

平成23年度「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」 (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) の研究開発目標・成果と今後の研究計画

1. 実施機関・研究開発期間・研究開発費

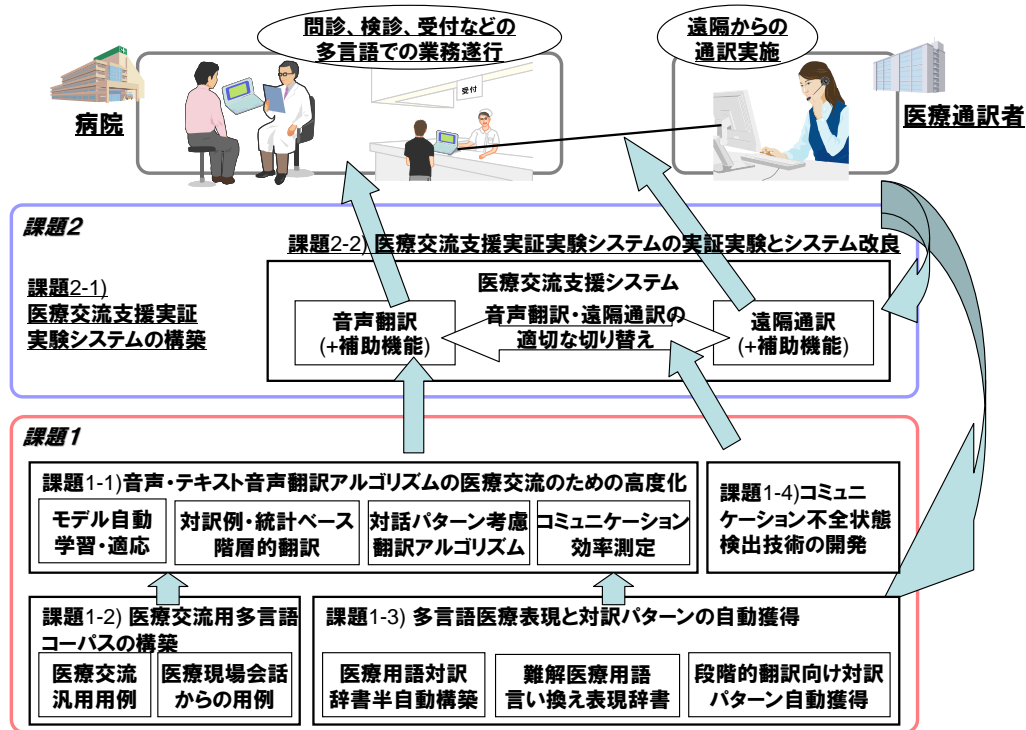
- 実施機関 日本電気株式会社(幹事者)、国立大学法人 奈良先端科学技術大学院大学
- 研究開発期間 平成23年度から平成25年度(3年間)
- 研究開発費 総額 169百万円(平成23年度60百万円)

2. 研究開発の目標

NICT開発の知識・言語情報グリッド基盤を利用し、さらに種々のアルゴリズムの高度化を行って、医療交流支援分野に特化した多言語コミュニケーション支援技術を確立する。

さらに、それらを活用した音声翻訳システム、および医療通訳者による遠隔サポートシステムを用いた実証実験システムを開発構築し、これらを組み合わせて利用するシステムについて、医療現場などでの検証を繰り返し行うことで、外国人患者、日本人の医師・看護師、医療通訳者のそれぞれにとってスムーズなコミュニケーションをサポートできるシステムを実現する。

平成23年度の目標としては、実際の医療従事者、患者、医療通訳者が多言語コミュニケーション支援技術を使用していくために必要な、基本的な技術研究開発、実験システムの試作を行う。試作においては、実際の医療現場の協力者を対象にした模擬評価などを通じて、医療業務現場での実証実験に具体的に適用していくための課題を抽出する。



3. 研究開発の成果 (NAIST)

