

## 3.5.2 ユニバーサルコミュニケーション研究所 多言語翻訳研究室

室長(兼務) 隅田英一郎 ほか12名

### 多言語翻訳システムの実現に必要なアルゴリズムと対訳データの構築法の研究開発

#### 【概要】

本研究室は、人と人との言葉の壁を克服するため、日本語と英語の対を主たる対象として研究を進めている。対訳データ(原文と訳文の対を集積したもの)に基づいて翻訳する手法を採用し、同手法の基盤になる大規模な対訳データを構築してきている。

その研究の背景は以下のとおりである。

- 2020年東京オリンピック・パラリンピック競技大会の開催に向け、多言語対応の強化のため、国の関係行政機関、関係地方公共団体、関係機関、民間団体及び企業等が相互に連携・協働して取り組むことを目的として2020年東京オリンピック・パラリンピック競技大会に向けた多言語対応協議会<sup>\*1</sup>が設置され、自動翻訳など多言語対応への関心が広がっている。
- 専門分野向け高精度自動翻訳システムを実現できる技術は、社会経済的に日本国にとって不可欠である。例えば、特許の自動翻訳の必要性については、産業構造審議会平成24年6月25日配布資料『知的財産立国に向けた新たな課題と対応』<sup>\*2</sup>のP.36「急増する外国語文献への対応」において、明記されている。
- 日本語と英語の対は、文法や語彙においてほとんど類似性・関連性がなく、互いに翻訳が難しい言語対の代表であることから選定している。この言語対を対象とした技術開発は、より翻訳が容易な言語対も含んだ多言語翻訳にも適用できることから、科学的・技術的意義が大きい。

#### 【平成26年度の成果】

翻訳システムの高精度化・多言語化・多分野化・話し言葉翻訳の高度化を進めた。

##### 【高精度化】

- 単語アラインメントのRNN(Recurrent Neural Networks)による高精度化の提案を行った。機械翻訳の性能を向上させるため、ノンパラメトリックベイズ法により、単語間の係り受け構造から、原言語と目的言語の両言語を考慮したバイリンガルな品詞を推定する手法を提案、自動翻訳精度の向上を確認した。

##### 【多言語化】

- ある言語から別の言語に言語知識を移植する手法による翻訳の多言語化の1つとして、特に、単語分割について進めた。
- 英語等を仲介する多言語化の例として、中日の特許翻訳を、中英と英日を組み合わせる方法で実用化した。

##### 【多分野化】

- (災害、医療分野) 災害、医療分野を含む対訳データの構築。医療翻訳については、翻訳性能を向上し、倫理審査(承認番号10704)を経て、東大病院で実証実験<sup>\*3</sup>(図1)を実施し、分析結果を次年度に反映予定。
- (特許分野) 翻訳の要素技術の分野適応の例として、汎用の構文解析を特許用に適応し、精度向上を実現し、翻訳システムに組み込んで実用化した。

##### 【話し言葉翻訳の高度化】

- 従来技術で実現できる逐次通訳にはなく同時通訳にある効用は応答時間の短縮であり、コミュニケーション促進に寄与すると

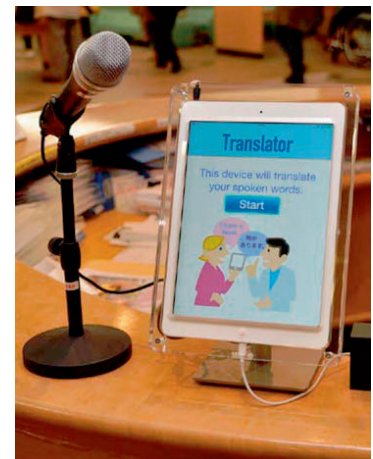


図1 東大病院の総合受付での実証実験  
(「東大病院だより No.83」から転載)

\*1 <http://www.2020games.metro.tokyo.jp/multilingual/index.html>

\*2 [http://www.jpo.go.jp/shiryoutou/toushin/shingikai/pdf/tizai\\_bukai\\_18\\_paper/siryoutou\\_01.pdf](http://www.jpo.go.jp/shiryoutou/toushin/shingikai/pdf/tizai_bukai_18_paper/siryoutou_01.pdf)

\*3 1月19日～3月13日の期間、外国人患者の方48人にご協力いただいた。

期待される。同時通訳の要となる「入力の途中から出力する」漸次翻訳システムを実装し、複数言語対で評価し、課題を明確にした。

- 書き言葉について今中期計画期間での目標としていた20語程度の文を越えて、平均35単語と文が長い特許文を対象として長文翻訳技術の実用性を証明し、実証実験にとどまらず英日翻訳システムを日本発明資料株式会社に、中日翻訳システムを一般財団法人日本特許情報機構(JAPIO)に技術移転した実用化実績が評価されて、第9回AAMT長尾賞を受賞した(図2)。
- ミャンマー語と日本語・英語の旅行会話用の自動翻訳システムを構築し、約9割の翻訳率を達成した(図3)。さらにミャンマー語の音声処理に不可欠な「ミャンマー語の書記素から音素への変換手法の提案」で13th International Conference on Computer Applications 2015で論文賞を受賞した。平成26年4月に開始した国のグローバルコミュニケーション計画<sup>\*4</sup>に沿った音声翻訳の研究開発も効率的に進めた。
- ACLやEMNLPなどの論文採録が困難とされている国際会議で9件採録。



図2 AAMT長尾賞表彰式の様子

エンジン	文章
原文	ဇွန်လငါးရက် နေ့ နာရီတာကိုသွားမယ့် လေယာဉ်အမှတ် နှစ်သုညငါးကို ကြိုတင်စာရင်းပေးထားပါတယ်။
ミャンマー	六月五日発の二零五便を予約してあります。 ဇွန်လငါးရက် နေ့ထွက်မယ့် လေ ကြောင်း အမှတ် ၂၀၅ ကို ကြိုတင်စာရင်းပေးထားပါတယ်။
エンジン	文章
原文	တဆိတ်လောက် ဆီယက်တဲလ်ကိုသွားတဲ့လေယာဉ်ကို ကြိုတင်စာရင်းပေးသွင်းချင်ပါတယ်။
ミャンマー	すみませんシアトル行きの便を予約したいのですが。 ကျေးဇူးပြုပြီး ဆီယက်တဲလ်ကို သွားတဲ့ လေယာဉ်ကို ကြိုတင်စာရင်းပေးသွင်းချင်ပါတယ်။

図3 ミャンマー語から日本語への自動翻訳画面(ミャンマー語原文、日本語の翻訳結果、確認のための逆方向の翻訳結果)

- ①アジア言語の自動翻訳にかかわるコンペ型国際会議 Workshop on Asian Translation (WAT、10月)を科学技術振興機構(JST)、京都大学と連携して創設し、また、②音声翻訳にかかわるコンペ型国際会議 International Workshop on Spoken Language Translation (IWSLT、12月)を開催し、多言語翻訳分野の発展への貢献。
- Berkeley Parser を並列化した。本並列化拡張プログラムは、オープンソースの構文解析プログラムとして著名な Berkley Parser に並列化拡張を施すものである。本拡張を用いることにより、OSにもよるが、2倍程度の高速化が可能である。本プログラムは、Berkeley Parser プログラムが GNU GPL v2 ライセンスであることから、同じライセンスで一般に公開した。

<sup>\*4</sup> [http://www.soumu.go.jp/main\\_content/000285578.pdf](http://www.soumu.go.jp/main_content/000285578.pdf)