

平成 27 年度研究開発成果概要書

課 題 名 : 自治体向け音声翻訳システムに関する研究開発

採 択 番 号 : 1800101

副 題 : 自治体窓口業務に対応した音声翻訳システムとサービス化に向けた研究

(1) 研究開発の目的

本研究開発は、NICTが開発した音声翻訳システムをベースに自治体向けの音声翻訳システムの基盤技術の確立を行うため、実施するものである。NICTが開発した音声翻訳システムは、対訳コーパスから翻訳システムに必要な統計的な知識を自動的に構築する技術であり、音声認識/翻訳できる語彙の不足等は音声翻訳システムの性能に直接関わる。現在、NICTでは、訪日・在留外国人を対象にした生活会話等のコーパス整備を進めており、実用場面・局面毎のコーパスが拡張されているところである。

本研究開発では、音声翻訳システムの導入により窓口業務が大きく改善が図られるものと、音声翻訳システムの導入した場合においても改善が図られないものを把握し、自治体窓口業務の構造の把握と体系化を行う。コーパスデータの整備では、各種窓口等における生活会話文等を参考に対訳コーパスの基礎整理を図るとともに、音声コーパスでは、シナリオ策定に向けたシーンの設定等を行う。これらの成果を踏まえ、自治体窓口向け音声翻訳システムの開発と実証実験を行い、翻訳精度とユーザビリティの向上を図る。あわせて、研究開発成果の社会展開に向けて、サービス開発にかかる基盤の確立を行う。

これらから、本研究開発では、NICTで整備されたコーパス情報等も積極的に活用し、自治体窓口における音声翻訳システムの精度向上と研究開発の効率化を図る。

(2) 研究開発期間

平成 27 年度から平成 31 年度 (5 年間)

(3) 実施機関

凸版印刷株式会社
株式会社フィート

(4) 研究開発予算 (契約額)

総額 749 百万円 (平成 27 年度 150 百万円)
※百万円未満切り上げ

(5) 研究開発課題と担当

課題 1 : 自治体窓口業務に関する研究 (凸版株式会社)

課題 2 : コーパスデータの収集に関する研究

- i. 音声・対訳コーパス収集研究 (凸版株式会社)
- ii. クラウドソーシングによるコーパス収集研究 (凸版株式会社)

課題 3 : 音声翻訳システムに関する研究

- i. 音声翻訳システム研究 (株式会社フィート)
- ii. ユーザーインターフェース開発 (凸版株式会社)

課題 4 : 音声翻訳システムの実証実験及び検証

- i. 音声翻訳システムの実証実験 (株式会社フィート)
- ii. 音声翻訳システムの評価 (凸版株式会社)

課題 5 : ビジネスモデル等統括研究 (凸版株式会社)

(27-1)

(6) これまで得られた成果（特許出願や論文発表等）

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

(7) 具体的な実施内容と成果

(7-1) 自治体窓口業務に関する研究

①目標

本課題では、自治体窓口サービスに提供可能な音声翻訳システム及び当該システムのユーザーインターフェースの開発に向けて、自治体窓口の業務の把握を行い、音声翻訳システムを導入することによって業務の改善が見込まれる自治体窓口業務を選定する。

②実施内容

平成 27 年度は、i) 業務の把握・体系化、ii) 業務体系確認と課題の抽出、iii) 対象業務の抽出、iv) 業務フローの作成を行った。平成 27 年度は、在留外国人の利用頻度の多いと想定される業務（例：戸籍住民課、市民課等）に焦点を当て、窓口業務の分析を行った。

③成果

自治体窓口業務の把握、体系化をもとに、外国人対応の多い自治体窓口業務、それら業務における会話例等を把握することができた。

- 自治体窓口の対応業務の全体像と外国人対応が多い優先業務の抽出（板橋区・つくば市・白馬村へのヒアリングと、各自治体が発行する外国人対応に関する資料より）
- 優先業務A（戸籍住民課、市民課関連）の代表的な手続きと会話例の収集（板橋区協力によるロールプレイ実施により）
- 自治体における専門用語集（対訳付）の作成（a：自治体国際化協会の多言語生活情報からのリスト：806 語、b：板橋区生活情報からのリスト：581 語）
- 優先業務Aに関連する会話バリエーションの作成（約 760 文）
- 優先業務B（子育て等）の代表的な手続きと会話例の収集（自治体業務経験者 9 名へのインタビューより）
- 今後のコーパス収集のための窓口会話モジュール表（案）の作成

