

[研究助成 (S)]

手話言語等の多文化共生社会に向けた ICT を活用したコミュニケーション基盤構築

2238002



研究代表者
(助成金受領者)

共同研究者

電気通信大学
大学院情報理工学研究科

筑波技術大学

筑波技術大学

電気通信大学

教授 高 橋 裕 樹

教授 加 藤 伸 子

教授 白 石 優 旗

教授 内 海 彰

[研究の目的]

現社会システムは無意識にきこえる人を前提として構築されており、きこえない人は日常生活や社会参画において多大なコミュニケーションの障壁に直面している。本課題では、視覚情報処理技術を用いた日本手話翻訳の研究開発を通して、これまでコミュニケーションが困難であった場面でも円滑なコミュニケーションの補助が可能な基盤構築を目指す。きこえない人には視覚言語である日本手話が母語となるが、きこえない子供にとって音声言語である日本語を習得することには、課題があることが知られている。また、日本語の音声の語順に合わせて手話単語や指文字を表示する日本語対応手話から「てにをは」を省いた方法では、日本手話の文法の重要部分が抜け落ち、論理的な表現を十分に伝えることが困難である。きこえない人の母語である日本手話や中間型手話では、時制、様態、疑問、あるいは格関係といった重要な文法構造が、顔の表情や頭の動きなどの非手指要素(Non-Manual Signals: NMS)、および手の動く方向や位置といった空間要素によって表現されるためである。

したがって、文化的所産である日本手話の普及を促し、きこえない人ときこえる人の双方向の円滑なコミュニケーションを可能にするため

には、これら非手指・空間要素を正確に解釈可能な日本手話の文法に対応した翻訳技術が不可欠である。しかし、こうした高度な日本手話翻訳の実現には、非手指要素や空間軌跡を詳細にタグ付けした大規模かつ高品質なコーパスが求められるが、そのアノテーション付与は極めて多大な労力と時間コストを要するという課題がある。

本課題では日本手話と日本語対応手話の双方の特徴を持ち日常会話で多用される中間型手話に類出する一致動詞、指差し、表情、首・頭の動作などの非手指・空間要素を自動的に解析・アノテーション支援するシステムを構築すること、および翻訳技術の基礎を検討し、きこえない人が社会において不自由なくのびやかに自己実現できる「共生的かつ共創的な多文化共生社会」の基盤を構築することを目的とする。

[研究の内容、成果]

1. はじめに

日本語の語順に手話単語を並べるだけの日本語対応手話では、格関係や重要な文法規則が欠落し、論理的なニュアンスが伝わりにくい。日本手話や、日常会話で多用される中間型手話では、手指表現に加え、顔の表情や頭の動きなどの非手指要素、および空間を立体的に活用する

空間要素が極めて重要な文法役割を担う。

本課題は、これら非手指要素と空間要素を考慮した手話翻訳コミュニケーションシステムの構築に向け、開発のボトルネックである手話コーパス構築を劇的に高速化する Interactive and Automated Annotation Assistance for Pidgin Sign Japanese (IA3-PSJ) の提案と、これまでデータが存在しない中間型手話を含む論理的表現の手話文収集を行った。各シーン 13 名の話者多様性を保持し、260 文のデータを収集した。また、手話話者である筑波技術大学の学生の協力を得ることで適切なアノテーション付与作業を行った。さらに、空間的格関係を捉える翻訳アルゴリズムの基礎検討を行った。

2. 中間型手話における文法要素解析とシーン選定

手話には、日本手話、中間型手話、日本語対応手話がある。本課題では、まず、それぞれの手話をより厳密に定義する。中間型手話の一端を日本語対応手話、もう一端を日本手話として、その中間のすべてを「中間型手話」とする [1]。ここで言う「日本語対応手話」とは、「てにをは」などの助詞をすべて指文字で表すなど、日本語を完全に手指で表す手指日本語である。一方、助詞がないものは中間型手話の一つと考えることができ、また、空間を使う一致動詞が含まれるなど、日本手話と共通する性質が見られる場合もある [2]。このように中間型手話は日本語対応手話に近いものから日本手話に近いものまで多岐にわたり、普段の手話による会話ではこの中間型手話が用いられている。まず、統語情報が失われると文意が伝わらない可能性がある文を選定し、日常的に手話を使っている筑波技大の学生 4 人に普段の会話と同じように手話表現をしてもらった。その手話表現に含まれる日本手話文法要素の種類、頻度を調べ、主要な文法要素を明確にした。中間型手話では、単に手指動作からなる手話単語だけでなく、空間や表情を用いて表される日本手話と共通する性質が

