

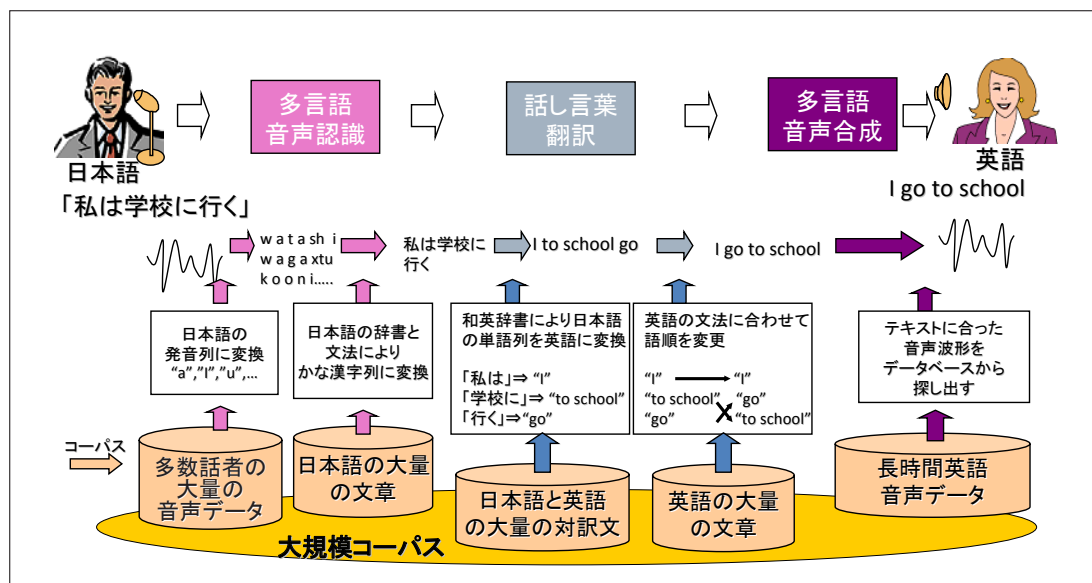
言葉の壁を越える音声翻訳技術

多国間の言語の壁はお互いの意志疎通にとって、今なお大きな問題であり、場合によっては、より深刻な問題となっている。話した言葉をその場で相手の言語に翻訳する自動音声翻訳技術を確立することは、経済活動・多言語観光ビジネス・外国人滞在者へのサービス向上を通じて、我が国のグローバル化にも大きく貢献する。

音声翻訳は、音声認識技術・テキスト翻訳技術・音声合成技術の統合技術であり、近年、それぞれの技術の発展とともにデータベースが拡充し、音声翻訳の精度は飛躍的に向上している。今世紀に入り、種々の言語的補助情報を付与した音声やテキストのデータベースである「コーパス」に基づく技術研究が急速に進み、現在では、日常の旅行会話に対し、一文ごとに日英中の双方向逐次音声翻訳を実現できる段階に達した。しかし、音声翻訳は発話者への依存性、表現の多様性が大きく、新しい語彙や概念が社会変化に応じ次々と創造されるため、特に多国間の音声翻訳には多くの研究課題が残されている。

音声翻訳技術は、基本技術の研究開発も大事だが、実世界でのコーパス収集と自動学習が不可欠であり、使われてこそ性能があがるという側面がある。今後、種々の場において、フィールド実験を重ねながら、可能なところから導入していくことが重要である。また、多くの国の言葉の壁を越えるべく、多言語多国間の連携スキームを進めていく必要がある。音声翻訳は日本が進んでおり、標準化などの場では他国をリードできる技術である。

音声翻訳のメカニズム



科学技術動向研究センターにて作成

言葉の壁を越える音声翻訳技術

中村 哲
客員研究官

1 はじめに

異なる言語を話す人々のコミュニケーションを可能にすることは、経済活動のグローバル化やボーダーレス化に伴い極めて重要になっている。話した言葉をそのまま相手の言語に自動で翻訳する音声翻訳技術は、人類にとって長年の夢であり、世界を変える10の技術の中のひとつとしても選ばれている。特に日本では、地理的条件や日本語の孤立性などに起因する外国語習得の困難さがあり、日常の話し言葉を自動翻訳する音声翻訳システムに対する期待が大きい。この音声翻訳技術は、ますます国際化する日本人と日本という国にとって恩恵の大きい技術である。

自動音声翻訳技術は、音声を認識する技術、認識した話し言葉を

翻訳する技術、相手の言語で音声を合成する技術の3つで構成される。最近の技術の発達により、日本語、英語、中国語の旅行会話の自動音声翻訳が実用可能なレベルまで到達しており、文が短く単純な会話の一文ずつを逐次翻訳できるまでとなった(日英翻訳では、TOEICで600点以上)。

しかし、より多くの言語への対応や、実用上必要となる場所名や人名などの種々の固有名詞の自動獲得など課題も多く、今なお実用化への挑戦が続く。さらには、「五月雨式」に音声翻訳する同時通訳のような技術の確立も望まれる。また、音声翻訳に使われる個々の技術は、音声情報検索・対話型ナビゲーション・口述筆記と要約・アー

カイビングなどにも幅広く適用が可能な技術であり、その新しい使い方にも期待される。

本レポートでは、まず音声翻訳技術の意義を確認し、これまでの研究開発状況や自動翻訳技術の歴史について概観する。さらに、音声翻訳システムの構成や現状のシステム性能について述べる。また、世界の研究開発動向について触れ、音声翻訳技術の実用化についても述べ、アジア言語への展開および接続標準化活動についても紹介する。最後に、音声翻訳技術の課題と展望をまとめ、音声翻訳技術を推進するにあたっての課題を解決すべき方策について提言する。

2 音声翻訳技術の歴史

2 - 1

音声翻訳研究の意義とこれまでの歴史

音声翻訳はある言語で発話された音声をも別の言語の音声に翻訳して出力する技術である。音声翻訳技術の意義は、異なる言語を話す世界の人々とのコミュニケーションを

可能とし、グローバルビジネスや異文化交流、ランゲージデバインドの解消などを実現することである。音声翻訳技術の実現が、人類にもたらす科学的価値、文化的価値、経済的価値は、非常に大きいといえる。An MIT Enterprise Technology Review 誌の2004年2月号の特集「10 Emerging Technologies That Will Change Your World」において

は、世界を変える10の技術のひとつとしてUniversal Translationが取り上げられており、種々の翻訳技術の中でも特に音声翻訳技術に焦点をあてて紹介されている。