(7-2) コーパスデータの収集に関する研究

①目標

本課題では、音声・対訳コーパスを質量ともに充実させるための効率的な手法を確立するとともに、研究計画書の必須要件であるデータ収集を行う。

②実施内容

【a. 音声・対訳コーパス収集研究】

音声・対訳コーパス収集研究では、生活会話対訳コーパス等の活用し、日本語シナリオを作成し、ブラジル・ポルトガル語のコーパス収集を優先して進行した。

- 平成 27 年度は生活会話対訳コーパス等を活用し、日本語シナリオを作成する

(27-1)

- 日本語シナリオを対象言語に翻訳し、対訳シナリオを作成する。作成したシナリオを評価・検証の上、改善点があれば、次年度に反映する
- 対訳シナリオを参考にして、音声収録を行い、実際に求められている音声データを収集できているかを関係者と評価・検証の上、改善点があれば次年度以降の作業方法や手順に反映する
- 日本語シナリオと対訳シナリオから対訳コーパスを作成する

【b. クラウドソーシングによるコーパス収集研究】

本項目では、必要となる用語や表現を効率的に収集する手法の確立を目的として、いわゆる「クラウドソーシング」を活用したコーパス収集手法について検討する。

- クラウドソーシングによる多言語の音声コーパス収集システムを構築する
- 音声品質が低いデータのクリーニング手法を検討する
- 対応言語の多言語化を実験する

③成果

当該研究項目の成果は下記のとおりである。

- 対訳コーパス（日伯：4万文、日英：265文）：日伯は伯語の発話シナリオ用に生活コーパスから和文を流用し作成した。日英は課題1で実施した板橋区ロールプレイシナリオを活用し作成した。
- 音声データ（伯語：2,400名）：国内は大泉町（21名）、海外はリオデジャネイロにて収録した。
- 発話付対訳辞書（英語277語）：板橋区固有名詞辞書から作成した。
- クラウドソーシング収録環境の構築（伯語）

(7-3) 音声翻訳システムに関する研究

①目標

本課題では、旅行会話向けの多言語音声翻訳システムをベースに、対象分野を自治体窓口等に拡張する技術の確立と、自治体窓口に適したユーザーインターフェースを含む音声翻訳システムの構築を目的として研究開発を実施する。

②実施内容

【a. 音声翻訳システム研究】

- 自治体窓口コーパスを利用した翻訳モデルの構築、評価を行う（コーパスはサブセット）
- 自治体窓口向け音声翻訳システムの構築（実証実験用システム1）

【b. ユーザーインターフェース〈UI〉開発】

- 実業務の行動観察
- 業務の課題と利用環境の把握
- デモアプリの仮要件定義

③成果

- 翻訳モデルの構築・評価／システム構築：翻訳精度の向上に向けた3段階のテストセットの作成
- UI開発：ロールプレイの実施によるUI仮要件定義の作成

(7-4) 音声翻訳システムの実証実験及び検証

①目標

平成28年度に実施する自治体実証実験に向けた準備として、研究項目1で得た会話例の音声翻訳について検証を行う（課題3で構築する実証実験用システム1を用いて、1言語を対象に実

(27-1)

施)。併せて、平成 28 年度に実施する実証実験の評価手法及び評価項目の検討を行う。

②実施内容

優先業務 A（戸籍住民課、市民課関連）で得た会話例を用いた音声翻訳精度等の検証を行う。

③成果

- 翻訳モデルの構築・評価／システム構築：翻訳精度の向上に向けた 3 段階のテストセットの作成（※本課題は課題 3 と連携し行うことから、課題 3 と同じ）

(7-5) ビジネスモデル等統括研究

①目標

基礎調査として、自治体が本システムの導入するにあたり、その市場環境について調査・分析を行う。調査は、在留外国人の動向、国・地方自治体の外国人定住化に関する施策等について、文献、Web 等のオープン情報の収集・整理を図るとともに、在留外国人が急増している自治体の状況調査を行う。

②実施内容

基礎調査として、自治体が本システムの導入するにあたり、その市場環境について調査・分析を行う。調査は、在留外国人の動向、国・地方自治体の外国人定住化に関する施策等について、文献、Web 等のオープン情報の収集・整理を図る。

③成果

- 音声翻訳システムの潜在的な導入候補自治体を把握するための自治体マップを作成した（在留外国人等数の定量情報と自治体ヒアリングを踏まえた定性的情報からなる整理表）