

FEATURE

世界の言葉の壁をなくす 多言語音声翻訳技術

Interview

音声翻訳技術で世界の最先端に行く



FEATURE

世界の言葉の壁をなくす多言語音声翻訳技術

Interview



- 1 **音声翻訳技術で世界の最先端に行く**
木俣 豊
- 4 **グローバルコミュニケーション計画の実現に向けた多言語音声認識技術の開発**
河井 恒
- 6 **グローバルコミュニケーション計画の実現に向けた多言語翻訳技術の開発**
隅田 英一郎
- 8 **研究成果を社会実装するためのソフトウェア開発とサービス運用**
葦苈 豊
- 10 **グローバルコミュニケーション計画の実現に向けた多言語音声翻訳技術の社会展開**
内元 清貴

TOPICS

- 12 **光出力500mWを超える極めて高出力な265nm 帯深紫外LEDを実現** 井上 振一郎
- 13 **NICTのチャレンジャー File 9** 丁 塵辰
計算言語学・自然言語処理のアジアを中心とした多言語化

INFORMATION

- 14 **パーマネント研究職・総合職 採用2021**
- 14 **Interop Tokyo 2020 出展のお知らせ／NICTER 観測レポート2019の公開**
- 14 **NICTオープンハウス2020 in 小金井 開催のお知らせ**

表紙写真：計算機室

言葉のデータベースであるコーパスを大規模に蓄積し、そのコーパスから、深層学習などの機械学習手法を用いて、音声認識、機械翻訳、音声合成等の音声処理や自然言語処理のモデルを作成するための大規模計算等を行っています。

左上写真：ユニバーサルコミュニケーション研究所の正面玄関

けいはんな学研都市に位置するNICTの関西拠点のひとつ、ユニバーサルコミュニケーション研究所の建物の正面外観。先進的音声翻訳研究開発センターの研究室、開発室、そして企画室の分室があります。

Interview

音声翻訳技術で
世界の最先端を行く

先進的音声翻訳研究開発推進センター (ASTREC)

木俵 豊 (きだわら ゆたか)

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

機械による自動通訳は、つい数年前までは夢のように思われていたが、今はスマホのアプリやAI翻訳機が急速に普及し始めている。

以前は外国語の翻訳を機械にやらせることは困難を極めていたのだが、なぜここ数年でこんなに性能が向上したのだろうか。日本の機械翻訳及び音声認識・音声合成技術をリードしているNICTの先進的音声翻訳研究開発推進センター (ASTREC) の木俵豊センター長に話を聞いた。

■先進的音声翻訳技術とは

— ASTRECの先進的音声翻訳技術とはどういうものでしょうか。

木俵 ASTRECでは多言語の音声翻訳技術を研究開発しています。多言語音声翻訳には、音声認識・翻訳・音声合成という3つの技術で構成されています。

まず、音声で入力された言語を多言語音声認識技術でテキスト化します。これを多言語翻訳技術によって目的の言語に翻訳します。さらに、これをテキストにして読み上げるのが多言語音声合成技術です。

私たちは、この3つの技術の研究開発を行っています。さらには、開発した技術を世の中で使っていただける品質のソフトウェアとしてまとめあげ、社会に提供するのが仕事です。

— その成果のひとつが^{ボイストラ}VoiceTraというわけですね。

木俵 そうです。これはスマートフォンの無料アプリとして提供しています。現在、31の言語について、機械翻訳による出力ができます。このうち、音声入力ができるものは、18言語。音声合成で出力できるのが16言語あります。

— アジアの国の言語が多いですね。

木俵 日本にとってアジアの国々は経済・政治ともに重要な国々です。観光客もアジアからの方々が多いためアジアの言語は重要なのです。欧米の言語についてはグーグル翻訳が強いので、それとの差別化ということもあります。

2014年から総務省によって、「世界から言葉の壁をなくす」というグローバルコミュニケーション計画が推進されてきて、それを実現するために研究開発を加速しているところです。今年は、東京オリンピック・パラリンピックがあり、多くの外国人が日本にやってくるの

で、言葉の壁をできる限り取り除くことが目標です。現在、VoiceTraは日常会話・旅行会話が中心ですが、長期滞在の外国人にも生活のサポートができるようにしていきたいと考えています。

