

平成25年度研究開発成果概要書

知識・言語グリッドに基づくアジア医療交流支援システムの研究開発 (156)

副題 音声翻訳技術と遠隔通訳者を用いた医療現場における
多言語コミュニケーション支援技術に関する研究開発

(1) 研究開発の目的

日本での外国人登録者は2007年時点で200万人を超えており、外国籍住民の医療機関の利用は増大の一途を辿っている。さらに近年は、高度医療を受けるなどの目的で訪日する外国人患者を受け入れて医療交流を推進することが、国内の今後の医療産業を活性化させる意味で非常に重要な国家戦略ともなっている。こうした外国人は日常会話以上の日本語が話せない人も少なくない。特に、医療機関利用の場面では、言語の問題が深刻になってきており、患者が日本語を話せないために医師に症状をきちんと伝えられず、適切な治療を受けられないケースが出てきている。また医療者側も、患者と意思疎通が図れないために診療に想定外の時間を要したり、誤診の可能性が生じるケースも出てくる。こうした問題解決のために、医療通訳者が外国人患者と医師の懸け橋となって支援を始めているが、医療に関する知識はもとより、異文化の理解、患者への配慮などが求められるため心理的負担も大きく、熟練を要する業務となっている。今後ますます医療交流が盛んになる中、検診ツアーなどでやってくる外国人を医療通訳者がサポートすることになるので、言語コミュニケーションに関わる医療通訳者の負担軽減が一層求められてくる。

このように、医療交流の分野では、医師、医療通訳を交えて、外国人患者との意思疎通をスムーズにしていくことが求められており、それを解決するためにはICTを活用した多言語コミュニケーション支援技術の実現が非常に重要となる。

このため、本研究では、情報通信研究機構（NICT）が開発する知識・言語グリッドに基づいて、アジア医療交流を支援する多言語音声翻訳などの技術と実証システムを研究開発することを目的とする。

(2) 研究開発期間

平成23年度から平成25年度（3年間）

(3) 委託先

日本電気株式会社（NEC） <幹事>

国立大学法人 奈良先端科学技術大学院大学（NAIST）

(4) 研究開発予算（契約額）

平成23年度	60（契約金額）
平成24年度	56（ 〃 ）
平成25年度	53（ 〃 ）

(5) 研究開発課題と担当

課題1 多言語コミュニケーション要素技術の開発

- 課題1-1 音声・テキスト音声翻訳アルゴリズムの医療交流のための高度化（NAIST）
- 課題1-2 医療交流用多言語コーパスの構築（NEC）
- 課題1-3 多言語医療表現と対訳パターンの自動獲得（NAIST）
- 課題1-4 コミュニケーション不全状態検出技術の開発（NEC）

課題2 医療交流支援実証実験システムの構築と実証

- 課題2-1 医療交流支援実証実験システムの構築（NEC）
- 課題2-2 医療交流支援実証実験システムの実証実験とシステム改良（NEC）

(6) これまで得られた研究開発成果

		(累計) 件	(当該年度) 件
特許出願	国内出願	3	2
	外国出願	0	0
外部発表	研究論文	0	0
	その他研究発表	23	13
	プレスリリース	0	0
	展示会	0	0
	標準化提案	0	0

(7) 具体的な成果実施内容と成果

課題1 多言語コミュニケーション要素技術の開発

課題1-1 音声・テキスト音声翻訳アルゴリズムの医療交流のための高度化

医療用音声翻訳のための、音声認識、音声合成、翻訳システムの構築を行った。さらに、NICT及びNAISTのサービスのクライアントを構築した。医療用音声翻訳のために、医療表現を記述した表現集、対訳辞書、及び、後述する現場コーパスから、医療用新規2,500単語を抽出し、日本語、英語、中国語の発音を付与し、医療用音声認識、合成のための基本発音辞書を構築した。また、一般生活、旅行用の音声認識、言語翻訳の基本言語モデル、翻訳モデルを構築した。さらにコミュニケーション

