

# Proper Name Machine Translation from Japanese to Japanese Sign Language

**Taro Miyazaki, Naoto Kato, Seiki Inoue,  
Shuichi Umeda, Makiko Azuma, Nobuyuki Hiruma**  
NHK Science & Technology Research Laboratories  
Tokyo, Japan  
{miyazaki.t-jw, katou.n-ga, inoue.s-li,  
umeda.s-hg, azuma.m-ia, hiruma.n-dy}@nhk.or.jp

**Yuji Nagashima**  
Faculty of Information,  
Kogakuin University  
Tokyo, japan  
nagasima@cc.kogakuin.ac.jp

## Abstract

This paper describes machine translation of proper names from Japanese to Japanese Sign Language (JSL). “Proper name transliteration” is a kind of machine translation of proper names between spoken languages and involves character-to-character conversion based on pronunciation. However, transliteration methods cannot be applied to Japanese-JSL machine translation because proper names in JSL are composed of words rather than characters. Our method involves not only pronunciation-based translation, but also sense-based translation, because kanji, which are ideograms that compose most Japanese proper names, are closely related to JSL words. These translation methods are trained from parallel corpora. The sense-based translation part is trained via phrase alignment in sentence pairs in a Japanese and JSL corpus. The pronunciation-based translation part is trained from a Japanese proper name corpus and then post-processed with transformation rules. We conducted a series of evaluation experiments and obtained 75.3% of accuracy rate, increasing from baseline method by 19.7 points. We also developed a Japanese-JSL proper name translation system, in which the translated proper names are visualized with CG animations.

## 1 Introduction

Sign language is a visual language in which sentences are created using the fingers, hands, head,

face, and lips. For deaf people, sign language is easier to understand than spoken language because it is their mother tongue. To convey the meaning of sentences in spoken language to deaf people, the sentences need to be translated into sign language.

To provide more information with sign language, we have been studying machine translation from Japanese to Japanese Sign Language (JSL). As shown in Figure 1, our translation system automatically translates Japanese text into JSL computer graphics (CG) animations. The system consists of two major processes: text translation and CG synthesis. Text translation translates word sequences in Japanese into word sequences in JSL. CG synthesis generates seamless motion transitions between each sign word motion by using a motion interpolation technique. To improve the machine translation system, we have been tackling several problems with translating in JSL. In this paper, we focus on the problem of proper name translation, because proper names occur frequently in TV news programs and are hard to translate with conventional methods.

Proper name translation is one of the major topics of machine translation. In particular, there are many methods that work with spoken language, such as “proper name transliteration,” which means character-to-character conversion based on pronunciation (Knight et al., 1998; Goto et al., 2003; Virga et al., 2003; Li et al., 2004; Finch et al., 2010; Sudoh et al., 2013). However, transliteration methods cannot be applied to Japanese-JSL proper name translation because proper names in JSL are not composed of characters but rather of sign words. To translate proper names using sign words, sense-based translation is required. Sense-based translation trans-

