

## 1. 研究課題・受託者・研究開発期間・研究開発予算

- ◆ 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
- ◆ 実施機関: 凸版印刷株式会社、マインドワード株式会社
- ◆ 研究開発期間: 令和元年度～令和2年度 (2年間)
- ◆ 研究開発予算: 総額120百万円 (令和元年度60百万円)

## 2. 研究開発の目標

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。

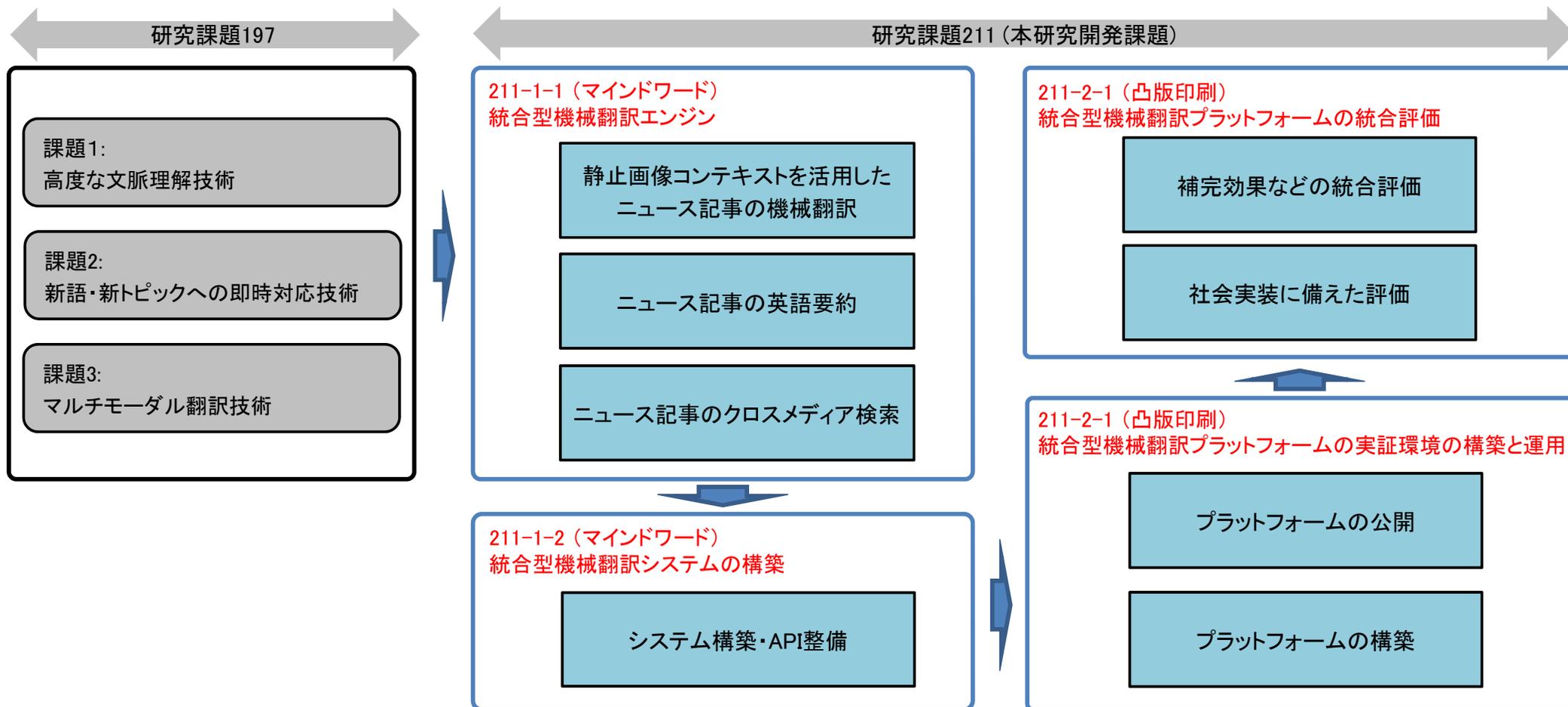


図1 研究開発の役割分担

### 3. 研究開発の成果

- 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、図2の通り、カテゴリ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、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。
- 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。
- 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。
- 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。

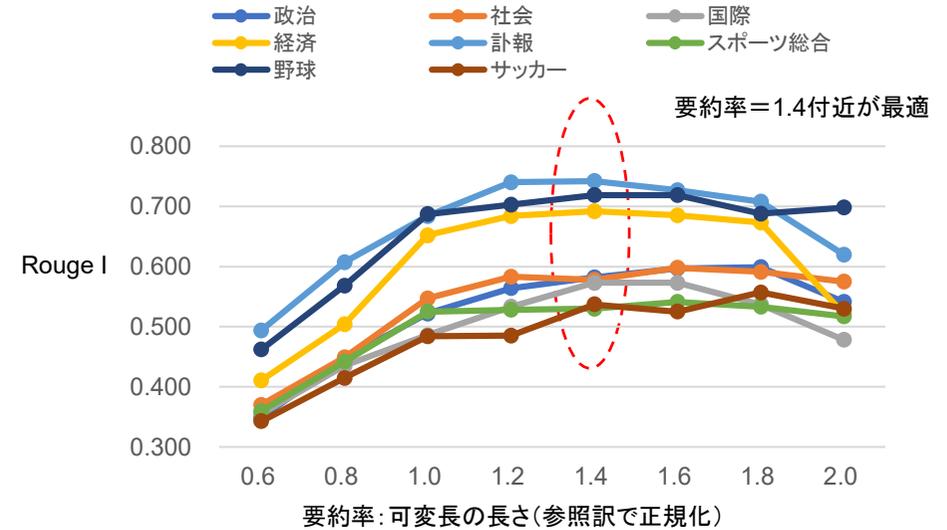


図2 要約率と性能(Rouge I)の関係



図3 実証予定のサービスイメージ

4. 特許出願、論文発表等、及びトピックス

国内出願	外国出願	研究論文	その他研究発表	プレスリリース 報道	展示会	標準化提案
0 ( 0 )	0 ( 0 )	0 ( 0 )				

※成果数は累計件数、( )内は当該年度の件数です。

5. 今後の研究開発計画

- 研究開発の最終年度である令和2年度は、商用の日本語記事から作られる英語ニュース配信の実証実験を行う。ホテルなどにおける外国人向けの情報サービスに英文ユース配信を盛り込んで、社会実装を見据えた社会実証を行う。
- ニュース配信フローの中での研究開発成果の主な検証項目は、①人手のチェック(ポストエディット)の効率化、②読者アンケートによる機械翻訳の許容性など社会実装に必要となるコストなどのデータを取得する予定。
- 商用のニュース記事を利用することから、高い品質のタイトルや翻訳が求められるので、研究課題197で策定した基準に準拠するチェックツールを利用した精度と効率化の確認を行う。
- 配信には、①記事本文からの要約技術によるタイトル生成、②記事本文の機械翻訳、③テキストに加えて、画像から生成されるメタデータを利用した検索の各技術を利用する。
- 要約技術は、その精度向上とともに、湧き出しのフィルタリング機能を実装する。
- 機械翻訳は、ポストエディット、基準準拠のチェッカーを含めた最適化を行う。
- 商用版の記事に付与された画像で画像認識エンジンを再学習して、検索精度を向上する。