

法務省の発表によると、在留外国人数は平成29年末に256万人を超え、過去最高を記録しました。また、日本政府観光局の統計では、同年の訪日観光客数が2,869万人を超え過去最多となり、日本を訪れる外国人も今後、増加するものと考えられます。

このような背景から、自治体の業務においても外国の方との円滑な対応が求められる場面が増えてきています。

今回は、昨今、自治体において実証・実用化が始まっている「音声翻訳技術」について、国立研究開発法人情報通信研究機構の河井恒氏にご寄稿いただきました。

実用化が始まった音声翻訳技術

国立研究開発法人情報通信研究機構
先進的音声翻訳研究開発推進センター
先進的音声技術研究室 室長 河井 恒



音声翻訳アプリは救世主か？

ここ数年、スマートフォン用の翻訳アプリや携帯型翻訳機を報道、広告などで目にする機会が増えている。日々の業務の中で外国語対応を課題としている人の中には、そのような製品で業務を効率化できないかと考える人は多いだろう。

答えを先に言うと、

- ・ 救急搬送用、自治体窓口業務用など用途に特化したアプリを使う。
- ・ 役に立つ場面とそうでない場面があることを理解して使う。
- ・ アプリの弱点をカバーして使いこなせるように、普段から練習をしておく。

以上に留意すれば、必ず役に立つはずである。

国立研究開発法人情報通信研究機構(NICT)が実証実験のために無料公開している音声翻訳アプリVoiceTraTMは、スマートフォンにインストールすると誰でも簡単に使えるので、是非試していただきたい。

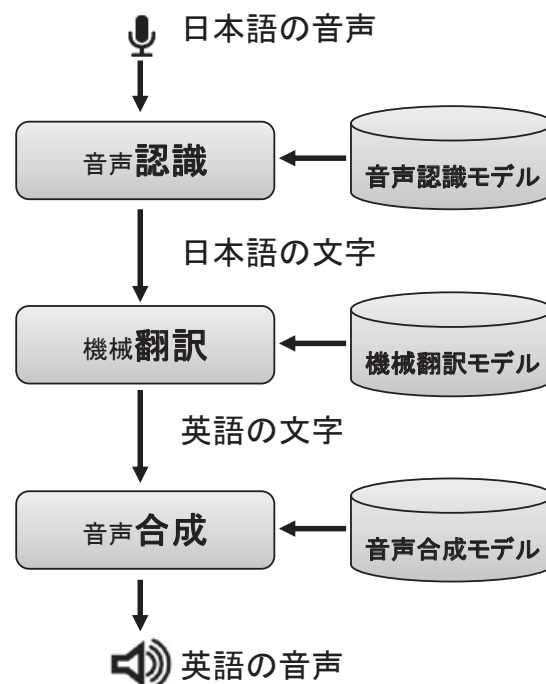
音声翻訳＝認識×翻訳×合成

音声翻訳アプリは、①人間が話した音声を変換する音声認識、②ある言語（文字）を別の言語（文字）に変換する機械翻訳、③翻訳結果（文字）を音声に変換する音声合成、の3要素を直列につないだものである(図1)。音声認識で誤りが生じると、正しい翻訳は期

待できないので、3つの要素を独立に行うのに比べると、音声翻訳は難しい問題である。VoiceTraTMでは、このほかに逆翻訳という機能を付けている。これは、いったん相手先言語に翻訳した結果を再度自分側の言語に翻訳する機能であり、自分が話した意図と異なる結果になっていれば、誤訳と推定して別の言い方で再試行できる。

図1の中でモデルとあるのは、辞書と文法書を合わせたようなものであり、大量の音声、文字、対訳の事例（コーパスと言う）から膨大な計算を行って自動生成される。

図1 音声翻訳システムの仕組み(日→英翻訳の場合)



研究開始後30年でようやく表舞台に

日本における最初期の音声翻訳研究の1つが、1986年に京都府、大阪府、奈良県が境を接する地域にある「けいはんな学研都市」の国際電気通信基礎技術研究所（ATR）で始まった（図2）。当時の性能は、あらかじめ厳選したシンプルな文を丁寧に発話すると数十秒後に翻訳結果が出る程度であり、当時大学院生であった筆者は、自分が現役の間に音声翻訳サービスが実用化されるとは思ってもよらなかった。

