

令和 5 年度研究開発成果概要書

採択番号 22501  
研究開発課題名 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  
副 題 マルチモーダル情報理解と制御可能なテキスト生成の研究開発

(1) 研究開発の目的

人間同士および人間とコンピュータ間の円滑なコミュニケーションの実現に向けて、コンピュータが文脈・状況を考慮しながら、言語、音声、画像、動画、表などによるマルチモーダルな情報を統合的に理解する手法を探求する。また、長さや焦点、スタイル、難易度などを指定したり、外部知識を与えることによって、コンピュータが人間のニーズに合わせてテキストを生成する技術を確立する。さらに、これらの成果を統合することで、自然言語生成の新たな応用を開拓するとともに、自動翻訳や同時通訳の研究開発の高度化につなげる。

全ての研究・開発において、深層学習を共通基盤として採用し、タスクやモーダルの垣根を超えた人工知能の実現を目指す。また、自然言語処理、画像処理、深層学習、報道などの分野で最先端の取り組みを進めているグループでチームを結成し、人材やデータ、技術の交流を促進する。研究と並行して言語資源の開発に注力し、その成果物を研究コミュニティに還元する。これにより、日本語の自然言語処理に関する研究で世界トップレベルを維持するとともに、マルチモーダル情報理解や制御可能な自然言語生成などの研究分野で、世界に先駆けた研究を展開する。

(2) 研究開発期間

令和 3 年度から令和 7 年度 (5 年間)

(3) 受託者

国立大学法人東京工業大学<代表研究者>  
国立大学法人東京大学  
国立大学法人愛媛大学  
国立大学法人一橋大学  
日本放送協会  
株式会社時事通信社

(4) 研究開発予算 (契約額)

令和 3 年度から令和 5 年度までの総額 230 百万円 (令和 5 年度 100 百万円)  
※百万円未満切り上げ

(5) 研究開発項目と担当

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

- 1-1 マルチモーダル動画対訳コーパスに関する研究開発 (国立大学法人東京大学)
- 1-2 マルチモーダル機械翻訳に関する研究開発 (国立大学法人愛媛大学)
- 1-3 マルチモーダル情報理解に関する研究開発 (国立大学法人東京大学)

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

- 2-1 自動要約の制御に関する研究開発 (国立大学法人東京工業大学)
- 2-2 翻訳の制御に関する研究開発 (日本放送協会)
- 2-3 スタイルの制御に関する研究開発 (国立大学法人一橋大学)
- 2-4 データ整備に関する研究開発 (株式会社時事通信社)

## (6) 特許出願、外部発表等

		累計 (件)	当該年度 (件)
特許出願	国内出願	0	0
	外国出願	0	0
外部発表等	研究論文	5	4
	その他研究発表	83	38
	標準化提案・採択	0	0
	プレスリリース・報道	1	1
	展示会	0	0
	受賞・表彰	11	5

## (7) 具体的な実施内容と成果

### 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

#### 1-1 マルチモーダル動画対訳コーパスに関する研究開発

前年度に引き続き、マルチモーダル動画対訳コーパスの拡張作業を行った。本年度は主にプレゼンテーション動画対訳コーパスの作成に注力し、YouCook2 データセットをベースとして新たに約17時間・三千文対のデータ作成（翻訳作業）を行った。これと前年度作成した部分を合わせ、バージョン1としての公開準備を完了した。また、コーパスを活用するマルチモーダル機械翻訳手法の検討も進め、構文木情報の画像情報へのグラウンディングや、画像をピボットとする少リソース言語の翻訳手法などを開発し、基礎的な評価を行った。