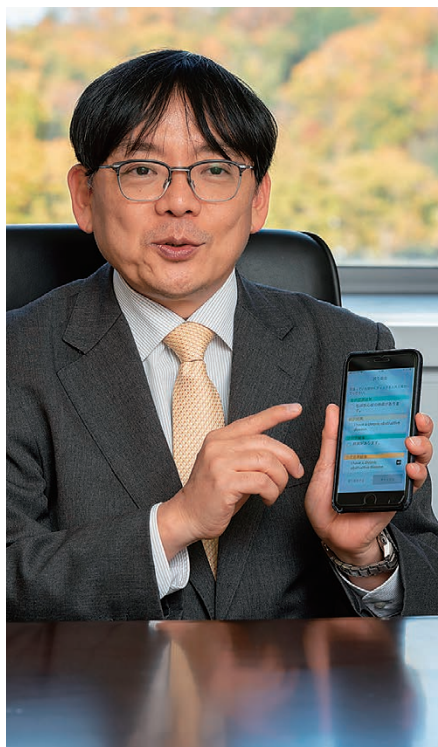
■ニューラルネットを使った機械翻訳の成果

— 機械翻訳とはどういう技術？

木俵 機械翻訳は文字通り機械で翻訳する技術です。現在の機械翻訳技術にはニューラルネットの技術を使っています。対訳コーパス (対訳文を集めたもの) のデータを大量に集め、学習することによって鍛えていきます。基本部分は人手で作りますが、ユーザからの誤り報告のデータを大量に集めることで翻訳精度を改善しています。

また、最近の新しい成果として、VoiceTraに言語の自動識別機能を付け

音声翻訳技術で世界の最先端に行く



たことを挙げることができます。この機能を使うと、あらかじめ言語を設定することなく、相手の話す言語に応じて自動的に切り替わります。この技術は、発話の最初の1.5秒を捕まえて何語かを認識するもので、ほぼ90%程度の確かさで識別できます。

この機能はサポートセンターなどで重宝されています。多言語の翻訳システムを用意していても、いきなり話されると何語かわからない場合がありますから、そういった場合に便利です。また、最近、普及が進んでいるAIスピーカーなども、何語で話されるかわからないので自動識別機能が役に立ちます。

一般向けの成果の一例としては、日常

会話・海外旅行会話に便利なポケット翻訳機「ポケトーク」があります。これにはNICT開発の翻訳エンジンが搭載されています。

——音声翻訳にはほかにどのような技術開発があるのでしょうか。

木俣 雑音の中から人の声を分離する技術もあります。最近ではディープラーニングの能力が向上したことで、雑音と音声とを切り分けるよりも、雑音混じりのデータを大量に学習させた方が認識精度が高くなるのが分かってきました。

もうひとつの研究テーマは、遠隔発話の音声認識です。スマホのVoiceTraでは、マイクの近くで発話するのが基本です。しかし、AIスピーカーのように、日常生活の中で離れた場所から、自然な感じで話している声も正確に認識する必要のあるものもあります。遠隔発話の認識技術は現在、研究を進めているところです。

また、大勢の人が話している場所で、特定の人の声だけを識別して認識する技術もこれから研究開発をしていかなければならないと考えています。

——音声合成による発話も素晴らしくリアルですね。

木俣 音声合成技術もどんどん向上していて、現在は音声合成でしゃべっているのか生身の人間がしゃべっているのか分からないまでになっています。

実はけいはんな拠点の電話の不在応答の音声は、音声合成を使っているのですが、最先端の研究成果を応用しているので、音声合成かどうかはまず分からないと思います。

また、VoiceTraの翻訳結果の音声も、入力した人の声でしゃべってくれるようになる面白いです。将来的には話者の声を解析し、それと同じ声で話すというのも夢ではないでしょう。

——音声翻訳の技術は最近急速に発達しましたがその理由は？

木俣 音声翻訳は開発に30年くらいかかりましたが、ようやく実用レベルになってきたと言えます。この間に、2つのパラダイムシフトがありました。

最初はルール（文法）ベースで文を解析していました。しかし、人が書けるルールは限られているので精度の高い翻訳ができませんでした。2000年ごろになると、統計翻訳の技術が出てきました。対訳コーパスで語の並び順を確率的に計算していくものです。これが第1のパラダイムシフトで、少し実用に近づきましたが、まだ本格利用には難しいものでした。

ところが2016年ごろに、ニューラルネットワークのディープラーニングを使った翻訳技術がグーグルから登場しました。第2のパラダイムシフトです。ルールベースでは1万クラスのルール、統計翻訳では数十万文レベルの対訳データを使用していましたが、ニューラルネットワークでは数百万文以上の対訳データを必要とします。

翻訳だけでなく、音声認識・音声合成もニューラルネットを使って大量のデータを使うことでぐんぐんと精度が向上しました。また、Seq2Seqという系列(sequence)を別の系列へ変換するモデルで、これによって前後の文脈の長期的な依存関係を学習させるLSTM*1等の技術が発明されています。このモデルをベースに世界中の研究者が開発を行っています。

自然言語処理については、2018年10月にグーグルがBERT*2という自然言語処理の新しいモデルを発表しました。これはもう力技と言えます。膨大な数のCPUとGPUを使って巨大なニューラルネットを構築して学習させていきます。これで意味抽出や翻訳を一気に高精度のものに変えつつあります。

ともかく機械翻訳には、膨大な量の計算機リソースが必要になります。今のAIの世界はある意味、物量作戦とも言えます。強力なコンピュータを持つ者がAIの勝者になれると言っても過言ではありません。

■多言語音声翻訳技術の社会展開

——ASTRECが開発した技術の社会との関わりについては。

木俣 ニューラルネットによる機械翻訳はデータが多ければ多いほど翻訳精度が上がります。そこで、ASTRECでは「翻訳バンク」というプロジェクトを立ち上げています。例えばグローバル企業なら、

