

令和元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（令和元年度 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	0	0
	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	0	0
	プレスリリース・報道	0	0
	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目標

- ニュース記事の要約文生成の検討、実装、評価を行う。今年度は、研究課題 197 の要約技術を使って生成した英語ニュース記事のタイトルとダイジェストについて評価を実施する。
- 当該技術の API を提供するための統合型機械翻訳システムの設計、連携研究課題の成果の統合、API の試作を行う。今年度は、統合型機械翻訳システムの設計と試作を実施する。具体的には、システム構成を設計し、また、現段階における研究課題 197 の課題 2 と課題 3 を疎結合で統合して評価を実施する。更に API を試作し、統合型機械翻訳プラットフォームで使用するための課題等を抽出する。

## ②実施内容

- ニュース記事の要約文生成の検討、実装、評価を行った。具体的には、日本語ニュース記事から研究課題 197 の要約技術を使って、日本語タイトルと日本語ダイジェストを生成し、それを機械翻訳して生成した、ニュース記事の英語タイトルについて評価を行った。評価データは商用の日本語ニュース記事データを利用した。評価の結果抽出された課題等を研究課題 197 へフィードバックした。
- 当該技術の API を提供するための統合型機械翻訳システムの設計、連携研究課題の成果の統合、及び、API の試作を行った。具体的には、システム構成を設計し、また、現段階における研究課題 197 の課題 2 と課題 3 を疎結合で統合して評価を実施した。更に API を試作し、統合型機械翻訳プラットフォームで使用するための課題等を抽出した。

## ③成果

- 英語記事タイトル自動生成のための要約技術を評価して、カテゴリー毎の性能の違いを明らかにし、要約率の最適化を行った。
- 研究課題 197 と連携して、ニュース記事翻訳エンジンの環境を構築して、要約結果や記事本文の翻訳評価及びレスポンス評価を実施した。
- 翻訳結果の湧き出し対策として、ポストエディットの必要性を確認した。
- 画像情報の利用のために、委託研究課題 197 と連携して、商用データの画像 3,000 枚を使い、シーン認識実験を行った。シーンカテゴリーや学習データの見直しが必要であることを明らかにした。
- 要約、翻訳、記事配信プラットフォームなどの連携に必要な API を試作した。

## 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

### ①目標

- 実証環境の設計を実施する。今年度は、「日本語ニュース記事の英語配信」のデモに向けて、配信先の連携プラットフォームを決定して、結合方法の設計を行う。設計にあたっては、実証環境の設計、及び、運用方法の設計を行う。
- 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。ニュース記事の評価手法は研究課題 197 で検討されているので利用する。今年度は、正規表現を用いて実装するスタイル違反検出器を開発し、人手のチェックと比較する。研究課題 197 の課題 2 の翻訳エンジンの翻訳結果に含まれるスタイル違反をカウントする。

### ②実施内容

- 実証環境の設計を実施した。具体的には、「日本語ニュース記事の英語配信」のデモに向けて、配信先の連携プラットフォームを決定して、結合方法の設計を行った。設計にあたっては、実証環境の設計、及び、運用方法の設計を行った。
- 統合型機械翻訳プラットフォームの統合評価を実施した。具体的には、正規表現を用いて実装するスタイル違反検出器を開発し、人手のチェックとの比較を実施した。実施にあたっては、研究課題 197 の課題 2 の翻訳エンジンの翻訳結果に含まれるスタイル違反をカウントした。

### ③成果

- 商用の 800 記事を選び、そのコストや訳質の改善効果などの有効性を定量的に確認するための翻訳作業、及び、機械翻訳出力を用いたポストエディット作業を実施した。

- 市場調査を行い、実証を「旅道ホテル向けインフォメーションサービス」と定めた。既存サービスの改修計画を策定し、実証実験場所の交渉を開始した。
- 記事配信プラットフォームの試作を行った。日本語記事の選定、記事配信許可、翻訳実施、タイトル生成機能を盛り込んだ。
- 記事配信プラットフォーム（クライアント側）を試作した。翻訳された記事の閲覧に加えて、アンケート機能を盛り込んだ。
- 商用記事を利用するので、高い翻訳品質が要求される。日時表現や固有名詞などが準拠すべきスタイル基準をチェックするチェッカーを試作した。