

2 多言語コミュニケーション技術

2 *Multilingual Communication Technology*

2-1 多言語コミュニケーション技術の概要

2-1 *Overview of Multilingual Communication Technology*

内元 清貴 隅田 英一郎 河井 恒 香山 健太郎

UCHIMOTO Kiyotaka, SUMITA Eiichiro, KAWAI Hisashi, and KAYAMA Kentaro

我が国では、国家プロジェクトとして2014年から言葉の壁のない社会を実現させるためにグローバルコミュニケーション計画が推進されている。NICTは、この計画の目標を達成するために、産学官から研究者が集結したオールジャパン体制で研究開発を進めてきた。その結果得られた研究開発成果は、産学官の力を結集した成果として大きく社会に広がっている。さらに、2020年、産学官連携体制による新たな研究開発がスタートした。大阪・関西万博が開催される2025年を目標に、ビジネスでも使える低遅延のAI同時通訳の実現を目指す。

Since 2014, Japan has been pursuing the national project, Global Communication Plan, with the aim of realizing a society free from language barriers. NICT has been promoting research and development with an “all-Japan” line-up of researchers from industry, academia, and government to achieve the goals of the Plan. As a result, the outcomes of these research and development activities, which are the fruit of industry-academia-government collaboration, are being widely disseminated throughout society. Furthermore, in 2020, a new research and development project has been launched under an industry-academia-government partnership. We aim to realize a low-latency AI simultaneous interpretation that is capable of supporting business scenes by the year 2025 when the Expo 2025 Osaka, Kansai will be held.

1 まえがき

世界中の人々が言葉や能力の違いを意識せずにコミュニケーションできる、言葉の壁のない社会は人類の大きな夢のひとつである。我が国では、国家プロジェクトとして2014年から言葉の壁のない社会を実現させるためにグローバルコミュニケーション計画が推進されている。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「グローバルコミュニケーション計画」[1]が最初に総務省から発表されたのは、2014年4月であった。同年、NICTは、この目標を達成するために、研究開発の中心的な役割を担う研究拠点として、先進的音声翻訳研究開発推進センター（ASTREC）を設立した。そして、東京オリンピック・パラリンピック競技大会の開催が設定されていた2020年を目標として、ASTRECに産学官の研究者が集結する形のオールジャパン体制で多言語音声翻訳技術の研究開発と実証実験・社会実装を

進めてきた。その結果、国内の鉄道などの交通機関やショッピング施設、観光地、医療の現場などで活用される実用性の高い多言語音声翻訳技術や、企業などにおいて他国の特許や医薬文書などを自動で翻訳できる多言語テキスト翻訳技術などの開発において、目覚ましい成果が得られた。これらの研究開発成果は「グローバルコミュニケーション開発推進協議会」での活動を通じて、産学官の力を結集した成果として大きく社会に広がっている。

さらに、2020年3月、総務省より「グローバルコミュニケーション計画2025」[2]が発表され、ASTRECと音声翻訳の民間サービスを展開する企業を中心とした産学官連携体制による、革新的多言語翻訳技術の新たな研究開発がスタートした。これに基づき、大阪・関西万博が開催される2025年を目標に、高精度と低遅延を両立する実用レベルのAI同時通訳の実現を目指すとともに、多言語翻訳システムの更なる普及・発展及び同時通訳システムの社会実装の推進に取り組んでいる。

本稿では、上記の複数の技術、つまり、多言語音声翻訳技術、多言語テキスト翻訳技術及び AI 同時通訳実現のための革新的多言語翻訳技術を「多言語コミュニケーション技術」と総称し、その概要を中心に紹介する。

2 多言語コミュニケーション技術の研究開発

2.1 多言語コミュニケーション技術のパラダイムシフト

我が国の多言語コミュニケーション技術の研究開発は1986年頃に、けいはんな(京都、大阪、奈良の3府県にまたがる)の地で始まった(図1)。当時は、ルールベースと呼ばれる手法が主流で、人がルールを書き、そのルールに従って入力音声を機械処理し、対象言語の音声に変換して出力していた。しかし、人が整合性を保って書けるルールの数は、例えば1言語対あたりせいぜい1万規則程度が限界であり、入力発話の多様性に対し網羅的に対応することは難しかった。そのため、当時は、特定話者のみ、文節に区切って発話した定型的な文のみ、静かな室内での発話のみ、かつ、会議室予約といった限られた話題の範囲でのみ音声翻訳

