JSL Colloquial Corpus:

Data Collection, Translation and Annotation Conventions

Mayumi Bono, Yutaka Osugi, Kouhei Kikuchi, Tomohiro Okada, and Mami Kaneko

Digital Content and Media Sciences Research Division, National Institute of Informarics, Japan 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430

Contact: jsl-corpus-mem@nii.ac.jp

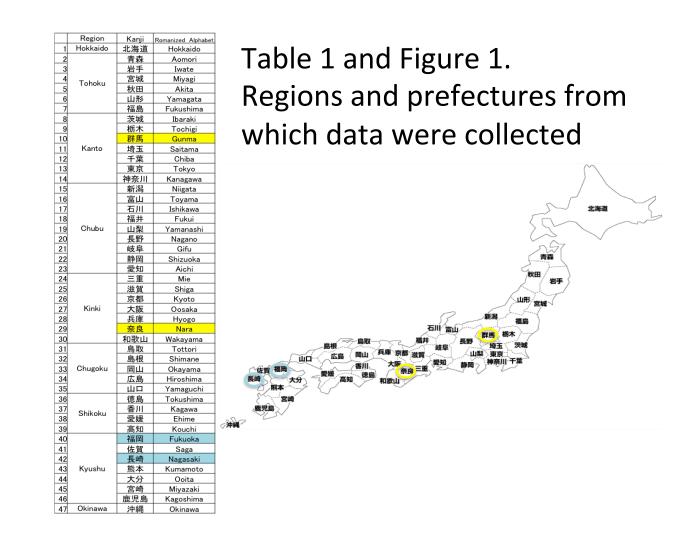
Introduction

This document describes the data collection, translation and annotation conventions used to build and share the JSP Colloquial Corpus. In addition, it presents our linguistic motivations toward sign language corpora by explaining several work procedures. 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. This is the first JSL corpus developed under the purpose of academic and public use.

In 2011, 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.

Data Collection

The first stage of this project was funded as Category B, 1,911,000JPY (10,462GBP) (PI: M. Bono) from 2011-2014 by the Japan Society for the Promotion of Science (JSPS). From May to July 2012, we videoed 40 deaf subjects in two prefectures, Gunma and Nara (colored in yellow in Fig. 1), each of the prefectures has one school for the deaf. We obtained data from an age-balanced sample of individuals aged 30–70 in each prefecture, and each age group was divided into same-sex pairs. Our participants from Gunma and Nara, were in their 30s, 40s, 50s, 60s and 70s, both male pairs and female pairs. We used three methods to collect data:



interviews, in which field workers, assistants of native signers living in the same area, and who knew the procedures in advance, asked participants about their language life, environment and so on (for introductory purposes only, not open access); dialogues about animation, in this procedure, one participant memorized the story of "Canary Row," and explained it to other participants; and lexical elicitation, in which participants showed correspondent signs for 100 slides of pictures and words shown on a monitor.

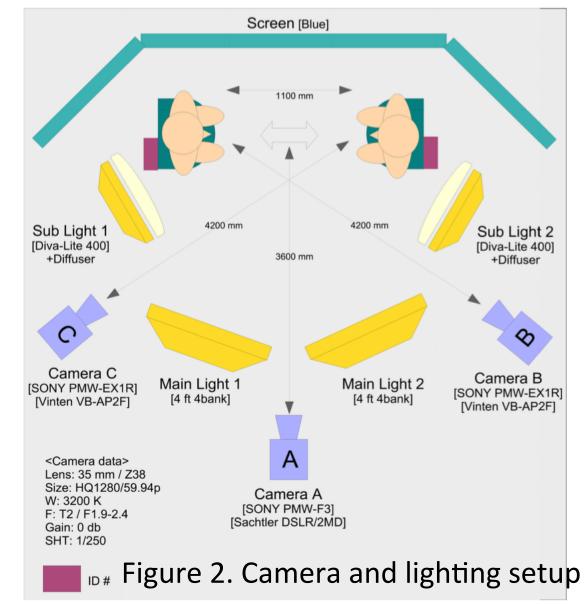
The second stage of this project currently is being funded as Category B, 1,320,000JPY (7,200GBP) (PI: Y. Osugi) from 2013-2016 by the Japan Society for the Promotion of Science (JSPS). From December 2014 to January 2015, we filmed 32 deaf subjects in two prefectures, Nagasaki and Fukuoka (colored in blue in Figure 1), each of the prefectures has several schools for the deaf. We obtained data based on school district. These were also age-balanced and divided into same-sex pairs. Our participants from Nagasaki and Fukuoka were in their 30s, 40s, 60s and 70s, both male pairs and female pairs. We used five tasks to collect data: *interviews*; *dialogues of their own curry recipes* (new), in which participants explained how each family cooked curry; *dialogue of okuni-jiman (proud of your country)* (new), in which participants talked about the points of pride in their local area; *dialogues about animation*; and *lexical elicitation*.