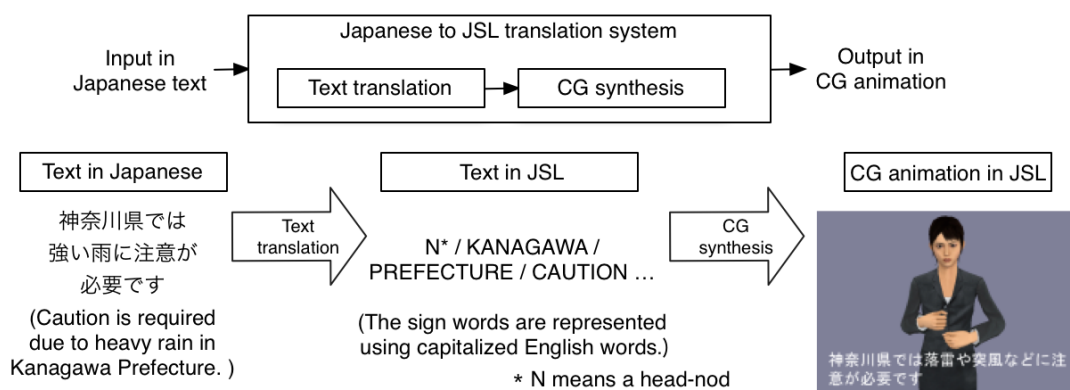


Figure 1: Japanese-JSL translation system overview

lates kanji, which are ideograms that compose most Japanese proper names, into closely related JSL words. Moreover, although several methods have been proposed to translate sentences in sign language, there is as yet no method to translate proper names (Massó et al., 2010; San-Segundo et al., 2010; Morrissey, 2011; Stein et al., 2012; Mazzei, 2012; Lugaresi et al., 2013).

This paper describes proper name translation from Japanese into JSL. The method involves sense-based translation and pronunciation-based translation. Both conversions are based on a statistical machine translation framework. The sense-based translation is a sense-based character-wise translation learned from phrase pairs in a Japanese-JSL corpus. The pronunciation-based translation is a pronunciation-based character-wise translation learned from a Japanese proper name corpus and is post-processed with transformation rules. We conducted a series of evaluation experiments and obtained good results. We also developed a proper name translation system from Japanese to JSL, in which the translated proper names are visualized with CG-animations.

## 2 Proper Names in JSL

### 2.1 Types of proper name in JSL

In JSL, proper name representations are classified into four types, as follows.

#### Type 1: sense-based case

Here, each character in Japanese proper names is translated into sign words in JSL. Most characters that make up Japanese proper names are kanji. Kanji are ideograms, i.e., each kanji representing concept, so they can be translated into words with the concepts in JSL.

For example, in the Japanese place name “香川 (Kagawa),” the kanji-characters “香 (aroma)” and “川 (river)” are respectively translated into sign words “AROMA<sup>1</sup>” and “RIVER.” Accordingly, the translation of “香川 (Kagawa)” is “AROMA / RIVER” in JSL.

#### Type 2: Pronunciation-based case

Here, the pronunciations of the kanji are transliterated into the Japanese kana alphabet. The kana are visualized by fingerspelling<sup>2</sup>. The transliteration in this case is not a spelling-based transformation from the source language because kanji are not phonograms<sup>3</sup>.

For example, in the Japanese personal name “茂木” (Motegi, written in kana as “モテギ”), the two kanji “茂” and “木” are respectively transliterated into the kana “モテ (mote)” and “ギ (gi).”

Each of the three kana, “モ (mo),” “テ (te)” and “ギ (gi),” is fingerspelled in JSL.

#### Type 3: Mixed case

This type includes Type 1 and Type 2. That is, some of the characters in the proper names are translated into sign words and the others are transliterated into kana and then visualized by fingerspelling. For example, regarding the Japanese place name “長野” (Nagano, written in kana as “ナガノ”), the kanji “長” is translated into the sign word “LONG” and “野” is transliterated into the kana “ノ (no).”

<sup>1</sup>The words in JSL are represented using capitalized English words. This notation method is called “glosses” in the sign language research community.

<sup>2</sup>All of the kana can be visualized by fingerspelling in JSL.

<sup>3</sup>For example, the character “木” is pronounced “ki,” “gi,” “moku,” “boku” etc. The decision as to which pronunciation should be used is by context, meanings or idiom.

Table 1: Analysis of proper name types

Place name	Type 1	43%
	Type 2	3%
	Type 3	10%
	Type 4	44%
Persons' name	Type 1	60%
	Type 2	14%
	Type 3	21%
	Type 4	21%

#### Type 4: Idiomatic case

These proper names are traditionally defined as fixed representations in JSL.

### 2.2 Analysis of Proper Name Types in Corpora

To investigate the frequencies of these four types in corpora, we analyzed a geographical dictionary (JFD, 2009) of place names and our corpus (mentioned in section 4.2.1) of persons' names. Table 1 shows the results of the analysis.

