

Evaluation on Sign Language CG Animation for Breaking News

Tsubasa Uchida†, Naoki Nakatani†, Takeshi Kajiyama†† and Hiroyuki Kaneko†

Abstract: NHK is developing a system for generating CG for any given sentence using machine translation and CG generation technology, with the aim of expanding information security services for the hearing impaired. An evaluation experiment was conducted on the hearing impaired to assess the comprehension of sign language CG generated from breaking news text, and the factors that hinder comprehension were analyzed for both machine translation and CG generation. Sign language CG generated by machine translation had a lower accuracy rate for comprehension than sign language CG generated by manual translation, indicating the need for improvement in translation accuracy. In generating CG, it was shown that the comprehension of proper nouns is affected by supplementary information such as finger spelling and mouth shapes, as well as the speed of presentation, and that facial expressions that do not match the context can also hinder comprehension.

Keywords: Sign Language, Accessibility, Assistive Technology, Deaf and Hard of Hearing People, Computer Graphics

1. Introduction

NHK is expanding its sign language broadcasting as one of its information security services for the hearing impaired. Native sign language speakers have difficulty understanding subtitles provided in Japanese, and there has been a demand for expanded provision of sign language information for high-priority news. However, the number of sign language anchors and interpreters who can appropriately convey the technical content of news in sign language is limited, and it is difficult to consistently secure experts well-versed in sign language expressions for news at broadcasting stations. Therefore, NHK is conducting research and development into an arbitrary sign language CG generation system that automatically generates sign language CG from Japanese text in breaking news using sign language presentation technology based on computer graphics (CG) animation (hereafter, sign language CG) 1).

Sign language is a language with its own grammar that differs from spoken language, and when conveying Japanese content in sign language, simply arranging the sign words in accordance with the Japanese word order does not accurately express the content. Therefore, it is necessary to translate the Japanese content by replacing it with the word order and grammar of the sign language. In this paper, we define errors in the selection of each word to be translated during translation and errors in the word order of the entire sentence as translation errors. It is said that sign languages used in Japan include "Japanese Sign Language," which has word order and grammar distinct from Japanese, and "Japanese-based Sign Language," which follows Japanese word order and grammar.2)3) In the Japanese news articles studied here, new words appear frequently. Consequently, even for deaf individuals whose native language is Japanese Sign Language, many expressions cannot be fully conveyed using Japanese Sign Language alone. Instead, they become expressions that possess characteristics of both Japanese Sign Language and Japanese-based Sign Language.

There are many expressions that cannot be expressed using this sign language alone, and the resulting expressions have the properties of both Japanese Sign Language and Japanese-compatible sign language.

This system is composed of two component technologies: machine translation from Japanese text to information representing sign language word order (sign language word strings), and CG generation that connects multiple sign language word actions according to the translated word order. Therefore, in this study, we conducted evaluation experiments to clarify the issues with the two component technologies that make up the generation of arbitrary sign language CG. In this study, we conducted evaluation experiments to clarify the issues with sign language CG generated from breaking news. For the evaluation, we used two types of sign language CG: one that was manually translated by a sign language newscaster, and the other that was machine translated by inputting the breaking news text into the system and translating it using a translation model. To improve the comprehension of the sign language CG, we considered issues that needed improvement through evaluation experiments using a web page and face-to-face interviews (depth interviews). The following points

were clarified in this paper: (1) Sign language CG generated by machine translation had a lower accuracy rate for content comprehension than sign language CG generated by manual translation, confirming the need for improvement in translation (2) In generating CG, we confirmed the necessity of adding finger spelling to proper noun expressions, using appropriate speed and mouth shapes, and reproducing appropriate facial expressions that match the context.

2. Related Research

Various methods employing different approaches have been proposed for automatic translation from spoken language to sign language. Many of these methods utilize avatar animation, employing 3DCG avatars for output. In sign language automatic translation, the input spoken language is translated sentence-by-sentence into a sequence of sign language words, and sign language utterances are generated by linking these sign language words together.4)-6)

2025年1月14日受付, 2025年4月9日再受付, 2025年5月9日採録

† 日本放送協会

(〒150-8001 東京都渋谷区神南 2-2-1, TEL 03-3465-1111)

†† 一般財団法人 NHK 財団

[Copyrights to Machine-Translated Content]

The copyright of the original papers published on this website belongs to the Institute of Image Information and Television Engineers. Unauthorized use of original papers or translated papers is prohibited. Please be sure to cite the original publication when referencing. For details, please refer to the copyright regulations of the Institute of Image Information and Television Engineers.

ニュース速報を対象とした手話 CG アニメーションの評価

Evaluation on Sign Language CG Animation for Breaking News

正会員 内田 翼[†], 中谷 真規[†], 正会員 梶山 岳士^{††}, 正会員 金子 浩之[†]Tsubasa Uchida[†], Naoki Nakatani[†], Takeshi Kajiyama^{††} and Hiroyuki Kaneko[†]

あらまし NHK では聴覚障害者への情報保障サービスの拡充に向けて、機械翻訳と CG 生成技術を用いた任意文手話 CG 生成システムの開発を進めている。聴覚障害者を対象として、ニュース速報文から生成した手話 CG の内容理解に関する評価実験を実施し、機械翻訳と CG 生成のそれぞれについて内容理解を妨げる要因を分析した。機械翻訳によって生成した手話 CG は、人手翻訳によって生成した手話 CG より内容理解の正答率が低く、翻訳精度の向上が必要であることが示された。CG 生成においては、指文字や口型の提示による情報の補完や提示速度が固有名詞の理解に影響するほか、文脈と合致しない顔表情が内容理解を妨げる問題があることが示された。

キーワード：手話、情報保障、障害者支援、聴覚障害者、Computer Graphics

1. ま え が き

NHK では聴覚障害者への情報保障サービスのひとつとして、手話放送の拡充を進めている。手話母語者にとって、日本語で提供する字幕の内容理解は難しく、緊急性の高いニュースにおける手話による情報提供の拡充について要望が寄せられている。しかし、ニュースの専門的な内容を手話で適切に伝えることができる手話キャスターや手話通訳者の数は限られ、ニュースの手話表現に精通する専門家を常に放送局に確保することは困難である。そこで、CG(Computer Graphics) アニメーションによる手話提示技術(以下、手話 CG)を用いて、ニュース速報の日本語テキストから手話 CG を自動生成する任意文手話 CG 生成システムの研究開発を進めている¹⁾。

手話は音声言語とは異なる独自の文法を持つ言語であり、日本語の内容を手話で伝える際には、日本語の語順に合わせて単純に手話単語を並べるだけでは、正確な内容を表現することができない。そのため、日本語の内容を手話の語順・文法に合わせた置き換え、すなわち翻訳をする必要がある。本論文では、この翻訳時の訳出する各単語の選択と、文全体の語順の誤りを翻訳誤りとして定義する。日本で使われている手話には、日本語と異なる語順・文法を持つ「日本手話」と、日本語の語順・文法に従う「日本語対応手話」があると言われている²⁾³⁾。本研究が対象とする日本語のニュースにおいては新語が頻出するため、日本手話を母語とするろう者であっても日

本手話だけでは表しきれない表現が多く、日本手話と日本語対応手話の双方の性質を有するような表現となる。

本システムは、日本語テキストから手話の語順を表す情報(手話単語列)への機械翻訳と、翻訳した語順に従って複数の手話単語動作を接続する CG 生成の2つの要素技術で構成される。そこで、本研究では、任意文手話 CG 生成を構成する2つの要素技術について、ニュース速報から生成した手話 CG における各々の課題を評価実験により明らかにする。評価には、ニュース速報テキストを手話ニュースキャスターにより正しく翻訳した人手翻訳と、ニュース速報テキストを本システムに入力し、翻訳モデルにより翻訳した機械翻訳の2つの手話 CG を用いた。手話 CG の理解度向上のため、Web ページによる評価実験と対面の聞き取り調査(デプスインタビュー)を通じて改善すべき課題を考察した。本論文で明らかにした項目を以下に示す。

- (1) 機械翻訳によって生成した手話 CG は、人手翻訳によって生成した手話 CG より内容理解の正答率が低く、翻訳精度向上の必要性を確認した。
- (2) CG 生成において、固有名詞表現における指文字の付与や適切な速度・口型の使用や、文脈に合った適切な表情を再現する必要性を確認した。

2. 関 連 研 究

音声言語から手話への自動翻訳については、異なるアプローチを用いた様々な手法が提案されている。それらの手法の多くは出力に 3DCG のアバターを用いたアバターアニメーションを採用している。手話の自動翻訳では、入力となる音声言語を手話単語が並んだ手話単語列

