sequence of glosses. While there is internal consistency in gloss labels used within our project, we have not employed a comprehensive system of "ID-glosses" like those of (Johnston, 2009). However, we may further standardize and edit our gloss notations in a future release of our corpus.
- Row 2: SRP#1 Establishment: This row indicates when the first spatial reference point ("SRP #1") is established by the signer somewhere in the signing space. When an SRP is established, then an annotation is added to this line with start- and end-times that align to the sign or phrase that established the existence of this SRP. The label of the annotation is meant to be a brief gloss of the entity referenced by this SRP. If there is a second SRP established in the signing space, then a new annotation row is added to the file for that additional SRP. Note that the integer after the colon at the end of the gloss "IX-1-s:1" indicates that the pointing sign is referring to SRP #1. A pointing sign directed at SRP #2 (if one were established) would appear as "IX-1-s:2". In this manner, each SRP is assigned an index number, and the gloss of each pronominal or verb-inflection reference to an SRP is marked with this index number (following a colon at the end of a gloss in the transcription).
- Row 3: SRP#1 References: This row indicates whenever a gloss or phrase in the passage references an SRP that has already been established in the signing space. Specifically, this row corresponds to SRP#1. On the first reference to a location for an SRP, this row receives an annotation with a label "e" (for "establishment"), and subsequent references to this SRP during the passage are indicated with an annotation added to this row with a label "r" (for "reference").

Figure 8 and Figure 9 illustrate the average number of SRP establishments (how many unique SRPs are established per passage) and the average number of SRP references per passage for each signer in our released corpus.

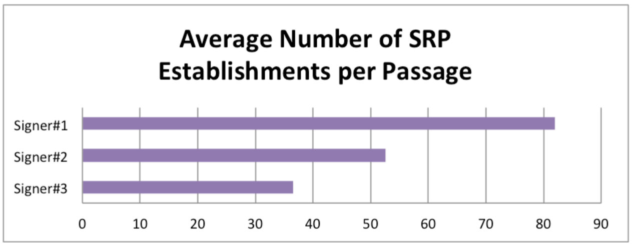

Figure 8: Average number of SRPs established per passage for each signer in our released corpus.

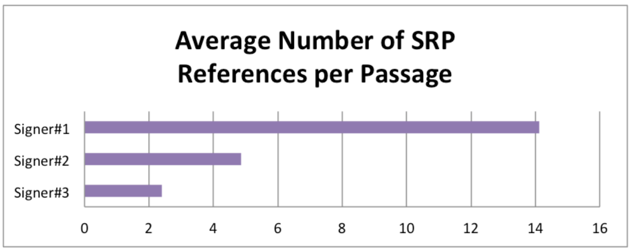

Figure 9: Average number of SRP references per passage for each signer in our released corpus.

In this first release of the corpus, we are distributing the time-aligned glosses, the annotation of the establishment of spatial reference points, and the English translation for each of the passages collected. We anticipate publishing additional layers of annotation in future corpus releases.

Text files of annotation information can be exported from the SignStream software, and we are including these plaintext versions of the annotation in this corpus release. The text files consist of all the annotation information and the file name of the video being annotated; each line in the file contains one type of annotation, such as glosses, SRP establishment, or SRP references. Each item of annotation is followed by its start and end frame numbers, corresponding to the video.

7. Sub-corpus of Inflected Verbs

A goal of our research on ASL animation is to design mathematical models of the movements of signers' hands during the production of inflected verbs (whose motion path and orientation is affected by how the SRPs for their subject and object have been set up in the signing space). In prior work (Lu & Huenerfauth. 2010b; Lu & Huenerfauth. 2011b), we needed larger numbers of examples of specific inflected verbs for all possible arrangements of subject and object in the signing space. This would not be possible to extract from our corpus because it is small in size and we would not be able to find all the possible combinations of each verb; so, we had to collect a special corpus of ASL verb movements.

For this corpus, we were only interested in obtaining information about the location and orientation of each of the signers' hands, not the information about head movement, eye gaze movement, or handshape. Thus, we used a motion-capture equipment configuration which was faster to set up, easier for the signer to put on, faster to calibrate, and easier to post-process. The trade-off is that less specific human body movement information was recorded, but this was sufficient for our ASL inflected verb research (Lu & Huenerfauth. 2010b; Lu & Huenerfauth. 2011b). Thus, we used our Intersense IS-900 system alone to record both head and hand data for our verb corpus. Previously, the IS-900 system was used only for head-tracking as part of our more complex equipment set-up for our unscripted multi-sentence corpus (section 5). The acoustical/inertial IS-900 system uses a ceiling-mounted ultrasonic speaker array (Figure 10) and a set of directional microphones on a small sensor to record its location and orientation.

For each verb, the signer was recorded performing it for different arrangements of the subject and object in the surrounding signing space. A set of color-coded squares were placed around the recording studio at various angles in a 180-degree arc in front of the signer; these targets were used as the subject and object SRPs for the various performances of inflected verb signs, for example, a white color target on the left could be the subject, and an orange color target on the right could be the object. We found this use of color targets in the room to be less error-prone than other approaches for collecting many samples of ASL inflecting verbs.