状態測定の基礎検討を開始した。また、システム統合を実行するプラットフォームの検討を進めた。

課題 1-2 医療交流用多言語コーパスの構築

昨年度までの構築分に加え、医療交流用に日英中の3言語で1,000文例以上の対訳コーパスを構築するため、実際の医療交流場面で発声され得る文例を考案することで各言語600文ずつの文例(計1,800文)を収集した(業務想定コーパス)。具体的には、初年度及び平成24年度に構築した参考図書コーパスより文例を抜き出し、その文例を直前の相手の発声と想定したうえで、その後自分がどのように発声するかを被験者に考案してもらうことにより収集した。また、上記文例に対して対訳文を作成し、初年度からの合計で12,054文の対訳コーパスを構築した。

課題 1-3 多言語医療表現と対訳パターンの自動獲得

多言語表現辞書を構築するための多言語医療表現辞書システムのデータベーススキーマの設計と基本機能の構築を完了した。さらに、多言語表現辞書を構築するための多言語医療表現辞書システムのデータベースシステムのプロトタイプを構築した。ライフサイエンス辞書中の単語(日本語表記と英語の対訳語表記)に対してGoogle n-gramを用いて頻度計算を行った。ライフサイエンス辞書の日英対訳用語を抽出し、日本語約3万語、英語約1万7千語(複数の日本語用語に対して同一の英語の用語が対応しているため見出し語数が英日で異なっている)からなる対訳辞書を対訳辞書システムに格納した。

また、類似の内容が書かれた多言語の文書に対して、文書中に現れる用語に関連するトピックをアラインする手法を設計し、トピック毎にアラインされた文書集合を構築することによって、単語アラインメントを学習し、対訳辞書を自動構築する手法を考案した。人手による評価により、従来手法の精度74%から85%への向上を確認した。

さらに、段階的翻訳に利用可能な対訳パターン獲得のために、英語文型パターンの調査を行い、大規模な英文データを統語解析することによって頻度の高い100パターンの抽出を行った。このようにして取り出された複文パターンによって翻訳対象の文の分割を行うことにより、統計的機械翻訳の評価指標BLEUスコアで2ポイント以上向上することを確認した。

課題 1-4 コミュニケーション不全状態検出技術の開発

NICT の音声翻訳システムを利用して日英双方向の音声翻訳結果を対象に本技術の評価を行ったが、昨年度の NEC の音声翻訳システムを利用した評価で得られた検出性能に及ばなかった。この改善のために課題を抽出した。さらに MFCC 特徴量を用いた同一発声による繰り返し検出の技術開発を行い、比較的高精度に検出可能である結果を得るとともに、同一発声以外の繰り返し現象についての課題を抽出した。これらの 2 つの技術を実証実験システムに組み込み、収集した対話コーパスを基に会話の流れの中で不全状態を検知する方式の予備評価により有効性を確認した。

課題 2 医療交流支援実証実験システムの構築と実証

課題 2-1 医療交流支援実証実験システムの構築

人間ドックの業務会話のうち、問診シーンに加えて検査シーンにおける多言語コミュニケーション支援も対象に利用要件を定義した。システムは実際の医療現場の従事者、患者、遠隔の医療通訳者の間での評価データを収集できるようクラウド型として仕様化した。また、NICT と NAIST の音声翻訳システムを切替接続できるようインターフェースの共通化と、前年度までの改善内容を踏まえた端末機能・アプリケーション機能の強化開発を行った。本システムの医療従事者らを対象にした実験への適用を通じて、システムの有用性と課題を確認できた。

課題 2-2 医療交流支援実証実験システムの実証実験とシステム改良

昨年度までの成果を踏まえ、適用業務は問診シーンに加えて検査シーンも対象にし、日英と日中の業務コミュニケーション支援を評価する実験実施計画を策定した。医療実務経験者、医療機関（En 女医会など）、医療通訳コールセンター業者の協力を得て、実機を使った評価データを収集した。音声翻訳と補完機能（類似文例提示）による会話タスク達成評価、遠隔通訳者への切替利用を通じてのシステム機能の有用性（①）、医療分野に特化した音声翻訳機能（NAIST）の有用性（②）を確認した。①では、問診シーンと検査シーンとも、定型的な会話タスクの例では日英で 87% から 94% 成立しており、効果があることが確認できた。②では、昨年実施した NICT の音声翻訳機能での同様な会話タスク（問診シーン会話、日英）との比較において、同等以上の評価結果が得られた。