2025 年 1 月 14 日受付, 2025 年 4 月 9 日再受付, 2025 年 5 月 9 日採録

[†] 日本放送協会

(〒 150-8001 東京都渋谷区神南 2-2-1, TEL 03-3465-1111)

^{††} 一般財団法人 NHK 財団

(〒 158-8711 東京都世田谷区用賀 4-10-1, TEL 03-5797-6700)

Sign language utterances are generated by translating sentences into a single sentence and connecting the sign words together. 4)-6) In avatar animation, sign language actions are reproduced by controlling an avatar using keyframe animation. 7)8) or motion capture data. 9)-12) Regarding avatar control, various approaches have been reported to reproduce natural, easy-to-understand sign language actions in avatar animation. 13) Sign language, as a visual language, has various unique grammatical elements that differ from spoken language, such as nodding, pointing, pauses in movements, and facial expressions. When generating sign language avatar animation, reproducing these grammatical elements presents a major challenge for accurate content transmission. 14)-16) Therefore, quantitative evaluation using BLEU alone is insufficient for evaluating the performance of machine sign language translation. Because accuracy must be evaluated, including the reproduction of grammatical elements unique to sign language, the quality of the final output avatar animation is crucial. As quantitative evaluation criteria for evaluating the quality of sign language avatar animation, such as content comprehension and naturalness, have not yet been established, many researchers have used their own evaluation methods, such as subjective evaluation experiments¹²⁾16)-18).

To evaluate the performance of our system for generating arbitrary sign language CG, we conducted evaluation experiments with hearing-impaired people to compare the performance of the system with that of machine translation.

We verified the effect that each elemental technology of CG generation has on the comprehension of the content of the final output sign language CG.

3. Arbitrary Sign Language CG Generation System

3.1 System Overview

This system translates input Japanese text into sign language, connects sign language motions based on the translation results, and automatically generates sign language CG animation¹⁾. An overview of this system is shown in Figure 1.

First, the Japanese text is machine translated into a sequence of sign language words arranged in chronological order. A neural machine translation model using Transformer is applied for the translation¹⁹⁾. A Japanese-sign language bilingual corpus based on NHK Sign Language News broadcast footage is used to train the machine translation model²⁰⁾. The bilingual corpus is bilingual data accumulated by annotating the Japanese announcement speech in the program footage into Japanese text, and the sign language expressions used by sign language news anchors into sign language word sequences.

Next, the motion data of the sign language words is read according to the sign language word sequence. The motion data of the sign language words was recorded in advance using an optical motion capture system. In order to reproduce facial expressions, which are an important grammatical element in sign language, in the sign language CG, the movements of the hands and fingers as well as the movements of the face were recorded as motion data. Sign language words published in sign language dictionaries²¹⁾22), sign language expressions that frequently appear in NHK Sign Language News, and the expressions required for the service were recorded as motion data.

The motion data of the signs was recorded and a motion database of sign language words was constructed. All motion data for the sign language words contained in the sign language word string was read from the motion database, and by interpolating and connecting the motion data, sign language motions for each sentence were automatically generated, which were then played on a CG avatar to render CG animation. Specifically, by connecting the motion data of pre-recorded sign language words according to the word order of sign language, such as [Tokyo, tomorrow, weather, bright, forecast] (The weather in Tokyo tomorrow will probably be sunny), it is possible to generate a variety of Japanese sentences as sign language CG.

3.2 Functions for reproducing grammatical elements specific to sign language

Previous research has shown that when sign language CG is presented in which word motions are simply connected in the correct word order, the level of comprehension is low.¹⁴⁾ Therefore, in this system, we developed functions to reproduce grammatical elements specific to sign language for each of the component technologies of machine translation and CG generation.

In sign language CG created by interpolating word motions, proper nouns expressed as compound words made up of multiple words and the division of phrases become unclear due to the constant time interval between sign language words. Therefore, we developed a system that outputs auxiliary information to reproduce the division of proper nouns and phrases along with the translation results at the CG generation stage¹⁹⁾.

Additionally, while we previously manually corrected errors in word selection and word order in machine translation and edited motions to reproduce grammatical elements, we have now incorporated a mechanism to automate some grammatical elements. We have developed a function to clarify the gaps between sign language words by synthesizing nodding and pointing motions while maintaining the spatial position of the arm in the previous sign. We have also developed a function to emphasize the unity of sign language words by narrowing the time intervals connecting multiple word motions that make up proper nouns. 1) 25) The word motions used in this system are recorded in dictionary form, the basic form of sign language actions listed in sign language dictionaries. 21) 22) Regarding facial movements, which are one of the grammatical elements, words whose meaning changes depending on the mouth shape (mouth shape) even if the hand and finger movements are the same are problematic. The pre-recorded word motions are fixed to dictionary-form mouth shapes, and if the mouth shape differs from the context of the sign language sentence, the content will change significantly from the original Japanese information. Therefore, we developed a function to replace and synthesize appropriate mouth motions for proper nouns¹⁾25).

4. Experimental Overview

Using an arbitrary sign language CG generation system that implements a function for reproducing grammatical elements specific to sign language, an evaluation experiment was conducted on hearing-impaired people to verify the comprehension of sign language CG generated from breaking news text. 4.1 Experimental conditions The experiment involved a

total of 101 people (88 deaf, 11 hard of hearing, and 2 people who had recently lost their hearing), with an age distribution of 20 people under 20, 20 people in their 30s, and

へ文単位で翻訳し、手話単語間を繋ぎ合わせることで手話の発話文を生成する^{4)~6)}。アバターアニメーションでは、キーフレームアニメーション⁷⁾⁸⁾やモーションキャプチャデータ^{9)~12)}などを用いてアバターを制御することで手話動作を再現している。アバターの制御に関しては、伝わりやすく自然な手話動作をアバターアニメーションで再現するために、様々なアプローチが報告されている¹³⁾。

視覚言語である手話には、顔や指差し、動きの間や表情といった音声言語とは異なる様々な特有の文法要素がある。手話のアバターアニメーションを生成する場合も、それら文法要素の再現が、正確な内容伝達の大きな課題となる^{14)~16)}。そのため、手話の自動翻訳の性能評価では、BLEU による定量評価だけでは不十分である。手話特有の文法要素の再現を含めた正確さの評価が必要となるため、最終出力であるアバターアニメーションの品質評価が重要となる。手話のアバターアニメーションに関する内容理解や自然さといった品質評価においては、定量的な評価指標が確立されていないため、主観評価実験などにより、多くの研究者が独自の評価手法で検証している^{12)16)~18)}。

我々は、任意文手話 CG 生成システムの性能を評価するために、手話単語列の情報から生成した手話 CG について、聴覚障害者による評価実験を通して、機械翻訳と CG 生成それぞれの要素技術が最終出力となる手話 CG の内容理解に与える影響を検証した。

3. 任意文手話 CG 生成システム

3.1 システム概要

本システムでは、入力した日本語テキストを手話に翻訳し、翻訳結果にしたがって手話モーションを接続し、手話 CG アニメーションを自動生成する¹⁾。本システムの概要を図 1 に示す。

まず、日本語テキストを、手話の単語が時系列に並んだ手話単語列へ機械翻訳する。翻訳には Transformer を利用したニューラル機械翻訳モデルを適用した¹⁹⁾。機械翻訳モデルの学習には、NHK 手話ニュースの放送番組映像をもとにした日本語と手話の対訳コーパスを用いた²⁰⁾。対訳コーパスは、番組映像中の日本語のアナウンス音声日本語テキストに、手話ニュースキャスターが表出する手話表現を手話単語列にそれぞれアノテーションして蓄積した対訳データである。

次に、手話単語列にしたがって手話単語のモーションデータを読み込む。手話単語のモーションデータは、光学式モーションキャプチャシステムによりあらかじめ収録したデータを用いた。手話における重要な文法要素のひとつである表情を手話 CG で再現するために、手指と同時に顔の動きもモーションデータとして収録した。手話辞典などに掲載されている手話単語²¹⁾²²⁾や NHK 手話ニュースに頻出する手話表現、サービスで必要となるフ

レーズの動作を収録し、手話単語のモーションデータベースを構築した。モーションデータベースから、手話単語列に含まれる手話単語のモーションデータをすべて読み込み、モーションデータ間を補間接続することで、文単位の手話モーションを自動生成し、それを CG アバターで再生して CG アニメーションをレンダリングする。具体的には、あらかじめ収録した手話単語のモーションデータを、[東京、明日、天気、明るい、予想]（東京の明日の天気は晴れでしょう）のように手話の語順に従って接続することで、様々な日本語の文章を手話 CG として生成することが可能となる。

3.2 手話特有の文法要素を再現するための機能

先行研究では、正しい語順に従って単語モーションを単純に接続した手話 CG を提示しても理解度が低いという評価結果が示されている¹⁴⁾。そこで、本システムにおいて、機械翻訳と CG 生成それぞれの要素技術について手話特有の文法要素を再現するための機能を開発した。