Proper names of Types 1, 2 and 3 needed to be translated, while those of Type 4 needed to be registered in an idiomatic translation dictionary of proper names. Furthermore, the proper name translations of Type 1, 2 and 3 reduce to sense-based translations and/or pronunciation-based translations.

Our translation method performs sense-based translation and pronunciation-based translation on the basis of statistical machine translation (SMT) methods. The next section describes this method.

## 3 Our translation method

### 3.1 Sense-based translation

#### 3.1.1 Basic method (baseline)

The sense-based translation uses SMT, and the translation probabilities (i.e. a lexicon model in SMT) are trained on our news corpus consisting of sentence pairs in Japanese and JSL. The basic method of training the lexicon model uses the corpus in a sentence-by-sentence manner (Figure 2-(a)). It segments the sentences into characters in Japanese and into words in JSL. Then, the model is trained on the characters of the Japanese sentences and the words of the JSL sentences. Regarding Sentence 1 below, the method segments it into Sentence 2 in Japanese and trains the model.

#### Sentence 1

**JP** 香川は朝から晴れるでしょう

(It will be fine from the morning in Kagawa)

**JSL** AROMA / RIVER / MORNING /  
FROM / FINE / DREAM

#### Sentence 2

**JP** 香川は朝から晴れるでしょう

(It will be fine from the morning in Kagawa)

**JSL** AROMA / RIVER / MORNING /  
FROM / FINE / DREAM

We took the basic method above to be the baseline method for the evaluations.

#### 3.1.2 Our method

Our method uses the corpus in a phrase-by-phrase manner. To use the phrase-segmented corpus, the method is composed of two steps. The first step aligns Japanese phrases to JSL phrases in each of the sentence pairs in the corpus by using many-to-many word alignment. Using the results of the alignment, each sentence pair is divided into phrase pairs. The second step segments the phrases into characters in Japanese and trains the sense-based translation part on the phrase pairs (Figure 2-(b)).

Let us illustrate our method using Sentence 1. The first step is dividing a sentence into phrase pairs. We use alignment pairs, the result of the many-to-many word alignment, as the phrase pairs. The alignment pairs are combined into phrase pairs, as shown in Phrase 1 below.

#### Phrase 1

**JP1** 香川 / は (in Kagawa)

**JSL1** AROMA / RIVER

**JP2** 朝 / から (from the morning)

**JSL2** MORNING / FROM

**JP3** 晴れる / でしょう (it will be fine)

**JSL3** FINE / DREAM

Alignment pairs that consist of many more or fewer sign words than Japanese words are discarded as alignment errors. In this paper, we regard the alignment pair as the alignment error when  $n_{sign} > (N_{JP} + \alpha)$  or  $(n_{sign} + \alpha) < n_{JP}$ . Here,  $n_{sign}$  means the number of sign words in

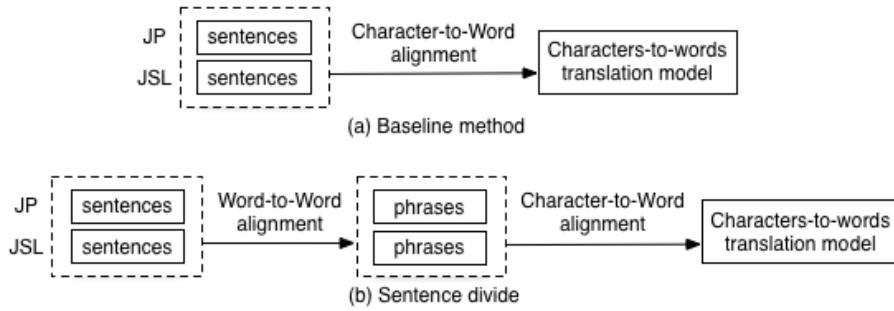


Figure 2: Two ways of learning translation models

the alignment pair, and  $n_{JP}$  means the number of Japanese words in the alignment pair. We chose  $\alpha$  to be 5, on the basis of preliminary experiment.