#### 1-2 マルチモーダル機械翻訳に関する研究開発

本年度は、(i) 音声・画像付きスピーチ対訳を用いたマルチモーダル機械翻訳、(ii) 画像生成を用いたマルチモーダル機械翻訳、(iii) 漫画を対象としたマルチモーダル機械翻訳を行った。(i)について、TED Talks の原ビデオ(英語)から、音声認識文(日本語)を作成し、WSLT2017 英日対訳コーパスとアライメントをとることで、約8万文対に対する音声・音声認識文・英語文・日本語文からなる対訳コーパスを開発した。開発したマルチモーダル対訳コーパスを用いて、音声認識誤り訂正を用いた機械翻訳手法、および動画からより適切な画像を抽出するマルチモーダル機械翻訳手法の開発を行い、それぞれ機械翻訳の性能が改善されることを確認した。(ii)について、潜在拡散モデルを用いて、入力画像を入力文により適した画像に変換する手法を開発し、マルチモーダル機械翻訳の性能を改善することができた。この成果はACL SRW 2023に投稿し採択された。(iii)について、漫画を対象としたマルチモーダル機械翻訳の研究を行い、コマを超えた吹き出し間の文脈情報を扱う機械翻訳手法、および著者情報、出版社情報、ジャンル情報などの書誌情報を用いた機械翻訳手法を開発し、この研究成果論文はLREC-COLING 2024に採択された。

#### 1-3 マルチモーダル情報理解に関する研究開発

本年度は、シーングラフの抽出精度向上や、シーングラフを介したマルチモーダル情報の統合に関する具体的な手法を複数開発した。まず、画像からのシーングラフ抽出において、教師データの頻度に起因するバイアスを低減するデータ拡張手法を開発し、電子情報通信学会誌へ論文投稿を行った。また、画像から得られるシーングラフを外部知識グラフにより拡張すると同時に、トランスフォーマを用いて言語情報と統合するメカニズムを開発した。これを応用し、与えられた画像の未来に起こるイベントを予測し説明する予知的画像キャプションングを実現し、コンピュータビジョン分野のトップ国際会議であるCVPR 2023で発表した。さらにこれを拡張し、大規模言語モデルに外部知識として画像辞書を接続したオープンワールド画像キャプションング手法を開発し、CVPR 2024へ採択された。

### 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

#### 2-1 自動要約の制御に関する研究開発

本年度は、クエリ推薦付き要約のためのデータセットとして、ビジネスシーン対話コーパスに対して

クエリと要約文を付与したコーパス（2 万文規模）を構築した。さらに、今年度に入ってから大規模言語モデルを活用した研究が増えているため、その生成の制御や高度化に関する研究を進めた。まず、GPT-3 を文法誤り訂正タスクに応用する際に、最小限の編集、流暢な編集、学習者レベルといった出力文の制御に与える側面を調査し、その研究成果を言語学習応用に関するワークショップである 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA) において発表した。敵対的事例を用いた文脈内学習 (in-context learning) により、大規模言語モデルが生成したテキストを検出するタスクを制御できることを示し、38th AAAI Conference on Artificial Intelligence (AAAI 2024) で発表するとともに、本成果は日本経済新聞に掲載された。また、大規模言語モデルにテキストを評価させるときに生じるバイアスを制御する手法や、敵対的発言を取り入れて言語モデルの推論力を制御する研究を進め、これらの成果を言語処理学会第 30 回年次大会で発表したところ、若手奨励賞をそれぞれ受賞した。

## 2-2 翻訳の制御に関する研究開発

ニュースには学習データに含まれない新しい語や表現が日々出現する。これらの翻訳に対応するために、随時人手で学習データを作成して追加することは難しい。そこで、本年度は、日々制作される日本語ニュースと英語ニュースから新語の翻訳知識を学習する手法を開発した。英語ニュースは日本語ニュースの直訳ではないため、日英ニュースから抽出した対訳文対は内容に差異があることが多い。内容に差異がある対訳文対をそのまま機械翻訳で学習すると、新語の翻訳知識とともに、入力とは異なる内容を出力するような翻訳パターンも学習してしまい翻訳品質が低下する。そこで、対訳データを学習する際に、出力側の新語とその直後の文脈のみから誤差逆伝播させる手法を考案した。評価実験により、既存手法と比較して、翻訳品質が低下することなく新語を正しく翻訳することを確認した。

## 2-3 スタイルの制御に関する研究開発

本年度は、テキストの難易度を制御する手法の開発のための、日本語の文書単位のテキスト平易化に関する新聞記事および Wikipedia 記事のデータセットの構築と整備を継続し、国際会議に論文を投稿し、LREC-COLING 2024 に採択された。そして、構築したデータセットを用いた国際コンペティションの開催に関する準備と調整を行なった。また、Wikipedia 記事の文単位の品質推定に関する研究を継続して発展させ、国際会議に論文を投稿し、AAAI 2024 に採択され、ポスター発表を行なった。