多く含まれる。このような例として、動詞を表す手話表現がある位置から別の位置へ移動することによって行為者と対象者が誰であるかという情報も表現可能な一致動詞などがある。今回の調査では、代表的な例として、一致動詞、指さし、首の動作（うなづき、首振り）、眉の動作（肯定的・否定的などの態度的極性）、程度を表す段階的形容詞、口形の有無を調べた。調査結果から一致動詞と指さしが中間型手話で重要な役割を果たすことが明らかになった。一致動詞の例として、図 1 に示すように「男が女を見る」を「てにをは」を使わず表現する場合には、男と女はそれぞれ異なる位置におかれ、男の位置から女の位置の方向に、「見る」という動作をすることで表現できる。そのため、単語が提示された位置や、動詞の方向の情報が必要となる。本課題では、主要文法要素として、一致動詞、指差しを含む表現の翻訳を具体的な目標とし、首や眉の動作、口型の情報も同時に収集可能な会話文を選定した。会話文で用いる一致動詞としては、見る、見せる、伝える、尋ねる、送る、メールする、助ける、与える、もらう、教える、話す、呼ぶ、貸す・借りる、怒る・叱る、殴る、行く・来る、歩く、車で行く・来るの 18 語を選出した。

3. Automated Annotation Assistance for Pidgin Sign Japanese (A3-PSJ) [成果 1]

非手指要素である表情や空間のジェスチャを自動アノテーションする初期プロトタイプシステム A3-PSJ を構築した。

1) 非手指要素の自動検出

a) 身体・手先・顔骨格の抽出：BlazePose [3]



図 1 「男が女を見る」という手話を、空間を用いて表した例

による身体姿勢推定, HDR [4] による遮蔽に頑健な手指姿勢推定形状, および dlib ベースアルゴリズム [5] による顔の高精度な特徴点検出。時間的な連続性を維持するため TAPIR による頑健な特徴点追跡。

- b) 表情・頭部動作の識別: Facial Action Unit (FAU) の自動識別 [6] によって, 眉の上下, ひそめ, うなずき, 首の左右の動きなどの非手指要素を自動タグ付け。
- 2) 空間アノテーション: 一致動詞や指差し等の解釈に必要な空間情報を得るため, 図 2 に示す話者の両肩中心と鼻を結ぶ安定した参照点から「動的な中央領域」および「周囲 8 領域」を定義し, ジェスチャが実行された空間領域と 3 次元の指差し方向ベクトルの自動取得。

予備データセットとして YouTube から公開手話動画 30 本を収集し, 94 個の非手指要素インスタンスを評価対象とした。2 クラスの顔の表情, 6 クラスの頭部の動きの分類を行った。図 3 に特徴点検出と空間アノテーション例を示す。従来は作業者がフレーム毎に手動でタグ付

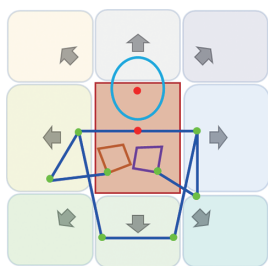


図 2 空間アノテーション

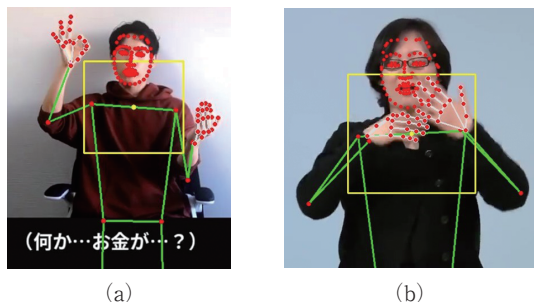


図 3 特徴点検出と空間アノテーション例

図 3 は (a) <https://youtu.be/zy7cZeCuwwE?si=nK7wN2ZtMYIAXBP> (b) <https://youtu.be/jh0xc4IoTrU> の動画を利用

けしていた高度な非手指表現や手の空間軌跡に対して, システムが初期タグを自動付与してアノテータが微調整する A3-PSJ を実現し, 作業負担の軽減を実現した。一致動詞の開始点・終了点および方向ベクトルを数値データとして自動抽出可能にしたことで, 誰が誰にアクションを行っているかという手話特有の空間的文法特性を数理的に解釈し, 高精度な自然言語翻訳システムへ連携させるための基盤を作成した。

4. Interactive and Automated Annotation Assistance for Pidgin Sign Japanese (IA3-PSJ) [成果 2]

指ドラッグや手先補間などの四つのインタラクティブ機能を A3-PSJ に導入し効率的かつ高精度なアノテーション支援システムを提案した。図 4 に IA3-PSJ のユーザインタフェースを示す。