The second step segments Phrase 1 into characters in Japanese, as in Phrase 2 below.

### Phrase 2

**JP1** 香/川/は (in Kagawa)

**JSL1** AROMA / RIVER

**JP2** 朝/か/ら (from the morning)

**JSL2** MORNING / FROM

**JP3** 晴/れる/で/し/よ/う (It will be fine)

**JSL3** FINE / DREAM

Then, as shown in Example 1, the sense-based translation is trained on the corpus of phrase pairs.

### Example 1

香 → AROMA

川 → RIVER

は → (null)

⋮

Our method can reduce the combinations of alignments between Japanese characters and JSL words, because it segments sentences into phrases in which the number of words is less than that in the sentences. Therefore, it improves the alignment accuracy.

### 3.2 Pronunciation-based translation

The pronunciation-based translation is not transliteration but translation, because kanji do not represent their pronunciation. Therefore, the translation probabilities are also trained on a Japanese proper name corpus as a lexicon model in the SMT training step.

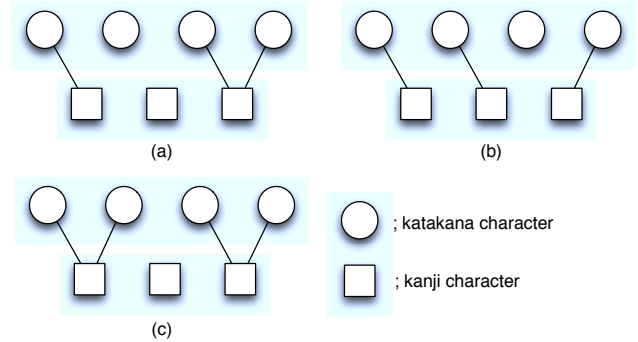


Figure 3: Patterns that cannot be aligned

Using the trained lexicon model, a decoder aligns the kana with the kanji. However, some of the kanji and kana are not aligned because of the sparse data problem. Such non-aligned cases are as follows.

**Pattern (a)** Aligned on neither the kanji nor the kana side (Fig.3-(a)).

**Pattern (b)** Insertion occurred (Fig.3-(b)).

**Pattern (c)** Deletion occurred (Fig.3-(c)).

The kanji-to-kana alignment is generally many-to-many, but we restricted the alignment to one-to-many.

To improve the result of these cases, we devised transformation rules that use the word's context, as follows.

**Rule (a)** Align all of the non-aligned kana with the non-aligned kanji.

**Rule (b)** Align the non-aligned kana to the kanji with the lower probability by comparing the translation probability of the left aligned kanji with the translation probability of the right aligned kanji.