ができるというものであった。また、日本語音声が入力されてから英語の合成音声出力されるまで、3文節程度の短い文でも20秒程度かかっていた。その後、2000年代頃になって、コーパスベースと呼ばれる手法が主流となり、対応できる範囲が格段に広がった。この手法は、コーパスと呼ばれる言葉のデータベースから統計的にルールを自動獲得して利用するというものである。例えば1言語対あたり必要とされるのは10万文規模のコーパスで、そこからルールを自動獲得し、ルールの優先度を同じくそのコーパスから機械学習により自動学習する。これにより、ルールの網羅性と信頼性が高まり、多様な表現に対応できるようになった。また、不特定の話者にも対応できるようになり、丁寧に発話すれば、一定の条件下であれば屋外でも音声認識可能で、対応できる話題の範囲も生活会話全般にまで広がった。NICTではこの手法に基づき、2009年の全国5観光施設での大規模実証実験を経て、2010年にネットワーク型の多言語音声翻訳アプリ「VoiceTra®」(ボイストラ)を世界に先駆けて公開した。さらに、2010年代半ばにニューラルネットベースのものが主流となり、飛躍的に高精度かつ自然な音声翻訳が可能となった。必要となるコーパスの規模は例えば1言語

多言語翻訳技術の進展

- 我が国の多言語翻訳技術は、30年の研究開発を経て実験室から社会へ、これからは日常生活からビジネスへ
- 「グローバルコミュニケーション計画」に基づく取組により、AI技術も導入し、翻訳精度を向上(12言語で実用レベルを達成)
- NICTから民間への技術移転も進め、多様なサービスが実用化・普及