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

・対訳例検索のための基盤となる表現の類似度計算法の検討とその評価を行うために、言語解析手法の性能向上の検討を行う。そのための基礎データとなる解析済みコーパス構築のためにアノテーション作業を行う。

・実際の会話データの一部を抽出し、複数の第三者の評価者により、コミュニケーションのスムーズさについての主観評価アノテーションを行う。

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

モデル自動
学習・適応

対訳例・統計ベース
階層的翻訳

対話パターン考慮
翻訳アルゴリズム

コミュニケーション
効率測定

課題1-3で収集するライフサイエンス辞書3万語を用いた音声認識・合成の発音辞書を構築するとともに、課題1-2で構築するコーパスとWeb上の関連文章から医療用表現をカバーする言語モデル、および翻訳モデルを構築する。

対訳に有効な表現パターンを同定するため、大規模言語コーパス中の表現のバリエーションに関する調査を行う。コーパス中で高頻度に現われる表現の意味的曖昧性の調査を行うために曖昧性判定作業を行う。

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データベースシステムを構築するために、必要な機能の洗い出しを行う。

課題1-3) 多言語医療表現と対訳パターンの自動獲得

多言語医療表現
辞書の構築

段階的翻訳向け対訳
パターン自動獲得

・段階的翻訳に利用可能な対訳パターン獲得のために、日英の文型パターンの調査を行う。

研究開発成果: 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

【課題】

NICTの知識・言語グリッドを用いた音声認識・合成用初期モデル構築、および翻訳モデル初期構築を行う。初年度収集分コーパス、および汎用コーパスを用いた評価を行う。

【成果】

音声認識・合成アルゴリズムの医療分野への特化・適応、およびコミュニケーション最適化技術

・医療用新規2,500単語に対し日英中発音辞書を構築した。さらに、医療用日本語音声認識、合成のための基本音響モデルを構築した。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の評価用の生体計測装置の利用を開始した。

対訳例検索・統計ベース翻訳からなる階層的翻訳技術

・対訳例検索結果をもとに、階層的翻訳を実行するプラットフォームの検討を行った。

対話パターンを考慮した翻訳アルゴリズム

・対話パターンを利用するため課題1-3で抽出したパターンの分析を進めた。

研究開発成果: 多言語医療表現と対訳パターンの自動獲得

【課題】

日本語ライフサイエンス辞書を対象にし、頻度情報等を考慮して3万語程度の医療用語を選定する。これらの用語を対象に、英語、中国語の対訳用語の自動構築手法について考察する。日本語および英語の一般分野の解析済みコーパスを対象に、複雑な文の構文パターンの抽出を行なう。構文パターンの複雑度と頻度を考慮して、100程度の構文パターン抽出を行う。

【成果】

多言語医療表現辞書の構築

・平成23年度は、多言語表現辞書を構築するための医療用語対訳辞書システムのデータベーススキーマの設計、構築を完了した。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。

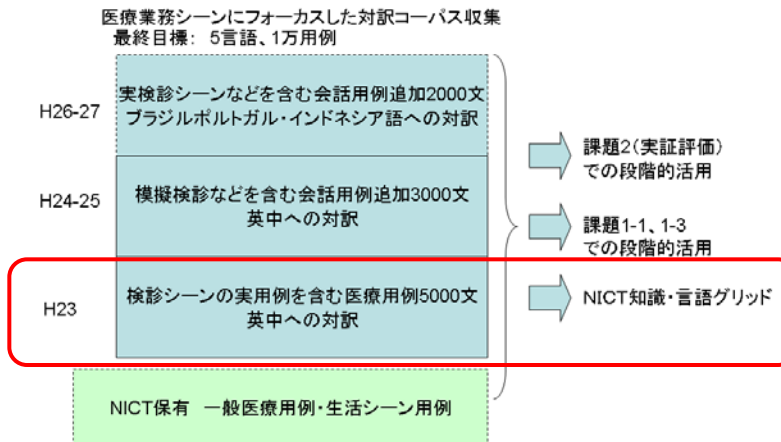
段階的翻訳のための対訳パターン獲得

・平成23年度は、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、100パターンの抽出を行った。

3. 研究開発の成果(NEC)