単語モーションを補間接続した手話 CG は、手話単語間の時間間隔が一定となることに起因して、複数単語で構成される複合語として表現される固有名詞や、文節の区切れが不明確となる。そこで、CG 生成の段階で固有名詞や文節などの区切れを再現するための補助情報を翻訳結果と併せて出力する仕組みを開発した¹⁹⁾。

また、機械翻訳における単語選択と語順の誤りの修正や、文法要素を再現するためのモーションの編集を手作業で実施していたが²³⁾、一部の文法要素については自動化する仕組みを取り入れた。直前の手話の腕の空間位置を保持して顔や指差しのモーションを合成することで手話単語間の間を明確にする機能²⁴⁾や、固有名詞を構成する複数の単語モーションを接続する時間間隔を狭くすることで手話単語のまとまりを強調する機能を開発した¹⁾²⁵⁾。

本システムで使用している単語モーションは、辞書形と呼ばれる手話の辞書²¹⁾²²⁾に掲載されている手話動作の基本形で収録する。文法要素のひとつである顔の動作については、手指動作が同じでも口の形（口型）で意味が変わる単語が問題となる。あらかじめ収録した単語モーションは辞書形の口型に固定され、手話文の文脈と異なる口型であった場合に、元の日本語の情報から内容が大きく変わってしまう。そこで、固有名詞については適切な口型のモーションを置換して合成する機能を開発した¹⁾²⁵⁾。

4. 実験概要

手話特有の文法要素を再現するための機能を実装した任意文手話 CG 生成システムを用いて、ニュース速報のテキストから生成した手話 CG の理解度を検証するために、聴覚障害者を対象にした評価実験を実施した。

4.1 実験条件

実験では全 101 名（ろう者 88 名、難聴者 11 名、中途失聴者 2 名）、年齢分布は 20 代以下 20 名、30 代 20 名、

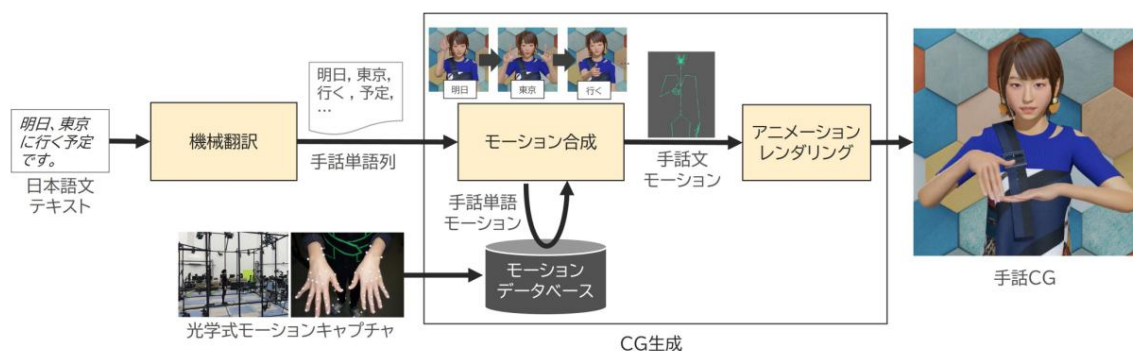


Figure 1. Overview of the system for generating arbitrary sign language CG

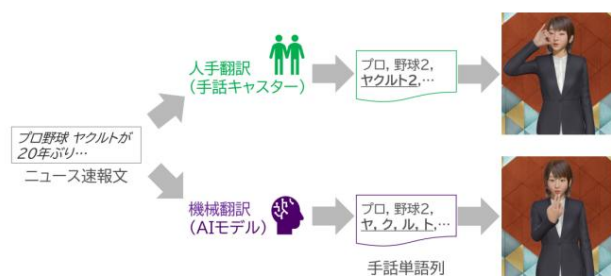


Figure 2: Illustration of sign language CG animation generation using manual and machine translation



Figure 3 Example of evaluation web page

The participants in the experiment were 30 people in their 40s , 20 people in their 50s ,

and 11 people over 60. Six categories of breaking news sentences (evacuation/ safety, election news, power shortages, missiles, sports, and medal news) were selected, and a total of 24 breaking news sentences from each category were used for evaluation. As shown in Figure 2 , two types of sign language CG videos were prepared for each breaking news sentence: a hand-translated video and a machine-translated video. The hand-translated video was a sign language CG video generated by translating the Japanese text of the breaking news into a string of sign language words by a sign language news anchor and connecting the word motions . Since it can be assumed that the translated individual sign language words and the word order of the entire sentence are correct and there are no translation errors, it was used to evaluate the impact on comprehension due to the CG generation alone . The machine-translated video was a sign language CG video generated by inputting the Japanese text of the breaking news into the system and translating it using a translation model . It was used to evaluate the impact on comprehension due to both machine translation and CG generation.

Two sets of 24 videos , including both manually translated and machine translated videos , were created and presented to two groups of participants. The first set of videos was selected from 12 manually translated videos and 12 machine translated videos to avoid overlapping breaking news content, while the second set consisted of the remaining 12 manually translated videos and 12 machine translated videos . In other words, each of the two groups ultimately evaluated all of the sign language CG videos related to the 24 breaking news sentences , with the only difference between the groups being the type of machine translated and manually translated videos. 4.2 Experimental

4.2 Experimental

The experiment was conducted on a webpage built for evaluation purposes, using sign language CG videos and multiple-choice questions about the content of each video.

Questions were displayed one by one on a web page after the video was played, with answer options selected using checkboxes. Figure 3 shows an example of the evaluation web page. Each video was viewed three times, and two types of questions were presented for each breaking news sentence : one to check comprehension of the entire sentence, and one to check proper noun reading. Proper nouns are particularly important when conveying breaking news. However, many compounds are composed of multiple sign language words, and the accuracy of the sign word connection and mouth shape are thought to significantly affect readability. For this reason, proper noun reading was evaluated separately from comprehension of the entire sentence. Table 1 shows example questions.

Table 1. Examples of questions regarding

内容理解	内容は 何でしたか？	☐ ① 安全を確保してくださいという呼びかけ ☐ ② 天気が悪くなるので注意してくださいという呼びかけ ☐ ③ 寒くなるので体調管理に気を付けてくださいという呼びかけ
固有名詞 読み取り	誰が呼びかけて いますか？	☐ ① 政府 ☐ ② 首相 ☐ ③ 気象庁

[Copyrights to Machine-Translated Content]

The copyright of the original papers published on this website belongs to the Institute of Image Information and Television Engineers. Unauthorized use of original papers or translated papers is prohibited. Please be sure to cite the original publication when referencing. For details, please refer to the copyright regulations of the Institute of Image Information and Television Engineers.

