A Colloquial Corpus of Japanese Sign Language: Linguistic Resources for Observing Sign Language Conversations

Mayumi Bono[1, 2], Kouhei Kikuchi[1, 3], Paul Cibulka[1, 3, 4], Yutaka Osugi[5]

[1] Digital Content and Media Sciences Research Division, National Institute of Informatics,
[2] Department of Informatics, School of Multidisciplinary Sciences, Graduate University of Advanced Studies
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 JAPAN
[3] Research Fellowship for Young Scientists, Japan Society for the Promotion of Science
5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083 JAPAN
[4] Department of Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg
Box 100, S-405 30, Gothenburg SWEDEN
[5] Division of Information Accessibility, Graduate School of Technology and Science, National University of Corporation Tsukuba University of Technology
4-3-15 Amakubo, Tsukuba, Ibaraki 305-8520 JAPAN
E-mail: {bono, vadise}@nii.ac.jp, paul.cibulka@gu.se, osugi@a.tsukuba-tech.ac.jp

Abstract

We began building a corpus of Japanese Sign Language (JSL) in April 2011. The purpose of this project was to increase awareness of sign language as a distinctive language in Japan. This corpus is beneficial not only to linguistic research but also to hearing-impaired and deaf individuals, as it helps them to recognize and respect their linguistic differences and communication styles. This is the first large-scale JSL corpus developed for both academic and public use. We collected data in three ways: interviews (for introductory purposes only), dialogues, and lexical elicitation. In this paper, we focus particularly on data collected during a dialogue to discuss the application of conversation analysis (CA) to signed dialogues and signed conversations. Our annotation scheme was designed not only to elucidate theoretical issues related to grammar and linguistics but also to clarify pragmatic and interactional phenomena related to the use of JSL.

Keywords: Japanese Sign Language (JSL), Annotation scheme, Conversation analysis for sign language

1. Introduction

We began building a corpus of Japanese Sign Language (JSL) in April 2011 with the support of the Japan Society for the Promotion of Science. The purpose of this project was to increase awareness of sign language as a distinctive language in Japan. This corpus is beneficial not only to linguistic research but also to hearing-impaired and deaf individuals, as it helps them to recognize and respect their linguistic differences and communication styles. This is the first JSL corpus developed for academic and public use.

The Deafness, Cognition and Language (DCAL) Research Centre,[1] based at University College London, conducted the research on which the British Sign Language Corpus (BSL corpus) is based.[2] We invited the principal investigator (Prof. Adam Schembri) of the BSL corpus project at that time to Japan to help us create such a corpus. Our initial steps in building a JSL corpus were based on advice from him and his colleagues.

However, no large-scale corpus such as the BSL corpus exists in Japan, although the Annotated JSL corpus, developed by the Central Research Laboratory, Hitachi, Ltd. (Koizumi et al., 2002) and the JSL Dialogue Corpus (KOSIGN) can be used for specific purposes (Kanda et al., 2001).

[1] http://www.ucl.ac.uk/dcal
[2] http://www.bslcorpusproject.org

2. The Corpus

2.1 Data collection

During the first stage of this project, from May to July 2012, we filmed 40 deaf subjects in two prefectures, Gunma and Nara, which are located about 50–100 km from Tokyo and Osaka, respectively. Each prefecture has one school for the deaf. We obtained data from an age-balanced sample of individuals 30–70 years of age in each prefecture, and each age group was divided into same-sex pairs. We used three approaches to collect data: interviews (for introductory purposes only), dialogues, and lexical elicitation. Each session, including our explanation of the ethical considerations and subjects' provision of written consent, lasted 1.5 h.

We streamed all videos of the dialogue task and lexical elicitation tasks on our website (Fig. 1).[3] The webpage was designed to be accessible to all people interested in JSL, as it provided simple explanations and included Japanese- and English-language pages. The main Japanese page featured a movie clip in JSL, and the main English page featured one in International Sign (IS).