音声翻訳が初めて注目されたのは1983年の世界電気通信展示会(テレコム'83)であり、日本電気株式会社(以下NECと表記)がコンセプト展示として音声翻訳のデ

モを行い注目を集めた。この後、音声翻訳実現のためには長期的な基礎研究を行う必要があるという認識のもとに、1986年に(株)国際電気通信基礎技術研究所(以下ATRと表記)が設立、音声翻訳の研究プロジェクトが開始され、国内外から様々な研究機関の音声言語研究者が参画することになった¹⁾。1993年には、ATR、カーネギーメロン大学(以下CMUと表記)、シーメンス社による世界3地点を結んだ音声翻訳実験が行われた。ATRのプロジェクト開始の後、世界でも音声翻訳のプロジェクトが開始された。ドイツでは「Verbmobil」プロジェクト、欧州共同体で「Nespole」、TC-Star」、米国では「TransTac」、GALE」プロジェクトが開始された。この中でも「GALE」プロジェクトは、2006年からアラビア語と中国語から英語へと自動翻訳するためのプロジェクトであり、これまで人間が行っていた多言語重要情報の抽出作業の自動化を目的にしており、バッチ型テキスト出力のシステムとして構成されている。これに対し、ATRやNECは、これまで対面・非対面のリアルタイム異言語コミュニケーションを達成する音声翻訳を目標にしており、音声から音声へのオンライン翻訳が前提となっており、処理の即時性が重要なファクターとなっている。

音声翻訳は、音声認識・自動翻訳・音声合成の3つのコンポーネントとそれらを統合する技術から構成され、それぞれの技術の困難さが存在する。特に、話し言葉の音声を認識し、翻訳する必要があるが、話し言葉の文には非文法的な口語表現が含まれること、疑問符や感嘆符、引用符などの記号は含まれないことから、テキスト翻訳よりも翻訳が困難である。また、音声の誤認識も重大な翻訳誤りを起こす。したがって、最初からあらゆる会話を対象とするのではなく、特定の比較的容易な会話を対象を絞り込むことにより、精度を利用可能なレベルまで向上させるという開発手法がとられた。図表1に音声翻訳技術の変遷を示す。比較的容易な翻訳からだんだんと高度な翻訳へと研究開発が進められ、対象とする会話は、会議予約、ホテル予約、旅行会話へと順を追って進められてきたが、今後はさらに多様な日常会話や高度なビジネス会話へと対象分野を拡げていく必要がある。

2 - 2

自動翻訳の歴史

3つのコンポーネントのうち、

テキスト翻訳技術の最近の進歩が、自動音声翻訳技術の実現に大きな貢献を果たしている。このテキスト翻訳技術の研究に関しては、半世紀を越える長い歴史がある。

最初のコンピュータが誕生して間もない1946年に、米国の科学技術政策に大きな影響力を持っていたロックフェラー財団のW.ウィーバーがテキストの自動翻訳技術の研究を提唱している。そして、1953年に、初めてIBM社が開発した商用コンピュータ701を利用して、ジョージタウン大学とIBM社が自動翻訳の共同研究を開始した。1954年には、このコンピュータで世界初の自動翻訳システムを構築し、露英翻訳が可能なことを実証した。このシステムは250語の辞書と6個の規則からなる極めて限定的な翻訳能力しかなかったものの、社会に与えた衝撃は大きく、当時の人々はすぐにでも言葉の壁は解消すると感じた。また、この後、米国政府はスプートニク・ショックへの対応の一環として、自動翻訳の研究にも2千万ドルもの資金を投入している。

ところが、1965年に、自動言語処理諮問委員会(ALPAC)は重大な報告書を米国科学アカデミーに提出した。自動翻訳は当面実用化できないので、むしろ基盤となる言語理論や言語理解の研究を進めるべきだと

図表1 音声翻訳に関わる研究開発の推移

研究フェーズ	実現性の確認 80年代	技術の拡張 90年代	実利用への挑戦 2000年代
対象分野	単純予約 (ATR-phase1)	予約とスケジュールリング (ATR-phase2, Verbmobil)	・ 日常旅行会話 (ATR-phase3) ・ 基調講演翻訳 (TC-Star) ・ 軍用会話 (TranTac) ・ 情報収集 (Gale)
言語的特徴	文法的に正しい表現	状況依存したり、非文等を含む日常的な表現	広範囲な話題や固有名詞を含む表現
音響的特徴	明瞭な発声	不明瞭な発声	雑音を含む発声
翻訳方式	規則に基づく翻訳 人工的中間言語翻訳	用例に基づく翻訳 英語中間言語翻訳	統計に基づく翻訳 多対多言語直接翻訳

(注: ATR-phase1: 1986~1992年, ATR-phase2: 1993~1999年, ATR-phase3: 2000~2005年, その他のプロジェクト略名は本文参照。)
科学技術動向研究センターにて作成

いう趣旨の報告書であった。以後、米国においては、自動翻訳に予算はつかなくなり、研究は基礎に向き、意味や理解というキーワードが重視された。その中では、1970年のヴィノグラードの世界知識を使った言語理解が有名な成果である。しかし、このような研究は、基礎となる知識ベースの不足から、汎用で実用的な意味での自動翻訳の性能向上には直接には結びつかなかった。

日本では1980年代に、ルールベース翻訳、用例ベース翻訳、統計ベース翻訳という3つの大きな技術革新の波が訪れた。日本では、1982年に科学技術庁の科学技術文献の要約を自動翻訳するプロジェクト（Muと呼ばれる）が成功した。この結果によって、辞書とルール（解析文法規則、変換規則、生成文法規則）に基づくルールベースの自動翻訳の研究開発が普及し始めた。ベンチャーのブラビス社による商用翻訳ソフトが発売された。これを機に、富士通(株)、(株)東芝、NEC、沖電気工業(株)など大手IT各社の自動翻訳システムも商用化された。現在までに世界で商用化されたパッケージソフトの全て、およびWEBで公開されているソフトのほとんどは、このルールベース技術を基本としたものとなっている。翻訳品質の改善に専門用語辞書の充実が有効で

あったために、地道な努力が積み重ねられ、辞書の規模は数万から数百万まで拡大した。

一方、1981年に長尾真京都大学教授（当時）が人間の行う翻訳過程にヒントを得て、入力文に類似した文とその翻訳（併せて対訳用例と呼ぶ）を活用する用例ベース翻訳方式を提唱した。この用例ベース翻訳が、1990年前後に京都大学とATRで行われた研究をきっかけに、二つ目の波として、日本から世界へ広がっていった。この方式は、ルールベースの商用システムに一部取り入れられ、さらに、(独)情報通信研究機構（以下NICTと表記）が中心になり実施している科学技術文献の日中翻訳プロジェクトの基本方式として現在も採用されている。

また、1988年にIBM社が、文法などの知識を排除した純粋に統計的処理と対訳データとを組み合わせた統計ベース翻訳という手法を提唱した。しかし、その論文は難解であり、計算機的能力不足、対訳データの不足、実行方法が特許明細書でしか開示されなかったこと、英語とフランス語のような類縁言語間以外には有効でなかった、などの理由のために、長らく重要視されなかった。しかし、2000年前後に「句に着目した統計ベース翻訳方式」が提案され、対訳データの充実や

計算機能力の向上を追い風に、第3の大波となり、現在では、統計ベース翻訳の研究に関する論文が9割を占めるに至っている。この研究領域がまだ伸びるのかは現在は判断しにくい状況である。

今はちょうど、上記の3つの大波が重なっている。ルールベース、用例ベース、統計ベースの自動翻訳の長所と短所とが次第に分かってきた。どれか単一の方式ではなく、3方式をうまく融合できたときに最高の性能が実現できるというのが現時点での見解である。しかし、3つの方式には共通の課題もあり、3方式とも文単位の翻訳である。文脈情報が利用できていない。すなわち、前後の文章の関係を使っておらず、結束性を担保できていない。特に、統計翻訳は入力文の意味の解釈を行わずに自動翻訳しているため、ナンセンスな訳文が生じることもある。