## 2-4 データ整備に関する研究開発

時事通信社の編集、配信システムから、コーパス構築など本研究に資するデータ（日本語と英語の記事セット、写真と書誌情報のセット、こどもニュース）の抽出などを行い、各機関に提供した。本年度は、研究者にとってより利用しやすいデータとすることを念頭に、過去に抽出したデータについて見直しを行い、フォーマット変更に取り組んだ。具体的には、HTML タグの構造の一元化などを行い、特に大容量、長期間のデータを一度に投入するような研究における利便性、汎用性の向上を図った。

時事通信社では、本年度から英文記事の作成に NICT の機械翻訳エンジンを導入している。現場の意見、感想を探ることは、ひいては翻訳精度の向上につながるとの認識から、実際にエンジンを利用している記者、翻訳者、編集者からの意見聴取、機械翻訳の利用率の調査などを行った。記事データから著作物であるコーパスを構築する場合には、特に個人情報扱いに難しさがあり、公開されている匿名化ツールなどにも限界がある。そこで、通常業務の過程で記事に付与しているフラグを活用し、匿名化作業などを効率化できないか可能性を探った。

## (8) 今後の研究開発計画

### 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

#### 1-1 マルチモーダル動画対訳コーパスに関する研究開発

まず、可能な限り早い段階でマルチモーダル動画対訳コーパスを完成させる。特に、プレゼンテーション動画対訳コーパスについては既に当初の目標量を達成しているが、より大規模かつ使いやすいものとするために YouCook2 データセット全体の翻訳完了を目指す。また、作成完了したコーパスを用いた実験・論文投稿を行うと同時に、最終年度のコンペティション実施へ向けた準備を進める。

## 1-2 マルチモーダル機械翻訳に関する研究開発

引き続き講演動画から抽出することでマルチモーダル対訳コーパスの開発を行い、それらを用いたマルチモーダル機械翻訳の研究開発を進める。既存の音声付き対訳コーパスや本プロジェクトで開発した音声・映像付きマルチモーダル対訳コーパスを用いて、音声・画像・映像を用いたマルチモーダル機械翻訳の研究・開発を進める。

## 1-3 マルチモーダル情報理解に関する研究開発

引き続き、シーングラフや知識グラフを活用したマルチモーダル情報の統合手法や、応用アプリケーションについて研究を進める。特に、近年の発展が著しいマルチモーダル大規模言語モデルへの接続に注力し、提案手法の貢献について明らかにしたい。また、開発した要素技術のマルチモーダル機械翻訳への適用についても検討を進める。

## 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

### 2-1 自動要約の制御に関する研究開発

引き続き、自然言語生成に関するコーパスを構築し、自然言語生成の制御に関する研究を進める。また、これまでに開発した自然言語生成技術をマルチモーダル情報理解の研究と統合する研究、例えば、会議のプレゼンテーションから字幕や要約を生成するタスクに取り組む。

### 2-2 翻訳の制御に関する研究開発

引き続き翻訳出力制御に関する研究開発を進める。文書のつながりや文書スタイルを考慮して機械翻訳結果を制御する手法の研究を行い、一貫した書式での翻訳性能の改善を行う。外部公開可能な翻訳制御のテストデータを新たに作成し、国際技術コンペを企画する。

### 2-3 スタイルの制御に関する研究開発

引き続き日本語の文書単位のテキスト平易化データセットを用いた国際コンペティションの開催に向けた調整を行い、2025年度中を目処にリーダーボードを作成する。また、文書単位のテキスト平易化の手法の研究を行いつつ、テキスト平易化のデータセットのマルチモーダル拡張についての検討を行う。最後に、文単位の品質推定のデータセットを用いたテキストの自動評価の研究を進める。

### 2-4 データ整備に関する研究開発

引き続き、時事通信社の編集、配信システムから、コーパス構築など本研究に資するデータ（日本語と英語の記事セット、写真と書誌情報のセット、こどもニュース）の抽出などを行い、各機関に提供する。社内に導入した機械翻訳の利用状況をモニターするとともに、利用者からの意見を吸い上げる。個人情報については、自社データを活用した独自の仕組みを構築できないか、検討と試行を続ける。