- 1) フレーム単位検出による軽量化: 従来の複雑な時間軸追跡戦略を排し, あえてフレーム単位の検出に特化するアーキテクチャへと変更した。これによって, 検出精度を同等に維持したまま, 計算オーバーヘッドを大幅に削減し, アノテーション作業中の極めて滑らかで応答性の高い操作画面を実現した。
- 2) 4 インタラクティブ編集機能の導入: A3-PSJ では, オクルージョンやモーションブラーによってエラーを起こすことがあった。これを解決するため, アノテータがその場で

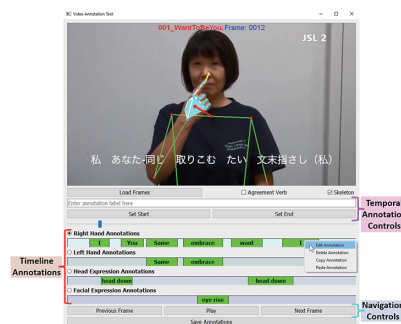


図 4 IA3-PSJ ユーザインタフェース

微調整できる「ランドマーク直接編集」,「指全体のドラッグ移動」,「手先形状のコピー&ペースト」,「手先位置の自動補間」の四つの手動補正機能を実装した。

3) 空間的格関係の自動アノテーション: A3-PSJ の処理を継承。

実用性と効率性を評価するため, A3-PSJ で作成した予備データセットを用いた比較実験を行った。

- a) 作業時間の削減: 完全に手動でアノテーションを行う場合平均 40.62 秒と比較し, 自動検出の初期結果を提案ツールでインタラクティブに微調整する手法では平均 24.37 秒へと大幅に短縮され, アノテーションに必要な時間を一律で 40.0% 削減した。
- b) アノテータの負担軽減とデータ品質向上: 計算負荷の低いフレーム単位処理と柔軟な手動補正ツールの統合は, 作業者の疲労を最小限に抑えつつ, 手先の欠落など自動アノテーション技術の限界を高いユーザビリティでカバーすることが可能になった。これにより, 手話翻訳システムの学習に必要な不可欠な, 大規模かつ高品質な日本手話アノテーションコーパスの構築プロセスを劇的に加速させた。

5. 自動アノテーション結果を用いた手話日本語変換方法の検討

手話の映像や動作を, 対応する単語 (形態素) のラベル (グロス) の列として 1 次元的に表現する際に, グロス列が「女」「男」「見る」だった場合には, 女が男を見る, 男が女を見る, 女と男が見る, など, 複数の日本語文の候補が考えられる。本研究では, 中間型手話でも多く使われていた手の位置と一致動詞の動きの方向等の情報から, 統語情報の抽出を検討する。

1) 手の位置について

図 5 の例において, 右手を RH, 左手を LH とし, 自動アノテーションにより空間位置として右 (R), 左 (L), 中央 (C) や動作方向が付

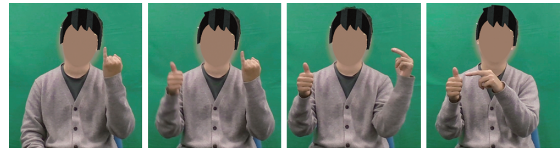


図 5 空間位置を使った表現の例



図 6 指差し例

図 7 空間位置の利用例

与できれば, 「女 (LH:L)」「男 (RH:R)」「見る (LH:L>R)」という情報を統合し「女が男を見る」を訳の第一候補として提示できる可能性がある。

現行ツールでは, 各フレームの手座標や指差し情報が保存される。ただし遮蔽等により検出できない場合もあり, 文法的に機能する位置をどの時点で抽出するかという課題がある。代表的な 3 つの例について実際のデータを元に検討する。

(1a) 指差しがあった場合

指差しがあった場合には, その位置が参照されていることを表す。自分を指差す PT1, 相手を指差す PT2, 他の特定の空間を指差す PT3 がある。図 6 は, 左手を指差して参照している例である。図 6 の例では PT3 として, 取得された指差し情報とグロスを対応付けることで, 指差し先を「女」に対応する空間位置の候補としてあつかえる可能性がある。

(1b) 左右の空間を使いわけている場合

指差しは行っていないが, 左右の空間を使い分けている例は, 中間型話者に多く見られた。ただし, 図 7 右側の例のように両手共に顔の近くに配置されると, 両手が同じ領域にあると判断される場合があり, 領域の設定には調整が必要である。