用例ベースおよび統計ベースを用いた手法を「コーパスベース翻訳手法」と呼ぶが、本稿では、主として統計ベースを用いた手法を紹介する。コーパスとは、読み、すなわち品詞情報や係り受け情報などの言語的な補助情報を付与したテキストのデータベースのことである。次章以降の記述は、主にコーパスベース翻訳手法についての記述である。

3 音声翻訳技術の概要と性能

3 - 1

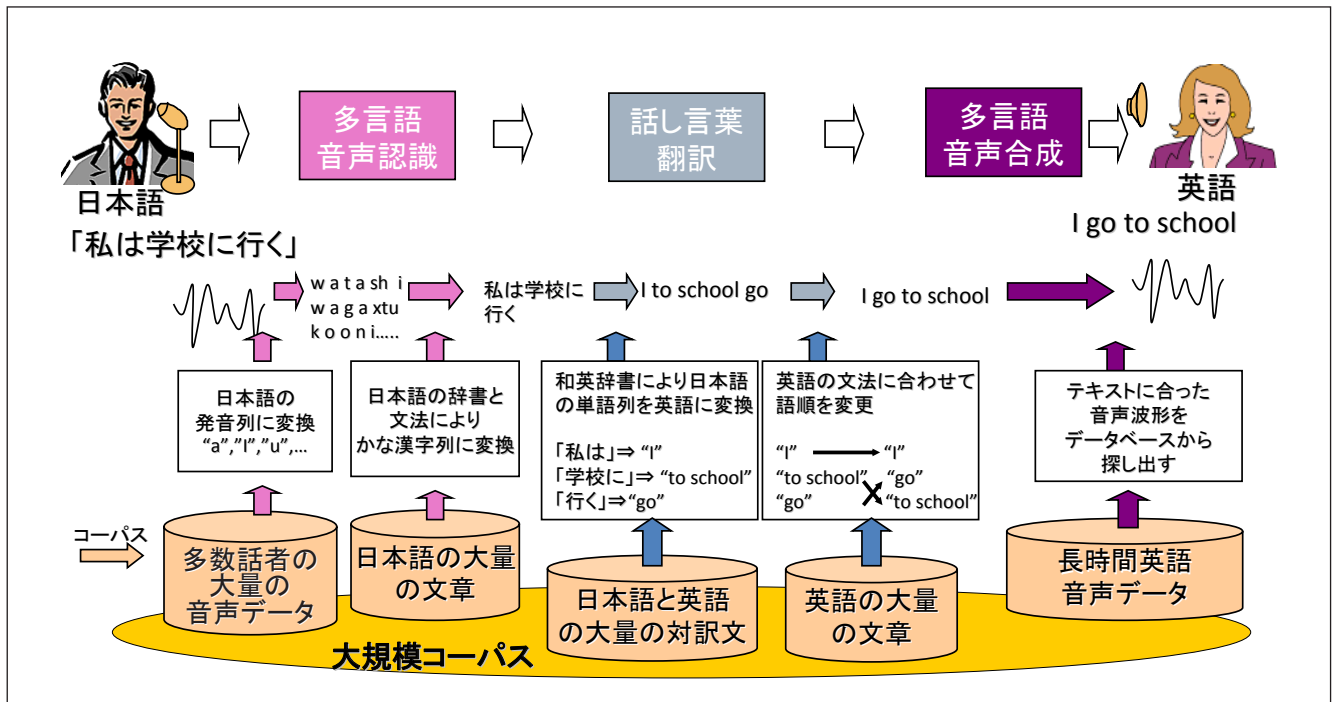
多言語音声翻訳処理の構成

音声翻訳システムの全体構成を図表2に示す。図表2は、発話者が話した日本語音声認識されて日本語文章となり、さらに英語文章に翻訳され、英語音声に合成さ

れる例を表している。多言語音声認識のモジュールで、多くの話者の多量の音声データから構成された音のモデル（モデルは音声を構成する音素ごとに構成される）と、入力音声との照合が行われて、カタカナ表記の音素列に変換される。次に、この音素の列は、日本語のかな漢字で表記される単語列確率を最大化するように変換される。

この変換では、日本語の大量のテキストから学習された、3つ組の単語列の生起確率をもとに、日本語として適切な単語列を確率付きで求める。これをさらに話し言葉翻訳のモジュールで、日本語の単語列が対応する英語の適切な単語との入れ替え、および語順の入れ替えが行われる。日本語の単語列を対応する英語の単語列に入れ替えるために、

図表2 音声翻訳システムのメカニズム



科学技術動向研究センターにて作成

同じ意味を持つ日本語と英語の対訳文から学習された翻訳モデルを用いて単語の入れ替えを行う。次に、語順を英語に合わせるため、大量の英語のテキストから学習された、3つ組の単語列の生起確率から英語として適切な単語列を求める。それを音声合成部に送る。音声合成部では、英語の単語列にあわせて発音、イントネーションパターンを推定し、それにあう波形を長時間音声データベースから選択、接続し、高品質な音声合成を行う。大量の音声コーパスを基に、統計モデルと機械学習を用いる音声認識、音声合成手法を「コーパスベース音声認識・音声合成手法」と呼ぶ。

ATRが開発した音声翻訳システム^{1,2)}では、旅行会話の音声翻訳を実現するため、一般の口語旅行会話コーパスが収集された。これまでに、旅行会話基本表現集BTEC (Basic Travel Expression Corpus)として、日英100万文対、日中・日韓それぞれ50万文対が構築されている。この対訳文データは、多言語の旅行会話コーパスとしては、世界最大規模のものである。このコーパスに格納されている文章は

英語の単語数の意味で平均7単語の長さであり、挨拶・トラブル・買い物・移動・宿泊・観光・レストラン・コミュニケーション・空港・ビジネスなどの日常旅行会話を網羅している。日本語1文に対して、話し言葉の英語対訳文の例を示す。

日本語:「窓をあけてもいいですか」
英語:

1. may i open the window
2. ok if i open the window
3. can i open the window
4. could we crack the window
5. is it okay if i open the window
6. would you mind if i opened the window
7. is it okay to open the window
8. do you mind if i open the window
9. would it be all right to open the window
10. i'd like to open the window

この例のように、音声翻訳で取り扱う文では、主語、固有名詞の一文字目が大文字にならず、疑問

文でも疑問詞がつかない。また、非常に口語的な表現も取り扱う必要がある。

BTECのほかに、MAD (Machine Aided Data)と呼ばれる音声翻訳システムを介した、実環境下での対話を記録した約10000発話のコーパスの収集データ、および、FED (Field Experiment Data)と呼ばれる、2004年12月から2005年1月にかけての計5日間に渡り、大阪府の協力を得て、関西国際空港において公開実験を行い、外国人(英語話者39人、中国語話者36人)と観光案内所のガイドが、音声翻訳システムを介して行った会話を合計約2000発話収集したデータを用いて評価を行った。

