

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ディープラーニング技術の研究開発
- ◆副題：深層学習によるマルチモーダル文脈理解と機械翻訳の高度化
- ◆実施機関：(大)東京工業大学、(大)東京大学、(大)愛媛大学、日本放送協会、(一財)NHKエンジニアリングシステム、(株)時事通信社
- ◆研究開発期間：平成30年度～令和2年度(3年間)
- ◆研究開発予算：総額780百万円(令和元年度240百万円)

2. 研究開発の目標

大規模な対話・対訳コーパス、文脈理解技術、シーン画像認識技術を開発し、それらを深層学習の枠組みで統合することで、マルチモーダルな文脈理解、および対話やニュースに対応した高度な機械翻訳の基盤技術を確立する。ニュース記事の日英翻訳において80%の翻訳成功率、対話・雑談の翻訳において70%の翻訳成功率を達成し、ニュース配信の実務や音声翻訳アプリにおいて実証実験を行う。

3. 研究開発の成果

①インテリジェント翻訳技術の研究開発

山本さん	もしもし、山本と申します。	Ms. Yamamoto	Hello, this is Yamamoto.
田中さん	販売部門の田中と申します。	Mr. Tanaka	This is Tanaka from the Department of Sales.
田中さん	輸出に関してご助言いただきたくお電話しました。	Mr. Tanaka	I called you to get some advice from you concerning export.
山本さん	はい、どのようなご用件でしょうか?	Ms. Yamamoto	Okay, what's the matter?
田中さん	イランの会社から遠視カメラの引き合いを受けているのですが、イランに対しては輸出制限があると新聞で読んだことがある気がして...	Mr. Tanaka	We got an inquiry from an Iranian company about our far-sight cameras, but I think I read in the newspaper that there are export restrictions against Iran.
田中さん	うちで売っているようなカメラなら、特に問題にならないのでしょうか?	Mr. Tanaka	Is there no problem with cameras like the ones we sell?
山本さん	恐れ入りますが、イランへの輸出は、かなり制限されているのが事実です。	Ms. Yamamoto	I'm afraid that the fact is, exports to Iran are highly restricted.
山本さん	また、遠視カメラのような高性能なデバイスですと、中には軍事利用可能なものもあります。	Ms. Yamamoto	Besides, some high-spec devices like far-sight cameras can be diverted to military use.

- A) 4000人を超える人々と非常にたくさんのロボットが、(そこで)働いています。(彼)在庫商品の保管、選別、こんごう発送を行っている、時には即日配達ということもあるのです。
- B) それですが、郊外にある昔ながらのショッピングモールが衰退している1つの理由です。(商品を)すぐに配達してもらおうという「即座に得られる満足感」に(私たちは)なれてしまいました。
- A) More than 4,000 people and many, many robots work there. They store, select, pack and ship the company's inventory – sometimes for same-day delivery.
- B) That's one reason for the decline of the traditional shopping mall in the suburbs. We've become used to the instant gratification of having products delivered quickly.

ビジネスシーン
日英対話対訳
コーパスの構築

文脈に応じて適切な
訳を選択

日英対話対訳コーパス

翻訳に必要な情報の
補完

訳語の一貫性
の保持

ただ、料金には、他国の携帯電話会社の通信回線などの利用料が上乗せされる。

新設する協議体では、3か国の携帯電話会社が設定している回線利用料を引き下げの方策について検討する。合意が得られ次第、各国政府が自国の携帯電話会社に対し、回線利用料を安くするよう要請する。

国内の携帯電話会社に料金引き下げ交渉を促している。

However, fees for communication lines of other countries' mobile phone carriers will be added to the fees.

The new panel is expected to discuss measures to lower the fees to be used by cell phone carriers in the three countries, the sources said. As soon as the agreement is reached, the governments will ask their own cell phone companies to lower their fees for the services they use, the sources said.

urged domestic cell phone companies to lower rates.

研究開発成果:ビジネスシーン日英対話対訳コーパスの構築

- 累計84,800文の対話対訳コーパスを構築(目標は10万文)
- シーン情報付きの日英対話対訳コーパスは世界初
- 既存の英語会話およびニュース累計約15万文を日本語へ翻訳