**Rule (c)** Align the non-aligned kanji to the kana with the lower probability and un-align the

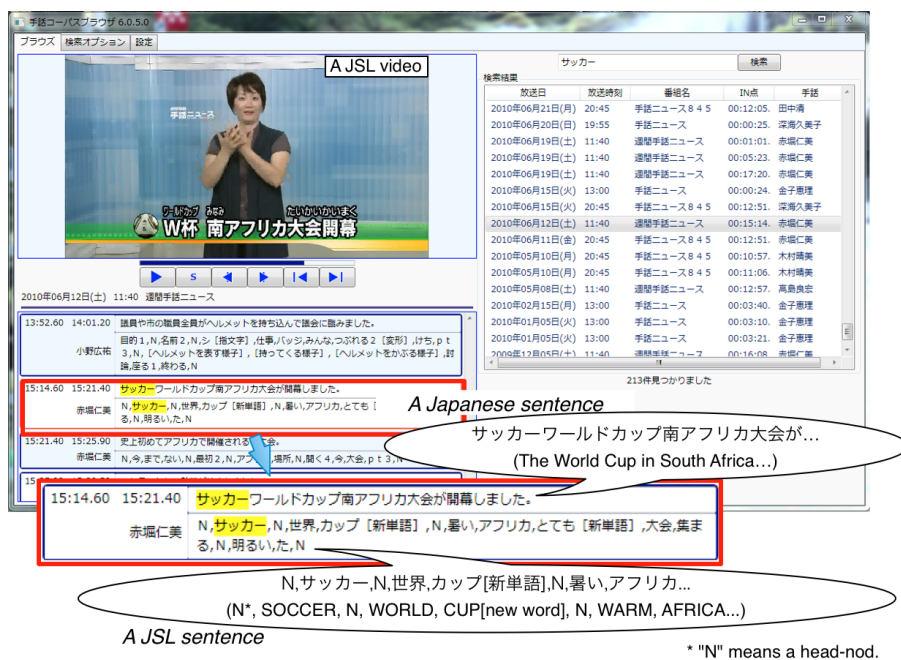


Figure 4: Japanese-JSL news corpus

old aligned kanji with the lower one by comparing the translation probability of the left aligned kana with the translation probability of the rightaligned kana.

Using these rules, our methods can align kanji to kana even if the kanji and/or kana are not in the training data. It has the advantage of robustness to the data sparse problem unlike conventional transliteration methods such as in (Finch et al., 2010; Knight et al., 1998). There are many different family names in Japan<sup>4</sup>, so these characteristics are important for translating Japanese proper names.

Our method applies these rules to the non-aligned kanji and kana from the beginning character in the sentences after the sense-based translation.

### 3.3 Combining sense-based and pronunciation-based translation

In our proper name translation, sense-based translation is first applied to a Japanese proper name and then pronunciation-based translation is applied to the characters that were not converted into sign words. Such characters occur in the following cases.

- The character does not appear in the training data of the sense-based translation.

<sup>4</sup>There are over 300,000 family names in Japan(Power, 2008).

- The character is translated into kana because the character is often translated into Kana in the training data of sense-based translation.

In these cases, our system translates the character into kana by using pronunciation-based translation.

## 4 Experiments and Results

### 4.1 Experimental setting

Our method uses GIZA++ and “grow-diag-final-and” (Och et al., 2003) as the model training and Moses (Koehn et al., 2007) as the decoding; it does not use a language model because word context and reordering are useless in proper name translation from Japanese to JSL.

The training sets were our Japanese-JSL news corpus (including 21,995 sentence pairs) for sense-based translation and a human-name corpus (including 34,202 personal names) for pronunciation-based translation. These corpora are described below.

The test set consisted of persons’ names and place names. Regarding the persons’ names, the candidates for the test set were first randomly sampled from a Japanese family name database<sup>5</sup>. The 100 sampled names were translated by three native signers and if two or three of the signers gave the same translation, the sample was added to the test

<sup>5</sup><http://www.douseidoumei.net/prof.html>

Table 2: Results of evaluation

	Person	Place	Total	Type 1	Type 2	Type 3	Type 4
# in the test set	96	82	178	123	16	32	7
Baseline	61 (63.5%)	37 (46.3%)	99 (55.6%)	86 (69.9%)	2 (12.5%)	9 (28.1%)	2 (28.6%)
Pialign	75 (78.1%)	41 (51.3%)	118 (66.3%)	97 (78.9%)	3 (18.8%)	15 (46.9%)	3 (42.9%)
Proposed (sense-based)	77 (80.2%)	43 (53.8%)	121 (68.0%)	95 (77.2%)	3 (18.8%)	20 (62.5%)	3 (42.9%)
Baseline + pronunciation-based	69 (71.9%)	44 (55.0%)	114 (64.0%)	86 (69.9%)	5 (31.3%)	21 (65.6%)	2 (28.6%)
Pialign + pronunciation-based	74 (77.1%)	47 (58.8%)	123 (69.1%)	97 (78.9%)	5 (31.3%)	18 (56.3%)	3 (42.9%)
Proposed (sense-based) + pronunciation-based	80 (83.3%)	53 (66.3%)	134 (75.3%)	95 (77.2%)	8 (0.50%)	28 (87.5%)	3 (42.9%)

set. This procedure produced a test set consisting of 96 names. The test set for place names was produced in the same way and amounted to 82 names. The total number of names used in our evaluation experiments was thus 178.