3 - 2

人間の音声翻訳能力との比較調査

音声翻訳の正確さの評価は、原理的には非常に困難である。音声合成部を評価に入れない場合には、音声翻訳の評価法は、いくつかの評価文をシステムに与え、この出力が

どの程度の品質かを評価する点で、テキスト自動翻訳の評価法と基本的には同じとなる。ただし、音声翻訳の場合は、評価文が文字列ではなく音声で与えられる。

翻訳品質の評価法には人手で5段階評価などを行う主観評価法やあらかじめ参照訳を用意してこの参照訳とシステム出力との類似度で評価する自動評価法が用いられる。後者はBLEU、NIST、WER (Word Error Rate) などの評価尺度が提案され、最近はこちらが広く用いられるようになってきた⁴⁾。これらの結果は単なる数値で、2つのシステムを比較することはできるが、スコアを達成したシステムが現実世界でどの程度の実力を持つのかという問いには答えられない。

この問題に対して、翻訳システムの能力が人間でいうとTOEICスコア何点に対応するかを推定する方法も提案されている⁵⁾。まず、TOEICスコアが既知の複数の日本語母語話者（ここではTOEIC被験者と呼ぶ）に、評価用の日本語文を聞かせて、英文に音声翻訳させ、次に各TOEIC被験者の翻訳文と自動翻訳システムの出力とを対にして、両者を日英バイリンガルの評価者が比較する。試験文全体の中で被験者の翻訳の方が優れている文の割合を示す被験者勝率を計算する。全ての被験者に対する被験者勝率の計算が完了した段階で、回帰分析により自動音声翻訳システムのTOEICスコアを計算する。性能をTOEICスコアに換算すると、図表3のようになる。基本旅行会話のような比較的短く表現も簡単なもの（BTEC）であれば、ほぼ正解に近い性能が出ているが、音声翻訳システムを介して行った会話に現れるような文（MAD、FED）では、TOEIC 600 点の日本人と同等の性能となっている。

さらに、長文やめったに現れない表現を含む複雑な文に対しては、著しく性能が劣化する。未だ性能向上のための余地が残されている。

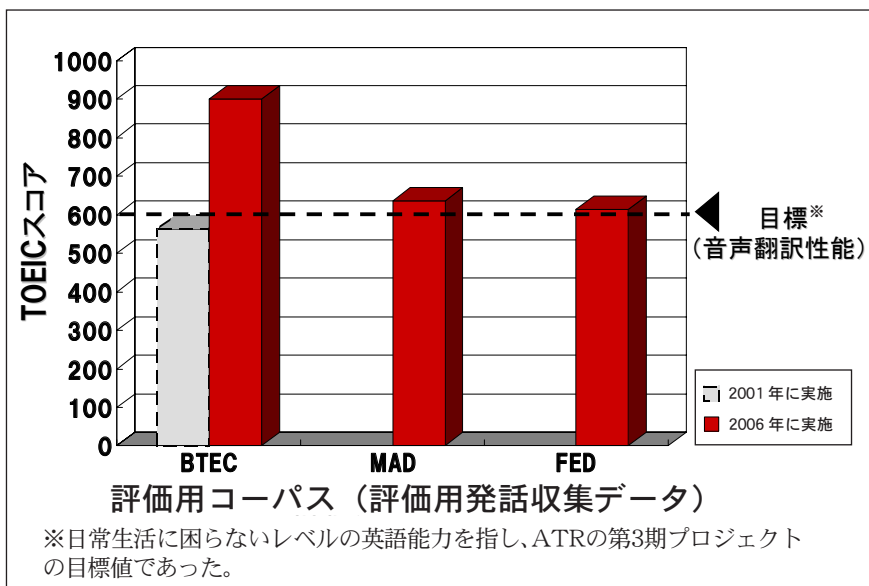
音声翻訳機を用いたフィールド実験

システム手帳大のスタンドアロン型音声翻訳機により、音声翻訳機を介した情報伝達の特徴や音声翻訳機の使用性の評価を目的としたフィールド実験が、2007年7月30日から8月24日にかけて、京都市内の繁華街で実施された⁶⁾。フィールド実験では次のように、被験者への制約をできるだけ排除した設定とする。①移

動・買物・飲食などの現実の旅行場面における音声翻訳機利用時の表現の多様性を収集するため、対話相手は事前に準備しない。②対話の目的はあらかじめ与えるものの、具体的な移動先や購入品の固有名詞に制限を加えない。③対話の流れによって被験者が課題を自由に変えることを許容する。④課題に応じて場所を適宜移動できる。⑤一対話あたりの制限時間を設けない。

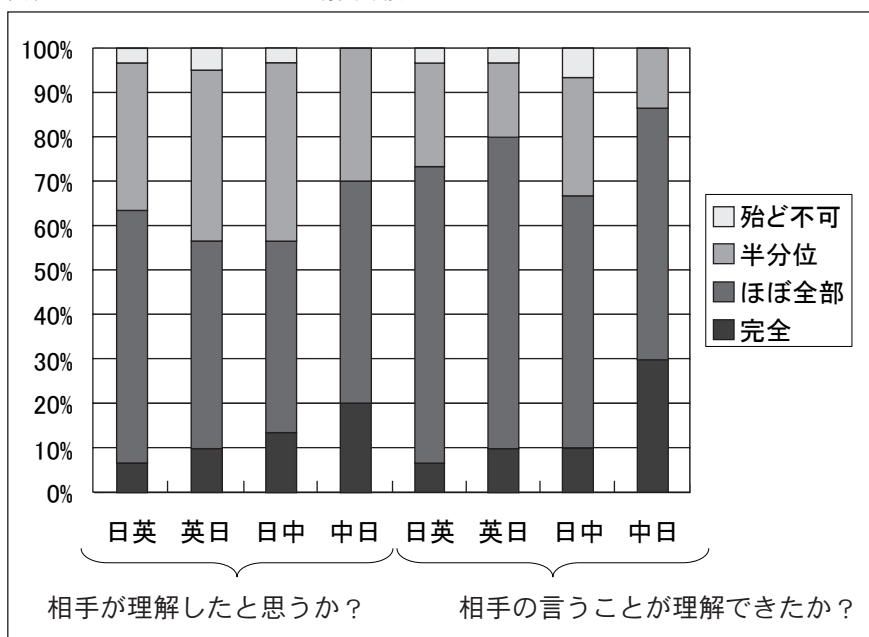
移動であれば移動先に関する情報が得られた場合、あるいは実際に移動できた場合、買物や飲食で

図表3 音声翻訳の正確さをTOEICスコアで評価した例



出典：参考文献¹⁾

図表4 アンケートに基づく理解度評価



出典：参考文献⁶⁾

あれば商品の購入や飲食が完了し領収証を受領した場合に、目的達成とした。

実験では、音声認識率・対話の応答率・翻訳率を定量的に評価し

たほか、アンケートに基づく理解度評価も行った。図表4に示されるように、英語ネイティブ話者50人の理解度評価では、相手がほぼ全部理解したと回答した割合は約

80%に達し、相手の言うことが半分以上理解できた割合は80%を超えた。この結果は、音声翻訳機を介したコミュニケーションが十分成立しうることを示唆している。

4 世界の研究開発動向

音声翻訳の技術進展を強力に後押ししたものに「評価型国際ワークショップ」がある。評価型国際ワークショップとは一種のコンテストであり、主催者が共通のデータを提供し、ワークショップに参加する研究機関に競争的にシステムを作成させ、各システムを定量的に評価するものである。評価結果から、提案された様々なアルゴリズムの優劣が定まり、優秀なアルゴリズムが以降の研究開発で広く採用されるようになる。これにより、世界の研究機関が競争的かつ協調的に研究することができ、効率的な研究が推進されてきた。ここでは、評価型国際ワークショップの代表例として、IWSLTとGALEを取り上げ、さらに、評価型ワークショップによる競争的研究スタイルを支える自動評価技術について述べる。

