

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 隅田英一郎 ほか 10 名

多言語翻訳システムの構築に必要な対訳データと翻訳アルゴリズムの研究開発

【概要】

本研究室は、人と人との言葉の壁を克服するため、日本語と英語のような異なる言語間の翻訳の研究を実施している。特に、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、自動化やコミュニティとの協業など新たな手法によって対訳データの構築を効率化し、同手法の基盤になる大規模な対訳データを構築した。さらに、この対訳データを用いて旅行と特許の分野において高精度翻訳を実現した。

【平成 24 年度の成果】

音声翻訳技術（音声コミュニケーション研究室と共同の研究成果）

事業化と他分野展開

社会還元加速プロジェクト「言語の壁を乗り越える音声コミュニケーション技術の実現」を 5 年計画（平成 24 年度末終了）で実施していたが、研究計画を上回る成果が得られたため、1 年前倒しで平成 23 年度末に成功裏に終了することができ、平成 24 年度は同技術の民間移転と他分野展開に注力した。累計で 5 社にライセンスし、各社での事業化が進んでいる。ネットワーク型では、2 社が実績を積み上げており、その他、ノート PC 向けのスタンドアロン型の開発を進めている会社もある。

また、NICT では、平成 22 年 8 月から、スマートフォン用の音声翻訳アプリケーション VoiceTra を公開し、実証実験を実施してきた。

- (1) VoiceTra は、公開以来、多くの方々にご利用いただいた（アクセス数が 11,142,589 件、ダウンロード数が 808,326 件）。これまでに、本実証実験の当初の目的である、音声翻訳システムの性能向上技術の創出やユーザーインターフェースの改良を実現できたことから、平成 25 年 3 月 31 日をもって、VoiceTra の運用を終了した。
- (2) VoiceTra が大変好評であったことを踏まえ、NICT は、音声翻訳の技術を（株）フィート（<http://www.feat-ltd.jp/>）にライセンスし、同社は、平成 24 年 12 月 17 日（月）から、VoiceTra + という名称で同機能のアプリケーションの無料サービスを開始している（図 1）。



図 1 VoiceTra+ へ

また、VoiceTra は各方面から注目を集め、その結果、他分野展開が始まった。NICT は情報バリアフリーに向けて事業支援や研究開発に取り組んできているところであるが、当研究室においても VoiceTra の技術を応用した「聾啞者と健聴者のコミュニケーション支援ツール」を開発し、聾学校と協力して実証実験を経て、平成 25 年度前期のアプリの公開を目指して準備をしている。旅行会話のシステムに、専門用語を追加したり、医療現場で多用される表現を追加したりして、病院での実証実験も準備され、今後、他の分野も含め幅広い活用が期待できる。

国際連携 U-STAR

アジア・ヨーロッパを中心とした音声・言語処理に関する代表的な研究機関から成る国際研究共同体 U-STAR を拡張した（23 カ国 26 機関）。U-STAR は平成 24 年 7 月から、世界人口の 95% と対話可能な音声翻訳アプリ VoiceTra4U-M を開発し全世界規模での音声翻訳の実証実験を実施している。

●テキスト翻訳技術

専門分野向け高精度自動翻訳システムを多分野で実現できる技術は、波及性も高く、社会経済的に我が国にとって不可欠である。この背景の下、特許などの長文のテキストの翻訳の基礎技術の研究に注力した。

特許翻訳の高精度な自動翻訳

文法が異なる言語間の翻訳の際、NICT の自動翻訳方式では、自動学習した語順変換規則に従って、語順を変換した後、自動学習した確率付きの対訳辞書を用いて訳語を選択する。今年度は、入力文を構文解析して、予め自動抽出した規則に従って、出力言語の語順に変換する語順制御の新技术を創出した（図 2）。また、対訳

辞書の規模は約 10 億に上り、従来ソフトの専門用語収録数の約百～千倍に相当する大規模なものになっている。中日、英日の平均約 25 語の特許データで翻訳実験を行い、同手法は翻訳率で従来法に大きく優る 80%、85% の高精度を実現した。

特許の自動翻訳の事業化

1. 日本特許情報機構（Japio）との連携による中日翻訳の事業化

日本特許情報機構と共同研究し、自動翻訳システムの訳文に対して、特許のプロの目で見えた「通じる翻訳」に向けた評価によって効率的な改良を行うことができた。新技術に基づいて開発した、「中日自動翻訳ソフトウェア」の翻訳者が判定した精度は、従来技術の 3 倍以上の値を達成している。