○コーパスベース方式により急速に精度向上

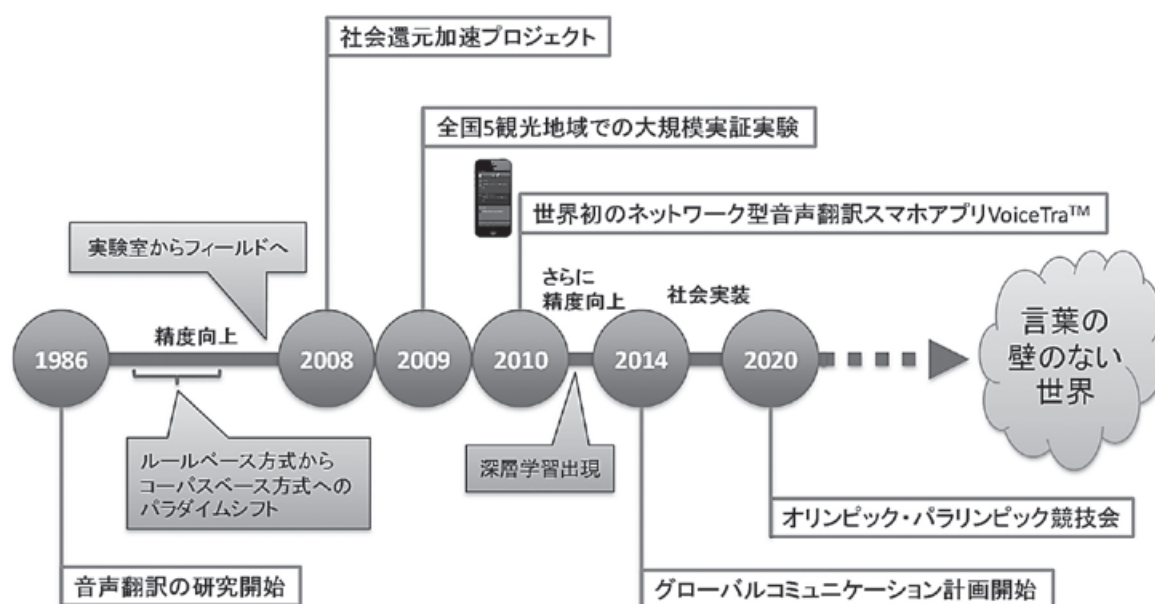
音声翻訳技術における第1のパラダイムシフトが2000年前後に起こった。機械翻訳の基本方式が、それまでのルールベース方式（翻訳元言語の文を翻訳先言語の文に変換する詳細な規則をプログラム化する方式）から、大量の対訳事例から統計的な翻訳モデルを自動的に作成するコーパスベース方式に移行した。コーパスベース方式で研究が始まっていた音声認識と相まって、これ以降音声翻訳精度が大きく向上し、2009年頃には実験室から持ち出しでも使える可能性が見えてきた。この頃音声翻訳の研究開発業務がATRからNICTに

移管された。折しもマイク、スピーカー、表示装置を備え、高速移動体通信網に接続された装置、すなわちスマートフォンが普及し始めていたことから、クラウド型の音声翻訳アプリが現実的となっていた。これらを背景として、2009年、総務省の委託事業として全国5カ所の観光地域で実証実験が行われた。結果としてわかったことは、ペンションでの団欒など、正確性、迅速性への要求レベルが高くない場面では有用である、つまり、この時点では実用化は時期尚早ということであった。その後、NICTは実証実験で得たデータと知見をもとに音声翻訳の精度を改善し、クラウド型音声翻訳アプリVoiceTra™を開発し、2010年7月に世界に先駆けて無料で一般公開した。

○ディープニューラルネットにより実用性へ

第2のパラダイムシフトは、2012年頃始まった。まず、音声認識にディープニューラルネット技術（人間や動物の層状の脳神経回路をモデル化したもののうち、特に多層構造のもの）が導入されたことにより、音声認識誤りが3割程度削減され、改善を体感できた。同じ頃、2020年のオリンピック・パラリンピックの開催地が東京に決定したことを弾みに、総務省

図2 NICTにおける音声翻訳技術の研究開発の歩み



が、世界の「言葉の壁」をなくし、グローバルで自由な交流を実現することを目標とする政策「グローバルコミュニケーション計画」を発表し、音声翻訳技術の一層の強化と社会実装をNICTとともに推進することとなった。

機械翻訳へのディープニューラルネットの導入（NMT: Neural Machine Translation）も、音声認識にやや遅れて学会レベルでは研究が進められていたが、2016年になりGoogle社がいち早く翻訳アプリに組み込み、一般公開した。NICTでも鋭意研究開発を進めており、一部の言語については試験的に実証実験システムで運用中である。旅行会話、生活会話など十分な量の対訳コーパスが利用可能な分野では、NMTの翻訳誤り率（意味が通じない程度の誤訳となる文の比率）は、従来手法の2/3から1/2である。従来手法では、誤訳が生じないように注意して使う必要があったが、NMTでは、特に配慮なく使用しても誤訳となることは少なく、安心して使えるようになった。