(a) IWSLT ワークショップ⁷⁾
(IWSLT: International Workshop on Spoken Language Translation)

は、日本のATR、米国のCMU、イタリアの科学技術研究所（以下IRSTと表記）、中国の中国科学院（以下CASと表記）、韓国の電気通信研究所（以下ETRIと表記）などが組織した音声翻訳研究の国際的なコンソーシアムであるC-STARが主催するもので、2004年から開催されている。毎年、参加機関数も増え、現在では世界の音声翻訳研究の中核的イベントとなっている。日本語、中国語、スペイン語、イタリア語等の言語から英語への旅行会話の音声翻訳を対象としている。対象が旅行会話という平和利用であること、コンパクトなタスクのためかなり精度の高い翻訳が可能であることなどがIWSLTの特徴である。

(b) GALE プロジェクト⁸⁾ (GALE: Global Autonomous Language Exploitation) は、米国防総省の高等研究計画局 (DARPA) のプロジェクトであり、公開されずクロードで行われる。年間50MUSドルの資

金が入入されている。アラビア語と中国語のテキストおよび音声を英語に翻訳し、情報を抽出することを目的としている。多数の機関が3チームに分かれ性能を競い合う。目標値が与えられた単年度で運営され、毎年、性能が外部機関により評価される。現在の米国では、自動翻訳研究はこのDARPAの予算に強く依存しており、米国防総省の意向が強く反映される。

これらのワークショップにおいては、翻訳の品質評価法が大きな議論のポイントとなっている。翻訳品質は、流暢さ (fluency) や適切さ (adequacy) などの様々な観点があり、高度に知的な作業と考えられてきた。近年提案されたBLEUと呼ばれる評価手法は、人間による主観評価との相関が高く自動で計算できるため、時間も費用もかからず、システムの開発と評価を短いサイクルで繰り返すことを可能にし、翻訳の研究開発に大幅な効率向上をもたらした⁴⁾。

5 音声翻訳技術の実用化

計算機の処理能力の向上とメモリの大規模化、および、ネットワークの普及により、携帯できる音声翻訳機器の実装が可能になり始めている。小型のハードウェアへ実装する単体方式と、携帯電話などの端末とネットワークを介して高性能サーバと接続して実装する分散方式の開発が進んでいる。

単体としては、重量・大きさ・バッテリー寿命などからパソコンを携帯して利用するのは現実には困難であること、一方で無線などインフラのない状況での利用も需要が見込めることを考慮して、音声翻訳機能を内蔵した専用の携帯機器での実用化に向けた努力がなされている。2006年に、NECは世

界で初めて携帯端末 (400MHzのMPUと64MBのRAMというハードウェア) 上に日英音声翻訳を搭載した製品を開発した。

一方、携帯電話・ネットワーク・サーバを利用した分散型実装については、2007年11月にドコモ905iシリーズの携帯電話向け音声翻訳システムがATRにより開発

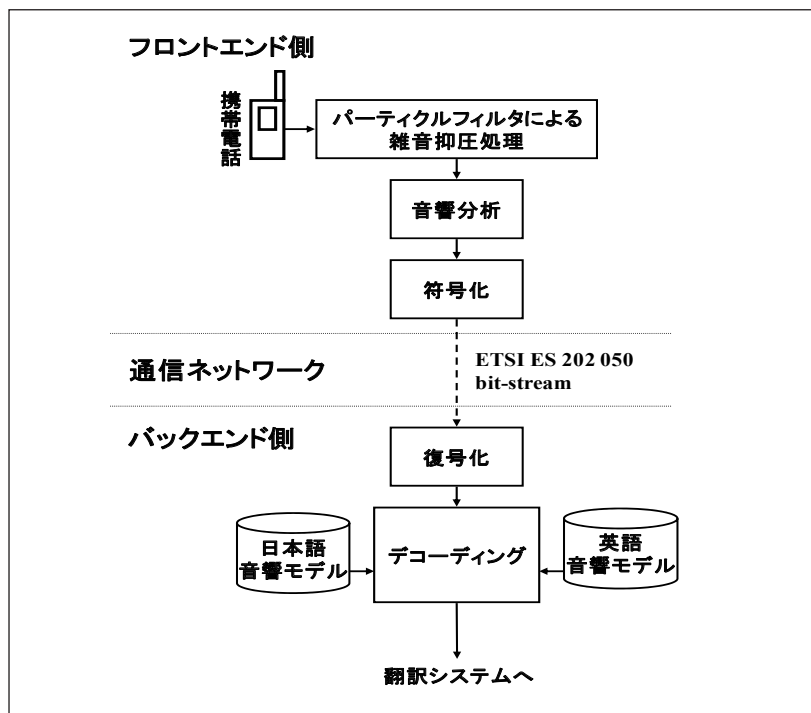
された。これは、(株)ATR-Trekよりリリースされた世界初の携帯電話による音声翻訳サービス「しゃべって翻訳」である(図表5参照)。さらに、2008年5月には、ドコモ906iシリーズから日中音声翻訳のサービスも開始されている。図表6に分散型音声翻訳の音声認識部の構造を示す。携帯電話側(フロントエンド)において、雑音抑圧および音響分析、ETSI ES 202 050に準拠した符号化が行われ⁹⁾、bit-streamデータのみが音声認識サーバに送信される。音声認識サーバ側(バックエンド)では、受信したbit-streamを展開し、音声認識および単語信頼度の計算処理が行われる。このようなシステム構造を採用することの利点は、携帯電話の情報処理能力の限界に縛られず、大規模かつ精密な音響モデルや言語モデルが利用可能な点が挙げられる。これらの各々のモデルは携帯電話ではなくサーバ側に存在するため、更新作業が容易であり、つねに最新の状態が維持可能である。2008年6月の時点で累積アクセス数は500万を超えており、すでに多くの利用実績がある。

図表5 世界初の携帯電話による音声翻訳サービス「しゃべって翻訳」の利用シーン



提供：(株)ATR-Trek

図表6 分散型音声翻訳の音声認識部の構造



参考文献^{9, 12)}等を基に科学技術動向研究センターにて作成

6 音声翻訳の多言語化に関わる標準化の状況

音声翻訳技術は、言語の壁を越える技術であり、多くの国の研究者および研究機関が共同研究を進めていくのが望ましい。国際間の共同研究としては、これまでATRとCMUが中心となって組織した国際音声翻訳共同研究コンソーシアムC-STARが活発な活動をしてきた。