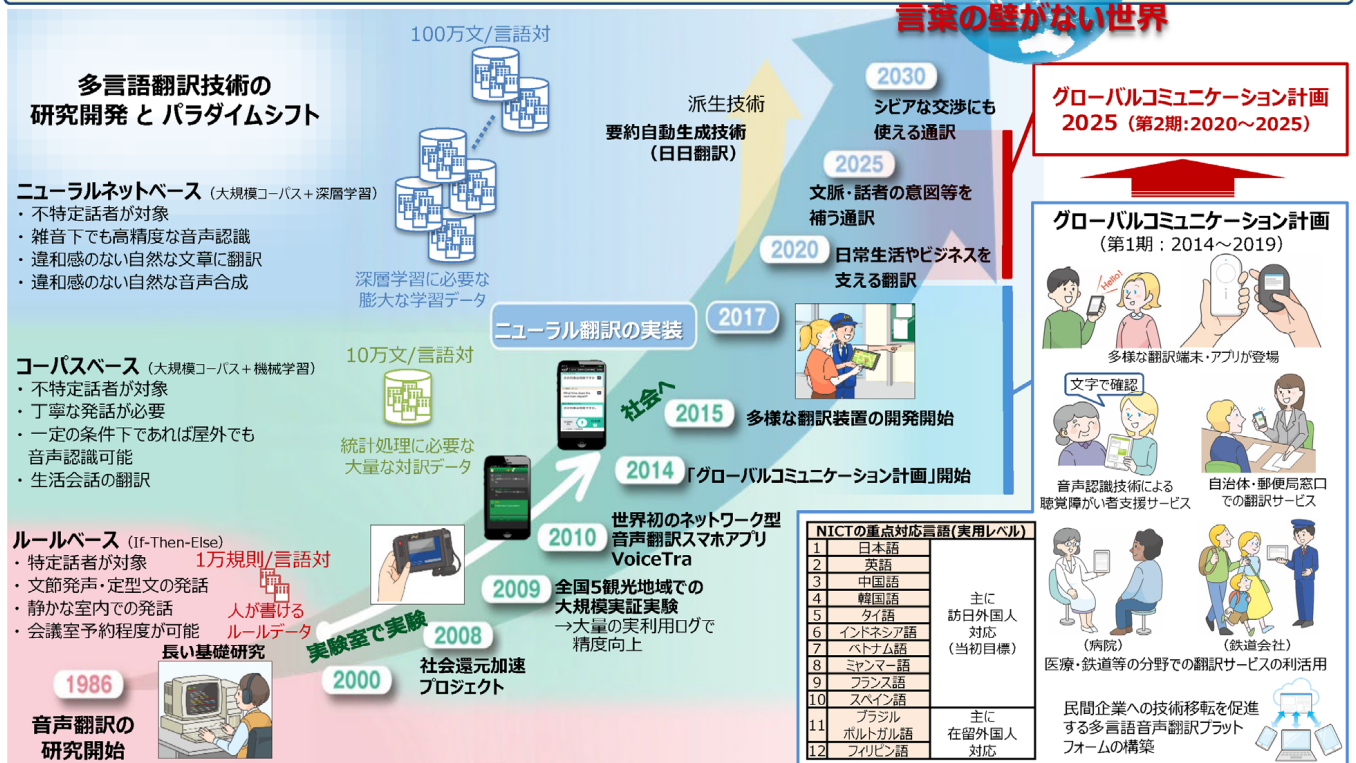


図1 多言語翻訳技術の進展

対あたり 100 万文規模と一桁大きくなったが、この手法では、深層学習により、ルールの獲得と優先度付けをまとめて自動学習できるようになったことから、効率的かつ精度良く自動学習できるようになった。ニューラルネットベースの手法の効果が画像認識など様々な応用分野で確認されつつある中、NICT の音声翻訳も 2017 年にニューラルネットベースに移行し、VoiceTra[®] に実装・公開した。これにより一気に NICT の多言語コミュニケーション技術の性能が向上し、商用サービスとして広く採用されるようになった。我が国で多言語コミュニケーション技術の研究開発が開始されてから、ここまで使える技術に到達するまで、実に 30 年以上かかっている。

2.2 NICT における多言語コミュニケーション技術の研究開発

NICT は独自の多言語コミュニケーション技術をネットワーク型の音声翻訳アプリ“VoiceTra[®]”に実装し App Store や Google Play で公開している。VoiceTra[®]

は旅行会話に好適で、テキストでは 31 言語間の翻訳に対応している。スマートフォンから入力された音声は、図 2 のようにネットワークを介してサーバに送信され、サーバ内で音声認識、機械翻訳、音声合成の処理がなされた後、翻訳された音声はそのサーバから再びネットワークを介してスマートフォンに返送され再生される。サーバ内の各処理では、音声とその書き起こしテキストとの対応関係や、日英対訳間の対応関係などをコーパスから自動学習した結果に基づいて処理している。

現在、自動学習には、2.1 で述べたニューラルネットベースの深層学習を採用しており、以前の手法に比べて、音声翻訳の精度は基盤となるコーパスの量と質にますます大きく依存するようになってきている。現在の VoiceTra[®] は話し言葉の音声翻訳を得意としているが、それは、コーパスベースが主流の頃より、大規模・高品質の話し言葉のコーパスを地道に収集・構築してきたからである。NICT では、これまで、旅行会話など対象分野を絞り、VoiceTra[®] の利用ログ情報も活用し