課題1 多言語コミュニケーション要素技術の開発

課題1-2 医療交流用多言語コーパスの構築



研究開発成果:医療交流用多言語コーパスの構築

【課題】

医療交流現場での音声・テキスト翻訳実用化していくには、医療交流に特化した多言語コーパスを収集構築する必要がある。具体的には、対訳の言語数としては5言語(日本語、英語、中国語、インドネシア語、ポルトガル語)、文例数としては1万文以上のコーパスを構築することを最終目標とする。

【成果】

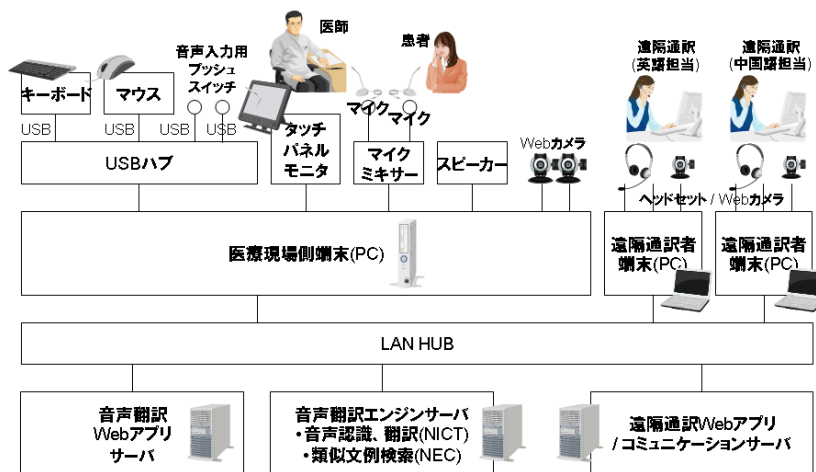
コーパスの収集と構築

平成23年度は、医療交流用に日英中の3言語で5,000文例以上の対訳コーパスを構築するため、一般的な医療専門用語や用例を広く収集・翻訳したものと、実際の医療交流の場面で交わされる実対話を収録・書き起こしたものの両面についてどう整備すべきかを検討し、計60,136文例を整備した。

- 図書を参考にした対訳コーパス(日英中)5,129文例
医療現場で利用が多いと想定される一般用例を書籍など参考に選定した。
- 医療現場の実業務会話コーパス(日英中)1,007文例
人間ドック検査(日本人2名、日本語)の実会話を収録を行い対訳を作成した。

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



研究開発成果:医療交流支援実証実験システムの構築

【課題】

医療分野での円滑な多言語コミュニケーションの支援のために、業務要件や技術課題を明確にする必要がある。特に音声翻訳、遠隔通訳から成る医療交流支援システムをどう使えば最適なコミュニケーションを実現できるかについて、プロトタイプを試作し、実証実験(課題2-2)に適用する。

【成果】

平成23年度は、音声翻訳機能と遠隔通訳支援機能の使い分けを実現する基本システムを試作した。課題2-2の実証実験適用で改善課題を洗い出した。

音声翻訳機能および補助機能(全体統合部分を含む)の実装・改良

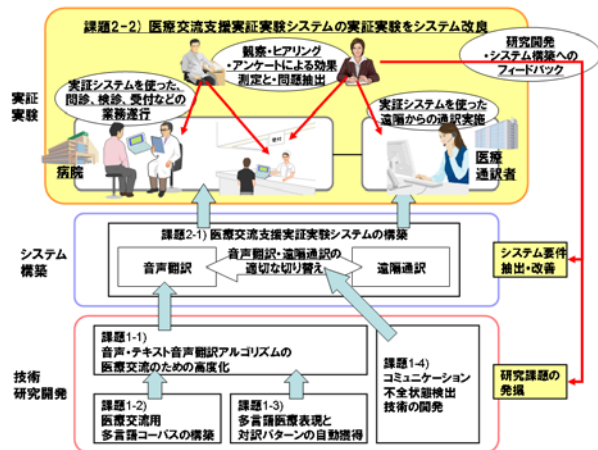
医療従事者(日本語)と患者(英・中)間での問診業務を想定し、どのようなユーザインタフェースが適しているか、どのような基本機能がインタラクションに必要なかを評価するためのソフトウェア(クライアント、NICT音声認識・翻訳エンジンとI/Fするサーバなど)を開発し、業務シミュレーションシステムを構築した。