図 3 のように、この「中日自動翻訳ソフトウェア」によって、Japio は中国の特許文献を日本語に翻訳及びデータベース化し、平成 25 年 4 月に、有償サービス「Japio 世界特許情報検索サービス (Japio-Global Patent Gateway)」の拡張版として事業化する予定である。企業の知財部や弁理士の知財調査や特許庁審査官の先行技術調査に役立つほか、日本の企業の特許侵害のリスクの軽減に役立てることができる。

2. 日本発明資料（ニッパツ）との連携による英日翻訳の事業化

NICT とニッパツが共同で開発した「英日自動翻訳ソフトウェア」では、(特許要約 1 件あたりの) 訳語誤り数を従来技術と比べて、約 12 分の 1 に削減するという高い品質を実現した。今回開発した「英日自動翻訳ソフトウェア」によってニッパツは、①英語特許文献を日本語に翻訳するサービス、②日本語に自動翻訳した英語特許文献データベースなどを開始する予定である。

●アカデミアへの貢献

共通の対訳データに基づくコンペ型の国際会議を主催したり、多数の論文を執筆したり、翻訳研究に関するコミュニティで主導的役割を果たした。

- ① 平成 24～25 年に NICT 主催で実施中の特許翻訳に関するコンペ型ワークショップ NTCIR10/PatentMT には、米国 SRI と BBN、ドイツ RWTH、SYSTRAN に加え、中国 7 チーム、日本 7 チーム、台湾 1 チーム、カナダ 1 チーム、アイルランド 1 チームで合計 21 チームが参加し、現時点の最高の技術を比較している。多数の手法を一度に並行して実験し比較できるので、単独機関ではできない研究加速が可能になっている。そのため、本会議は、日本国特許庁 (JPO)、欧州特許庁 (EPO)、中国国家知識産権局 (SIPO) などの各国政府機関からも注目されている。
- ② トップレベルの国際学会で本年度も論文が 7 本採択され、翻訳研究分野の世界最強のグループの 1 つとなっている。

●特記事項

音声翻訳プロトコルを ITU-T で国際標準化して多言語化の基礎を整え、23 カ国 26 機関と U-STAR を組織し、世界人口の約 95% をカバーするシステム「VoiceTra4U-M」で世界実験を開始し、活発に民間へライセンス供与してきたという一連の音声翻訳に関わる業績が評価され、音声コミュニケーション研究室と当研究室と共同で第 58 回前島密賞を受賞した。同賞史上初めて受賞者に外国籍の研究者 (Dixon Paul Richard 研究員、Finch Andrew Michael 主任研究員、Paul Michael 主任研究員) が含まれていたことも大変注目された。

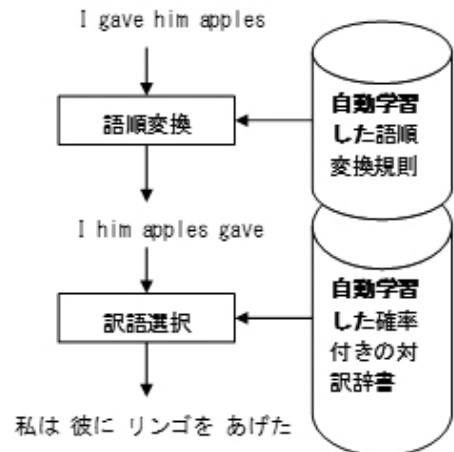


図 2 新しい翻訳手法

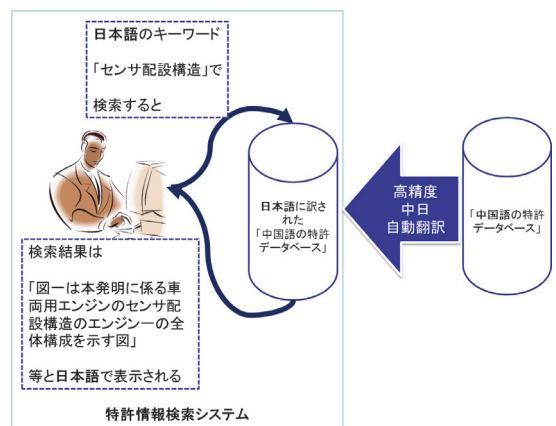


図 3 翻訳の検索システムでの利用