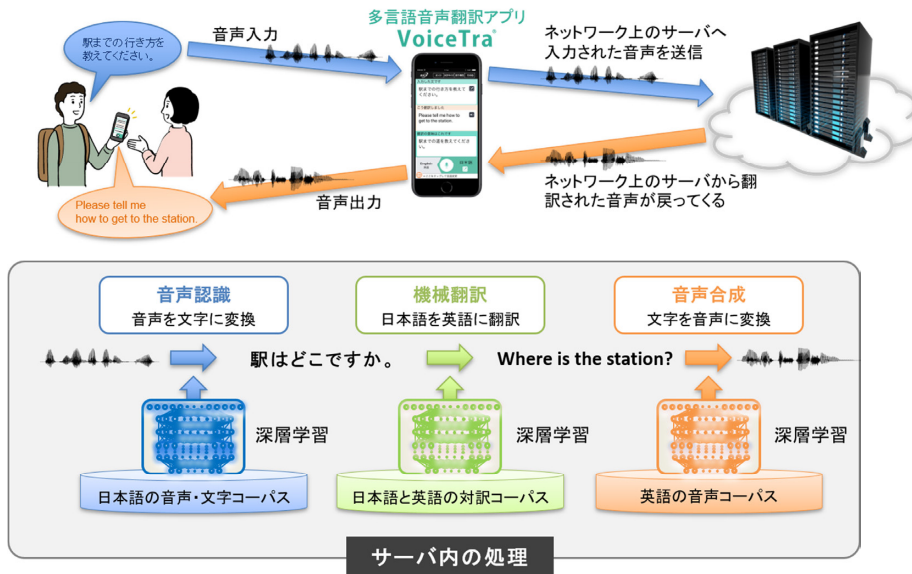


図 2 ネットワーク型多言語音声翻訳の仕組み

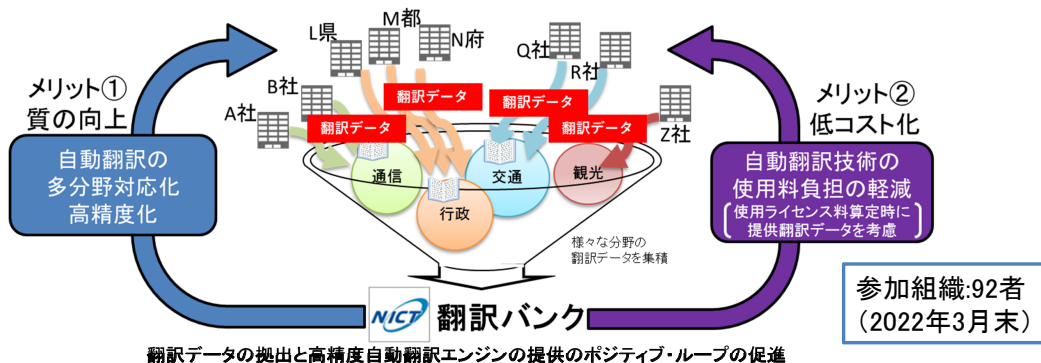


図 3 翻訳バンクの枠組み (<https://h-bank.nict.go.jp/>)

2 多言語コミュニケーション技術

て効率良くコーパスを構築することにより、高い精度を実現してきた。さらに対応範囲は、当初の旅行会話から、日常会話を対象とした生活、災害、医療の分野へと広げてきた。

しかし、言語や分野、発話スタイルによっては十分な性能が得られない場合がある。性能向上の鍵となるのは大規模・高品質のコーパスの収集・構築である。そこで、NICTでは、技術の高精度化だけでなく低コスト化や多分野化を加速するため、産学官の協力により翻訳データ（日英など複数言語のテキストが対になった対訳のコーパスや辞書）をNICTに集積する「翻訳バンク[®]」（<https://h-bank.nict.go.jp/>）の運用を2017年に開始した。この翻訳バンクの枠組みでは、提供いただいた対訳コーパスを用いて翻訳システムを高精度化し、提供元の組織に対し、高精度化した翻訳システムを安くライセンス供与している（図3）。これはまさに、みんなで国産の翻訳システムを育てるプロジェクトと言える。これまでに、中央官庁、地方自治体、民間企業、各種団体など、データを提供してくださった組織は90者を超え、更に増え続けている。例えば、医薬業界については、アストラゼネカ株式会社から提供いただいた対訳コーパスを用いて翻訳精度を向上させ、その高精度化した翻訳システムを下記に活用することにより、アストラゼネカ社内での翻訳業務の労力を半分にすることができるといった画期的な効果が確認された。この結果を受けて、製薬分野の他の大手企業約8社からも翻訳データを提供いただけることになり、製薬分野の翻訳精度は更に高精度なものとなった。対訳コーパスの提供は、医薬だけでなく、特許や金融など様々な分野に広がっている。これに伴い、対応できる言語や分野も増えつつある。各分野でのコーパスが充実すると、各分野における翻訳精度が向上し、新しいサービスが創出されるという好循環が生まれている。NICTは公的研究機関であるため、民間企業などに比べてデータ提供を受けやすい立場にある。更に安心してデータを提供していただけるようにするために、情報セキュリティマネジメントシステムの国際規格である「ISO / IEC27001」の認証も取得した。今後、データ提供に協力いただける分野が更に広がることを期待している。

