

2021 年度 委託研究

課題 225

自動翻訳の精度向上のための「マルチモーダル情報の
外部制御可能なモデリング」の研究開発

研究計画書



1. 研究開発課題

『自動翻訳の精度向上のための「マルチモーダル情報の外部制御可能なモデリング」の研究開発』

2. 本研究開発の目的及びアウトカム目標

講演・プレゼン・打合せ・会議等のビジネス・コミュニケーションのリモート化がコロナ後のニューノーマルとして定着し、コミュニケーションのDX・低コスト化・高頻度化・国際化が一気に進むことに疑念を挟む余地はない。

国立研究開発法人情報通信研究機構（以下、「機構」という。）は「グローバルコミュニケーション計画 2025」の下、低遅延の同時通訳技術を研究開発し、ビジネス・コミュニケーションを効率化することを目指している。コミュニケーションのDXや同時通訳の自動化によって生産性は向上し国際化は加速する。このことによって、感染症リスク・少子高齢化による労働力不足・少資源等の日本が抱える課題に妨げられることなく、経済力の向上を可能にする。また、同計画のロードマップにあるように、2030年により高いレベルである「シビアな交渉にも使える同時通訳技術」を目指している。

人間の通訳は、第一に、テキストを越えた情報、例えば、映像情報に内在する話者や状況などを利用することにより一文の理解を深化させ翻訳精度を向上させている。第二に、人間の通訳は、例えば、情報の過不足なく直訳で／ポイント中心に意識で（あるいは、一般向け用語で／専門用語を活用して、等）のように外部からの制御に従って、通訳の仕方を変更している。上記の同時通訳技術のロードマップを実現していくためには、このような制御が可能な計算機上のモデリングが望まれているところである。つまり、通訳者が行っているように、人間の知覚機能をフル活用し、かつ、状況に応じて訳出内容を柔軟に制御して通訳する手法を創出する必要がある。この課題に対し、本委託研究では、例えば、①映像情報や②直訳／意識の使い分けが必要な状況のように、通訳対象のテキストの情報に留まらず、通訳者が活用／考慮している情報／状況（以下、「一文を越えた情報」という。）を利用する手法の確立を目指す。そのため、

研究開発項目 1 マルチモーダル情報理解技術の研究開発

研究開発項目 2 制御可能なテキスト生成技術の研究開発

の2つの研究開発を並行して進め、自動翻訳の精度向上を目指す。ただし、本委託研究課題では、同時通訳の漸次性自体はスコープ外とし、他のモジュールと連携することによって同時通訳を実現するものとする。

機構は、本委託研究開発期間の途中から、自ら実施する研究開発の成果と本委託研究の成果を組み合わせ、活用、拡張することにより同時通訳を高精度化し、社会実装の加速を目指す。そのためのマイルストーンとして、

- 2024年 2023年度までの本委託研究のアウトプットを活用し、限定した場面においてマルチモーダル同時通訳を実現する。
- 2026年 本委託研究の最終成果を活用し、制御可能な同時通訳を実現する。

と設定している。

アウトカム目標として、本委託研究の成果として期待されるマルチモーダル情報と制御を導入したモデリング技術を自ら実施する研究開発の成果である同時通訳技術に統合することにより、

2024年頃から2030年にかけて、コミュニケーションにおける相互理解を強力に支援する製品・サービスが創出される、とする。より詳細には、研究開発項目ごとに以下をアウトカム目標とする。

- 研究開発項目 1 本マルチモーダル情報理解技術の研究開発の成果を用いて、自ら実施する研究開発の成果である同時通訳基本システムが扱うテキストと音声に加えて映像を統合することによって、意味内容や話題に関連した静止画（写真、表、グラフ等）や動画を付加するなど出力の多様化が可能になる。マルチモーダル情報を駆使することによってコミュニケーション効率を向上させる、イノベティブな製品が生み出される。
- 研究開発項目 2 本制御可能なテキスト生成技術の研究開発の成果を用いて、例えば、要約する際に（すなわち、要点を絞り込む際に）、新聞とその見出し、会議議事録とそのEXECUTIVE SUMMARY、あるいは、同時通訳したときに通じやすくなるような絞り込み、さらには、研究開発項目 1 の成果と合わせて要点に関連した映像などを含むリッチな情報について、相手やニーズに合わせた出力制御が可能となり、応用範囲の広い競争力のある商品として上市される。

