

# An Annotated Japanese Sign Language Corpus

Atsuko Koizumi , Hirohiko Sagawa and Masaru Takeuchi

Central Research Laboratory, Hitachi, Ltd.

1-280, Higashi-koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan

E-mail: {koizumi, h-sagawa, mtakeuch}@crl.hitachi.co.jp

## Abstract

Sign language is characterized by its interactivity and multimodality, which cause difficulties in data collection and annotation. To address these difficulties, we have developed a video-based Japanese sign language (JSL) corpus and a corpus tool for annotation and linguistic analysis. As the first step of linguistic annotation, we transcribed manual signs expressing lexical information as well as non-manual signs (NMSs) - including head movements, facial actions, and posture - that are used to express grammatical information. Our purpose is to extract grammatical rules from this corpus for the sign-language translation system underdevelopment. From this viewpoint, we will discuss methods for collecting elicited data, annotation required for grammatical analysis, as well as corpus tool required for annotation and grammatical analysis. As the result of annotating 2800 utterances, we confirmed that there are at least 50 kinds of NMSs in JSL, using head (seven kinds), jaw (six kinds), mouth (18 kinds), cheeks (one kind), eyebrows (four kinds), eyes (seven kinds), eye gaze (two kinds), body posture (five kinds). We use this corpus for designing and testing an algorithm and grammatical rules for the sign-language translation system underdevelopment.

## 1. Introduction

Linguistic annotation of video data is essential for linguistic analysis of signed languages. Several types of tools are now available or under development to allow linguistic annotation of video-based language data (Neidle, 2000; SignStream, 2001).

We are developing a method for recognizing Japanese Sign Language (JSL) and translating into Japanese (Sagawa, 2001; Sagawa, 2000; Xu, 2000). To translate JSL into Japanese, the relationships between signed words have to be analyzed based on JSL grammar. Since the linguistic phenomena in JSL have not been clarified enough for this purpose, we have developed a video-based JSL corpus and a corpus tool that provides functions for annotation and linguistic analysis.

## 2. Linguistics characteristics of JSL

Japanese sign language (JSL) has three main characteristics:

### (1) Use of non-manual signs

JSL uses almost all parts of the upper body. Along with manual signs expressing lexical information, non-manual signs (NMSs) - including head movements, facial actions, and posture - are used to express grammatical information.

### (2) Realistic description

A lot of JSL words are a realistic description of things or events. For example, a JSL word meaning “rain / to rain” is a description of a rain scene by moving both hands up and down. “Heavy rain” is expressed by strengthening hand movement accompanied by non-manual signs. Namely, a sentential expression often reflects a realistic image of a specific event in the signer’s mind.

### (3) Use of space

JSL makes good use of three-dimensional space. At the lexical level, a JSL word consists of hand-shape, relative position of a hand to the body, and hand movement.

These elements can be changed to express additional meaning such as number, aspect, and person. At the syntactic level, a noun followed by a pointing action is located at the pointed place and this spatial information is used to express coreference.

From the viewpoint of these characteristics of JSL, we have developed a video-based corpus of JSL utterances by native signers. For each utterance, a video segment is associated with a detailed transcription of manual signs as well as NMSs.

## 3. Collection of elicited data

### 3.1. Requirements

Spontaneous utterances in an interactive dialogue would provide naturalistic data. However, since our primary concern is to collect data in order to extract grammatical rules for our sign-language translation system, we need the following kinds of elicited data.

#### (1) Contrasting examples

We need to collect contrasting examples for the analysis of the basic structure of JSL sentences, for example, to observe linguistic phenomena regarding modification and juxtaposition.

(e.g.)

I nod CAMERA nod BOOK nod BUY  
*I'll buy a camera and a book.*

I nod CAMERA BOOK nod BUY  
*I'll buy a book on cameras.*

#### (2) Various forms

Various forms expressing the same meaning are also necessary.

(e.g.)

I nod CAMERA BOOK nod BUY  
*I'll buy a book on cameras.*

I nod CAMERA PT BOOK nod BUY  
*I'll buy a book on cameras.*  
("PT" indicates the manual sign of pointing.)

### (3) Variety of signers

Utterances signed in the same form by different signers are necessary in order to test grammatical hypotheses.