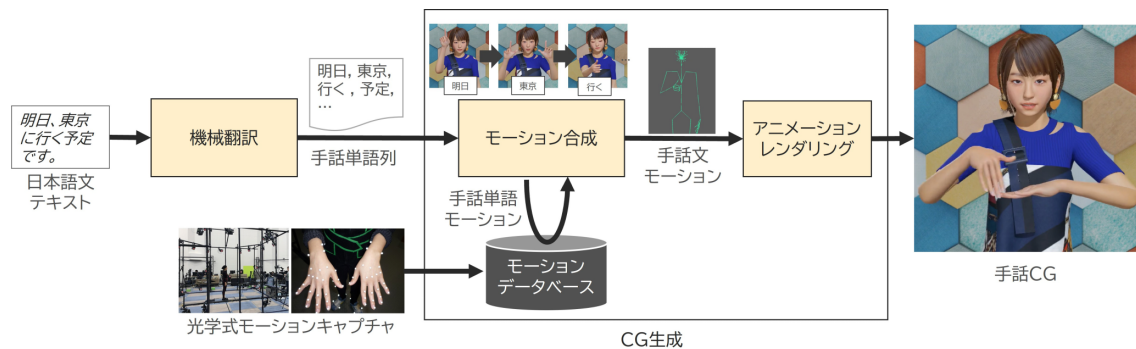


図 1 任意文手話 CG 生成システムの概要

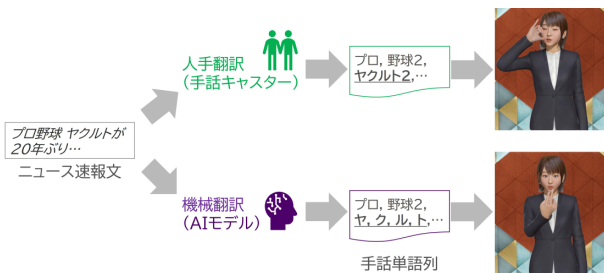


図 2 人手翻訳と機械翻訳による手話 CG 動画生成のイメージ

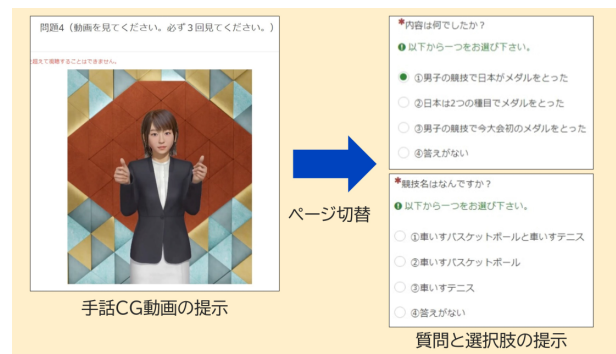


図 3 評価用 Web ページの例

40代30名、50代20名、60代以上11名を実験協力者とした。

ニュース速報文から6つのカテゴリ(避難・安全/選挙速報/電力不足/ミサイル/スポーツ/メダル速報)を選定し、各カテゴリから選定した合計24文の速報文を評価に用いた。図2に示すように、各速報文について人手翻訳動画と機械翻訳動画の2種類の手話CG動画を用意した。人手翻訳動画は、ニュース速報の日本語テキストを手話ニュースのキャスターにより手話単語列に翻訳し、単語モーションを接続して生成した手話CG動画である。訳出された各手話単語と文全体の語順が正しく翻訳誤りがないと仮定できることから、CG生成のみに起因する理解度への影響の評価に用いた。機械翻訳動画は、ニュース速報の日本語テキストを本システムに入力し、翻訳モデルにより翻訳した結果をもとに生成した手話CG動画である。機械翻訳とCG生成の両方に起因する理解度への影響の評価に用いた。

人手翻訳動画と機械翻訳動画を含む全24動画の動画セットを2セット作成し、2つに分けた実験協力者グループにそれぞれ提示した。1つ目の動画セットは速報文の内容が重複しないように人手翻訳12動画と機械翻訳12動画を選定し、2つ目の動画セットは残った人手翻訳12動画と機械翻訳12動画で構成した。つまり、2つのグループそれぞれが最終的に24文のニュース速報文に関する手話CG動画をすべて評価し、グループ間で機械翻訳と人手翻訳の動画の種類のみが異なる形であった。

4.2 実験手順

実験は評価用に構築したWebページで実施し、手話

CG動画の提示と各動画の内容理解に関する選択式の質問により評価した。質問は動画を再生した後にWebページ上に1問ずつテキスト表示し、回答の選択肢をチェックボックスで選ぶ形とした。図3に評価用Webページの例を示す。1つの動画に対する繰り返し視聴回数は3回とし、ニュース速報1文ごとに、文全体の内容理解を確認するための質問と、固有名詞の読み取りを確認するための質問の2種類を提示した。固有名詞については、ニュース速報を伝える際に特に重要である。一方で、複数の手話単語で構成される複合語が多く、手話単語の接続精度と口型が読み取り易さに大きく影響を及ぼすと考えられる。このため、固有名詞の読み取りは、文全体の内容理解と分けて評価することとした。表1に質問項目の例を示す。

表 1 内容理解に関する質問項目の例。

内容理解	内容は 何でしたか？	☐ ① 安全を確保してくださいという呼びかけ ☐ ② 天気が悪くなるので注意してくださいという呼びかけ ☐ ③ 寒くなるので体調管理に気を付けてくださいという呼びかけ
固有名詞 読み取り	誰が呼びかけて いますか？	☐ ① 政府 ☐ ② 首相 ☐ ③ 気象庁

5. Results and Discussion

5.1 Correct Answer Rate for Questions Regarding

Content Comprehension The evaluation results, Table 2 shows the correct answer rate for questions regarding content comprehension. The correct answer rate was calculated as the ratio of the number of correct answers to the total number of questions in each category, broken down by the attributes of the participants. The evaluation results showed that the correct answer rates for all categories of sign language CG videos for both manual and machine translation were 66.5% (standard deviation 15.2%) and 65.5% (standard deviation 15.9%) for all participants. The difference in correct answer rate between categories for machine translation was larger than for manual translation. This indicates that the current machine translation model has variation in translation accuracy for each category. Furthermore, in categories other than sports and medal news, which are frequently included in the training data for the machine translation model and are expected to have high machine translation accuracy, the correct answer rate for machine translation was lower than for manual translation, indicating that machine translation accuracy affects the comprehension of the sign language CG content.

5.2 Reading Proper Nouns The

accuracy rates for questions regarding the reading of proper nouns only are shown in Table 3. For proper nouns, the accuracy rates for manual translation and machine translation across all participants were 82.1% and 76.1%, respectively. For manual translation, the accuracy rate for some organization names, such as sports team names and political party names, was 69.1% , lower than the 74.0 % for machine translation. When expressing proper nouns that are unfamiliar to the speaker, sign language often emphasizes mouth shapes or uses finger spelling.2 However, in addition to the issue of not being able to reproduce mouth shapes in CG generation, as discussed in Section 5.4 , the manual translation did not use finger spelling and instead used only sign language words, which is thought to have affected comprehension. On the other hand, the machine translation often used finger spelling to signify organization names, which is thought to have resulted in a higher accuracy rate than the manual translation. 5.3 Analysis by participant

attributes (1) Analysis of the effect of sign language proficiency To verify

whether sign language proficiency affected the

evaluation results, participants were divided into two

groups: those with hearing loss or late deafness whose first language is not sign language, and deaf participants who have used sign language as their first language since childhood. As shown in Tables 2 and 3 , with the exception of a few categories, the deaf group had a higher correct answer rate than the hearing loss or late deafness group. The correct answer rate for all participants in the manually translated videos was an average of 62.8% (standard deviation 16.5%) for the hearing loss or late deafness group and 67.0% (standard deviation 11.4 %) for the deaf group . Although the difference was not significant, the deaf group tended to have a higher correct answer rate. On the other hand, the accuracy rate for machine-translated videos was 60.3% (standard deviation 11.9%) for the hearing-impaired/late-deafened group and 66.3% (standard deviation 16.8%) for the deaf group.

The correct answer rate tended to be higher in the deaf group than in the hearing-impaired/late-deafened group. Figure 5 shows a boxplot of the distribution of correct answer rates for all participants. The boxplot shows that the variance in correct answer rates was smallest when the deaf group was presented with the manually translated video. Excluding the outliers in the deaf group, the distribution of correct answer rates for the manually translated video showed a larger difference between the maximum and minimum values of correct answer rates and a larger variance in the hearing-impaired/late-deafened group than in the deaf group. The distribution of correct answer rates for the machine-translated video also showed a larger variance in the deaf group, but more participants in the deaf group had a higher correct answer rate than the hearing-impaired/late-deafened group. The overall correct answer rate was higher in

the deaf group whose native language is sign language than in the hearing-impaired/late-deafened group whose native language is not sign language, suggesting that this difference is due to differences in sign language proficiency. Furthermore, there was no significant difference in the trend in correct answer rates for each translation method between the two groups, and manual translation tended to have a higher correct answer rate than machine translation, consistent with the overall trend for all participants. Therefore, it was shown that the overall trend in correct answer rates does not depend on sign language proficiency. (2) Analysis of the effect of age To verify whether

age had an effect on correct answer rates,

we analyzed the correct answer rates by age group. Table 4 shows the correct answer rates by age group for participants in the experiment, including both deaf people and people with hearing loss or late-onset hearing loss. The analysis showed that only the group under 20 had a higher correct answer rate for machine translation than for manual translation, but the other four groups showed the same overall trend of a higher correct answer rate for manual translation than for machine translation. The accuracy rates of machine translation for the group under 20 years old were 92.5% and 73.3% for the sports and medal news categories, respectively, which were higher than the accuracy rates of machine translation for the same categories for other age groups. As mentioned above, in the sports and medal news categories, the machine translation results tended to frequently use fingerspelling for proper nouns such as sports team names and athlete names. While many older generations are not accustomed to Japanese, which is not their native language, and have difficulty reading fingerspelling, many younger generations have received sufficient education in written Japanese (reading and writing). As a result, younger generations tend to actively use fingerspelling frequently even in sign language and are also good at reading it26), and it is possible that this difference may have influenced the results.

5.4 Problems caused by CG generation Since

the accuracy rate of manual translation assuming no translation errors was 66.5% , it is thought that even if machine translation outputs a 100% correct sign language word sequence, content comprehension will decrease due to problems caused by CG generation. One of the problems caused by CG generation is that the inflection of sign language words cannot be reproduced by connecting dictionary-type sign language word motions. In addition, the sign language word sequence resulting from the translation is necessary to express a sign language sentence.