図 8 に手の動きの例を示す。左手が上に動いて静止し保持したまま, 右手が上に動いて静止

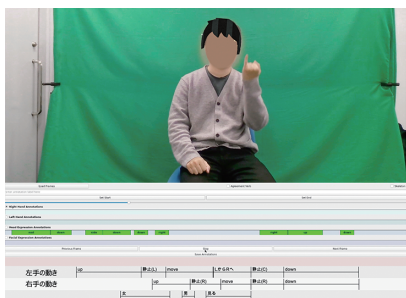


図8 自動アノテーションによる頭部の動きの抽出例と手の動きの例

している。この静止時の手座標とグロス区間を対応付け、名詞表現に対応する空間位置を推定する方法を検討した。両手表現の手話で空間に配置する場合も同様に静止時の位置情報で把握することが可能である。

(1c) 一致動詞の方向を用いる場合

「見る」を表出している際には、左手が「見る」の手形のまま右手の位置の近くに移動している。そのため、「見る」のグロス区間の開始時と終了時の手の座標から移動方向を推定する方法を検討した。

2) うなずきと頭部の動きについて

自動アノテーションツールを用いた場合、うなずきは「nod」と判定される場合の他、「down」と「up」等と判定される場合も見られた。頭部の動きの速さの個人差もあり判定の調整方法を検討する必要がある。

また手の動き始めと同時に頭部も動く事例があるなど、目視では気づきにくい細かな頭部動作が検出された。さらに映像上で視線を特定の方角へ向けた場面で、それに対応する頭部動作が検出された事例もあった。このような情報は統語情報を得るための有効な手掛かりになることが期待される。

中間型手話は話者の数と同じだけ変種があると言われているが [1], 指差し, 手座標, 動作方向, 頭部動作等を組み合わせることで, 日本語訳候補の絞り込みに利用できる可能性がある。

[今後の研究の方向, 課題]

1. 特徴点検出技術の高度化とロバスト性向上

高速動作に伴うモーションブラーや自己遮蔽が生じる複雑な会話シーンで, 既存手法では安定した特徴点検出が困難であるという課題も明らかになった。現在, 指先姿勢推定手法の適切な評価手法の検討と事後処理として推定値補正を行う手法の検討も行っており, より安定した特徴点検出を可能とし, アノテーション支援や手話翻訳性能の向上を目指す。

2. 空間格関係を表現する時空間グラフの改善

「一致動詞」と「指差し」は, 話者と対話者, 仮想対象の3次元的な空間配置, および動詞の移動方向ベクトルによって日本語の格関係を符号化している。この関係を「時空間グラフ構造」として適切なモデルを用いた翻訳ネットワークを構築し, 手話特有の空間文法に対応した日本手話翻訳の性能向上を目指す。

3. 手話日常会話コーパスの多様性拡大

手話話者ごとの表現の多様性や論理的会話の文脈の十分な学習のため, 話者多様性や会話文例を拡充するとともに, 多方向から捉えた「見る」といった空間的多様性を含む単語データも拡充する。

[参考文献]

- [1] 菊澤 (編), 吉岡 (編): “しゃべるヒト ことばの不思議を科学する”, 原大介: “影響する言葉,” pp. 266-276, 文理閣, 2023.
- [2] 松岡和美: “日本手話で学ぶ手話言語学の基礎”, くろしお出版, (2022) .
- [3] V. Bazarevsk, et al. “Blazepose: On-device Real-time Body Pose tracking”, arXiv: 2006.10204, 2020.
- [4] H. Meng, et al. “3D Interacting Hand Pose Estimation by Hand De-occlusion and Removal”, ECCV, 2022.
- [5] N. Feldman “81 Facial Landmarks Shape Predictor”. Original date: 2018-05-19.
- [6] C. Luo, et al. “Learning Multi-dimensional Edge Feature-based AU Relation Graph for Facial Action Unit Recognition”. IJCAI, 2022.

[成果の発表, 論文など]

- [1] G. Benitez-Garcia, N. Kato, Y. Shiraishi and H. Takahashi, “Automatic and Interactive Annotation of Non-Manual and Spatial Features in Pidgin Sign Japanese for SLR”, *Advanced Concepts for Intelligent Vision Systems (ACIVS) 2025*, 2025.
- [2] G. Benitez-Garcia, N. Kato, Y. Shiraishi and H. Takahashi, “Automated Annotation Assistance for Pidgin Sign Japanese in Sign Language Recognition”, *CyberWorlds 2024*, pp. 104–107, 2024.
- [3] 加藤, 白石, 他, “専門場面における手話コミュニケーション～中間型手話のパラメータの検討～”, *信学技報*, 125(279), ET2025-40, pp. 18–22, 2025.