### (4) Repeated utterances

Utterances repeated by the same signer are necessary in order to test our method of sign-language recognition.

## 3.2. First procedure for data collection

Our procedure for data collection consists of five steps as explained in the following.

### (1) Preparation

To analyze the basic structure of JSL sentences, we prepared 360 Japanese sentences that show basic grammatical phenomena, including complementation, modification and juxtaposition, tense and aspect, and modality. For each example sentence, we prepared a cue sentence for eliciting natural expressions. First, we prepared question sentences like "What did you buy?" to extract declarative sentences like "I bought a book". However, we found that cue sentences using meta expressions like "Explain what you bought" are more appropriate. Responding to the cue sentence "What did you buy?", informants tend to use expressions like "A book" or "What I bought was a book".

### (2) Filming

For each example sentence, we collected utterances by three or four native signers. As JSL makes good use of space and eye gaze, a signer can perform more natural utterances in the presence of another native signer who signs the cue sentence for each example sentence.

### (3) Checking

Native signers and interpreters compared utterances by different signers in order to find variations as well as ungrammatical or unnatural expressions. For judging grammaticality and naturalness, we sometimes had to consult more than two native signers.

### (4) Filming of variations

We asked signers to imitate expressions performed by other signers. Though our primary purpose was to collect data from different persons, this imitation process was also effective for finding ungrammatical or unnatural expressions.

### (5) Filming of repeated utterances

For some sentences, we collected repeated utterances for use as test data. Repeating utterances helps informants to be more sensitive to their own expressions.

Accordingly, we collected 2500 utterances corresponding to 360 example sentences.

## 3.3. Second procedure for data collection

In order to get more natural utterances, we collected utterances by presenting assumed scenes, situations, and intentions. We collected 300 utterances by two native signers and found characteristic expressions of JSL. For example, an utterance inspired by "SCENE: hospital, SITUATION: history taking, INTENTION: to explain that you sprained your ankle" was "While riding a bicycle, I fell off and sprained my left ankle." Realistic description in chronological order seen in the utterance is a typical JSL expression.

Because of the realistic descriptions, it is hard for a native signer to produce a sentence without a specific situation; namely, the following conditions are required:

- (1) Time (especially when the topic relates to time)
- (2) Place (especially when the topic relates to place)
- (3) Relationship between the signer and other people mentioned in the sentence
- (4) Location of objects

## 4. Annotation

It is impossible to completely transcribe a signed utterance. The degree of preciseness in annotation depends on the purpose. This section explains the annotation required for grammatical analysis.

### 4.1. Manual Signs

A manual sign consists of handshape, relative position of a hand to the body, and hand movement (Kanda, 1994). Though describing these elements is called phonological description, a set of phonemes in JSL has not been well defined. For grammatical analysis, phonological description is not always necessary. Information for identifying a manual sign and agreement inflection is essential for grammatical analysis. Information regarding other morphological changes would be useful for further analysis. We annotate the following information in our corpus.

#### (1) A label for identifying a manual sign

We use a Japanese equivalent<sup>1</sup> as a label for a manual sign because it is easy to describe and recognize. In the case of using a Japanese equivalent, labellers should fully understand that a Japanese label is used as a unique identifier, not as an equivalent word in the context. For example, a manual sign meaning "to teach" and "teacher" should be given the same label. In case that more than one manual sign correspond to the same Japanese word, we added a mnemonic character after the Japanese word so that they can be distinguished.

#### (2) Agreement inflections

An agreement inflection is indicated by suffix (e.g. The starting position and end position of a manual sign

---

<sup>1</sup> In this paper, we use an English equivalent word for readers' convenience.

“GIVE” inflects according to the persons of the actor and the recipient. An inflected form “GIVE<sub>RI</sub>” indicates that the movement is from the right (i.e. third person) to the signer (i.e. first person).

### (3) Aspectual inflections and classifiers

Aspectual inflections are described in parentheses (e.g. READ(repetition)). Classifiers are indicated by “CL” (e.g. CL(pile of books), CL(array of books)).

## 4.2. Non-manual Signs

Movements involving the head and upper body have the following possibilities.