3. 内容

同時通訳技術の要素技術である自動翻訳において一文を越えた情報を利用し精度を改善する二つのテーマを並行して研究する。ただし、研究開発項目 2 の制御は研究開発項目 1 にも適用可能とすること。

研究開発項目 1 マルチモーダル情報理解技術の研究開発

一文を越えた情報の典型例の一つである動画を参照する自動翻訳の精度の改善手法を創出する。具体的には、以下の（1）マルチモーダルコーパスの構築方法、及び（2）マルチモーダル情報を理解する手法に関する研究開発を行う。

（1）マルチモーダルコーパスの構築方法の研究開発

マルチモーダルコーパスの構築方法の確立を目指し、動画及び対訳テキストからなる動画・対訳コーパスを構築する。先行事例はほとんどなく、仕様決め、試作、仕様の改訂、再試作、と試行錯誤して、有用な構築方法を確立する必要がある。コーパスの元となる情報源としては、例えば、遠隔の観光ガイドやプレゼンテーションを収録した動画が考えられる。しかしながら、コーパス構築方法自体が未開拓であるため、これらに拘ることなく翻訳精度向上に有効な特性を明らかにすることを目指して多様な案を検討の上、下記（2）のアルゴリズムによる有意な精度向上を検証するために必要な規模のコーパスを構築し実証的に研究を進める。検証確認済みのコーパスを動画・対訳コーパスの成果物とする。

（2）マルチモーダル情報を理解する手法に関する研究開発

自動翻訳の精度向上に貢献するべく、マルチモーダル情報を理解するアルゴリズムを研究し、ソフトウェアとして実装し、上記（1）の動画・対訳コーパスを用いて検証する。

研究開発項目 2 制御可能なテキスト生成技術の研究開発

テキスト生成（例えば要約）の出力の変更を制御する情報という一文を越えた情報の活用手法を確立する。具体的には、以下の（１）テキスト生成の制御に関するコーパス構築方法の研究開発、及び（２）テキスト生成において内容や様式等を制御する手法に関する研究開発を行う。

（１） テキスト生成の制御に関するコーパス構築方法の研究開発

先行事例はほとんどなく試行錯誤して有用な構築方法を確立する。要約であれば、記事と見出し・リード文等の対が考えられるが、これらに拘らず、多様な案を検討の上、下記（２）で研究開発するアルゴリズムによる有意な精度向上を検証するために必要な規模のコーパスを構築し実証的に研究を進める。検証確認済みのテキスト生成コーパスを成果物とする。また、ここでは、日本語または英語を対象とする。

（２） テキスト生成において内容や様式等を制御する手法に関する研究開発

外部から与えられた文長という情報に従って要約することは、従来は出来なかった。これに対し、本委託研究では、（文長等の）外部情報に応じて要約を含むテキスト生成の出力を制御する基本手法を確立する。また、これらをパイプラインやマルチタスク的に自動翻訳システムと連携させることを前提とする。ソフトウェアとして実装し、上記（１）のテキスト生成コーパスを用いてテキスト生成精度向上につながることを検証する。

4. アウトプット目標

共通

- 同時通訳での活用を視野に、話者の発話長と同じ時間内で下記研究開発項目 1（2）と研究開発項目 2（2）の処理が完了するアルゴリズムであること。
- 1件1分間以上からなる室内と室外での発話、例えば、学会やビジネスにおけるプレゼンテーション、遠隔の観光ガイドを対象とすること。
- プレゼンテーションについては発表スライドも考慮できること。

研究開発項目 1 マルチモーダル情報理解技術の研究開発

（１） マルチモーダルコーパスの構築方法の研究開発

- 2種類以上の動画・対訳コーパスを試作すること。
- 動画・対訳コーパスは、合計 500 分以上の動画とその動画を説明するテキストの対訳対からなること。
- 動画・対訳コーパスの一部は公開し国際技術コンペを実施すること。
- 動画・対訳コーパスの一部は 2023 年度末までに 1 種類公開し、さらに 2025 年度末までに 1 種類追加すること。

