

令和2年度研究開発成果概要書

採 択 番 号 : 21101  
 研究開発課題名 : 多言語音声翻訳高度化のための統合的深層学習の研究開発  
 副 題 : 統合型機械翻訳技術の研究開発

(1) 研究開発の目的

本研究開発課題では、日本語ニュース記事の英語配信を実現する統合型翻訳技術の研究開発を行い、統合型翻訳システムを開発する。研究課題 197 と密接に連携しながら、研究課題 197 の研究課題の個別成果を組みあわせ、単独技術では達成できない高い精度を目指す。統合した技術を用いて「英語ニュース記事配信」プラットフォームを構築し、実証する。

(2) 研究開発期間

令和元年度から令和2年度(2年間)

(3) 実施機関

凸版印刷株式会社 <代表研究者>  
 マインドワード株式会社

(4) 研究開発予算(契約額)

総額 120 百万円(令和2年度 60 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: 統合型機械翻訳システムの研究開発

- 1. 統合型機械翻訳エンジンに関する研究開発 (マインドワード)
- 2. 統合型機械翻訳システムの構築に関する研究開発 (マインドワード)

研究開発項目2: 統合型機械翻訳プラットフォームの実証利用と評価に関する研究開発

- 1. 統合型機械翻訳プラットフォームの実証環境の構築と運用に関する研究開発 (凸版印刷)
- 2. 統合型機械翻訳プラットフォームの統合評価に関する研究開発 (凸版印刷)

(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	1	1
	外国出願	0	0
外部発表等	研究論文	0	0
	その他研究発表	0	0
	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	0	0

## (7) 具体的な実施内容と成果

### 研究開発項目1：統合型機械翻訳システムの研究開発

#### 1-1 統合型機械翻訳エンジンに関する研究開発

##### ①目標

- 当該項目では、ニュース記事の要約文生成の検討、実装、評価を実施する。令和2年度は、日本語記事を機械翻訳したあとに英語要約をする技術について評価を実施する。結果等は、研究課題197へフィードバックする。

##### ②実施内容

- 英語記事を機械翻訳したあとに日本語要約をする要約翻訳技術について評価を実施した。評価データは研究課題197のデータを利用した。

##### ③成果

- 英語ニュース記事本文から日本語要約文を生成する手法について、英語ニュース記事本文を要約してから日本語に翻訳する方式と、英語ニュース記事本文を日本語に翻訳してから要約する方式についてBLEUスコアによって評価を行ったところ、英語ニュース記事本文を機械翻訳によって日本語に翻訳した後に要約を生成する方式が最も優れていることがわかった。研究開発項目2の実証利用においても本方式を採用した。

#### 1-2 統合型機械翻訳システムの構築に関する研究開発

##### ①目標

- 当該項目では、統合システムの設計と、統合型機械翻訳システムの開発と実証利用を行う。令和2年度は、統合型機械翻訳システムの実装を実施する。具体的には、システム構成の設計に基づいて実装作業を実施する。更にAPIを実装し、統合型機械翻訳プラットフォームと接続する。あわせて、APIの利活用実験を実施する。

##### ②実施内容

- 統合型機械翻訳システムの実装を実施した。具体的には、システム構成の設計に基づいて実装作業を実施した。実装作業にあたっては研究課題197と連携し、研究課題197の進捗に応じて、ニュース機械翻訳、シーン認識及びキャプション生成機能を統合し、マルチモーダルに対応した統合型機械翻訳システムを構築した。

##### ③成果

- 各機能はNICTが開発する「みんなの自動翻訳」APIと互換性を有するAPIとして使用が可能となるように実装した。APIを利用して統合型機械翻訳プラットフォームと接続し、研究課題2-1の利活用実験を実施した。

### 研究開発項目2：統合型機械翻訳プラットフォームの実証利用と評価に関する研究開発

#### 2-1 統合型機械翻訳プラットフォームの実証環境の構築と運用に関する研究開発

##### ①目標

- 当該項目では、実証環境を構築して、期間を限り一般公開する。令和2年度は、統合型機械翻訳プラットフォームの設計に基づいてプラットフォームの実装を実施する。また、利用者と期間を定めて、実証システムを構築し運用する。システムの運用を通じて、利用者の記事選考性、翻訳に対する許容度測定、ワークフロー効率データを収集する。記事の閲覧規模は延べ10,000記事以上とする。

##### ②実施内容

- 統合型機械翻訳プラットフォームの設計に基づいてプラットフォームの実装を実施した。また、利用者と期間を定めて、実証システムを構築し運用した。システムの運用を通じて、利用者の記事選考性、翻訳に対する許容度測定、ワークフロー効率データを収集した。なお、記事

の閲覧規模は延べ 10,000 記事以上とした。

- ・ 統合型機械翻訳プラットフォームを活用し、実証実験を実施した。実証実験にあたっては、新型コロナウイルス感染拡大に伴う、訪日外国人での実証が困難になったことから英語のニュース記事を日本人向けに翻訳して配信する、英日翻訳の実証実験も追加で実施した。
- ・ 実証実験の閲覧記事数は、延べ 10,000 記事を目標とし、実閲覧記事数は記事別アンケートの回答記事数と同じと考えられ、40,739 記事であった。当初の目標を超えるサンプル数を確保した。被験者の選定にあたっては、ニュース記事のコンテンツの性質上、本研究開発が社会実装された場合にビジネスターゲットとなりうる層を対象とした。