[Copyrights to Machine-Translated Content]

The copyright of the original papers published on this website belongs to the Institute of Image Information and Television Engineers. Unauthorized use of original papers or translated papers is prohibited. Please be sure to cite the original publication when referencing. For details, please refer to the copyright regulations of the Institute of Image Information and Television Engineers.

5. 結果と考察

5.1 内容理解に関する質問の正答率

評価結果について、内容理解に関する質問の正答率を表2に示す。正答率は、カテゴリごとの全質問数に対する正解した回答数の割合とし、実験協力者の属性ごとに分けて算出した。評価の結果、実験協力者全体における人手翻訳と機械翻訳それぞれの手話CG動画の全カテゴリの正答率は66.5%（標準偏差15.2%）と65.5%（標準偏差15.9%）となった。機械翻訳において、カテゴリの違いによる正答率の差が人手翻訳に比べて大きい結果となった。したがって、現状の機械翻訳モデルにおいては、カテゴリごとに翻訳精度にばらつきがあることが示された。さらに、機械翻訳モデルの学習データに多く含まれ、機械翻訳精度が高いと考えられるスポーツとメダル速報以外のカテゴリでは、人手翻訳に比べ機械翻訳の正答率が低いことから、機械翻訳精度が手話CGの内容理解に影響することが示された。

5.2 固有名詞の読み取り

固有名詞部分のみの読み取りに関する質問の正答率を表3に示す。固有名詞については、実験協力者全体における人手翻訳と機械翻訳の正答率がそれぞれ、82.1%と76.1%となった。人手翻訳については、スポーツのチーム名や政党名といった一部の組織名の正答率が機械翻訳の74.0%よりも低い69.1%となった。手話では話者にとって馴染みのない固有名詞を表現する際に、口型を強調して表現したり、指文字を併用する場合が多い²⁾。しかし、5.4節で後述する口型が再現できていないというCG生成の課題に加えて、人手翻訳では翻訳の段階で指文字を併用せずに、手話単語のみで表現されていたことが内容理解に影響したと考えられる。一方で、機械翻訳では組織名の手話表現に指文字が使用されることが多かったため、人手翻訳に比べ正答率が高くなったと考えられる。

5.3 実験協力者の属性別の分析

(1) 手話の習熟度の影響に関する分析

手話の習熟度が評価結果に影響しているか検証するために、実験協力者を母語が手話ではない難聴・中途失聴者と幼少期から母語として手話を使用しているろう者の2つのグループに分けて分析したところ、表2と表3で示すとおりのいくつかのカテゴリを除いて、難聴・中途失聴者グループよりろう者グループの方が高い正答率となった。人手翻訳動画におけるすべての実験協力者の正答率は、難聴・中途失聴者グループが平均62.8%（標準偏差16.5%）で、ろう者グループが平均67.0%（標準偏差11.4%）であり、有意な差ではないが、ろう者グループの方が正答率が高い傾向が見られた。一方で、機械翻訳動画における正答率は、難聴・中途失聴者グループが平均60.3%（標準偏差11.9%）で、ろう者グループが平均66.3%（標準偏差16.8%）であり、こちらも有意な差ではないが、ろ

う者グループの方が正答率が高い傾向が見られた。また、すべての実験協力者の正答率の分布に関する箱ひげ図を図5に示す。箱ひげ図においては、ろう者グループに対して人手翻訳動画を提示した場合が最も正答率のばらつきが少ない結果となることが示された。人手翻訳動画の正答率の分布は、ろう者グループの外れ値を除くと、ろう者グループよりも難聴・中途失聴者グループの方が正答率の最大最小値の差やばらつきが大きい結果となった。また、機械翻訳動画の正答率の分布は、ろう者グループの方が正答率のばらつきは大きい、難聴・中途失聴者グループよりもろう者グループの方が高い正答率の実験協力者が多いことが示された。

母語が手話ではない難聴・中途失聴者グループより、母語が手話であるろう者グループの方が全体的に正答率が高くなったことから、手話の習熟度の違いによって生じた差であると考察できる。また、翻訳手法ごとの正答率の傾向は、2つのグループ間で大きな差はなく、実験協力者全体の傾向と同じく機械翻訳より人手翻訳の方が正答率が高い傾向となった。したがって、全体的な正答率の傾向に関しては、今回の評価結果が手話の習熟度に依存しないことが示された。

(2) 年齢の影響に関する分析

正答率に年齢の影響がないか検証するために、年代別に正答率を分析した。表4に、ろう者と難聴・中途失聴者をどちらも含めた実験協力者の年代別の正答率を示す。分析の結果、20代以下のグループのみ人手翻訳より機械翻訳の正答率が高くなったが、それ以外の4つのグループは機械翻訳より人手翻訳の正答率が高いという全体の傾向と同じ結果となった。20代以下のグループの機械翻訳の正答率について、スポーツとメダル速報のカテゴリの正答率がそれぞれ92.5%と73.3%で、他の年代の同カテゴリの機械翻訳の正答率と比べて高くなった。スポーツやメダル速報では、機械翻訳結果において、先述したようにスポーツのチーム名や選手名の固有名詞に指文字が多用される傾向があった。高齢の世代は母語ではない日本語に慣れておらず指文字の読み取りを苦手とする人が多い一方で、若い世代は書記日本語（読み書き）の教育を十分に受けている人が多い。そのため、若い世代の人は普段から手話でも指文字を積極的に多用し、読み取りも得意な人が多い傾向があり²⁶⁾、その差が結果に影響した可能性があると考えられる。

5.4 CG生成に起因する問題

翻訳誤りがないと仮定した人手翻訳の正答率が66.5%であったことから、機械翻訳で100%正しい手話単語列が出力されても、CG生成に起因する問題により内容理解度が低下すると考えられる。CG生成に起因する問題のひとつとして、辞書形の手話単語モーションの接続では手話の語形変化が再現できないことが挙げられる。また、翻訳結果の手話単語列は手話を表現するうえで必



Figure 4 Comparison of manual and machine-translated sign language CG

(Example of organization name "Yakult")

Another problem is the lack of necessary information.

is a sign language word that represents the sign language words that make up a sign language sentence and their word order.

In addition to word sequences, we also generated CG images such as nodding and pointing.

In order to automatically identify grammatical elements of sign language, proper nouns and phrases are

However, the system outputs auxiliary information to indicate divisions such as facial expressions.

It is not possible to reproduce the context-dependent changes in sign language word forms, such as the change of

Therefore, the content of the sign language CG automatically generated by machine translation is

To improve the solution, we first need to address the shortcomings in CG generation.

We have added a function to reproduce the inflections of sign language that is currently used.

Auxiliary information for automatic processing of sign language functions is combined with the sign language word sequence.

It is necessary to output the

Another factor is that the mouth shape cannot be reproduced.

There are also problems with controlling the mouth shape of sign language CG .

The mouth motion and hand motion are linked together for the famous words.

The avatar is not moving and the mouth motion

There are two issues with this method: the lack of sufficient facial control.

When replacing the mouth motion of a noun, the hand motion is

The data on the joints of the mouth of the character is used to derive the pronunciation of proper nouns.

Mouth shape motion data generated separately based on the information

When replacing 1)25) , the mouth movement and hand movement become asynchronous.

This creates a sense of incongruity.

The transformation method differs depending on the sequence of characters.

In this system, mouth motion is generated by combining characters.

As a result, the deformation of the mouth movement, which should differ depending on the position, is not reproduced.

As a result, the mouth shape becomes unnatural.

The problems caused by CG generation will be discussed later in Chapter 6.

This was further verified in detail through face-to-face interviews.

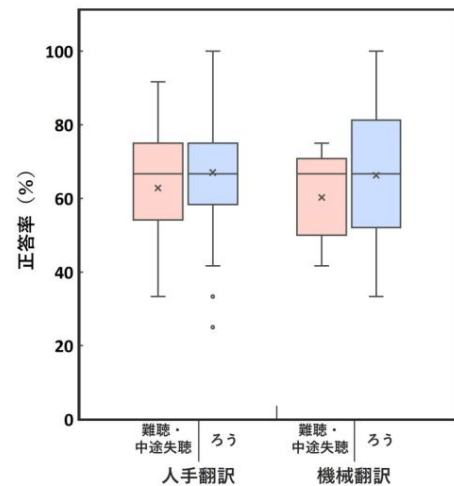


Figure 5. Distribution of correct answer rates among participants

Table 3. Correct answer rate for questions about reading proper nouns.