## 4.2 Training Corpora

### 4.2.1 Japanese-JSL corpus

We have been building up a Japanese-JSL news corpus to study Japanese-to-JSL machine translation. The corpus was collected from daily NHK Sign Language News programs, which are broadcast on NHK TV with Japanese narration and JSL signs.

The corpus consists of Japanese transcriptions, their JSL transcriptions, and their JSL movies. The Japanese transcriptions are transcribed by revising the speech recognition results of the news programs. The transcriptions are carried out by changing the sign gestures of the newscasters into sequences of JSL words. The JSL movies are manually extracted from the program by referring to the time intervals of the transcribed JSL transcriptions. The corpus currently includes about 22,000 sentence pairs taken from broadcasts running from April 2009 to August 2010. Our bilingual corpus is larger than other recent sign language corpora built in various sign language research projects (Bungeroth et al., 2006; Schembri, 2008; Johnston, 2009; Balvet et al., 2010; Matthes et al., 2012; Mesch et al., 2012). Figure 4 shows an example of our corpus.

### 4.2.2 Human Name Corpus

The human-name corpus was constructed by extracting personal names written in both kanji and kana from the IPADIC dictionary<sup>6</sup>.

## 4.3 Evaluation and Discussion

We conducted a series of experiments to evaluate our method. Table 2 shows the translation accuracies for proper names. The tested methods were as follows.

**Baseline** A simple baseline method (mentioned in 3.1.1)

**Pialign** The conventional character-based translation method (Neubig et al., 2012)

**Proposed (sense-based)** Our method for sense-based translation (described in 3.1.2)

**Pronunciation-based** Our method for pronunciation-based translation (described in 3.2)

Our overall method is “Proposed (sense-based) + pronunciation-based.” The upper row of each cell in the table shows the number of the correct words, whereas the lower row of each cell is the accuracy.

The table indicates that compared with the baseline, our method is higher in accuracy by 19.7 points in total, 19.8 points on persons’ name, and 19.6 points on place names. It is higher in accuracy than the baseline for each type of translation. The sense-based translation is effective at the raising total translation accuracy, whereas

<sup>6</sup><http://code.google.com/p/mecab/downloads>



the pronunciation-based translation increases the translation accuracy Types 2 and 3.

Each method had lower accuracy for place names than for persons' names. The reasons are as follows. One problem is that some of the characters in the place names are used only in place names, and though they appear in the test set, they do not appear in the training set. This is the out-of-vocabulary problem, which is a major issue with the corpus-based method. To tackle this problem, we will make our corpus larger by using Japanese-JSL place name dictionary. The other problem is that some of the place names have ambiguous Japanese-JSL translations. In this regard, the rate of agreement of the signers making was lower for place names (i.e. 82) than for personal names (i.e. 96).

The sense-based translation method is more accurate than pialign especially in translating type 2 and 3. This is because our discard process is able to delete infrequently used kanji in the corpus from the training data. Infrequently used kanji are often translated using their pronunciation because native signers cannot imagine the sign word that well represents the kanji.

Some of the type 4 words that occurred frequently in the training data were translated with the phrase-based method, however, the accuracy was low. An idiomatic translation dictionary is required for this purpose.

A Japanese-JSL place name dictionary would also improve the character-to-word conversion. For example, our method mistranslated the character “神 (god)” in a personal family name “神谷 (Kamiya)” into “KOBE (Kobe).” The cause of this error is that our method trains the character-to-word conversion “神 (god) → KOBE(Kobe)” from Phrase 3.

