K-026

YouTube-SL-25 における日本手話動画の再分類と段階的アノテーションの必要性
Reclassification of Japanese Sign Language Videos in YouTube-SL-25 and the Necessity of Multi-Stage Annotation

船山 滉介 [1*]　設楽 明寿 [2]　加藤 伸子 [3]　白石 優旗 [3]
Kosuke Funayama[1*]　Akihisa Shitara[2]　Nobuko Kato[3]　Yuhki Shiraishi[3]

1　はじめに

手話認識・翻訳モデルは，大規模言語モデルの進展と手話コーパスの充実により急速に向上している．YouTube-ASL [1] や How2Sign [2] など米国手話言語（ASL）を対象とした大規模コーパスに加え，SignGemma [3] のようなマルチモーダル学習の枠組みも登場しており，多様な映像データの拡充が一層重要となっている．日本手話言語（JSL）に関しては，KoSign [4] や日本手話話し言葉コーパス [5] など，高品質な資源の整備が進んでいる．ただし，これらのコーパスは主に経験豊富な手話使用者を対象としており，収集・アノテーションには多大なコストがかかる．

こうした中，多様な手話言語を収集した YouTube-SL-25 [6] は，JSL ラベル付き動画を 1,075 本（62 時間）含む新たな資源として注目される．しかし本データは，キーワード検索と 1 人のアノテーターによる分類に基づくため，JSL・SJ・中間型手話 [7]・手話歌などが混在しており，言語ラベルの信頼性には一定の懸念がある．

本研究では YouTube-SL-25 に “JSL” ラベルで収録された動画を対象に，手話に関する専門知識を用いず，視認可能な情報に基づき、出現形式の多様性を段階的に整理・記述することを目指す．

(i) 手話歌の有無，(ii) 同時出演者数，(iii) 話者属性，(iv) 日本語音声併用の有無の 4 分類軸を設計し，順序立てたアノテーションを実施した．このような段階的手法は，JSL ラベル下に混在する多様な出現形式を，観察負荷を抑えながら構造的に把握する上で有効である．特に各段階で焦点を絞って観察することで，分類の透明性と再現性が高まり，JSL ラベルの運用に含まれるばらつきを実証的に可視化できる．

こうした実践的な再分類の枠組みを提示することで，JSL ラベルの曖昧さに対し，観察に基づく実践的な分類手法の一端を示す試みである．

2　対象データとアノテーション設計・流れ

YouTube-SL-25 に “JSL” ラベルで収録された動画のうち，アクセス可能な 1,073 本（61 時間 14 分 54 秒）を分析対象とした．当初は 1,075 本がリストされていたが，2 本は非公開設定により除外した．本セクションでは，これらの動画に対して設計した分類軸と，段階的に実施したアノテーション手順を述べる．

2.1　分類軸の設計と選定理由

JSL ラベルのもとに含まれる動画の多様性を可視化するため，(i) 手話歌の有無，(ii) 同時に映る出演者の人数，(iii) 話者属性，(iv) 日本語音声の併用有無の 4 点を設計した．特に手話歌は構文や語順といった言語的分析とは異なるため，除外対象とした．出演者数や話者属性の把握は，個人性や関係性の分析にも資する．日本語音声の有無は、SJ 使用傾向を推定する補助的指標となる．これらの指標は，動画のメタ情報などから視認的に判断可能で手話に関する専門的知識がなくても記録できる．

2.2　アノテーション手順の流れ

アノテーションは以下の 4 段階で実施した：

1. **手話歌の除外**：音楽や構図，リズム表現から手話歌を判別し，分析対象から除外．
2. **同時出演者数と音声の確認**：出演者の人数（単独：同時に登場する話者が 1 人の場合／複数）と日本語音声の有無を記録．
3. **話者属性の分類**：全動画を対象にメタ情報を参照しつつ属性を分類．(判断困難な場合は「？？？」と明示)
4. **単独出演動画の抽出と集計**：単独出演動画に限定して話者属性を集計・可視化．