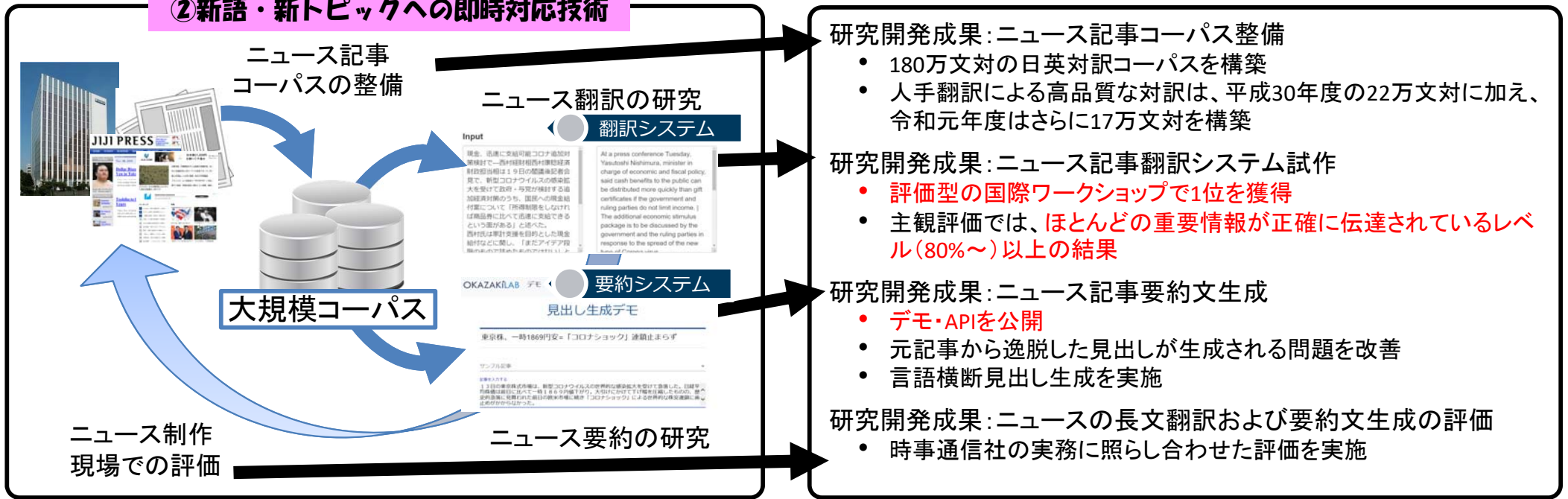
研究開発成果:翻訳に必要な情報の補完

- BERTを用いて、文脈として適切な文を選択する手法を開発
- 与えられた2文が連続する2文かの判定を行うタスクと、翻訳を行うタスクのマルチタスクでBERTをファインチューニング
- 翻訳対象文と、その1~5文前をそれぞれペアにして入力し、分類器による連続文判定の確率が最も高い文を文脈文として利用
- ベースラインと比較して、文脈文選択を行った場合は日英翻訳で0.10から0.67、英日翻訳で0.15から0.94ポイントBLEU値が向上したが、Oracleのスコアと比べるとまだ7ポイント程度の差がある

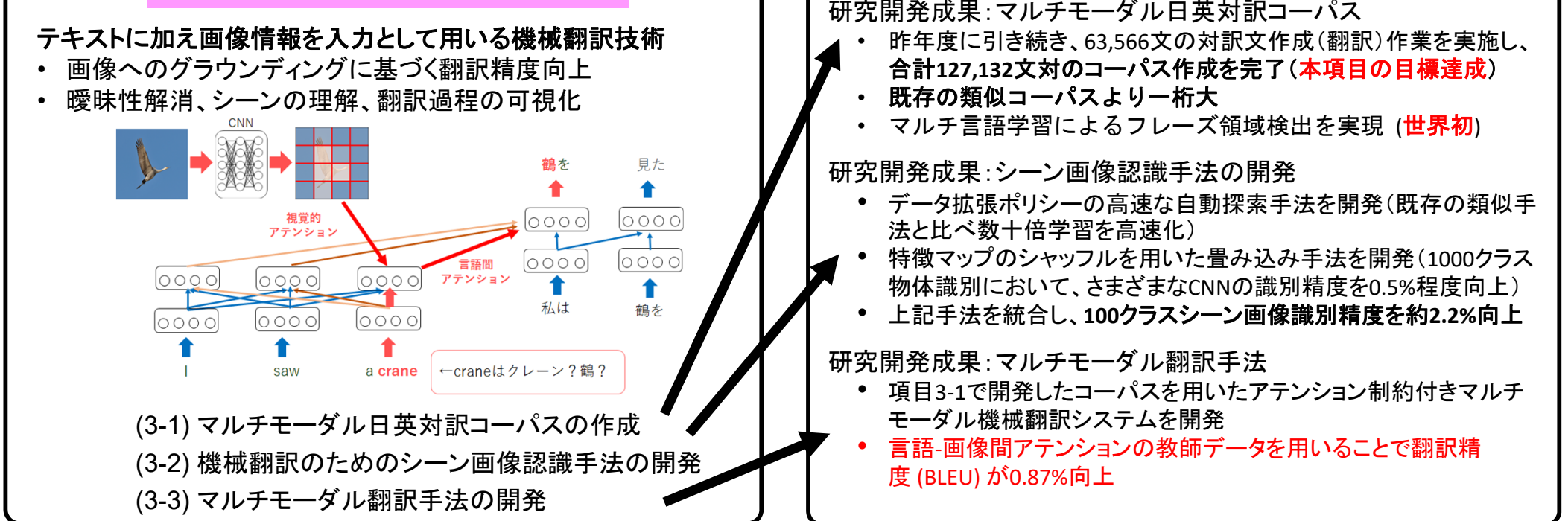
研究開発成果:適切な訳を選択・訳語の一貫性を保持

- 単文翻訳エンジンの結果をクラウドソーシングにより評価したところ、文脈付きでは日英で約15%、英日で約20%で翻訳誤りがあった
- 文脈付き機械翻訳では単文での翻訳よりBLEUスコアが向上するが、オラクルのスコアよりは低い
- 文脈情報が必要かどうかをあらかじめ判定する必要がある
- 対話対訳では直前文が相槌などで情報量が少ないものが多い

②新語・新トピックへの即時対応技術



③マルチモーダル翻訳技術の研究開発



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	標準化提案	プレスリリース 報道	展示会	受賞・表彰
1 (0)	0 (0)	1 (0)	54 (34)	0 (0)	1 (0)	0 (0)	8 (6)

※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

引き続き各研究項目の研究・開発を進め、最終目標の達成を確認する。活発な対外発表を継続するのはもちろんのこと、本プロジェクトの社会実装を見据えた取り組みを進める。例えば、これまでに開発したニュース対応日英機械翻訳エンジンや自動要約エンジンについて、報道の実務レベルでの有用性を検証する。また、本委託研究で構築したコーパス、機械翻訳のソフトウェア、学習済みモデルを研究者向けに公開する方法を検討し、委託研究の実験結果を後続の研究が再現できるように準備する。