一方で、邦人の海外旅行や移住・留学先の多様化、様々な国々からの日本への旅行者・留学者・就労

者の拡大などの変化は、英語圏以外の国々の人々との交流支援手段に対するニーズを高めている。

とりわけ、我が国は、社会経済面でロシアを含めてアジア諸国との幅広い地域的関係強化が進んでおり、草の根レベルの相互理解の増進や経済関係の強化も重要な課題となってきている。アジア諸国との関係は日本にとって今までにないほど重要となっている。したがって、英語だけでなく、中国語・

韓国語・インドネシア語・タイ語・ベトナム語・ロシア語といった、これまで日本で馴染みの薄かった近隣諸国の言語にまで対応できる必要性が生じている。

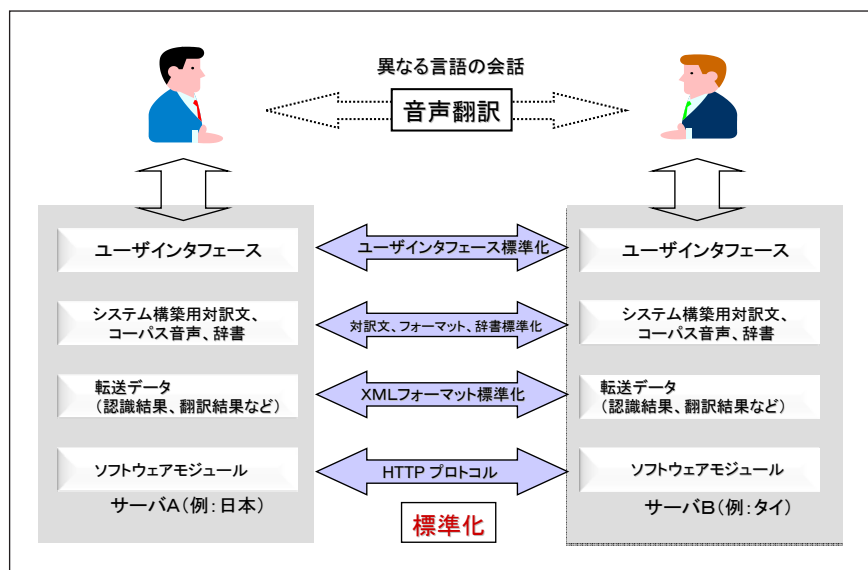
そのような背景で、アジア圏内で言語の壁を越えた音声言語コミュニケーションを実現するための基本インフラを整備する音声翻訳コンソーシアムとして、A-STARが発足した。本コンソーシアムでは、アジア圏における当

該分野の研究機関と共同で、技術の研究開発そのものではなく、研究開発を進めるために不可欠となる対訳文コーパスのフォーマット的设计、アジア圏の言語間での基本対訳文コーパスの设计・収集、音声翻訳のモジュールを国際的に接続するインタフェース、データフォーマット標準化の设计のための国際共同研究体制を確立することを目指している。このコンソーシアムの活動は、科学技術振興調整費事業「アジア科学技術協力の戦略的推進」の委託研究として進められている。この活動はさらに APEC TEL (Telecommunications and Information) のプロジェクトとしても提案、採択されている¹⁰⁾。さらに、音声翻訳のモジュールを接続するインタフェース・データフォーマット標準化については、標準化ドラフトの作成にむけて、アジア圏での通信に関する標準化フォーラムである APT ASTAP (Asia-Pacific Telecommunity Standardization Program) に Expert Group を設置

して活動が行われている¹¹⁾。図表 7 に、これらの活動で検討されている接続標準化のイメージを示す。音声翻訳を構成するモジュールが、インターネット上で接続可能になるようにインタフェース、データフォーマットの標準化を行う。さらに、音声認識、翻訳の辞書の共通化、標準化された対訳

コーパスの収集も必要となる。通信インタフェースとしては WEB ベースの HTTP1.1 による通信を基本とし、アプリケーションの接続におけるデータフォーマットとして音声翻訳用のマークアップ言語 STML (Speech Translation Markup Language) が現在開発中である¹²⁾。

図表 7 音声翻訳標準化のイメージ



出典：参考文献¹²⁾

7 音声翻訳技術の課題と展望

7 - 1

音声翻訳を進展させる上での課題

このように音声翻訳技術は異なる言語を話す人々のコミュニケーションを実現する技術である。しかし、発話者への依存性、特に表現の多様性が大きく、また、新しい語彙、概念が次々と社会の変化に応じて創造されるといった要因など、多くの研究課題が残されている。現在の音声翻訳技術は、旅行会話という一文あたり7単語程度の長さのシンプルな文章を対象にしているレベルである。したがって、新聞や講演などの長く複

雑な文の発話の音声翻訳は、未解決の課題が多く残されている。当面の技術的な課題をまとめると、以下ようになる。

1) 実応用におけるユーザビリティの評価と性能向上

人間の発話に内在する個人差、すなわち、発話様式の差・アクセント・表現様式の差は多様である。この差による音声翻訳性能のばらつきを押さえ、万人に同様の高い性能を目指す必要がある。また、実利用時には、音響的な雑音・残響・他の話者の音声も大きな影響を与える。これらの外的要因への対処も非常に重要である。一方、コミュニケーションツールとして

のユーザビリティという観点からは、音声認識から翻訳を経て音声合成が出るまでの時間をさらに短縮することが不可欠である。音声翻訳が用いられる場では、利用者は翻訳先の言語を知らない。そのため、翻訳結果が正解であるかどうかを確認する術がない。これについては、利用者の言語に再度翻訳し直す、あるいは逆翻訳をするなどして、翻訳結果が正しいかどうかを確認する方法を提供する必要がある。さらに、旅行中における情報獲得ツールとして考える際には、人に聞くだけでなく、多言語でインターネット上の情報を獲得するなどの手段の同時提供も不可欠となる。

このような種々の課題については、フィールド実験と技術開発を同時に進めて、データ収集・性能向上・ユーザビリティ向上・トライアルサービス提供の成長的ループを確立する必要がある。

2) 多言語化

実質的な世界共通語になりつつある英語への翻訳だけでなく、今後は世界中に 6000 あると言われている言語への直接の翻訳が必要になる。多言語の音声翻訳を実現するためには、これら言語のそれぞれの音声認識・翻訳・音声合成を構築する必要がある。すなわち、それぞれの言語で、大量の音声コーパス・対訳コーパス・テキストコーパスが必要となる。特に、音声コーパスの収集には大きな費用がかかる。また、このような技術は、利用者が減少し消えゆく言語の保存という観点からの価値も大きいと言える。

3) ネットワークにより世界の音声翻訳を接続するための標準化

現在、アジア圏でのモジュール接続の標準化が進められている。今後、さらに広く国際的に接続するための標準化と同研究体制の構築を進めていく必要がある。

4) 翻訳例として WEB 上のデータを利用するための著作権緩和

音声翻訳技術の構築には、翻訳元言語のテキストコーパス、翻訳先言語のテキストコーパス、それらの間の対訳文コーパス、そして、音声コーパスが必要となる。これらのコーパスは従来方法では作成・収集に大きなコストがかかる。現在、これらを、爆発的に規模が拡大しているインターネットの WEB 上のデータから収集する方法が注目されている。たとえば、音声翻訳の性能向上に、多言語で発信されているニュースなどの媒体の 2 次利用が有効である。しかし、現在のところ、