### ③成果

- ・ 2020 年度に構築した実証用の統合型機械翻訳プラットフォームでは、ニュース記事コンテンツの提供元である時事通信社の記事サーバよりニュース記事を収集し、課題 197 の自動要約技術と連携しながら機械翻訳のシステムで翻訳～配信まで実施した。
- ・ 機械翻訳の満足度の比較として、さらに人手翻訳のニーズを把握することを目的として機械翻訳で翻訳した記事を人手で翻訳補正する、ポストエディットについてもワークフローに組み込んだ。
- ・ 実証実験で得られた成果は、i) ニュース分野に機械翻訳が適用できること、ii) 速報性への価値が高く、誤訳や日本語の不自然さはおおむね許容できること、iii) 自動要約技術のレベルが高いこと、iv) 有償サービスの可能性等である。

## 2-2 統合型機械翻訳プラットフォームの統合評価に関する研究開発

### ①目標

- ・ 当該項目では、研究課題 197 の翻訳結果の評価によると、誤訳とスタイル違反が多い。誤訳の評価は、機械翻訳機の性能に大きく依存するため、令和 2 年度に集中的に実施する。一方、スタイル評価については、スタイル違反の種類が機械翻訳機の性能で大きく変わることはないと考えられるので、本評価を優先する。これらから、令和 2 年度は、研究開発項目 1-2 で構築する統合型機械翻訳システムの翻訳エンジンを利用し、翻訳結果をニュース記事としての評価基準で評価し、実データでの翻訳精度などの分析評価を実施する。評価を通じて補完の効果を明らかにする。評価記事数は 30 記事以上とする。

### ②実施内容

- ・ 機械翻訳エンジンの評価を実施した。また、統合型機械翻訳システムの翻訳エンジンを利用し、翻訳結果をニュース記事としての評価基準で評価し、実データでの翻訳品質などの分析評価を実施した。
- ・ 統合型機械翻訳プラットフォームの統合評価を実施するにあたり、ニュース記事データが必要のため時事通信社よりデータを借用した。データ数量は、日英記事 65,860 本、和文記事 4,405,974 本、英文記事 1,047,175 本、短文記事 210,457 本、ニュース写真データ（日英キャプション付き記事データ）約 34.5 万枚、AFP 英文記事 493,290 本、日本語翻訳文約 50 本である。

### ③成果

- ・ 1 記事に平均 1 か所以上の修正がある項目は、Aa（過剰な翻訳）、Ab（誤訳）、Ad（カタカナ表記）Ae（訳抜け）、Ah（日付/時刻/数値）、Aj（湧き出し表現）、Ca（文構成）、Cb（訳文解釈）となった。翻訳精度向上に向け、上記原因については内容を明らかにして社会実装に向けて改善を検討する。

## (8) 研究開発成果の展開・普及等に向けた計画・展望

### ①計画

凸版印刷は、本委託研究を通じて得た知見を活用して、2021年4月以降、時事通信社、マイソドワードと連携し、日本語ニュース記事の機械翻訳配信(英文)のサービス化を目指す。また、日本語ニュース記事の多言語翻訳配信や海外ニュース記事の機械翻訳による日本語配信サービスの検討も行い、多言語コミュニケーションの課題を解決する事業を拡げていく。

実証実験において、ニュース翻訳の機械翻訳システムの性能は実用レベル近くに達したことが示された。凸版印刷では、様々な業種にソリューションの提供をしており、それらのソリューションの1つのサービスとしてニュース配信の機能搭載を検討していく。

#### 【凸版印刷ソリューションとの連携候補】

- ・ ビジネス書要約読み放題サービス「flier (フライヤー) 法人版」  
(<https://www.toppan.co.jp/solution/service/flier.html>)
- ・ 旅道®  
(<https://www.toppan.co.jp/solution/service/tabido/>)
- ・ デジタルサイネージソリューション ReView  
(<https://www.toppan.co.jp/solution/service/digitalsignage.html>)
- ・ BotFriends® Vision (ぼっとふれんず ビジョン)  
(<https://www.toppan.co.jp/solution/service/botfriendsvision.html>)

社会実装に向けては、i)「ニュース機械翻訳配信実証サービス(仮)」のデモ用システムの構築、ii) デモ用システムを利用した実証実験(各業界での市場調査)、iii) サービス化に向けた事業計画の検討を進める。

#### ②展望

本研究開発成果の直接的な訴求先として、ビジネスシーンでの展開を検討している。本研究開発では、企業の経営層、知識層等を海外情報の摂取ニーズの高い属性を対象に実証実験を行った。速報性や要約技術を高く評価しており、機械翻訳のサービスレベルを把握することができた。このことから、次年度以降、研究開発成果の社会実装を図っていく中で、5年後には標準的なサービスとして確立している。

本研究開発成果は、ビジネスシーンにおけるニュース配信サービス以外にも、情報通信研究機構が自ら実施している研究開発活動と連携することにより、幅広い応用分野への展開が期待できる。

以上