In this paper, we focus particularly on data collected during a dialogue task to discuss the application of conversation analysis (CA) to signed dialogues and signed conversations. This task consisted of two parts. In the first part, a participant viewed an

[3] http://research.nii.ac.jp/jsl-corpus/en/

'animated narrative' ('Canary Row,' a 1950 cartoon featuring the characters Sylvester and Tweety) before filming began and then explained its content to another participant. In the second part, they confirmed their understanding by watching the animation together. To validate the breadth and significance of the data obtained in this dialogue task, we confirmed the presence of several expressions that naturally occur in sign language dialogue (e.g., classifiers (CL) in JSL (Ichida, 2005) and role shifts (RS) in narrative discourse (Padden, 1986; Morgan, 1999; Ichida, 2005)). We asked an expert sign language interpreter to count the CLs and RSs using her definitions of these phenomena. She identified 147 CLs and 24 RSs during the 9.5-min first part of this task. However, these data were not sufficient to form the scientific foundation of our corpus, and we therefore performed a more detailed analysis of these phenomena from the perspectives of conversation analysis and multimodal interaction studies. This analysis is discussed in section 3.

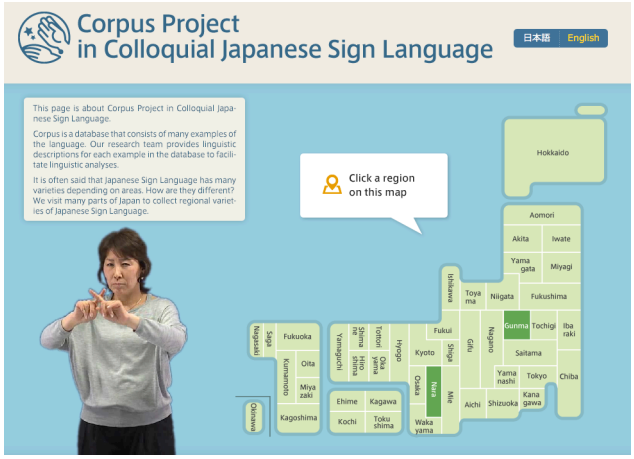


Figure 1: Web page

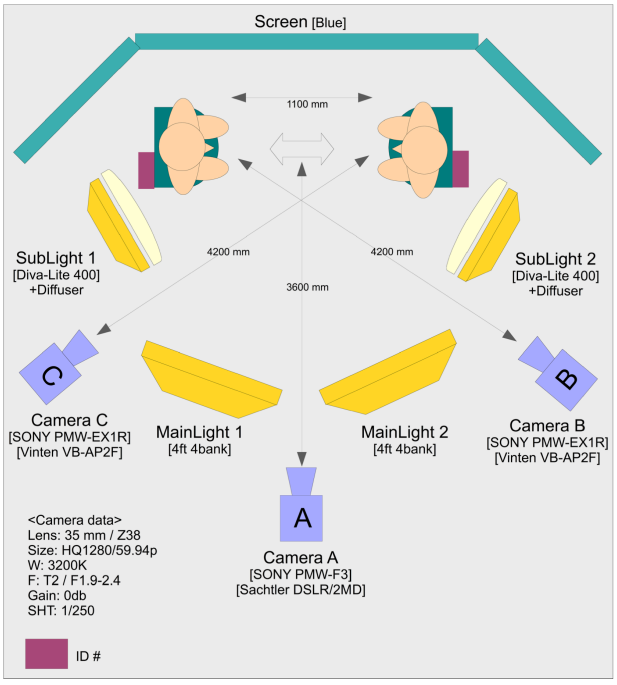


Figure 2: Camera setup

2.2 Procedure

We used three high-definition cameras, four lighting devices, blue panels, and blue chairs for the recordings (Fig. 2). During the dialogue task, camera A showed the two participants from the knees up; camera B focused on the participant on the left (hereafter, G-02), also showing the back of the other participant; and camera C focused on the participant on the right (hereafter, G-01), also showing the back of G-02. The camera angles and spatial configuration were designed to enable spatial reproducibility in the service of annotating gaze direction and pointing during the dialogues. We placed numbers on the back of each chair to identify each participant for purposes of data analysis.