We have made use of these recordings in our prior research (Lu & Huenerfauth, 2011b) to produce models of the motion-path of ASL verbs, and we decided to also release this data to the research community – to facilitate the work of ASL linguistics or animation researchers studying ASL verbs. This "verb" corpus contains a high-resolution video recording of the signer during the collection and the plaintext data files from the IS-900, which consists of a tab-delimited file with columns for: the time code (milliseconds) and for each sensor: the location coordinates (x, y, z) and orientation (yaw, pitch, roll). Sensors were placed on both of the signer's hands, the signer's torso, and the signer's head. We recorded this "verb" corpus from three signers (different people than those recorded in the unscripted multi-sentential discourse discussed in section 5). This small corpus contains several hundred instances of eight ASL inflected verbs: ASK (one-handed version), GIVE, MEET, SCOLD, TELL, EMAIL, COPY, and SEND. The fact that we were able to make use of these recordings to produce models of ASL verb movement, which native ASL signers judged to be of good quality in an experimental study we conducted (Lu & Huenerfauth, 2011b), is good evidence that the quality of the motion-capture data is sufficient for supporting computational linguistic research on these verbs.

Figure 10: Close-up views of the hand-mounted sensor used in the motion capture sub-corpus data collection.

8. Conclusion

To address the lack of linguistically annotated ASL corpora with sufficient 3D movement detail for animation research, we began a multi-year project to collect and annotate a motion-capture corpus of ASL. In this paper, we are releasing the first portion of the "CUNY ASL Motion-Capture Corpus," which has been collected and annotated at our laboratory at Queens College of the City University of New York (CUNY). Our goal is for the digital 3D body movement and handshape data we collect from native signers to become a permanent research resource for NLP researchers, ASL linguists, and sign language animation researchers. This corpus will allow researchers to create new ASL generation technologies in a data-driven manner by analyzing the subtleties in the motion data and its relationship to the linguistic structure.

Our initial research focus is to model where signers tend to place spatial reference points in the signing space. Another early goal of our research is to discover patterns in the motion paths of inflecting verbs and model how they relate to layout of SRPs. These models we develop could be used in ASL generation software or could be used to partially automate the work of humans using ASL-scripting systems.

Because we are still collecting, post-processing and annotating this corpus, we plan to provide additional releases of this corpus in future years. This paper has suggested various additional forms of annotation and motion-capture data that we intend to release in the future, and we welcome feedback from the research community about how this resource can be made more useful and accessible.

9. Acknowledgements

This material is based upon work supported in part by the US. National Science Foundation under award number 0746556 and award number 1065009, by The City University of New York PSC-CUNY Research Award Program, by Siemens A&D UGS PLM Software through a Go PLM Academic Grant, and by Visage Technologies AB through a free academic license for character animation software. Jonathan Lamberton assisted with the recruitment of participants and the conduct of data-collection sessions. Kenya Bryant, Wesley Clarke, Kelsey Gallagher, Amanda Krieger, Giovanni Moriarty, Aaron Pagan, Jaime Penzellna, Raymond Ramirez, and Meredith Turtletaub have also assisted with data-collection and have contributed their ASL expertise to the project.

10. References

Cormier, K. (2002). *Grammaticalization of Indexic Signs: How American Sign Language Expresses Numerosity.* Ph.D. Dissertation, University of Texas at Austin.

Huenerfauth, M., Lu, P. (2010a). Eliciting Spatial Reference for a Motion-Capture Corpus of American Sign Language Discourse. *Proceedings of the Fourth Workshop on the Representation and Processing of Signed Languages: Corpora and Sign Language Technologies, The 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valetta, Malta.

Huenerfauth, M., Lu, P. (2010b). Modeling and Synthesizing Spatially Inflected Verbs for American Sign Language Animations. In *Proceedings of The 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2010),* Orlando, Florida, USA. New York: ACM Press.

Huenerfauth, M., Lu, P. (2012). Effect of Spatial Reference and Verb Inflection on the Usability of American Sign Language Animations. *Universal Access in the Information Society.* Berlin/Heidelberg: Springer.

Johnston, T. (2009). Creating a Corpus of Auslan within an Australian National Corpus. In *Michael Haugh et al. (eds.), Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus,* Somerville, MA: Cascadilla Proceedings Project, pp. 87--95.

Lu, P., Huenerfauth, M. (2009). Accessible motion-capture glove calibration protocol for recording sign language data from deaf subjects. In *Proceedings of The 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009),* Pittsburgh, Pennsylvania, USA. New York: ACM Press, pp. 83--90.

Lu, P., Huenerfauth, M. (2010). Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research. *Proceedings of the First Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics,* Los Angeles, CA.

Lu, P., Huenerfauth, M. (2011a). Collecting an American Sign Language Corpus through the Participation of Native Signers. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction (UAHCI).* Orlando, Florida, USA.

Lu, P., Huenerfauth, M. (2011b). Synthesizing American Sign Language Spatially Inflected Verbs from Motion-Capture Data. *The Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT), The 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2011),* Dundee, Scotland, United Kingdom.

Liddell, S. (2003). *Grammar gesture and meaning in American Sign Language*. UK: Cambridge University Press.

Meier, R. (1990). Person deixis in American Sign Language. In *S. Fischer, P. Siple (Eds.) Theoretical issues in sign language research, Vol 1, Linguistics.* Chicago: University of Chicago Press, pp. 175--190.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., Lee, R. (2000). *The Syntax of American Sign Language: functional categories and hierarchical structure.* Cambridge, MA: MIT Press.

Padden, C. (1988). *Interaction of morphology and syntax in American Sign Language.* New York: Garland.

Traxler, C. (2000). The Stanford achievement test, ninth edition: national norming and performance standards for deaf and hard-of-hearing students. *Journal of Deaf Studies & Deaf Education* 5(4), pp. 337--348.