ジャンル	固有 名詞数	正答率 [%]					
		人手翻訳			機械翻訳		
		難聴・ 中途失聴	ろう	全体	難聴・ 中途失聴	ろう	全体
地名	10	75.0	85.3	84.1	66.2	78.9	77.1
人名	4	76.9	83.0	82.2	69.2	77.8	76.7
競技名	6	85.7	88.3	88.0	77.8	73.3	73.8
組織名	6	63.9	69.8	69.1	66.7	75.2	74.0
災害名	2	92.3	93.2	93.1	76.9	83.0	82.2
全ジャンル	28	76.9	82.9	82.1	69.6	77.1	76.1

6. Issue verification through in-depth interviews

Chapter 5 : Issues regarding CG generation obtained from the evaluation experiment results

To examine the issue in more detail, we used only the manually translated videos.

A face-to-face in- depth interview was conducted.

The in-depth interview is shown in Figure 1. The participants were in their 20s .

A total of six people (five deaf and one with hearing loss) in their 60s and 70s .

We provide hand-translated CG videos of breaking news in sign language.

After showing the information, the student will listen to what they have read in sign language.

The breaking news used for the evaluation consisted of nine sentences, as shown in Section 4.1 .

The six categories used in the web evaluation experiment were traffic information,

The information was selected from the seven categories to which it was added.

As a result of the interviews, the number of people who responded "unfamiliar" to proper nouns was

"Place names are difficult to read," "The speed is too fast," "There are no mouth shapes,"

Therefore, they mistook it for another word or could not read it.

Many comments were made.

Regarding the sign language representation of proper nouns that do not exist, the considerations in Chapter 5

As in the previous results, in addition to improving mouth control, supplementary explanations of sign language expressions were provided.

It was shown that it was necessary to add finger alphabets to facilitate this.

If the mouth shape is easy to read like a human sign language speaker, then it is possible to use finger spelling.

Although there is no need for further explanation, there are issues with the accuracy of mouth shape control.

In the sign language CG of this system , finger alphabets are added.

It is believed that providing additional information through this method can improve comprehension of the content.

can be obtained.

Other comments include, "The topic of missiles has become stiff.

【Copyrights to Machine-Translated Content】

The copyright of the original papers published on this website belongs to the Institute of Image Information and Television Engineers.

Unauthorized use of original papers or translated papers is prohibited. Please be sure to cite the original publication when referencing.

For details, please refer to the copyright regulations of the Institute of Image Information and Television Engineers.



図 4 人手翻訳と機械翻訳の手話 CG 比較
(組織名「ヤクルト」の例)

要な情報が不足していることも問題となる。本システムは、手話文を構成する手話単語とその語順を表す手話単語列に加えて、CG 生成において頷きや指差しといった手話の文法要素を自動付与するために、固有名詞や文節などの区切れを表す補助情報を出力している¹⁹⁾が、表情の変化など文脈に応じた手話の語形変化は再現できない。したがって、機械翻訳で自動生成した手話 CG の内容理解を向上するためには、まず CG 生成において不足している手話の語形変化を再現する機能を追加し、さらにその機能を自動処理するための補助情報を手話単語列と併せて出力する必要がある。

また、他の要因として口型が再現できていないという問題も挙げられる。手話 CG の口型制御においては、固有名詞について置換した口型モーションと手指動作が連動していないことと、口型モーションによるアバターの顔の制御が充分ではないことの2つの課題がある。固有名詞の口型モーションを置換する際は、手指動作のモーションの口の関節部分のデータを、固有名詞の読み仮名の情報をもとに別途生成した口型モーションのデータと差し替える¹⁾²⁵⁾が、口型動作と手指動作が非同期となることが違和感となる。また、人間の口の動きは発話する文字の並びによってその変形の仕方が異なる。しかし、本システムの口型モーション生成では、文字の組み合わせによって異なるべき口の動きの変形が再現されず、結果として口型が不自然になる。

CG 生成に起因する問題については、6 章で後述するデプスインタビューによってさらに詳細に検証した。

表 2 内容理解に関する質問の正答率。

カテゴリ	文数	正答率 [%]					
		人手翻訳			機械翻訳		
		難聴・中途失聴	ろう	全体	難聴・中途失聴	ろう	全体
避難・安全	5	64.5	72.1	71.2	58.8	65.2	64.3
選挙速報	2	50.0	55.8	55.2	50.0	44.4	45.3
電力不足	4	65.2	64.9	65.0	55.2	64.6	63.3
ミサイル	3	77.8	72.5	73.2	61.9	70.7	69.5
スポーツ	4	81.3	76.1	76.9	90.0	79.7	80.7
メダル速報	6	42.9	59.0	56.8	52.8	64.9	63.4
全カテゴリ	24	62.8	67.0	66.5	60.3	66.3	65.5

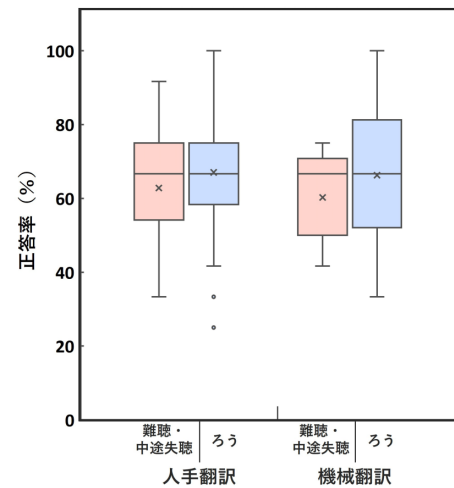


図 5 実験協力者の正解率の分布

表 3 固有名詞の読み取りに関する質問の正答率。

ジャンル	固有名詞数	正答率 [%]					
		人手翻訳			機械翻訳		
		難聴・中途失聴	ろう	全体	難聴・中途失聴	ろう	全体
地名	10	75.0	85.3	84.1	66.2	78.9	77.1
人名	4	76.9	83.0	82.2	69.2	77.8	76.7
競技名	6	85.7	88.3	88.0	77.8	73.3	73.8
組織名	6	63.9	69.8	69.1	66.7	75.2	74.0
災害名	2	92.3	93.2	93.1	76.9	83.0	82.2
全ジャンル	28	76.9	82.9	82.1	69.6	77.1	76.1

6. デプスインタビューによる課題検証

5 章の評価実験結果から得られた CG 生成に関する課題をより詳細に検証するために、人手翻訳動画のみを用いて、対面によるデプスインタビューを実施した。図 6 にデプスインタビューの様子を示す。実験協力者は 20 代から 60 代までの全 6 名 (ろう者 5 名, 中途失聴者 1 名) で、ニュース速報に関する人手翻訳の手話 CG 動画を提示した後に読み取れた内容について手話でヒアリングした。評価に用いたニュース速報は全 9 文で、4.1 節で示した Web の評価実験で使用した 6 つのカテゴリに交通情報を追加した 7 カテゴリから選定した。

ヒアリングの結果、固有名詞について、「馴染みのない地名は読み取り困難」、「速度が速すぎる」、「口型がなかったため他の単語と間違えた・読み取れなかった」といった指摘が多く得られた。地名など話者にとって馴染みのない固有名詞の手話表現に対しては、5 章で示した考察結果同様、口型制御の改善に加えて手話表現を補足説明するための指文字を追加する必要があることが示された。人間の手話話者のように口型が読み取り易ければ指文字による補足説明は不要であるが、口型制御の精度に課題のある本システムの手話 CG においては、指文字を追加して付加情報を伝えることで内容理解を向上できると考えられる。

そのほかの指摘として、「ミサイルの話題はこわばった

Table 4. Correct answer rate by age of

カテゴリ	文数	正答率 [%]									
		～20 代		30 代		40 代		50 代		60 代～	
		人手	機械	人手	機械	人手	機械	人手	機械	人手	機械
避難・安全	5	78.0	68.0	66.0	66.0	71.6	67.1	69.4	56.9	70.4	60.7
選挙速報	2	45.0	30.0	55.0	55.0	64.3	50.0	55.6	45.5	50.0	41.7
電力不足	4	55.0	72.5	80.0	60.0	67.2	53.2	65.8	66.7	47.6	73.9
ミサイル	3	70.0	66.7	63.3	73.3	81.8	69.6	75.9	67.7	68.8	70.6
スポーツ	4	80.0	92.5	80.0	72.5	75.0	83.9	77.3	83.3	70.8	60.0
メダル速報	6	61.7	73.3	58.3	70.0	56.5	62.5	58.1	63.8	67.6	50.0
全カテゴリ	24	66.7	70.8	67.5	67.1	68.3	65.0	67.1	64.6	64.4	59.8



Figure 6 : Depth interview

There were also many comments about facial expressions that hinder understanding of the content, such as "A smiling face is appropriate, but a bright expression is inappropriate," and "If service is suspended, it is better not to smile." These indicated that improvements are needed to control facial expressions in line with the context in order to properly convey information.