Because we found there were a lot of variations in the lexicons in home use and local use in our preliminary observation of the lexical elicitation task (e.g. vegetables) (Osugi and Bono, in press), we added two dialogue tasks: 'curry recipe,' and 'proud of your country,' to observe the use of lexicons related to private scenes (cooking, relaxing at home) and to collect local expressions in spontaneous dialogues in the second stage.

The PI of the first stage, Prof. Mayumi Bono, took the lead in drawing up translation and annotation conventions for the data set collected in dialogue tasks. The PI of the second stage, Prof. Yutaka Osugi, took the lead in analyses of lexical elicitation tasks to observe dialectic differences in each area. We streamed all videos of the dialogue task and lexical elicitation tasks on our website, http://research.nii.ac.jp/jsl-corpus/en/

Devices and Setting

We used three high-definition cameras, four lighting devices, blue panels, and blue chairs for the recordings (Fig. 2 & Fig. 3). During the dialogue task, camera A showed the two participants from the knees up; camera B focused on the participant on the left, also showing the back of the other participant; and camera C focused on the participant on the right, also showing the back of the other participant. 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 added a sound cue made by a clapperboard to each file for timing synchronization later.



Synchronizing and cropping of multiple video clips

Three independent files are synchronized using Final Cut Pro. The original combined-angles image includes the interlocutor's back recorded by cameras B and C; also there is dead space—showing as black areas. The cropped combined-angles image does not include the interlocutor's back and there is no dead space. Video images from all camera angles were enlarged to make then easy to see for detail analysis. All clips were made available on our website.

Camera B Camera A Camera A Camera A Camera A Camera A

Prefectures and Participants ID

There are 47 prefectures in Japan, and we assigned an ID to each prefecture using an abbreviated form by referring to the BSL corpus project.

In the animation task, the narrator who had watched "Canary Row" sat on the right (as we viewed the stage). The recipient who had not seen it sat on the left (as we viewed the stage). We placed numbers on the back of each chair to identify each participant for the purposes of data analysis. We set the abbreviation for the prefecture as shown in Table 2. The camera operator put the prefecture ID and the number of each participant on the back in order of appearance, e.g. GM01, narrator of the first pair in the animation task; GM02, recipient of the first pair in the animation task in Gunma Prefecture.

Figure 3. Two versions of the three camera angles used for data collection: Original (upper),
Cropped (lower)

Translation Conventions (1) Practical Steps

Basically, we perform two steps when making annotations: 1) translation into text: sign language interpreters translate sign language into written Japanese. They create Gloss, Word-order-translations, and Idiomatic translations in Microsoft Word. They then ask native signers who live in different regions of Japan to check the translations using their native sense. 2) Referring to the text in Word created in step 1, native signers annotate the features of hand movements for each unit of Gloss and units of utterances in ELAN to observe the temporal relationships between or within them.

After making the translation and annotations, we modify each file in a circulatory way using the findings noted in the working process of each, as shown in Fig. 3. We pay close attention to making the information in these files exactly the same.

Representation of Time

When translated into Word (Fig. 4), each line has an independent time axis. If line 01 and line 02 were to overlap, it would be difficult to represent their temporal relationships. Conversely, because annotations in ELAN (Fig. 5) have one time axis, it is easy to represent a micro-pause or show that two lines are overlapped. However, the annotation scheme of ELAN, due to the need to scroll to see the next and following lines, can make it difficult to find a sequential relationship, e.g. question and answer in speech act theory, conversation analysis (CA), and the discourse structure of narratives. We are trying to not only collect data but also build a hybrid notation system like a 'transcript' in CA (Jefferson, 1986) in this project.