(1) Grammatical markings expressing syntactic information and adverbial information

(2) Lexical markings associated with specific manual signs (e.g. pointing the signer’s mouth means “teeth” if the teeth are seen, while the same manual sign means “mouth” if the mouth is closed.)

(3) Affective markings

(4) Meaningless movements such as natural blink and involuntary tilt or movements physically influenced by manual signs

For grammatical analysis, only (1) and (2) are necessary and they should be annotated distinctively from each other. We annotated grammatical markings and lexical markings as NMSs in our corpus.

Since kinds and functions of NMSs in JSL have not been clarified, it is often difficult to judge if two similar movements should be treated as the same NMS or not. We treated them as different NMSs if they are recognized as different NMSs by native intuition, even if distinctive functions have not been found. Actually, we found some signers using head shaking as a negation marker and jaw shaking as an interrogative marker, while some did not use them distinctively.

## 5. Corpus tool

We developed a corpus tool for annotation and analysis of JSL utterances captured on video. This tool displays the following information (see Figure 5.1):

(1) Video images: three synchronized video images (upper body, face, and side view of face) can be viewed by splitting the field into four sections

(2) Sign language animation based on the data from a glove-based input device

(3) Japanese translation

(4) Time scale: clicking and dragging the mouse cursor along the time scale can replay video images from a specified period

(5) Manual signs: sign language words expressed using hands are annotated along the time scale

(6) Non-manual signs: NMSs expressing grammatical information are annotated in the relevant fields (head, jaw, mouth, cheeks, eyebrows, eyes, eye gaze, and body posture) set along the time scale

Video files were reviewed and annotated by native signers and interpreters. Linguistic tags were synchronized with video frames after the annotation.



Figure 5.1 Sign-language annotation tool

### 5.1. Functions for Annotation

This tool has the following functions for annotation.

(1) Display of candidates for word boundaries  
Notches on the time scale indicate segmentation boundaries detected by gesture segmentation method (Sagawa, 2000), which is used in our JSL recognition system. Segmentation boundaries are detected by analyzing gesture information inputted through a glove-based input device and a magnetic sensor device. This function helps to improve efficiency and consistency in deciding the manual sign boundaries.

(2) Application of annotated data as a template  
An existing annotation for a similar expression (especially, a repeated utterance) can be a good template. This function retrieves the annotation information for the specified data. Since the annotator shifts or changes labels only if necessary, this function leads to efficient and consistent annotation.

### 5.2. Functions for Linguistic Analysis

This tool has the following functions for linguistic analysis

(1) Retrieval of data including the specified label  
This function is used for finding examples of a manual or non-manual sign. Since the file names are listed on the screen, the user can view the annotated video data by selecting a file name. In accordance with the user’s instruction, transcription of manual signs (and non-manual signs, if specified) can be output to a text file.



Figure 5.2 Display of comparison data

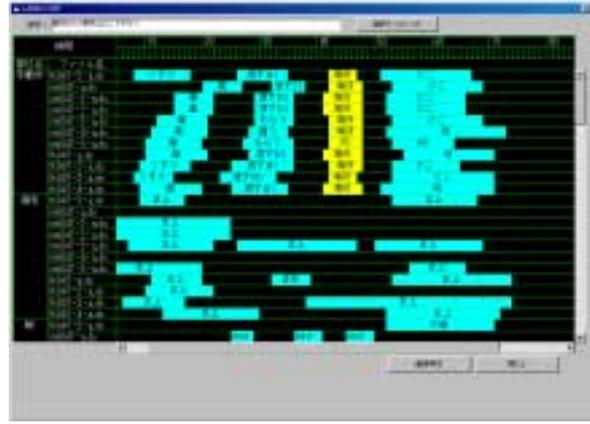


Figure 5.3 Alignment of comparison data

## (2) Comparison of data

This function is used for comparing utterances that correspond to the same example sentence. Comparison data is displayed in accordance with the user's specification on categories of signs (i.e. manual signs, eyebrows, eyes, eye gaze, mouth shape, jaw, cheeks, head, posture) and file names (see Figure 5.2). Labels can be shifted so that the specified labels are aligned. In Figure 5.3, labels are shifted so that the third manual signs are aligned. Video files can be viewed by clicking file names that appear at the left in Figure 5.2 and 5.3.