7. Conclusion

We developed a function to reproduce grammatical elements specific to sign language in an arbitrary sign language CG generation system and conducted evaluation experiments of the sign language CG generated by the system using breaking news. The results of a question-based evaluation experiment on a web page showed that the accuracy rate of questions regarding comprehension of the sign language CG generated by machine translation was lower than that of sign language CG generated by manual translation, indicating the need for improved translation accuracy. We also found that translation accuracy varied by news category, and that even when the translated sign language words and the overall word order of the sentence were correct and there were no translation errors, comprehension of the content decreased due to CG generation. Furthermore, in-depth interviews regarding the CG generation revealed issues regarding the addition of supplementary finger spellings to express proper nouns, the use of appropriate presentation speeds and mouth shapes, and appropriate facial expression control depending on the context. By expanding the training data to improve the accuracy of machine translation, and by overcoming the challenges of CG generation by introducing techniques such as finger spelling for proper nouns and controlling facial expressions according to the context, we believe that comprehension of the content of the sign language CG automatically generated by this system will improve.

- 高嶋由布子: “危機言語としての日本手話”, 国立国語研究所論集, **18**, pp.121-148 (2020)
- J.H. Kim, E. J. Hwang, S. Cho, D. H. Lee, J. Park: “Sign Language Production With Avatar Layering: A Critical Use Case over Rare Words”, In 13th International Conference on Language Resources and Evaluation (LREC 2022), pp.1519-1528 (2022)
- S. Ebling: “Automatic translation from german to synthesized swiss german sign language”, PhD Thesis (University of Zurich) (2016)
- S. Morrissey: “Data-driven machine translation for sign languages”, PhD Thesis (Dublin City University) (2008)
- J. C. McDonald, R. Wolfe, J. Schnepf, J. Hochgesang, D. G. Jamrozik, M. Stumbo, L. Berke, M. Bialek, F. Thomas: “An automated technique for real-time production of lifelike animations of american sign language”, Universal Access in the Information Society, **15**, 4, pp.551-566 (2016)
- M. Filhol, J. C. McDonald: “Extending the AZee-Paula shortcuts to enable natural proform synthesis”, Proceedings of the LREC 2018 8th Workshop on the Representation and Processing of Sign Languages, pp.45-52 (2018)
- S. Gibet, N. Courty, K. Duarte, T. L. Naour: “The signcom system for data-driven animation of interactive virtual signers: methodology and evaluation”, ACM Transactions on Interactive Intelligent Systems (TiiS), **1**, 1, pp.1-23 (2011)
- L. Naert, C. Larboulette, S. Gibet: “LSF-ANIMAL: A motion capture corpus in French Sign Language designed for the animation of signing avatars”, In 12th International Conference on Language Resources and Evaluation (LREC 2020), pp.6008-6017 (2020)
- H. Brock, K. Nakadai: “Deep JSLC: A multimodal corpus collection for datadriven generation of Japanese Sign Language expressions”, In 11th International Conference on Language Resources and Evaluation (LREC 2018), pp.4247-4252 (2018)
- H. Brock, S. Nishina: “Quantifying Sign Avatar Perception: How Imperfect is Insufficient?”, Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp.1-8 (2020)
- R. Wolfe, J. C. McDonald, T. Hanke, S. Ebling, D. V. Landuyt, F. Picon, V. Krausneker, E. Efthimiou, E. Fotinea, A. Braffort: “Sign language avatars: A question of representation”, Information, **13**, 4:206, WP-B-5 (2022)
- 内田翼, 佐野雅規: “手話 CG アニメーション生成における手話特有の文法要素の影に関する調査”, 映像学創立 70 周年記念大, 31D-2 (2020)
- L. Naert: “Capture, annotation and synthesis of motions for the data-driven animation of sign language avatars”, PhD Thesis (Southern Brittany University) (2020)
- M. Huenerfauth, P. Lu, H. Kacorri: “Synthesizing and evaluating animations of american sign language verbs modeled from motion-capture data”, Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2015), pp.22-28 (2015)
- A.L. Dimou, V. Papavassiliou, J. McDonald, T. Goulas, K. Vasiliaki, A. Vacalopoulou, S.E. Fotinea, E. Efthimiou, R. Wolfe: “Signing avatar performance evaluation within the EASIER project”, Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT), pp.39-44 (2022)
- F. Picon, D. V. Landuyt, R. Omardeen, E. Efthimiou, R. Wolfe, S.E. Fotinea, T. Goulas, C. Tismer, M. Kopf, T. Hanke: “The EASIER Mobile Application and Avatar End-User Evaluation Methodology”, Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources, pp.276-281 (2024)

[Copyrights to Machine-Translated Content]

The copyright of the original papers published on this website belongs to the Institute of Image Information and Television Engineers. Unauthorized use of original papers or translated papers is prohibited. Please be sure to cite the original publication when referencing. For details, please refer to the copyright regulations of the Institute of Image Information and Television Engineers.

表 4 実験協力者の年代別の正解率.

カテゴリ	文数	正答率 [%]									
		～20 代		30 代		40 代		50 代		60 代～	
		人手	機械	人手	機械	人手	機械	人手	機械	人手	機械
避難・安全	5	78.0	68.0	66.0	66.0	71.6	67.1	69.4	56.9	70.4	60.7
選挙速報	2	45.0	30.0	55.0	55.0	64.3	50.0	55.6	45.5	50.0	41.7
電力不足	4	55.0	72.5	80.0	60.0	67.2	53.2	65.8	66.7	47.6	73.9
ミサイル	3	70.0	66.7	63.3	73.3	81.8	69.6	75.9	67.7	68.8	70.6
スポーツ	4	80.0	92.5	80.0	72.5	75.0	83.9	77.3	83.3	70.8	60.0
メダル速報	6	61.7	73.3	58.3	70.0	56.5	62.5	58.1	63.8	67.6	50.0
全カテゴリ	24	66.7	70.8	67.5	67.1	68.3	65.0	67.1	64.6	64.4	59.8



図 6 デプスインタビューの様子

顔が適当, 明るい表情は不適当」, 「運転見合わせの場合は, 笑顔でない方が良い」といった内容理解を妨げる表情に関する指摘も多く, 情報を適切に伝える上で文脈に沿った表情制御へと改善する必要があることが示された.

7. む す び

任意文手話 CG 生成システムにおいて手話特有の文法要素を再現するための機能を開発し, ニュース速報を対象として, システムで生成した手話 CG の評価実験を実施した. Web ページによる質問形式の評価実験の結果, 機械翻訳で生成した手話 CG の内容理解に関する質問の正答率が人手翻訳で生成した手話 CG の正答率より低下したことから, 翻訳精度の向上が必要であることが示された. また, ニュースのカテゴリごとに翻訳精度にばらつきがあることや, 訳出された各手話単語と文全体の語順が正しく翻訳誤りがない場合でも CG 生成が原因で内容理解が低下することが示された. さらに, CG 生成に関するデプスインタビューにより, 固有名詞表現における補助的な指文字の付与および適切な提示速度・口型の使用に関する課題や, 文脈に応じた適切な表情制御に関する課題があることが示された. 学習データを拡充して機械翻訳精度を向上するとともに, 固有名詞への指文字付与や文脈に応じた表情制御手法の導入により CG 生成の課題を改善することで, 本システムで自動生成した手話 CG の内容理解が向上すると考えられる.

(文 献)

- 1) 中谷真規, 内田翼, 宮崎太郎, 金子浩之, 佐野雅規: “句や節のまとまりを考慮した任意文手話 CG 生成システムの開発”, 2023 映情学年大, 31B-5 (2023)
- 2) 文部科学省: “聴覚障害教育の手引”, ジアース教育新社 (2020)