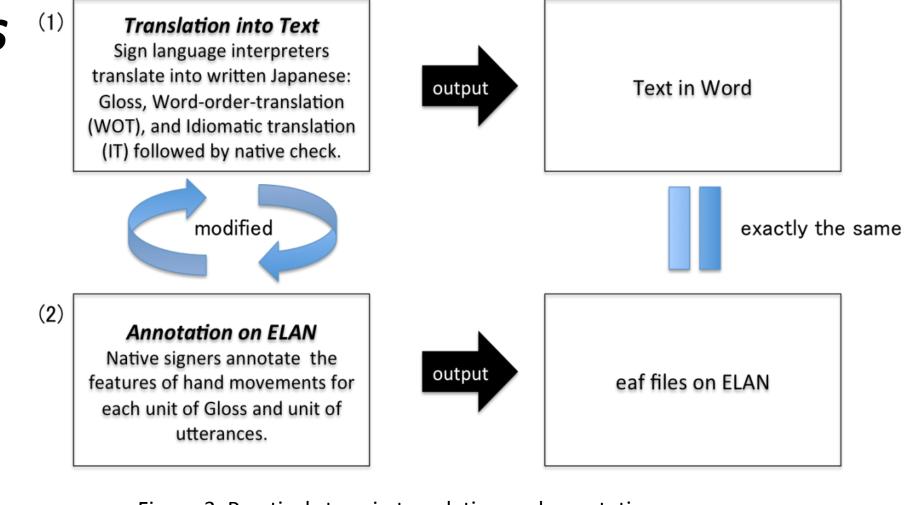
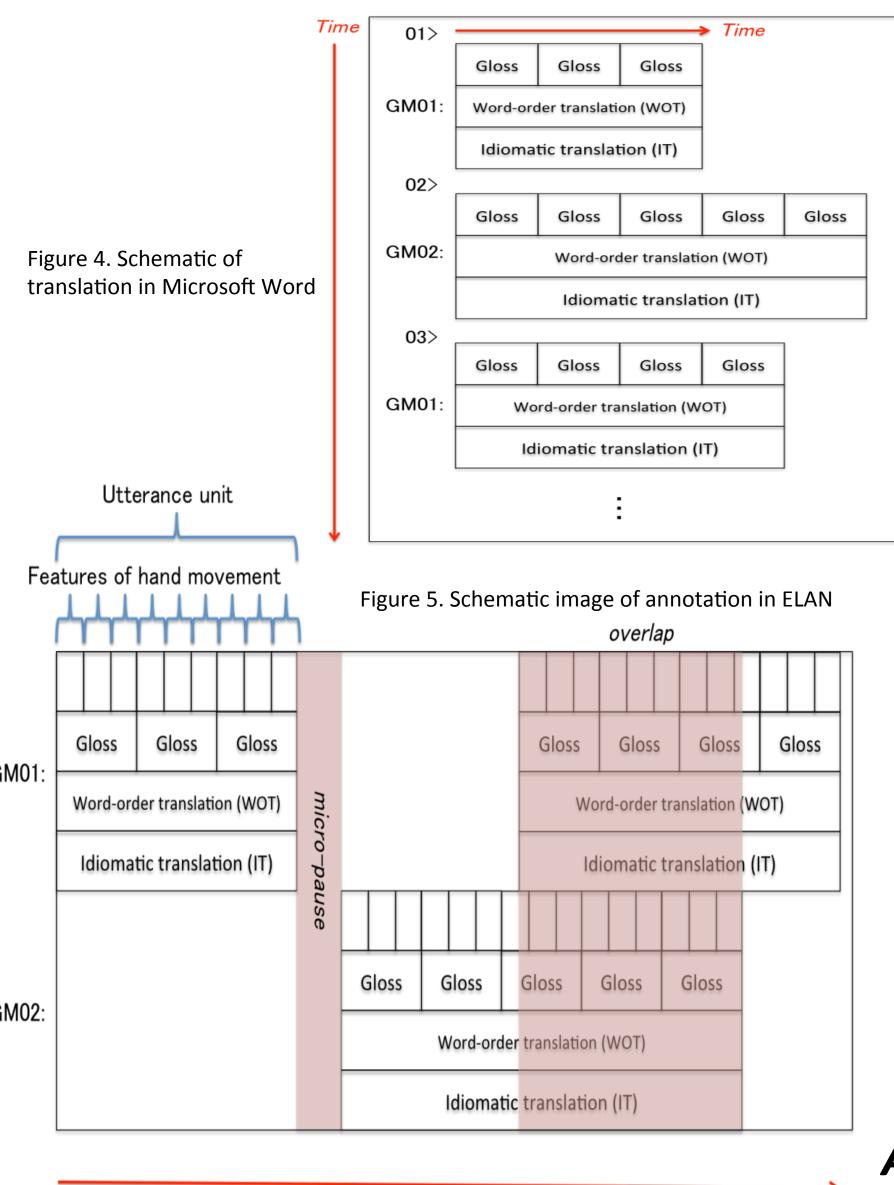