Figure 3: Screen shot: the three camera angles used for data collection

2.3 Field workers

The main purpose of this project was to elucidate local signed representations. To eliminate the influence of spoken Japanese, signed Japanese and Tokyo standard JSL, we asked two field workers (FWs)[4] to perform the a a . During the recordings, the FWs observed participants perform the tasks (Fig. 4)[5] and answered questions about the tasks using local signs.



Figure 4: Image of data collection process

[4] This is not a FW in the framework of anthropology. We just called them like this according to the BSL corpus project. In fact, they are collaborators for building the corpus.

[5] This picture was taken during a rehearsal 1 month before the study began, and the blue back screens were not yet available.

3. Conversation Analysis

As mentioned above, one purpose of this study was to apply the concepts of CA (e.g., turn-taking systems (Sacks et al., 1974), repair sequences (Schegloff et al., 1977), etc.) to signed dialogues and signed conversations.

CA is the study of naturally occurring speech in social interactions. Sacks, Schegloff, and Jefferson (1974; SSJ) proposed several concepts related to turn-taking systems to analyze spoken conversational data.

We argue that these theoretical and methodological frameworks can be applied to the analysis of signed conversations. SSJ proposed the concept of a turn construction unit (TCU) (Fig. 5), which is a fundamental unit that differs from a sentence. They assumed that the participants in a conversation are able to anticipate whether the ongoing TCU will be closed by the current speaker. One TCU sometimes has several possible completion points; phrasal boundaries, intonation units, and so on. At the end of some TCUs, considered transition-relevance places (TRPs), one of three turn-allocation techniques is used: 1(a) the current speaker selects the next speaker; 1(b) self-selection occurs; or 1(c) the current speaker continues (Fig. 6).

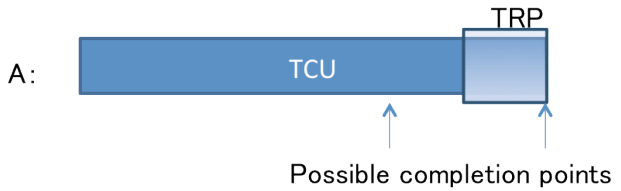


Figure 5: TCU and TRP (SSJ, 1974)

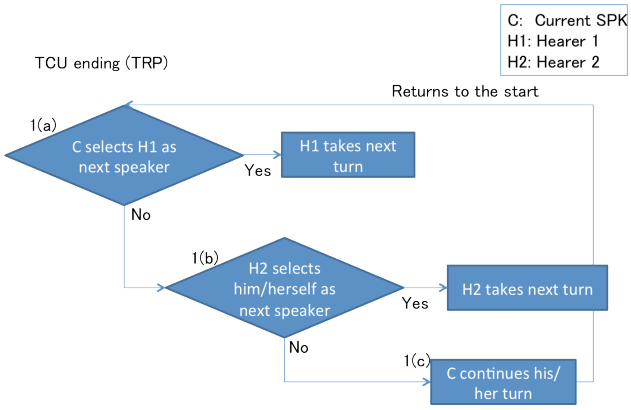


Figure 6: The rules of turn-taking systems (SSJ, 1974)

4. Annotation

Our annotation scheme was designed not only to elucidate theoretical issues related to grammar and linguistics but also to clarify the pragmatic and interactional phenomena involved in the use of JSL.

We used the annotation software ELAN[6], which was developed by the Max Planck Institute in Nijmegen, the Netherlands, to analyze TCUs and TRPs in signed conversations. Several tiers were used for each participant: TCUs (nearly equal phrases or sentences), words, movements of the right hand, movements of the left hand, gaze direction, mouth movements, non-manual signals (nodding), and non-manual signals (other than nodding) (Fig. 7). We separated nodding from other non-manual signals because a substantial amount of the nodding was synchronized with other non-manual signals in these data. When a non-manual signal was frequently synchronized with other non-manual signals, we increased the number of tiers accordingly.