（２） マルチモーダル情報を理解する手法に関する研究開発

- 動画を利用しないベースラインシステムと比べて、翻訳精度が有意水準 5% で優れていることを示すこと。
- 上記（１）の動画・対訳コーパスで精度向上が確認されたソフトウェアを公開すること。
- 論文がトップコンファレンスで採択されること。

- ・ ソフトウェアの公開は最低でも動画・対訳コーパスの公開に同期させること。

研究開発項目2 制御可能なテキスト生成技術の研究開発

(1) テキスト生成の制御に関するコーパス構築方法の研究開発

- ・ 2種類以上のテキスト生成コーパスを試作すること。
- ・ テキスト生成コーパスは、少なくとも10万文以上の本文とその要約の対からなること。
- ・ テキスト生成コーパスの一部は公開し国際技術コンペを実施すること。
- ・ テキスト生成コーパスの一部は2023年度末までに1種類公開し、さらに2025年度末までに1種類追加すること。

(2) テキスト生成において内容や様式等を制御する手法に関する研究開発

- ・ テキスト生成及び制御情報に関するテキスト生成精度を設定し、適切なベースラインシステムと比べて、テキスト生成精度が有意水準5%で優れていることを示すこと。
- ・ 上記(1)のテキスト生成コーパスで精度向上が確認されたソフトウェアを公開すること。
- ・ 論文がトップコンファレンスで採択されること。
- ・ ソフトウェアの公開は最低でもテキスト生成コーパスの公開に同期させること。

5. 採択件数、研究開発期間及び研究開発予算等

採択件数 : 1件

研究開発期間: 2021年度(契約締結日)から2025年度

研究開発予算: 各年度、総額100百万円(税込)を上限とする。

ただし、初年度は総額30百万円(税込)を上限とする。

(提案の予算額の調整を行った上で採択する提案を決定する場合がある。)

研究開発体制: 単独提案も可能であるが、産学官連携等による複数の実施主体からなる体制とすることを推奨する。

6. 提案に当たっての留意点

- 具体的目標に関しては、定量的に提案書に記載すること。
- 研究開発成果の情報発信を積極的に行うこと。
- 本委託研究の遂行過程で得られる研究データがあれば、広くオープンにするのが望ましい。公開できると想定する研究データの有無と、有る場合には公開計画(例: 公開するデータの種類、公開先、公開方法)を提案書に記載すること。
- 実施体制については、本研究開発の目的に則した実施体制を構築することとし、それぞれの役割を明記すること。

7. 運営管理

- 機構と受託者の連携を図るため、代表提案者は、プロジェクトオフィサーと定期的に連絡調整会議を開催すること。

- 複数の機関が共同で受託する場合には、代表提案者が受託者間の連携等の運営管理を行い、受託者間調整会議を定期的を開催すること。
- 社会情勢や研究環境の変化等、必要に応じて、プロジェクトオフィサーが研究計画書を変更する場合があるので、留意すること。

8. 評価

- 機構は、2023 年度に中間評価を実施する。本評価結果により、当該年度で本委託研究を終了する場合がある。
- 機構は、2025 年度に終了評価を実施する。また、機構は、本委託研究終了後に成果展開等状況調査を行い、追跡評価を行う場合がある。
- 機構は、上記以外にも本委託研究の進捗状況等を踏まえて、臨時にヒアリングを実施することがある。

9. 成果の展開・普及に向けた取組

「2. 本研究開発の目的及びアウトカム目標」にあるように、本委託研究の成果は、「グローバルコミュニケーション計画 2025」の下、研究開発期間の途中から、自ら実施する研究開発の成果と組み合わせ、機構が主体的に社会実装を推進する計画である。提案に当たっては、提案者による成果の展開・普及の計画を、1. 機構の計画への寄与、2. 提案者独自の研究開発成果の展開・普及に向けた取組の双方について、提案書に記載すること。「4. アウトプット目標」に記載されている内容以外にも、本委託研究で得られた成果のオープン化（例えば、成果発表やそれに留まらずコミュニティ先導のための国際ワークショップや国内特別セッション主催、展示、標準化等）を行う等、成果の社会実装等に向けて必要な取組を行うこと。

10. プロジェクトオフィサー

所属：先進的音声翻訳研究開発推進センター

氏名：隅田 英一郎

参考

- 「グローバルコミュニケーション計画 2025」（2020 年 3 月総務省）
https://www.soumu.go.jp/main_content/000678485.pdf