Figure 3. Practical steps in translation and annotation



Time

Annotation Conventions: two core concepts

Applying gesture phases to signing movements

One of our original points was to establish a physical and hand movement unit smaller than Gloss, called a Movement Unit. We applied the concept of the 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 preand 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 turntaking in interactions.

Movement phases

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.

tr: 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: Post-stroke hold. The phase in which the hand shape and the hand position are sustained after the previous stroke

ret: Retraction. The phase in which the hands are returned to the home position or rest.

hold: An independent holding phase.

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

- GM01-RH [123] GM01-LH [73] 中 GM01-Gloss-jp [55] GM01-Gloss-jp (roman) [50] - GM01-Gloss-en [50] Ф¬ GM01-Word_Order_Translation-jp [5] GM01-Word_Order_Translation-jp (roman) [4] GM01-Word_Order_Translation-en [2] GM01-Grammatical_Gloss-en [4] GM01-Idiomatic_Translation-jp [5] GM01-Idiomatic_Translation-jp (roman) [5] GM01-Idiomatic_Translation-en [5] GM01-gaze [1] GM01-Mouth [13] GM01-NMA(nodding) [1] GM01-NMA(others) [1] GM01-misc [1] GM02-RH [20] GM02-LH [9] GM02-Gloss-jp [9] GM02-Gloss-jp (roman) [9] GM02-Gloss-en [9] GM02-Word_Order_Translation-jp [2] GM02-Word_Order_Translation-jp (roman) [2] GM02-Word_Order_Translation-en [1] GM02-Grammatical_Gloss-en [2] GM02-Idiomatic_Translation-jp [2] GM02-Idiomatic_Translation-jp (roman) [2] GM02-Idiomatic_Translation-en [2] GM02-gaze [0] GM02-Mouth [4] GM02-NMA(nodding) [1] GM02-NMA(others) [3] GM02-misc [0]

Applying the turn constructional unit to signed interactions

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), which is a fundamental unit that differs from a sentence. SSJ 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).

An utterance consists of one or several Glosses. The annotators who are native signers who segment using their native sense. This is close to utterance and GU as mentioned above. *TUC* is represented by seven tiers, including three notations: Word_Order_Translation, Grammatical Gloss and Idiomatic_Translation on ELAN (Figure 3). Translations have multilingual environments as well as Gloss.

Translation

Translation at the Gloss level is a common method of learning sign language for beginners and training in interpretation skills at an advanced level.

EXAMPLE: 01> GM01: 今/PT/まんが1 (M:ma-n-ga)/劇 (M:a-ni-me)/PT/= (Gloss-JPN) =FS:ア(M:a)ニ(M:ni)メ(M:me)/手話/何/PT:G02/ NOW/PT/COMIC1(M:ma-n-ga)/THEATER-PLAY (M:a-ni-me)/PT/= (Gloss-ENG) =FS: A(M:a)NI(M:ni)ME(M:me)/SIGN-LANGUAGE/WHAT/PT:G02/ 今 まんが アニメを…アニメって手話は何 あなたは? (WOT-JPN) Now, comic, theater play, animation... Animation, how do you sign? (WOT-ENG) 今、アニメを見たんだけど、ねえ、アニメって手話はどうやる? (IT-JPN) (IT-ENG) I just watched a cartoon, uhm, how do you sign "cartoon"?