One of our original contributions is our method for annotating hand movements and gaze direction. We applied the concept of a gesture unit (GU) proposed by Kendon (1972, 1980, 2004) to annotate the beginning and end points of signed turns. The GU is the interval between successive rests of the limbs, rest positions, or home positions. A GU consists of one or several gesture phrases. A gesture phrase is what we intuitively call a "gesture," and it, in turn, consists of up to five phases: preparation (optional), stroke (obligatory in the sense that a gesture is not said to occur in the absence of a stroke), retraction (optional), and pre- and post-stroke hold phases (optional). When analyzing overlapping communications in conversations, it is important to note the timing of the expressions of both the signer and recipient. In signed conversations, articulation involves hand signs that appear in front of the participants; this process of articulation is comparable to the visible lip movements made by those involved in spoken conversations. Using this methodology, we can observe how participants engage in an articulation phase in which signers move their hands to the signing space from the home position as a signal for the start of turn-taking in interactions.

4.1 Annotation of hand movements

We developed the following annotation scheme for the hand movements involved in signing a word:

prep: Preparation phase of signing. Signers raise their hands from the home or rest position to the signing space.
pre-s-h: Pre-stroke hold. The phase in which the hand shape and the hand position are sustained before the next stroke phase.
str: Stroke. The phase in which the core part of a sign is presented, with the hand changing shape and moving within the signing space.
post-s-h:[7] Post-stroke hold. The phase in which the hand shape and the hand position are sustained after the previous stroke phase.
ret: Retraction. The phase in which the hands are returned to the home position or rest.
hold: An independent holding phase.

4.2 Annotation of gaze direction

We established the following annotation scheme for gaze direction:

[6] http://tla.mpi.nl/tools/tla-tools/elan/

[7] We sometimes use "p-s-h" instead given the limited space on a tier.