著作権の問題が解決されていない。

5) 自分の現在の居場所に応じた、最新の固有名詞の利用

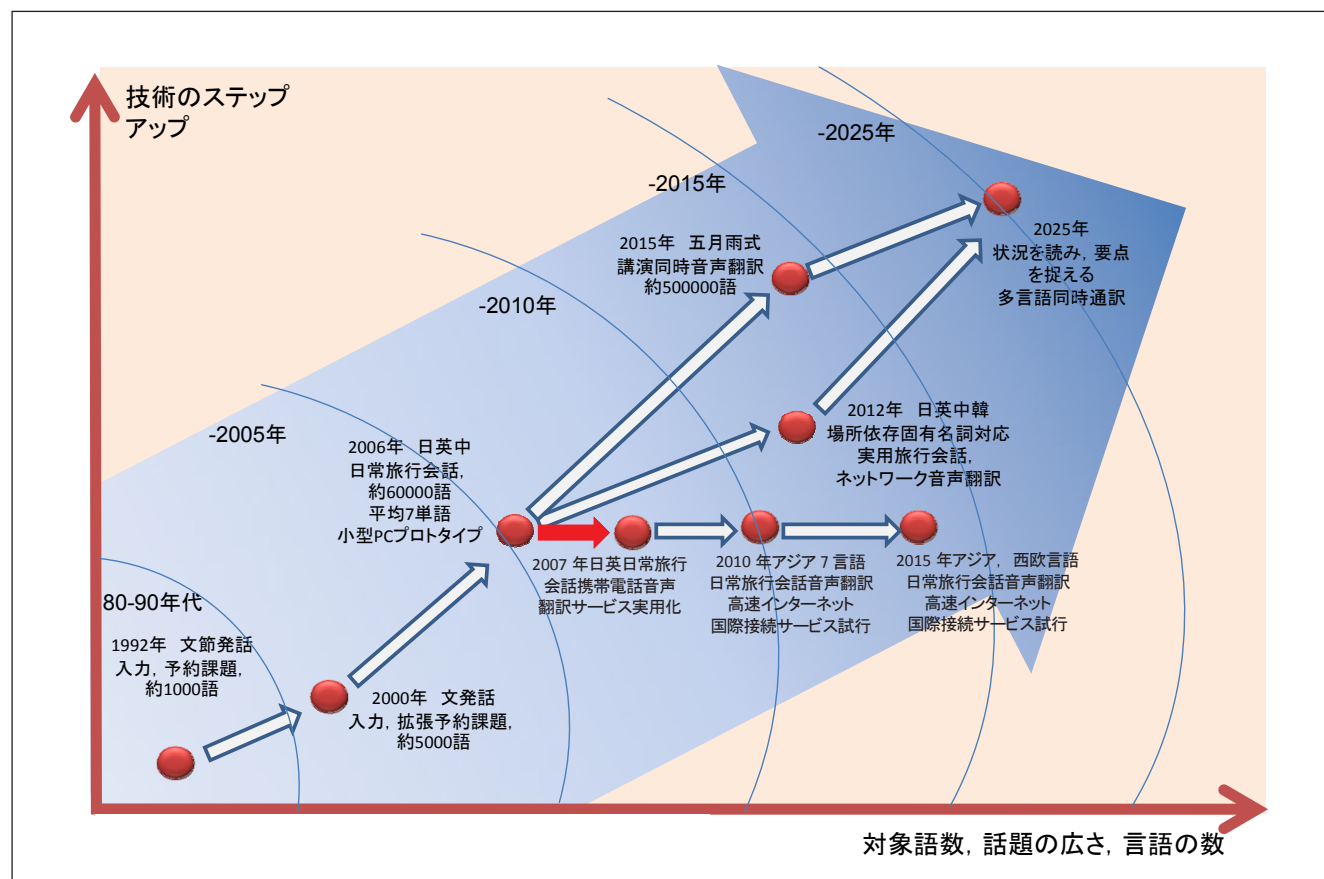
場所や物の名前数は膨大であり、これら固有名詞をすべて同時に音声翻訳することは、性能の点でも速度の点でも不可能に近い。このため、GPS などを用いて自分の現在の居場所に応じた固有名詞を自動的にネットワークから獲得し、その場に応じた音声認識・その場に応じた翻訳・その場に応じた音声合成を行うことが効率的である。

7 - 2

今後の研究開発 (ロードマップ)

図表 8 に、これまでの音声翻訳の開発経緯と、今後の研究開発動向を示す。2010 年に、アジア言語に関する国際研究コンソーシア

図表 8 音声翻訳技術研究開発動向予測



科学技術動向研究センターにて作成

ムが、インターネットを経由した音声翻訳を試行する予定である。これを拡張し、西欧言語を含む、接続標準化が進み、国際研究コンソーシアムによる接続試行が行われるのは2015年頃と推測され

る。また、日本の社会還元加速プロジェクト（次節で紹介する）では種々の実証実験を経て2012年に、ネットワーク音声翻訳としての技術を確認する予定になっている。中長期的には、2015年頃に

はビジネスや講演での五月雨式同時音声翻訳、2025年頃には状況を読み、要点を捉える多言語同時通訳が実現し、次第に夢の同時通訳へ近づいていくと予測される。

8 音声翻訳に関わる日本の施策

8 - 1

内閣府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」

総務省の音声翻訳プロジェクトを内閣府社会還元加速プロジェクトと認定して、2008年度から進めているプロジェクトである。国際化の進展の中で、国民レベルでの直接的な国際コミュニケーションにより、多国間の相互理解を深めることが目的である。アジア圏など海外の人々と言語の壁を乗り越えた直接会話による交流を可能にすることができる「自動音声翻訳システム」を目指している。当面の利用ニーズと今後5年程度で期待できる技術向上等を考慮して、観光やショッピング、国際交流イベントなどにおける実証試験を企画・推進する。また、プロジェクト終了後には、短期間で産業界における事業化ベースでのサービスにつなげ、その成果の社会還元を加速する。

このプロジェクトでは、技術開発としては、場所の地名・物の名前などの名詞だけでなく、幅広い話題に対応するためにネットワーク上に分散する翻訳知識を活用し、翻訳端末と組み合わせるネットワークベース音声翻訳技術の確

立に取り組む。それとともに、このシステムの円滑な実用化・事業化を図り、普及を促進するために、音声翻訳コミュニケーション技術というイノベーションの「見える化」にも取り組んでいる。ここで「見える化」とは、開発サイドと利用者サイドのミスマッチが実用化・事業化を妨げているとの認識のもとで、開発サイドが技術開発状況を適宜開示し、利用者サイドは利用場面などの利用環境条件を想定できるようにすることである。そして、その利用環境条件に適した機能とユーザーインターフェースを持ったシステムを構成して、具体的に実証を繰り返す。このような開発サイドと利用者サイドとの密接な連携による実証を積み重ねることにより、このプロジェクトは「日本、英語・中国語圏で、普通の旅行者がほとんど支障なく海外旅行を楽しむ」など草の根レベルの国際交流における重要な役割を果たすことができるようになる。さらに、新たなビジネスを創造し、地域における地場産業等の振興にも資することができると期待される。