### Phrase 3

**JP** 神戸 (Kobe)

**JSL** KOBE

Our method would be able to avoid such a conversion error by deleting from the training set phrase pairs such as Phrase 3 that are registered in the place dictionary.

## 5 Proper Name Translation System

Using our translation method, we developed a proper name translation system from Japanese to



Figure 5: Motion capture system

JSL. This system visualizes the translated proper names as computer graphics (CG) animations.

The CG animation is a high-quality 3D model of human hands and fingers, and the model is controlled using motion-capture (MoCap) data. The data is captured with an optical MoCap system in which many markers are attached to fingers to pick up their movements precisely. Figure 5 shows the MoCap system. The CG-model has about 100 joints with three rotation angles. The CG-animation is rendered from scripts written in TVML (TM program Making Language<sup>7</sup>), which is a scripting language developed by NHK to describe full TV programs (Kaneko et al., 2010).

Figure 6 shows an example of the Japanese-to-JSL proper name translation system. When a proper name in Japanese is entered, a corresponding sign language animation is created and shown in the system. The translation system will be used in subjective evaluation of proper name translations.

## 6 Conclusion

We presented a Japanese-JSL proper name machine translation method. The method involves sense-based translation and pronunciation-based translation, both of which are based on statistical machine translation. We conducted a series of evaluation experiments and obtained 75.3% of accuracy, increasing from baseline method by 19.7 points.

We will incorporate our method of proper name translation from Japanese to JSL in our machine translation system.

<sup>7</sup><http://www.nhk.or.jp/strl/tvml/english/player2/index.html>

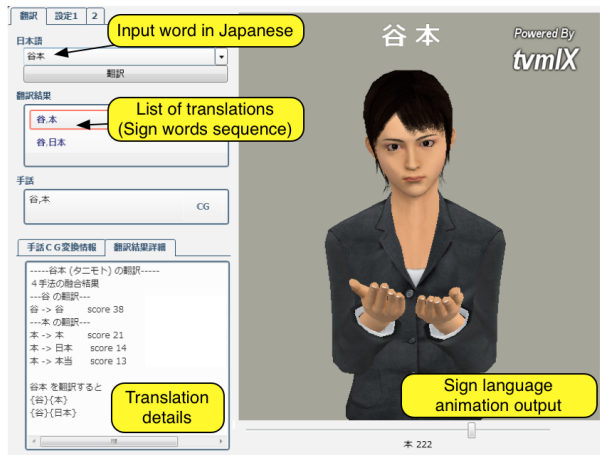


Figure 6: Japanese-JSL proper name translation system

## Acknowledgements

The authors would like to express our deep gratitude to Hirokazu Kosugi for implementing the experimental system for translating sign word sequences into sign language CG animations. They would also like to express their gratitude to Hideki Tanaka, Ichiro Yamada, Tadashi Kumano, Isao Goto, and the anonymous reviewers for their valuable comments and suggestions.