救急VoiceTraは人口普及率50%に

NICTは、総務省消防庁消防研究センターと共同で、札幌市消防局の協力を得て救急搬送業務に音声翻訳アプリを導入すべく研究開発と実証実験を行った。試行錯誤の結果、VoiceTraTMに定型文機能を実装し、救急現場で使用頻度の高い16言語46文の対訳を定型文として登録したものを「救急VoiceTra」として消防庁を通じて全国の消防本部に展開した。図3は、定型文を使用して翻訳を行う際のアプリ画面の例である。札幌市が2017年1月31日から運用を開始したのを皮切りに、2018年7月1日現在で728本部中345本部が利用開始済み（管轄エリアの人口カバー率48%）、33本部が利用開始予定となっている。

消防研究センターが各消防本部からの報告にもとづいて集計した2017年4月から12月の利

図3 救急VoiceTraのアプリ画面例



用実績は、1,187件であった。音声翻訳サーバーへのアクセス記録にもとづいて集計した2017年4月から2018年6月までの翻訳件数は15.3万件であり、入力手段の内訳は、音声が70.2%、定型文が21.2%、文字入力が8.6%であった。

○救急搬送に音声翻訳が適する4つの理由

救急VoiceTraが短期間で全国に普及した理由としては、以下のことが考えられる。

- ①外国人を搬送する機会は観光地を有する自治体などで増えてはいるが、多くの消防本部では、頻度はそれほど高くないため、外国語会話習得の費用対効果が不良である。
- ②救急搬送の目的は、傷病者を医師のもとに搬送することであり、診断・治療を行うのは医師であるので、多少の誤訳があったとしても、搬送時間を短縮できれば、傷病者の利益になる。
- ③救急搬送は、業務フローがおおむね定型化し、傷病者への質問もおおよそ決まってい

るため、必要十分な定型文集を整備することが比較的容易である一方、傷病者本人や付き添い者との会話で自由文入力による翻訳が必要になることもある。

- ④音声翻訳アプリは、騒音の大きい場所では口をマイクに近づけて話す、シンプルかつ明確な表現をするなど使い方の工夫によって高い翻訳精度を発揮する。救急隊員は、待機中は訓練に努めており、その中で音声翻訳アプリの習熟訓練を行えることから、実際の現場でも音声翻訳アプリを慌てずに利用できる。

自治体窓口への応用に向けて

NICTでは、凸版印刷株式会社および株式会社フィートに委託して自治体窓口業務に対応した音声翻訳システムの開発とサービス化に向けた研究を行っている。

○住民登録、国保、子育て、年金

まず市役所等の窓口において外国人居住者に対して行われている業務を分類・分析し、音声翻訳システム導入により業務改善が見込まれる業務として、住民登録・国保業務および子育て・年金業務を選定した。

次に、窓口担当職員へのヒヤリングおよび模擬会話を行って窓口で交わされる会話事例を収集・作成・翻訳し、対訳コーパスとした。並行して、自治体窓口での会話で使われる用語を収集し、対訳辞書を作成した。

○専用アプリを開発して実証実験を実施中

自治体窓口会話の対訳コーパスを用いて翻訳精度を改善するとともに、カウンターを挟んだ会話に適した画面構成をデザインし（図4）、行政用語の外国語解説文の表示機能を持たせるなど、自治体窓口での手続きに適したユーザインタフェースを有する専用の音声翻訳アプリを開発した。これを用いて東京都板橋区、群馬県前橋市、神奈川県綾瀬市、長野

図4 自治体窓口のカウンターを挟んだ会話向けにデザインしたアプリ画面



県塩尻市、山梨県甲府市等で実証実験を行っている。

おわりに

自治体窓口向け音声翻訳の研究では、小中学校教育の現場への展開も検討されている。また、多くの警察本部でVoiceTra™を交番等の業務に利用する実証実験が行われている。沖縄および岡山では、すでに本格導入されており、他の複数の警察本部等でも導入に向けて準備が進められている。一方、鉄道の駅での案内業務への適用など、民間部門への展開も進んでいる。

今後NICTでは、音声翻訳精度の改善、対応言語の拡大など引き続き技術の強化をはかる一方、音声翻訳技術の本格的な社会実装を目指して努力を続ける予定である。

著者略歴

河井 恒（かわい・ひさし）

1984年東京大学工学部電気工学科卒業、1989年東京大学大学院工学系研究科電子工学専門課程修了。工学博士。1989年4月KDD研究所（現KDDI研究所）に入社、現在に至る。この間、2000～2004年9月ATR勤務、2009年4月～2012年3月および2014年10月より現在までNICTに勤務。先進的音声翻訳研究開発推進センター先進的音声技術研究室室長。音声の認識・合成、音響信号処理、音声翻訳の研究開発に従事。