一方、音声コーパスの収集・構築に関しては、VoiceTra[®]などのスマートフォン上のアプリを介して収集できる実利用データを有効活用している。収集したデータは、音声コーパスの構築コスト削減に大きく貢献している。さらに翻訳バンクの枠組みを音声コーパスにも適用することにより、種々の発話スタイルの音声コーパスの収集、さらには、音声認識の精度向上にもつながると期待される。また、VoiceTra[®]の利用ログ情報などの実利用データは他の用途にも役立つこ

とが確認されている。NICTは、VoiceTra[®]の実利用データを活用することにより、複数の言語に対し、入力音声の言語がどの言語であるかを瞬時にかつ高精度で識別できる技術を開発した。音声を入出力のインターフェースとして利用するアプリやシステムの多くは、入出力の言語をあらかじめ指定する必要があるが、この自動言語識別技術を利用することにより、あらかじめ言語を指定する必要がなくなる。例えば、案内ロボットはタッチパネルなどでの言語選択が不要となり、コールセンターでは適切な言語で対応できるオペレータにスムーズにつながることができるようになる。この自動言語識別技術はVoiceTra[®]にも機能の一つとして組み込んでおり、手軽に試していただくことが可能である。

上述のNICTの多言語コミュニケーション技術は、本特集号の**2-4**「多言語コミュニケーション技術の社会実装」に述べるように、産学官連携による実証実験を経て、様々な形態で民間サービスなどに採用されるようになった。2020年以降のコロナ禍により、インバウンドの需要は一時的に減少したが、その一方で、在日外国人への対応やリモート観光、リモート会議での活用などの需要が増えつつある。VoiceTra[®]に代表される逐次翻訳は、今や、普通に使われる技術になったと言える。

2.3 多言語コミュニケーション技術の更なる展開

新型コロナウイルスがまん延するまでは、訪日外国人の数は増え続け、2019年には3,000万人を超えていた。そのような訪日外国人への対応ニーズに加え、在留外国人、特定技能人材への対応のニーズに対しては、ここ5年ほどの急速な技術の発展を踏まえ、VoiceTra[®]に代表される逐次翻訳の技術が様々なサービスに活用されてきた。一方、グローバル化が加速する中、ビジネス・国際会議での講演や議論の場面、企業での協業の場面などでのニーズが高まっており、従来の技術では対応しきれない場面が増えている。それに伴う課題は大きく二つある。第一に、対応が必要な分野が拡大している。この分野拡大の課題に対しては、上述の翻訳バンクの枠組みを最大限に活用して解決していきたい。第二に、国際会議での講演やビジネス会議でのやりとりなどに対して、長い発話を次から次へと翻訳し、自分の理解できる言語で内容を把握したいという要求が高まっている。しかし、これまでの逐次翻訳のように、交互に発話して翻訳結果をその都度相手に伝え合うという方法では対処しきれない。そこで必要となるのが同時通訳である。