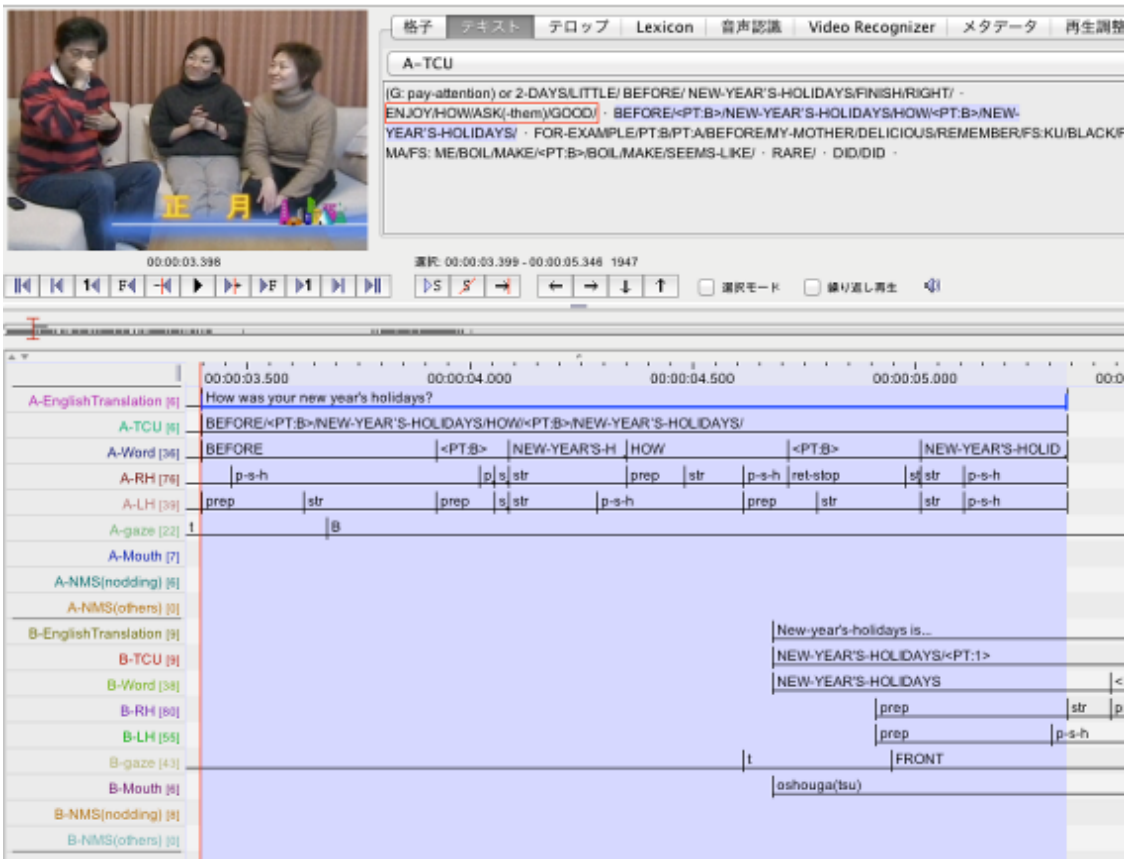


Figure 7: Annotated excerpt 1 according to ELAN

t: Transition phase of looking. The starting point of the gaze movement, which has not yet reached the target. During t, the gaze direction changes and has not settled.

(CAPITAL LETTER OR WORD): Settled phase of looking. For example, if signer A looks at interlocutor B, we assigned a code of "B" to the tier for A's gaze direction. If signer A looks toward the front, we assigned a code of "FRONT" to the tier for A's gaze direction (see B's gaze direction tier in Fig. 7).

5. Analysis

5.1 Analysis 1: What is a TCU in signed conversation?

Analysis 1 relies on a three-party signed conversation rather than on dialogue data (Fig. 7) because examining the three-party conversation offers a better way to explain a turn-taking system, especially turn-allocation techniques, in the CA framework. However, given the specific focus of this analysis, only the tiers for A and B are presented in Figure 7.

In the three-party conversation (Fig. 7), the male participant on the left is referred to as "A," the seated female participant in the middle is referred to as "B," and the female participant on the right is referred to as C.

In excerpt 1, one TCU includes two lines: the first is word information segmented by a slash and written in capital letters, and the second is an English translation written in italics.

Excerpt 1: Multi-unit turn

TCU1:
(G: pay attention) or 2-DAYS/ LITTLE/ BEFORE/ NEW-YEAR'S-HOLIDAYS/ FINISH/ RIGHT/
Recently, we have just finished New Year's holidays.

TCU2:
ENJOY/ HOW/ ASK(-them)/ GOOD/
I'll ask them how that was.

TCU3:
BEFORE/ <PT:B>/ NEW-YEAR'S-HOLIDAYS/ HOW/ <PT:B>/ NEW-YEAR'S-HOLIDAYS/
How were your New Year's holidays?

At the beginning of excerpt 1, participant A offers an explanation of the theme of this clip, new-years-day. In this utterance, A brings both his hands up from his lap at the beginning of the first TCU and returns them at the end of the third TCU. This utterance consists of three TCUs and is thus referred to as a multi-unit turn (Schegloff, 1981).

One of the most important points in the CA conception of a TCU is that all people who participate in a conversation can anticipate whether a TCU will close sooner or later based on projections about the actions performed by the current speaker. At the end of TCU1 in

this clip, interlocutors B and C nod their heads despite the fact that A is looking at the camera. We assume that B and C recognized some sort of boundary condition at this point, such as a phrase or sentence signed by A.

At the end of TCU3, A turns to B to address his utterance to B. B starts her answer before A finishes his turn. The microanalysis of this portion showed that he repeated the end of his question at this point; that is, he signed twice, <PT:B> and NEW-YEAR'S-HOLIDAYS, which is a part of 'How were your New Year's holidays?' Deaf people frequently repeat the end of questions; known as 'doubling' in sign language research, this is a common phenomenon in many sign languages across the world (Shimamura and Tieu, 2013).

Using our annotation, we can discuss the practical reason for this phenomenon. After A signs the stroke of HOW, B stops looking at A and turns to the front. While A is holding the sign for HOW, B starts to answer A's question. It seems as if the repeated signing of <PT:B> and NEW-YEAR'S-HOLIDAYS was ignored.