- 3) 高嶋由布子: “危機言語としての日本手話”, 国立国語研究所論集, **18**, pp.121-148 (2020)
- 4) J.H. Kim, E. J. Hwang, S. Cho, D. H. Lee, J. Park: “Sign Language Production With Avatar Layering: A Critical Use Case over Rare Words”, In 13th International Conference on Language Resources and Evaluation (LREC 2022), pp.1519-1528 (2022)
- 5) S. Ebling: “Automatic translation from german to synthesized swiss german sign language”, PhD Thesis (University of Zurich) (2016)
- 6) S. Morrissey: “Data-driven machine translation for sign languages”, PhD Thesis (Dublin City University) (2008)
- 7) J. C. McDonald, R. Wolfe, J. Schnepp, J. Hochgesang, D. G. Jamrozik, M. Stumbo, L. Berke, M. Bialek, F. Thomas: “An automated technique for real-time production of lifelike animations of american sign language”, Universal Access in the Information Society, **15**, 4, pp.551-566 (2016)
- 8) M. Filhol, J. C. McDonald: “Extending the AZee-Paula shortcuts to enable natural proform synthesis”, Proceedings of the LREC 2018 8th Workshop on the Representation and Processing of Sign Languages, pp.45-52 (2018)
- 9) S. Gibet, N. Courty, K. Duarte, T. L. Naour: “The signcom system for data-driven animation of interactive virtual signers: methodology and evaluation”, ACM Transactions on Interactive Intelligent Systems (TiiS), **1**, 1, pp.1-23 (2011)
- 10) L. Naert, C. Larboulette, S. Gibet: “LSF-ANIMAL: A motion capture corpus in French Sign Language designed for the animation of signing avatars”, In 12th International Conference on Language Resources and Evaluation (LREC 2020), pp.6008-6017 (2020)
- 11) H. Brock, K. Nakadai: “Deep JSLC: A multimodal corpus collection for datadriven generation of Japanese Sign Language expressions”, In 11th International Conference on Language Resources and Evaluation (LREC 2018), pp.4247-4252 (2018)
- 12) H. Brock, S. Nishina: “Quantifying Sign Avatar Perception: How Imperfect is Insufficient?”, Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp.1-8 (2020)
- 13) R. Wolfe, J. C. McDonald, T. Hanke, S. Ebling, D. V. Landuyt, F. Picron, V. Krausneker, E. Efthimiou, E. Fotinea, A. Braffort: “Sign language avatars: A question of representation”, Information, **13**, 4:206, WP-B-5 (2022)
- 14) 内田翼, 佐野雅規: “手話 CG アニメーション生成における手話特有の文法要素の影に関する調査”, 映情学創立 70 周年記念大, 31D-2 (2020)
- 15) L. Naert: “Capture, annotation and synthesis of motions for the data-driven animation of sign language avatars”, PhD Thesis (Southern Brittany University) (2020)
- 16) M. Huenerfauth, P. Lu, H. Kacorri: “Synthesizing and evaluating animations of american sign language verbs modeled from motion-capture data”, Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2015), pp.22-28 (2015)
- 17) A.L. Dimou, V. Papavassiliou, J. McDonald, T. Goulas, K. Vasiliaki, A. Vacalopoulou, S.E. Fotinea, E. Efthimiou, R. Wolfe: “Signing avatar performance evaluation within the EASIER project”, Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT), pp.39-44 (2022)
- 18) F. Picron, D. V. Landuyt, R. Omardeen, E. Efthimiou, R. Wolfe, S.E. Fotinea, T. Goulas, C. Tismer, M. Kopf, T. Hanke: “The EASIER Mobile Application and Avatar End-User Evaluation Methodology”, Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources, pp.276-281 (2024)

- 19) T. Miyazaki, N. Nakatani, T. Uchida, H. Kaneko, M. Sano: "Machine Translation to Sign Language using Post-translation Replacement without Placeholders", Eighth International Workshop on Sign Language Translation and Avatar Technology (SLTAT 2023), WP-B-5 (2023)
- 20) 加藤直人, 内田翼, 東真希子, 梅田修一: "ニュースを対象にした手話マルチメディアコーパスの構築", 言語資源活用ワークショップ発表論文集, **3**, pp.316-326 (2018)
- 21) 全日本ろうあ連盟: "わたしたちの手話 (1)~ (9)", 全日本ろうあ連盟出版局 (2004-2007)
- 22) 全日本ろうあ連盟: "新・日本語-手話辞典", 中央法規出版 (2011)
- 23) 梅田修一, 加藤直人, 東真希子, 内田翼, 金子浩之, 比留間伸行, 住吉英樹: "機械翻訳後に手修正可能な手話 CG 制作支援システム", 電子情報通信学会 HCG シンポジウム, A7-5 (2018)
- 24) T. Uchida, N. Nakatani, T. Miyazaki, H. Kaneko, M. Sano: "Motion Editing Tool for Reproducing Grammatical Elements of Japanese Sign Language Avatar Animation", Eighth International Workshop on Sign Language Translation and Avatar Technology (SLTAT 2023), WP-B-10 (2023)
- 25) 森田祐介, 内田翼, 佐野雅規: "ニューラル手話翻訳における固有名詞訳出手法の検討", 2022 映像学大会, 11B-1 (2022)
- 26) 福田友美子, 森本行雄, 四日市章: "聴覚障害者のコミュニケーション手段の使用に関する実態調査", AUDIOLOGY JAPAN, **37**, 3, pp.229-235 (1994)



うちだ つばさ
内田 翼 2011 年, 茨城大学大学院理工学研究科メディア通信工学専攻修士課程了。同年, NHK に入局。仙台放送局を経て, 2014 年から放送技術研究所にて, 手話 CG の研究開発に従事。現在, スマートプロダクション研究部に所属。正会員。



なかたに なおき
中谷 真規 2017 年, 同志社大学大学院生命医科学研究科修了。同年, NHK に入局。放送技術局を経て, 2019 年から放送技術研究所にて積層型撮像デバイス, 磁性記録デバイス, 2022 年より手話 CG の研究開発に従事。2023 年よりメディアイノベーションセンターで手話 CG の実用化に従事。



かじやま たけし
梶山 岳士 2003 年, 電気通信大学大学院電気通信学研究科修了。同年, NHK に入局。札幌放送局, NHK 放送技術研究所を経て, 2022 年より, NHK 財団に勤務。薄型光ディスク, スーパーハイビジョン記録装置, 手話 CG の研究に従事。正会員。



かねこ ひろゆき
金子 浩之 2000 年, 上智大・理工・電気電子卒。2002 年, 同大学院理工学研究科電気電子工学専攻修士課程了。同年, NHK に入局。仙台放送局を経て, 2006 から放送技術研究所においてコンテンツ記述・制作及び手話 CG 生成技術の研究に従事。正会員。

- 19) T. Miyazaki, N. Nakatani, T. Uchida, H. Kaneko, M. Sano: “Machine Translation to Sign Language using Post-translation Replacement without Placeholders”, Eighth International Workshop on Sign Language Translation and Avatar Technology (SLTAT 2023), WP-B-5 (2023)
- 20) 加藤直人, 内田翼, 東真希子, 梅田修一: “ニュースを対象にした手話マルチメディアコーパスの構築”, 言語資源活用ワークショップ発表論文集, **3**, pp.316-326 (2018)
- 21) 全日本ろうあ連盟: “わたしたちの手話 (1)～(9)”, 全日本ろうあ連盟出版局 (2004-2007)
- 22) 全日本ろうあ連盟: “新・日本語-手話辞典”, 中央法規出版 (2011)
- 23) 梅田修一, 加藤直人, 東真希子, 内田翼, 金子浩之, 比留間伸行, 住吉英樹: “機械翻訳後に手修正可能な手話 CG 制作支援システム”, 電子情報通信学会 HCG シンポジウム, A7-5 (2018)
- 24) T. Uchida, N. Nakatani, T. Miyazaki, H. Kaneko, M. Sano: “Motion Editing Tool for Reproducing Grammatical Elements of Japanese Sign Language Avatar Animation”, Eighth International Workshop on Sign Language Translation and Avatar Technology (SLTAT 2023), WP-B-10 (2023)
- 25) 森田祐介, 内田翼, 佐野雅規: “ニューラル手話翻訳における固有名詞訳出手法の検討”, 2022 映像学年大, 11B-1 (2022)
- 26) 福田友美子, 森本行雄, 四日市章: “聴覚障害者のコミュニケーション手段の使用に関する実態調査”, AUDIOLOGY JAPAN, **37**, 3, pp.229-235 (1994)



うちだ つばさ
内田 翼 2011 年, 茨城大学大学院理工学研究科メディア通信工学専攻修士課程了。同年, NHK に入局。仙台放送局を経て, 2014 年から放送技術研究所にて, 手話 CG の研究開発に従事。現在, スマートプロダクション研究部に所属。正会員。



なかに なおき
中谷 真規 2017 年, 同志社大学大学院生命医科学研究科修了。同年, NHK に入局。放送技術局を経て, 2019 年から放送技術研究所にて積層型撮像デバイス, 磁性記録デバイス, 2022 年より手話 CG の研究開発に従事。2023 年よりメディアイノベーションセンターで手話 CG の実用化に従事。



かじやま たけし
梶山 岳士 2003 年, 電気通信大学大学院電気通信学研究科修了。同年, NHK に入局。札幌放送局, NHK 放送技術研究所を経て, 2022 年より, NHK 財団に勤務。薄型光ディスク, スーパーハイビジョン記録装置, 手話 CG の研究に従事。正会員。



かね こ ひろゆき
金子 浩之 2000 年, 上智大・理工・電気電子卒。2002 年, 同大学院理工学研究科電気電子工学専攻修士課程了。同年, NHK に入局。仙台放送局を経て, 2006 から放送技術研究所においてコンテンツ記述・制作及び手話 CG 生成技術の研究に従事。正会員。