このように段階的に情報を整理することで，JSL ラベル下に含まれる動画の実態を把握し，その多様な内訳を可視化することが可能となった．

3　アノテーション結果の可視化と傾向

3.1　手話歌の分布と出現傾向

JSL とラベル付けされた動画本数のうち 15.2 %が手話歌であった（図 1）．この事実は，ラベルと実態の間に無視できない乖離があったことを示している．手話歌は音楽に合わせた演出的表現が特徴であり，日常的な JSL とは大きく異なる．投稿もごく少数のチャンネルに偏在しており，内容・構成の両面で特殊な位置づけにある．

本研究ではこうした特性を踏まえ，アノテーションの初期段階で手話歌を明示的に検出・除外した．

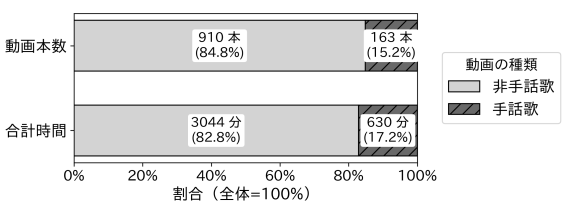

図 1: 手話歌／非手話歌の動画本数・合計時間および割合
（注：合計時間は分単位で統一するため，秒未満を切り捨てて表示．）

1）　筑波技術大学大学院 技術科学研究科.
Graduate School of Technology and Science, Tsukuba University of Technology.
2）　筑波大学大学院 図書館情報メディア研究科.
Graduate School of Library, Information and Media Studies, University of Tsukuba.
3）　筑波技術大学 産業技術学部 産業情報学科.
Department of Industrial Information, Faculty of Industrial Technology, Tsukuba University of Technology.
＊　a253104@a.tsukuba-tech.ac.jp

Copyright © 2025 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

3.2　同時出演者数の分布と構成傾向

出演者数の観点から見ると，非手話歌の動画の大多数は単独出演で構成されており，複数人が同時に登場する事例は極めて限定的であった（図 2）．

特に三人以上による出演は稀であり，集合的な対話や演出的構成は例外的であることが分かる．この傾向は，個人が主導する制作・発信が基本であること，あるいは収録や編集の実用的制約を反映している可能性がある．

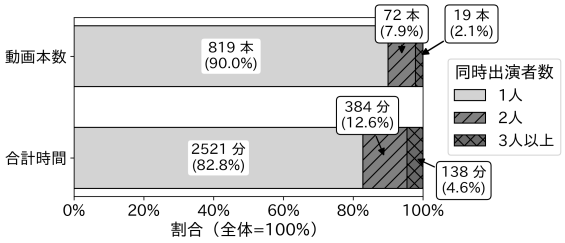

図 2: 同時出演者数ごとの動画本数・合計時間および割合
（注：分析対象は非手話歌の 910 本．合計時間は分単位で統一するため，秒未満を切り捨てて表示．）

3.3　話者属性の分布と構成傾向

手話歌を除いた全 910 本の動画について，出演者の属性を観察的に分類した結果が図 3 に示される．ろう者による発信が多数を占めている一方で，CODA や聴者，属性不明の動画も一定数含まれており，話者層には一定の多様性が見られる．

特に，単独出演が大多数を占めていたこと（図 2）を踏まえ，話者属性の構成をより明確に捉えるために，単独出演の 819 本に限定して再整理したものが図 4 である．ここでもろう者が主である傾向は変わらず，CODA，聴者の参加も確認された．なお，難聴者は確認されず，属性不明に含まれる可能性がある．