The annotated data reveal that B does not use A's repeated signing as a resource to continue the conversation. That is to say, doubling is a phenomenon related to sign production and is not a conversational resource for recipients because B was not able to observe A's repeated signing at this point.

5.2 Analysis 2: Signing and mouthing as resources of multimodal interaction

In analysis 2, we will use the dialogue data in our corpus. We chose the beginning part of dialogue between 50's female pair recorded in Gunma prefecture (Fig. 3). G-01 watched cartoon film before the recording, and G-02 did not.

Excerpt 2: Dialogue signing

LINE 01

G-01>

NOW/ pt1/ COMIC1/ THEATER-PLAY (m: a-ni-me) /pt1 /fs: A-NI-ME (m: a-ni-me) / SIGN-LANGUAGE/ WHAT/ pt2

Now, comic, theater play, animation... Animation, how do you sign?

I just watched a cartoon, uhm, how do you sign "cartoon"?

LINE 02

G-02>

COMIC1(m: a-ni=) /COMIC2(m: =-me) / COMIC1 (m: a-ni-me)/

Comic, comic, comic

Uhm, like this (I guess)?

LINE 03

G-01>

COMIC1/ COMIC2/ COMIC2/ COMIC1/ pt1/ WATCHED/ pt1/ WATCHED/ pt1/

Comic, comic, comic, comic⋯, (I) watched a comic, watched.

Okay, like this. So, I just watched a cartoon.

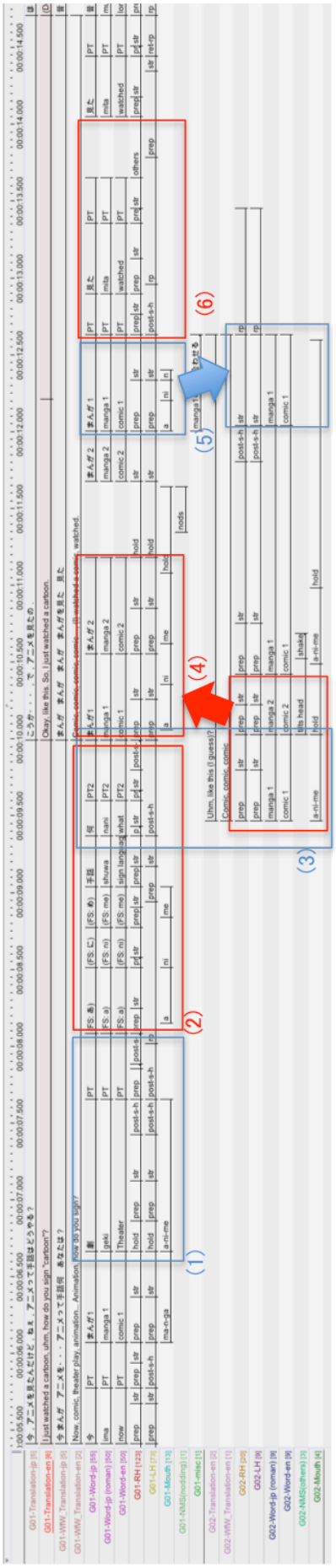


Figure 8: Annotated excerpt 2 according to ELAN

In excerpt 2, we inserted a line between the word line and the English translation. We refer to this line as a 'structured translation,' as it maintains the order in which the words were signed. However, although this line resembles a spoken sentence, it nonetheless differs from the translation.

Excerpt 2 shows that signers negotiate temporary signs when Japanese Sign Language (JSL) does not include an appropriate sign because the iconic nature of signed representations enables signers to create shared meanings. This analysis examines how signers create and negotiate the definition of a "word" during an ongoing conversation. We call these ongoing processes of defining a word "dialogue signing" in this analysis.

The use of ELAN with our annotation scheme underscored important observations related to the application of CA to signed dialogue and signed conversation. The following itemized numbers for observations coincide the focus places in Fig. 8.