8 - 2

ユビキタス特区

ユビキタス特区事業は総務省の

「ICT改革促進プログラム」および「ICT国際競争力強化プログラム」に基づいて創設されたものである。革新的なサービスの開発・実証実験を支援する事業として、2008年度より3年間の計画で実施される。(財)京都産業21を代表とした8法人((株)インテージ、(株)東映京都スタジオ、NICT、ATR、(株)JTB法人東京、(株)ウィルコム、NEC)からなるコンソーシアムによる共同提案が、「ユビキタス特区」事業に採択された。この事業は、京都府と連携し、京都を来訪する外国人旅行者を対象とした市場調査・多言語翻訳・観光情報提供を行うための携帯端末サービスの開発と、これを広く観光地に普及させるために高速モバイル通信を実現する次世代PHSに対応したユビキタス多機能サーバの実証試験を行う。また、外国人旅行者の満足度向上と観光産業の増進に貢献することを目指し、多言語翻訳とともに、次世代PHSに対応したウェアラブル動画配信サーバ等の先進的な技術を活用し、ユーザとなる観光施設の売店や飲食店で導入・利用しやすいサービスの開発が推進される。また、特区であることを活用すると、コーパス構築における著作権問題などの課題をクリアできる。

9 おわりに

音声翻訳は、音声・言語の研究も進展したため、利用価値が比較的明確で容易な課題については実用に近いレベルまでに至った。しかし、現在のレベルはまだ技術の核の部分ができただけにすぎない。より高度な音声翻訳を実現するために、さらに研究開発を加速させることが望まれる。以下が、今後の注目すべき点である。

まず、第一に、コーパスベース技術は使われてこそ磨かれていくという特質を有する。したがって、実証実験や社会実験の場を確保し、開発した技術を積極的に利用することが重要である。オリンピックや万博などの多言語の参加者が期待できるイベントは、多言語音声翻訳技術にとって絶好の実証実験の場である。このような場を利用して技術を進化させていくことが重要である。NICTは、北京オリンピックにおいて日本からの旅行者を主な対象としたモニター実験を行った。北京市内の固有名詞に対応した音声翻訳システムを開発し、モニターの人々に北京市内での移動・観光・ショッピング等におけるコミュニケーション手段として音声翻訳機能を利用してもらった。アンケートによるサービス利用満足度調査により、より実用に近い状況に対応可能な音声翻訳技術の実験となった。

日本が観光立国をめざす上で、言語音声翻訳による外国人観光客のための継続的な観光情報サービスは有効な施策となろう。一方、昨今の外国人滞在者や労働者の増加により、自治体・医療・警察・教育の現場における多言語音声翻訳は、意志疎通の手段としてもはや不可欠と言える。通訳者がいる場合にも、通訳者の負担軽減に役立つはずである。しかし、それぞ

れが個々に使用していたのでは、断片的な知識しか蓄積されず、研究開発へのフィードバックという点では効率が悪い。効率改善のためには、国・自治体・民間の協力体制が必要となろう。たとえば、小型翻訳機については、使用が想定される公的機関への配布、外国人労働者や観光客への無償貸与などが有効であろう。

第二に、音声翻訳は異なる言語の間の話し言葉の翻訳を行う技術である。英語への翻訳はもちろん重要だが、他の種々の国の母語と日本語の直接の音声翻訳にも大きな意味がある。このため、対象の言語を増やしていくことが重要となる。この研究開発では、特にコーパスの収集については、日本だけで進めるには限界がある。多言語多国間の連携スキーム、つまり、種々の国が連携しながら音声翻訳、音声・言語の研究開発を進めるしくみが必要である。連携研究開発のスキームとして、国際音声言語技術研究センターなどを設置することで、音声や方言の収集や言語構造の研究などの幅広い研究からのフィードバックが期待できる。

第三に、実際に音声翻訳を多くの国で研究開発するようになると、それぞれの言語の処理モジュールを接続するための種々の標準化を行う必要が出てくる。接続方式・データフォーマット・辞書などの標準化を意識しながら開発を進める必要がある。各国がばらばらに開発し、相互接続できなくなる事態は避けなければならない。日本は、音声翻訳技術が進んでいる国であり、標準化に関しても他国をリードすることができる。

最後に、著作権も注目すべき点である。音声処理や翻訳処理には、音声コーパスとテキストコーパス

が必要であり、音声翻訳の性能はこれらコーパスの量と質に大きく依存している。したがって、放送ニュース・新聞・インターネットのコーパスの利用は非常に有効である。現在の著作権法では二次利用としてのこれらのコーパスの利用は考慮されていない。新しい技術の研究開発のためには、柔軟な法改正や運用がなされてしかるべきであろう。このことに関しては、現在、文化審議会著作権分科会で議論されており、近々、結論がまとめられることになっている。この検討結果を待って、将来の音声翻訳の本格的なサービスに向けた制度的な課題を整理し、サービスモデルも含めて、改めて対応を検討していく必要がある。

参考文献

- 1) S.Nakamura, et al, "ATR Multilingual Speech-To-Speech Translation System", IEEE Trans. ASLP, vol.14, no.2 (2006)
- 2) A.Finch, et al, "The NICT/ATR Speech Translation System for IWSLT 2007", IWSLT (2007)
- 3) T.Takezawa, et al, "Toward a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World" Proc. LREC (2002)
- 4) 安田、他、「自動翻訳の研究・開発における翻訳時道評価技術とその応用」、人工知能学会誌、23巻1号 (2008)
- 5) 菅谷、他、「音声翻訳システムと人間との音声翻訳能力評価手法の提案と比較実験」、信学論 J84-D-II, 11 (2001)
- 6) 伊藤、他、「日英中音声翻訳機のフィールド実験とその評価」、1-Q-33、日本音響学会講演論文集 (春) (2008)

- 7) IWSLT : <http://www.slt.atr.jp/IWSLT2004/>
- 8) SLTC e-Newsletter, "DARPA's GALE Program to Get More Challenging in 2007" (GALE プロジェクトの一例) : <http://ewh.ieee.org/soc/sps/stc/News/NL0701/NL0701-GALE.htm>
- 9) ETSIES 202 050 ETSI ES 202 050 v1.1.1 Speech Processing, Transmission and Quality aspects (STQ); Distributed Speech Recognition; Advanced Front-end Feature Extraction Algorithm; Compression Algorithms, ETSI, April 2002.
- 10) APEC TEL WORKING GROUP : <http://www.apectelwg.org/>
- 11) ASTAP: <http://www.aptsec.org/Program/ASTAP/>
- 12) 木村、他、「多言語音声翻訳基盤のための通信インタフェースの検討」、3-Q-17、日本音響学会講演論文集（秋）（2007）

執 筆 者



中村 哲

客員研究官

(独) 情報通信研究機構 上席研究員

MASTAR プロジェクトリーダー、

ATR 音声言語コミュニケーション研究所 所長

ATR フェロー

<http://www2.nict.go.jp/x/x162/mastar>



工学博士。音声翻訳、音声認識を始めとした音声・言語処理に従事。ドイツ・カールスルーエ大学客員教授、けいはんな連携大学院教授。