遠隔通訳コミュニケーション機能および補助機能の実装・改良

音声翻訳ではカバーできないケースを遠隔地にいる通訳者がフォローする場合に、どのような支援機能が必要かを評価するための映像・音声通話/情報共有支援ソフトウェアを開発した。

3. 研究開発の成果 (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



研究開発成果:

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

【課題】

課題2-1の医療交流支援システムを問診、受付などの医療業務シーンに適用する実証実験を行うことで、システムの改良および多言語コミュニケーション支援技術の課題を抽出する。そのフィードバックにより研究開発、システム改良を進める。

【成果】

平成23年度は、多言語コミュニケーション支援技術を利用した医療交流支援システムの基本的な受容性について評価し、改善課題を洗い出した。

シナリオシミュレーション実証実験の準備と実施

- ・問診業務シーンを例に業務遂行や円滑なコミュニケーションに支障が出そうな点をシステムが支援できるかを評価する業務シミュレーションを設計した。
- ・システムが想定する目標精度や操作性を評価する実験を行い、医療従事者、医療通訳者に対するアンケートおよびヒアリングから、精度や操作性などの技術研究課題、システム改良課題を抽出した。

4. これまで得られた成果(特許出願や論文発表等) ※成果数は累計件数と()内の当該年度件数です。

	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	展示会	標準化提案
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	0 (0)	0 (0)	0 (0)	2 (0)	0 (0)	0 (0)	0 (0)

5. 研究成果発表会等の開催について

(1) 研究発表

- ・第1回コーパス日本語学ワークショップ(2012/3/5)にて、コロケーションの定義や代表的な検出指標、定量的に区別するために実施したアンケート調査について発表。
- ・音響学会春季研究発表会(2012/3/13)にて、分散共有フルコンテキストモデルを用いた尤度に基づくパラメータ生成法の提案、および実験的評価での有効性について発表。

6. 今後の研究開発計画

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

- ・NICT開発の知識・言語グリッドを用いた音声認識・合成用モデル構築、および翻訳モデルの精度向上、音声翻訳アルゴリズムの高度化の継続実施。
- ・初年度収集コーパス、および汎用コーパスを用いた評価と分析・コミュニケーションの効率測定技術について実測定と主観評価の関係分析に関し、基礎検討。

課題1-2 医療交流用多言語コーパスの構築

- ・初年度に引き続き、医療交流用多言語コーパスの追加収集と構築の継続実施。

課題1-3 多言語医療表現と対訳パターンの自動獲得

- ・頻度情報等より抽出した医療用語(日本語ライフサイエンス辞書を対象)の、対訳用語(英語、中国語等)の自動構築手法について検討。(日英対訳医療用語辞書のプロトタイプを作成、対訳用語辞書を格納する辞書管理システムの設計とプロトタイプシステムの構築、構文パターン(日英)の抽出)

課題1-4 コミュニケーション不全状態検出技術の開発

- ・平成23年度実施した業務シミュレーション評価の内容などを参考とし、音声翻訳誤りの分析などを基に、誤りを検出するための方式試作。

課題2-1 医療交流支援実証実験システムの構築

- ・平成23年度実施の業務適用要件を踏まえ、音声翻訳と遠隔通訳による医療従事者と患者の異言語コミュニケーションを円滑に補助する実証実験システムの開発。

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

- ・医療交流支援システムや課題1で研究開発する多言語コミュニケーション支援技術の問題点を洗い出すための実証実験を実施。