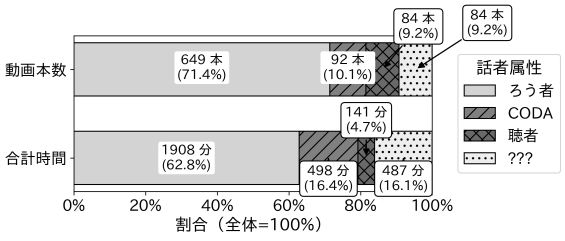

図 3: 話者属性ごとの動画本数・合計時間および割合
（注：分析対象は非手話歌の 910 本．合計時間は分単位で統一するため，秒未満を切り捨てて表示．）

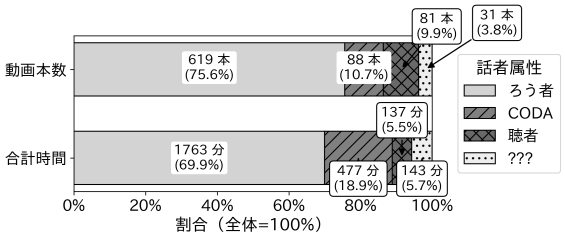

図 4: 単独出演動画における話者属性別の動画本数・合計時間および割合
（注：分析対象は単独出演の 819 本．合計時間は分単位で統一するため，秒未満を切り捨てて表示．）

3.4　JSL ラベル下に見られる出現形式のばらつき

日本語音声の併用が確認された動画も存在したが，その内容には通訳者による発話，冒頭のみのあいさつ，自然に漏れ出た声など使用形式のばらつきが見られた．

また，語順が日本語に近く，非手指文法（頷きや眉の動きなど）の使用が限定的な動画も一定数確認された．これらは中間型手話に該当する可能性があり，JSL ラベル下に文法的に多様な表現が混在していることが示唆される．

4　まとめと今後の課題

本研究では，YouTube-SL-25 に含まれる “JSL” ラベル付き全 1,073 本の動画を対象に，手話歌の有無，同時出演者数，話者属性，日本語音声の併用有無という 4 つの分類軸に基づき手話に関する専門知識を用いず，視認可能な情報による段階的アノテーションを実施した．

その結果，手話歌は JSL ラベル付き動画全 1,073 本中 163 本（15.2%）を占めており，出現形式の大きく異なる動画がラベル下に多数含まれていることが明らかとなった．出演者は単独話者が多く，話者属性にはろう者を中心に，CODA や聴者も一定数含まれていた．

また，語順が日本語に近く，非手指文法の使用が限定的な動画も複数見られ，中間型手話に該当する可能性のある表現が含まれていた．こうした混在は，JSL ラベルを前提とした分析や学習に影響を及ぼす可能性がある．視認可能な分類軸に基づく段階的アノテーションは，JSL ラベル下の出現形式を記述的に整理し，専門的なアノテーションに先立つ実践的な再分類の枠組みとなる．

今後は，より精緻な文法的基準の導入，中間型手話の体系的整理，話者の自己認識に基づく質的分析などを通じて，日本語音声の使用形式を含む分類軸のさらなる精緻化が求められる．

参考文献

[1] Dave Uthus, Garrett Tanzer, Manfred Georg, “YouTube-ASL: A Large-Scale, Open-Domain American Sign Language-English Parallel Corpus,” *Advances in Neural Information Processing Systems*, vol. 36, pp. 29029–29047, 2023.

[2] Amanda Duarte, et al., “How2Sign: A Large-Scale Multimodal Dataset for Continuous American Sign Language,” *Proc. CVPR*, pp. 2735–2744, 2021.

[3] Google DeepMind, “SignGemma: A Multimodal Foundation Model for Sign Languages,” *Google I/O 2025 Blog*, https://blog.google/technology/developers/google-ai-developer-updates-io-2025/, accessed 2025-06-11.

[4] 長嶋祐二 他, “多様な研究分野に利用可能な超高精細・高精度手話言語データベースの開発,” 言語資源活用ワークショップ発表論文集, 国立国語研究所, pp. 148–155, 2018.

[5] Mayumi Bono, et al., “A Colloquial Corpus of Japanese Sign Language: Linguistic Resources for Observing Sign Language Conversations,” *Proc. LREC*, pp. 1898–1904, 2014.

[6] Garrett Tanzer, Biao Zhang, “YouTube-SL-25: A Large-Scale, Open-Domain Multilingual Sign Language Parallel Corpus,” *arXiv preprint arXiv:2407.11144*, 2024.

[7] 原大介. 日本手話と日本語対応手話の混合言語（中間型手話）の言語的特徴の解明. 科学研究費助成事業 研究成果報告書, 2014, 課題番号 23520530.

Copyright © 2025 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.