## 6. Conclusion

As the result of annotating 2500 utterances collected by the first procedure (mentioned in 3.2), we confirmed that Japanese sign language has at least 50 kinds of non-manual signs using head, jaw, mouth, cheeks, eyebrows, eyes, eye gaze, and body posture. Our experiments confirmed the following kinds of non-manual signs.

- (1) HEAD (seven kinds): nod, tilt, shake, turn away, move backward, move forward, reverse nod
- (2) JAW (six kinds): raise, drop, pull in, move backward, move forward, move sideways
- (3) MOUTH (18 kinds): hold shape, closed, closed (indicating subject), open, open (indicating subject), closed pout, closed grin, open pout (shape of [u]), open grin (shape of [i]), round (shape of [o]), protruding grin, turn down angulus oris, poke tongue out, shape of [pa], shape of [pi], shape of [pu], shape of [pe], shape of [po]
- (4) CHEEKS (one kind): puff out
- (5) EYEBROWS (four kinds): turn up, turn down, strain sideways, knit eyebrows
- (6) EYES (seven kinds): blink, open, close, gaze, squint, squint in one eye, wink
- (7) EYE GAZE (two kinds): look at object, tilt away to think
- (8) BODY POSTURE (five kinds): move backward, move forward, tilt, move upward, move downward

Since the utterances collected by the second procedure (mentioned in 3.3) were expressed more spontaneously, they include a variety of modality expressions using NMSs. As the result of annotation, we confirmed that the above 50 kinds of NMSs are enough to annotate these utterances, though we found these NMSs have more functions when used on their own as well as in combination.

We have confirmed the effectiveness of the corpus and the tool by using for designing and testing an algorithm and grammatical rules for the sign-language translation system underdevelopment. In particular, we confirmed that the distinction of NMSs for head movements and jaw movements proposed in this paper is appropriate by applying to a recognition method of JSL sentences.

## 7. Acknowledgements

The research reported here was carried out within the Real World Computing Project, supported by Ministry of Economy, Trade and Industry.

## 8. References

- Sagawa, H., Koizumi, A., and M. Takeuchi, 2001. A Recognition Method of Japanese Sign Language Sentences Based on Head Movements. *Proceedings of HCI2001*.
- Sagawa, H. and M. Takeuchi, 2000. A Method for Recognizing a Sequence of Sign Language Words Represented in a Japanese Sign Language Sentence. *Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition (FG 2000)*.
- Xu, M., Raychev, B., Sakaue, K., Hasegawa, O., Koizumi, A., Takeuchi, M., and H. Sagawa, 2000. A Vision-Based Method for Recognizing Non-Manual Information in Japanese Sign Language. *Proceedings of the Third International Conference on Multimodal Interfaces (ICIM2000)*.
- Neidle, C., Kegl, J., JacLaughlin, D., Bahan, B., and R.G. Lee, 2000. *American Sign Language*, MIT Press
- SignStream, 2001. <http://www.bu.edu/asllrp/SignStream>
- Kanda, K., 1994. *Lectures on Sign Language Study*. Fukumura Publisher. [In Japanese]
- Ekman, P., and E. Rosenberg (Eds.), 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press.

Sign language is characterized by its interactivity and multimodality, which cause difficulties in data collection and annotation. To address these difficulties, we have developed a video-based Japanese sign language (JSL) corpus and a corpus tool for annotation and linguistic analysis. As the first step of linguistic annotation, we transcribed manual signs expressing lexical information as well as non-manual signs (NMSs) - including head movements, facial actions, and posture - that are used to express grammatical information. Our purpose is to extract grammatical rules from this corpus for t

Keywords: corpus linguistics, corpora, annotation, sign language, language documentation, Auslan (Australian Sign Language). 1 Introduction A modern linguistic corpus is something more than just a reference dataset of written or transcribed texts of a particular language on which a description of a language is based.Â The aim is to create an annotated signed language corpus, and not, contrary to the practice of many signed language researchers, a body of signed language texts which have been transcribed to a greater or lesser degree of detail. The reason is that one can now use multi-media annotation software to transform a video recording of signed language into a machine-readable text without it first being necessary to transcribe that text.