我々は、2025年までに同時通訳システムを実現し、その先に、言葉の壁がない世界を実現したいと考えて

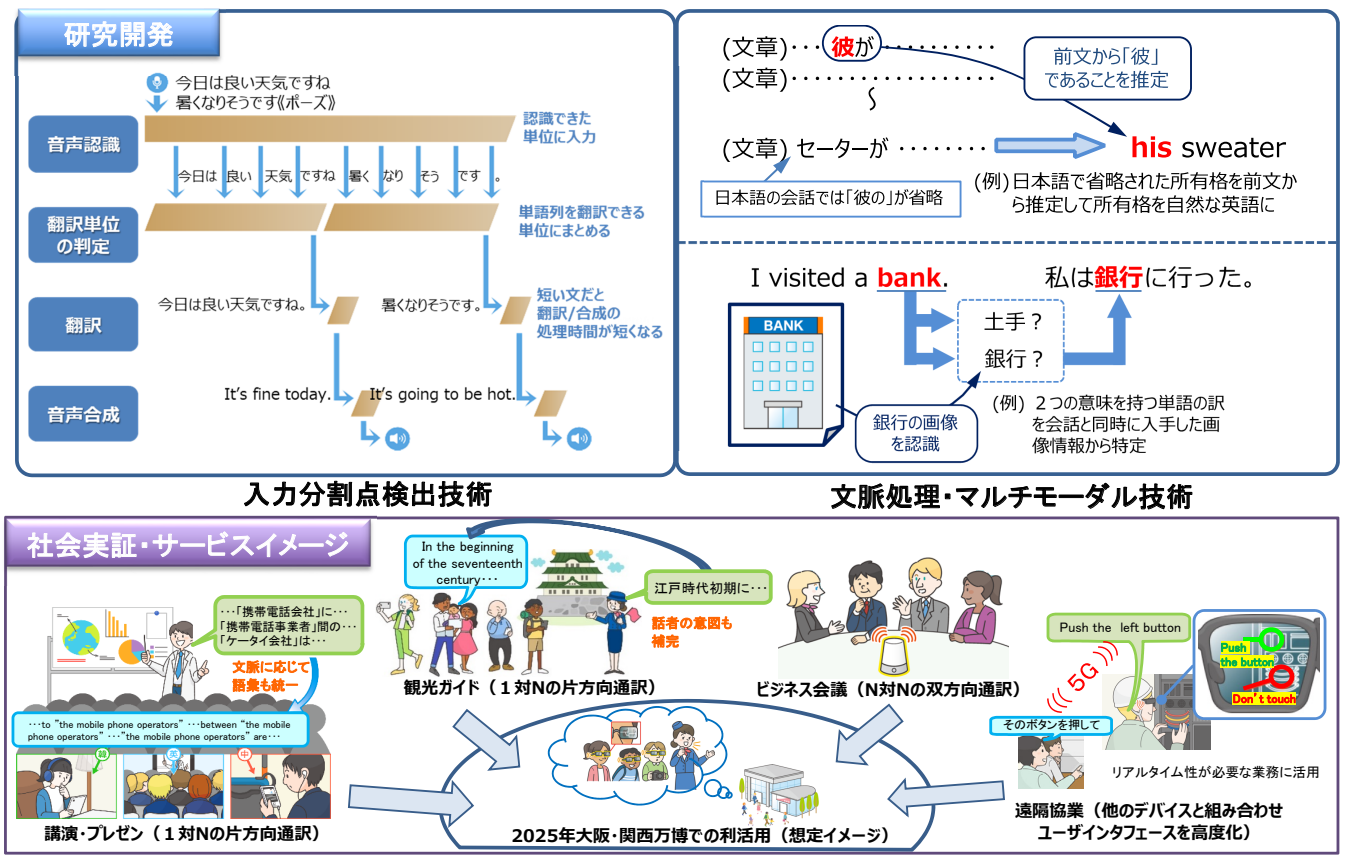


図4 AI同時通訳実現のための研究開発と社会実証・サービスイメージ

いる。この同時通訳システムを実現するために、総務省が2020年3月に発表したのが「グローバルコミュニケーション計画2025」である。この計画では、大阪・関西万博が開催される2025年を見据え、これまでの一文単位の「逐次翻訳」から、「同時通訳」へと技術を進化させ、ビジネスや国際会議などの場面においても利用可能な実用レベルのAI同時通訳を実現することを大きな目標の一つとしている。コロナ禍の影響により生活スタイルが大きく変化したことにより、リモートによる多言語対応のニーズも増え、計画の重要性もますます高まっている。

NICTはこの計画に基づき、民間企業とともにコア技術の研究開発と社会実証を推進している。これまでの「逐次翻訳」では、ビジネスシーンのように長い発話のやりとりがある場面では翻訳結果が得られるまでに時間がかかってしまう。これに対し、人間の同時通訳者のように発話の途中でも訳出できるところから次々に訳出するためには、意味的なまとまりを検出して切り出す必要がある。例えば、図4の左上の例のように「今日は良い天気ですね。暑くなりそうです。」という発話に対し、「今日は良い天気ですね」で切り出す技術が必要となる。これが入力分割点検出技術である。また、省略の多い日本語を外国語に翻訳する際に、話の