## References

- Antonio Balvet, Cyril Courtin, Dominique Boutet, Christian Cuxac, Ivani Fusellier-Souza, Brigitte Garcia, Marie-Thérèse L’Huillier and Marie-Anne Sallandre. 2010. The Creagest Project: a Digitized and Annotated Corpus for French Sign Language (LSF) and Natural Gestural Languages. *International Conference on Language Resources and Evaluation (LREC 2010)*: 469–475.
- Jan Bungeroth, Daniel Stein, Philippe Dreuw, Morteza Zahedi and Hermann Ney. 2006. A German Sign Language corpus of the domain weather report. *International Conference on Language Resources and Evaluation (LREC 2006)*: 2000–2003.
- Andrew Finch, Keiji Yasuda, Hideo Okuma, Eiichiro Sumita and Satoshi Nakamura. 2011. A Bayesian Model of Transliteration and Its Human Evaluation when Integrated into a Machine Translation System. *IEICE transactions on Information and Systems*: Vol. E94–D, No. 10, pp.1889–1900.
- Isao Goto, Naoto Kato, Noriyoshi Uratani and Terumasa Ehara. 2003. Transliteration considering context information based on the maximum entropy method. *The 9th Machine Translation Summit*: 125–132.
- Japanese Federation of the Deaf (JFD). 2009. Place names map in Japanese Sign Language in Japan (in Japanese, “全国地名手話マップ”) Japanese Federation of the Deaf Press.
- Trevor Johnston. 2009. Creating a corpus of Auslan within an Australian national corpus. *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages*.
- Hiroyuki Kaneko, Narichika Hamaguchi, Mamoru Doke and Seiki Inoue. 2010. Sign language animation using TVML. *9th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI 2010)*, ACM 2010:289–292.
- Kevin Knight, Jonathan Graehl. 1998. Machine transliteration. *Computer Linguistics*, 24: 599–612.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowen, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. *Annual meeting of the Association for Computational Linguistics (ACL 2007)*, demonstration session.
- Li Haizhou, Zhang Min, Su Jian. 2004. A joint source-channel model for machine transliteration. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL ’04)*. Article No. 159.
- Camillo Lugaresi and Barbara Di Eugenio. 2013. Translating Italian connectives into Italian Sign Language. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pp 270–280. Sofia, Bulgaria, August.
- Guillem Massó and Toni Badia. 2010. Dealing with sign language morphemes in statistical machine translation. *4th workshop on the representation and processing of sign language: interactions between corpus and lexicon at LREC 2010*: 154–157.
- Silke Matthes, Thomas Hanke, Anja Regan, Jakob Storz, Satu Worseek, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert and Eva Safar. 2012. Dicta-Sign – Building a multilingual sign language corpus. *5th workshop on the representation and processing of sign language: interactions between corpus and lexicon at LREC 2012*: 117–122.
- Alessandro Mazzei. 2012. Sign language generation with expert systems and ccg. *Proceedings of the Seventh International Natural Language Generation Conference (INLG ’12)*: 105–109.
- Johanna Mesch, Lars Wallin and Thomas Björkstrand. 2012. Sign language resources in Swedes: dictionary and corpus. *5th workshop on the representation and processing of sign language: interactions*



- between corpus and lexicon at LREC 2012: 127–130.
- Sara Morrissey. 2011. Assessing three representation methods for sign language machine translation and evaluation. *15th annual meeting of the European Association for Machine Translation (EAMT 2011)*: 137–144.
- Graham Neubig, Taro Watanabe, Shinsuke Mori and Tatsuya Kawahara. 2012. Machine Translation without Words through Substring Alignment. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL2012)* : 165–174.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*,29: 19–51.
- John Power. 2008. Japanese Names. *The Indexer*, Volume 26, No 2, pp. C4-2–C4-8.
- Rubén San-Segundo, Verónica López, Raquel Martín, David Sánchez, Adolfo García. 2010. Language resources for Spanish – Spanish Sign Language (LSE) translation. *The 4th workshop on the representation and processing of sign languages: corpora and sign language technologies at LREC 2010*: 208–211.
- Adam Schembri. 2008. British Sign Language corpus project: open access archives and the observer’s paradox. *3rd workshop on the representation and processing of sign languages at LREC 2008*.
- Daniel Stein, Christoph Schmidt and Hermann Ney. 2012. Analysis, preparation, and optimization of statistical sign language machine translation. *Machine Translation* 26: 325-357.
- Katsuhito Sudoh, Shinsuke Mori and Masaaki Nagata. 2013. Noise-aware Character Alignment for Bootstrapping Statistical Machine Translation from Bilingual Corpora. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*: 204–209.
- Paola Virga, Senjeev Khudanpur. 2003. Transliteration of proper names in cross-lingual information retrieval. *MultiNER ’03 Proceeding of the ACL 2003 workshop on multilingual and mixed-language named entity recognition* Volume 15, pp 57–64.