(1) When G-01 represents THEATER-PLAY by mouthing 'a-ni-me,' which means animation (cartoon), she enacts a hold phase before the preparation phase of THEATER-PLAY. This word was maintained for a period in a post-stroke-hold (p-s-h). Indeed, disfluency is possible in sign production. Following the referential pointing, pt1 also contains a post-stroke-hold. In CA, disfluency in speech production can be one starting point for a self-repair sequence. We can observe that G-01 has a something trouble to go forward her narrative here.
(2) Next, following the phase observed in (1), she uses fingerspelling to write 'a-ni-me' while mouthing 'a-ni-me.' She emphasizes the fingerspelling, aligning it to the mouthing on a mora-by-mora basis. At this point, she interrupts the flow of her previous narrative to ask a question of interlocutor G-02, *'how do you sign "cartoon"?',* fs: a-ni-me (m: a-ni-me) / SIGN-LANGUAGE / WHAT/ pt2.
(3) While G-01 is executing the stroke for WHAT, G-02 starts to answer the question. This temporal relationship shows that G-02 anticipated and understood that G-01 had already stopped her narrative and had a question. She gleaned this from G-01's preparation of WHAT or her facial expression (i.e., raising her eyebrow).
(4) G-01 does not stop signing while G-02 answers. At first glance, this excerpt is not consistent with the perspective of CA, which treats conversations as interactions in which only one person speaks at a given time. Indeed, it appears to exemplify characteristic that are specific to signed conversations. That is, G-01 moves her hand following G-02's answer, COMIC1/ COMIC2. However, we consider this to be an instance of mirroring an interlocutor's signing rather than an independent turn. Furthermore, while signing COMIC1/ COMIC2, G-01 used one mouthing of 'a-ni-me' to represent two signed words, COMIC1/ COMIC2. In general, mouthing a word that is also signed emphasizes that word and/or specifies its meaning. In this case, one mouthing covers two signed words, which is not usual use of mouthing. The use of mouthing gives us an impression that this signing is not an independent turn in interaction.
(5) On the other hand, at the end of G-02's turn, COMIC1, G-02 imitates G-01's COMIC1, which she had signed slightly before this imitation. At this point, G-02 does not demonstrate a preparation phase; instead, G-02 continues signing COMIC1. It seems that G-02 is confirming that G-01 understands her opinion, namely, 'G-02 prefers COMIC1 for representing animation (cartoon).'
(6) G-01 then returns to the main theme of this dialogue, namely her narration of her experience of watching the cartoon, from pt1/ WATCHED/ pt1/ WATCHED/ pt1/, '*I just watched a cartoon*'.

This application of our annotation scheme to the dialogue data in the corpus revealed the operation of 'dialogue signing' at the beginning of signed narrative discourses and signed dialogue. We assume that several candidate words could represent the same meaning, and this phenomenon may underlie the regional differences in JSL. When we asked native signers and expert interpreters about the reason for this phenomenon, they identified three possibilities:

(1) Native signers in Japan, especially those living in the same region, do not have a shared knowledge about the words used in standard JSL.
(2) Due to the iconic nature of sign language, deaf children growing up with hearing parents often create "home signs" to communicate within the family (Frishberg, 1987; Morford, 1996; Torigoe and Takei, 2002; Goldin-Meadow, 2003). Home signs provide a way for people who do not have a common language (e.g., deaf children and hearing parents) to communicate. As a result, each signer has his/her own words for specific phenomena that are frequently referenced at home (e.g., "vegetables"). When they asked to represent their signings in front of cameras, there is a possibility they could not have self-confidences toward representing their own signings.
(3) Sign language does not have its own written language, which means that deaf people live in a bilingual environment. That is, deaf people in Japan use written Japanese to read and write and use JSL in their conversations with deaf interlocutors. Mouthing is a useful tool for specifying the meanings of signed words.

6. Conclusion

Our observations in analysis 2 are very complex, and the interactional practices in signed dialogue and signed conversations discussed herein would go unnoticed in the absence of our annotation scheme. This annotation

scheme and our analyses elucidated the temporal and sequential structures of answering in sign language, including bodily expressions and culture-specific ways of using body parts (e.g., mouth), which are important for understanding everyday conversations.