流れを考慮して省略を適切に補うといったことが難しい。例えば、図4の右上の例のように日本語の「セーター」を英語に翻訳する際に、誰のセーターであるかを補う必要がある。その際、文脈を遡ることにより多様な情報を取り込んで曖昧性を解消し、正しく彼のセーターであると推定する技術が必要となる。これが文脈処理技術である。もう一つ重要なのがマルチモーダル技術である。例えば、図4の右中段の例のように、英語の“bank”は日本語では二つの訳語があるが、音声やテキストの情報だけでは正しく訳語を選択できない場合がある。その際、ほかに参照できる画像情報などを参考に、より精度良く訳語を選択する技術が必要となる。これらの技術を高度化し、低遅延と高精度を両立させることにより、講演やビジネス会議でのやりとりを次から次へと適切に翻訳することが可能となる。NICTでは、民間企業とともにこれらの技術の研究開発に取り組み、2025年を目途に実用レベルのAI同時通訳の実現を目指す。また、これとともに、例えば、図4の下段に示すような講演通訳、ガイド通訳、会議通訳、遠隔協業など、様々な場面でAI同時通訳が活用されるよう、技術の普及に努め、これらの技術が2025年開催予定の関西・大阪万博でも広く活用されるようにしたい。

3 まとめ

本稿では、誰もが分かり合えるユニバーサルコミュニケーションの実現に向けた取組の一つとして、言葉の壁をなくす多言語コミュニケーション技術にフォーカスを当てて概要をご紹介した。VoiceTra[®]でもお試しいただける NICT の逐次翻訳の技術は、様々な形態で民間サービスなどに採用され、産学官の力を結集した成果として大きく社会に広がっている。現在、ビジネス・国際会議での講演や議論の場面などで逐次翻訳では対応しきれないニーズに応えるため、2025 年を目途に AI 同時通訳を実現し、普通に使われる技術とすることを目指している。さらに、総務省が推進するグローバルコミュニケーション計画 2025 では、2025 年の5年先である 2030 年を目途に、シビアな交渉にも使える同時通訳を実現するという目標が掲げられている。この一段高い目標を達成するには、言葉の壁をなくすだけではなく、相互理解を支援する工夫も必要となる。そのために、本特集号の 1「緒言:ユニバーサルコミュニケーション研究所における研究開発について」で述べた3つのコア技術、すなわち、ビジネスで使える低遅延の AI 同時通訳を可能とする多言語コミュニケーション技術、仮想的人格を用いてユーザの興味・背景に合わせた対話を可能とする社会知コミュニケーション技術、パブリック/プライベートデータを連携させた実世界の状況分析・予測を可能とするスマートデータ利活用基盤技術及びユーザインタフェースの工夫などコミュニケーションの質を向上させる技術を有機的に結びつけ、その課題を克服したい。

【参考文献】

- 1 総務省, “グローバルコミュニケーション計画,” https://www.soumu.go.jp/main_content/000285578.pdf, April 2014.
- 2 総務省, “「グローバルコミュニケーション計画 2025」の公表,” https://www.soumu.go.jp/menu_news/s-news/01_tsushin03_02000298.html, March 31, 2020.



内元 清貴 (うちもと きよたか)
 ユニバーサルコミュニケーション研究所
 研究所長/
 先進的音声翻訳研究開発推進センター
 研究開発推進センター長/
 先進的リアリティ技術総合研究室
 室長
 博士(情報学)
 自然言語処理
 【受賞歴】
 2001年 情報処理学会 2001年度(平成13年度)
 山下記念研究賞



隅田 英一郎 (すみた えいいちろう)
 NICT フェロー/
 ユニバーサルコミュニケーション研究所
 先進的音声翻訳研究開発推進センター
 先進的翻訳技術研究室
 室長
 博士(工学)
 自然言語処理
 【受賞歴】
 2020年 内閣府 第2回日本オープンイノベーション大賞総務大臣賞
 2013年 内閣府 第11回産学官連携功労者表彰
 総務大臣賞
 2010年 文部科学大臣 科学技術賞



河井 恒 (かわい ひさし)
 ユニバーサルコミュニケーション研究所
 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
 博士(工学)
 音声情報処理
 【受賞歴】
 2015年 電気通信普及財団 第31回(2015年度)
 電気通信普及財団賞(テレコムシステム技術賞)
 2014年 電子情報通信学会 2014年度論文賞
 2010年 情報処理学会 2010年度喜安記念業績賞



香山 健太郎 (かやま けんたろう)
 ユニバーサルコミュニケーション研究所
 総合企画室
 室長
 博士(工学)
 人工知能