7. Appendix

Transcription conventions used in the excerpts:

pt1 pointing to signer-self, points at nose or breast
pt2 pointing to interlocutor/s
(m: a-ni-me) mouthing by mora rhythm unit
(fs: A-NI-ME) fingerspelling by mora rhythm unit
(m: a-ni=), (m: =-me) no gap between mora rhythm unit in mouthing (latching)

8. Acknowledgements

The authors would like to thank the fieldworkers in the local communities and the native signers for participating in our research as well as the staff of Rakuda Studio for helping us to record the data. This paper was supported by the Sign Language Corpus Project: For a comprehensive understanding of speech and sign language (Project ID #: 23320092, PI: Mayumi Bono); Grants-in-Aid for Scientific Research, Category B, General; the Japan Society for the Promotion of Science (JSPS), April 2011–March 2014, and the Documentation, preservation, and analyses of Japanese Sign Language lexical system from the perspective of deaf communities (Project ID #:25284075, PI: Yutaka Osugi); Grants-in-Aid for Scientific Research (B), General; and the Japan Society for the Promotion of Science (JSPS), April 2013–March 2016.

9. References

Frishberg, N. (1987). Home sign. In Cleve, J. V. (Ed.), *Gallaudet encyclopedia of deaf people and deafness*. Vol. 3, pp. 128--131. New York: McGraw Hill.

Goldin-Meadow, S. (2003). *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language*. NY: Psychology Press.

Ichida, Y. (2005). Shuwa no gengogaku (Linguistics of Sign Language). *Gengo (Language)* 34(1)-(12). Tokyo: Taishuukan-Shoten[In Japanese].

Kanda, K., Ichikawa, A., Nagashima, Y., Kato, Y., Terauchi, M., Hara, D. and Sato, M. (2002). Gesture and Sign Language in Human-Computer Interaction. *Lecture Notes in Computer Science*, 2298, pp.181--192.

Kendon, A. (1972). Some relationships between body motion and speech. In Siegman, A., and B. Pope. (Eds.), *Studies in Dyadic communication*, pp.177--210. Elmsford, NY: Pergamon Press.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In Key, M.R. (Ed.), *The relation between verbal and nonverbal communication*, pp. 207--227. The Hague: Mouton.

Kendon, A. (2004). *Gesture: Visible action as utterance. Cambridge*. UK: Cambridge University Press.

Koizumi, A., Sagawa, H. and Takeuchi, M. (2002). An Annotated Japanese Sign Language Corpus. In *Proceedings of 3rd International Conference on Language Resources and Evaluation (LREC)*. pp.927--930.

Morgan, G. (1999). Event packaging in Brithsh Sign Language Discourse. In Lucas, C. (Ed.) *Storytelling and Conversation: Discourse in Deaf Communities*. pp.27--58. Washington, DC: Gallaudet University Press.

Padden, C. (1986). Verbs and Role-shifting in ASL. In Padden, C. (Ed.) *Proceedings of the 4th National Symposium on Signing Research and Teaching*. Las Vegas, Nevada. Washington, DC: The National Association of the Deaf.

Sacks, H., Schegloff, E.A. and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, pp. 696--735.

Schegloff, E. A. Jefferson, G. and Sacks, H. (1977). The Preference for Self- Correction in the Organization of Repair in Conversation. *Language*, 53(2) pp.361--382.

Schegloff, E.A. (1981). Discourse as an Interactional Achievement: Some Uses of 'uh huh' and Other Things That Come Between Sentences, in Georgetown University Roundtable on Languages and Linguistics l98l; Tannen, D. (Ed.) (l982). *Analyzing Discourse: Text and Talk.* Georgetown University Press, pp. 7l--93.

Shimamura, K. and Tieu, L. (2013). When you can and can't see double: Revisiting focus doubling in ASL. Shwayder, K. (Ed.) In *Proceedings of the 36th Penn Linguistics Colloquium, University of Pennsylvania Working Papers in Linguistics*, 19(1), Article 22, pp. 189--196.

Torigoe, T. and Takei, W. (2002). A Descriptive Analysis of Pointing and Oral Movements in a Home Sign System, *Sign Language Studies*, 2(3), pp. 281--295.