ビジネス上のデータを英語やその他の様々な言語に翻訳したデータが保存されているはず。そこで、企業の皆様にデータの寄付を呼び掛けて、翻訳に関するデータを増やそうというのが翻訳バンクです。データを提供していただく方のメリットとしては、提供された翻訳データを勘案して、自動翻訳技術の使用ライセンス料の負担を軽減する仕組みです。

最初に成果を出したのが製薬会社の持つ対訳データです。製薬会社は、年に何度か英語の文書を厚生労働省に提出しないといけないそうです。しかし、英語に翻訳するのに1か月はかかる。そこで、過去の対訳データを提供いただいて、ニューラルネットで学習させてみたところ、高精度な翻訳ができるようになりました。最終的には人手で修正する必要がありますが、翻訳が従来の半分の2週間できるようになったと喜んでもらいました。

このほか、凸版印刷が開発した日本全国の郵便局の窓口での外国人向けの案内システムや、全国的に多くの救急隊員が使っている救急ボイストラ（救急隊用多言語音声翻訳アプリ）があります。これは外国人が病気や事故にあったときに役立っています。

■今後の展望

——次のターゲットは何でしょうか。

木俣 2025年の大阪万博までに、同時通訳を実現させたいと考えています。プ

ロタイプは既に動いていますが、前後の文脈を考慮したり主語や代名詞を推定したりしないとイケないので、更なる研究が必要です。

さらにその5年後の2030年には、外交やビジネスの場においてタフなネゴシエーションのできる高度な翻訳システムを作ることを考えています。

それにはまず優秀な人材を確保することが必須だと思います。私たちは、数多くの人材を投入しているといわれるグーグルと比べると非常に少ない人数で頑張っています。なんとか国立の研究機関という立場を活かして、優秀な学生を集めることができないものかと作戦を練っているところです。

当研究所は、大学にはないような大量の計算機資源とデータがあります。そこに魅力を感じてくれる学生が増えてくれば、優秀な人材に来てもらえるでしょう。

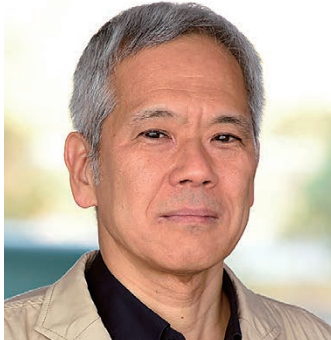
言葉の壁を取り除くというのは、人類の何百年にもわたる夢だったと言えるでしょう。それが、ようやく実現しつつあるのです。これほど面白くワクワクするような分野はないと思います。

ぜひ、多くの若い方々にNICTでの音声翻訳の研究開発に興味を持っていただきたいと思っています。

*1 LSTM : Long Short Term Memory

*2 BERT : Bidirectional Encoder Representations from Transformers

グローバルコミュニケーション計画の実現に向けた多言語音声認識技術の開発



河井 恒

(かわい ひさし)

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

1989年、大学院博士課程修了。同年4月、国際電信電話株式会社入社。音声合成、音声認識の研究に従事。2000～2004年、国際電気通信基礎技術研究所（ATR）に出向し、音声合成の研究に従事。2014年10月よりNICTに出向し、現職。博士（工学）。

先 進的音声技術研究室では、2020年東京オリンピック・パラリンピック競技大会での社会実装に向けた(1) 10言語*の実用的な音声認識技術と、(2) 10言語の実用的な音声合成システムの開発、及び2021年以降の世界を見据えた(3) 現場音声認識技術と、(4) 混合言語音声対話技術の研究を行っています。ここでは、文字数の制約から、(1)に絞って開発と成果の概要を紹介します。

■ 音声認識システムの全体構造

音声認識システムは、図1に示すように音声認識エンジン、音響モデル、言語モデル、発音辞書及び外部辞書から構成

されます。エンジンは、対象言語すべてに共通ですが、それ以外の要素は、対象言語ごとに用意します。

■ DNN 音響モデル

音響モデルは、音声信号の0.01秒程度の断片 x_t が音素 s_t である確からしさ $P(s_t|x_t)$ を計算するモデルです。音響モデルは、DNN (Deep Neural Network) を採用しており、大規模な音声コーパスから深層学習により作成します。音素とは、言語音声の最小単位であり、日本語では、母音、子音合わせて40個程度です。コーパスとは、言語の音声またはテキストを大量に集積・整理したデータです。

音響モデルの構造は、総務省のグローバルコミュニケーション計画（GC計画）が開始された2014年ごろは、図2（1）に示すような、現在の時刻のみの音声入力を受けて結果を出力するシンプルな構造でした。その後改良を重ねた結果、現在では、図2（2）のように前後の時刻の情報を利用する構造となっています。音声は時間構造を持つため、このような前後の時刻の情報を利用する構造が有効に働きます。音響モデルの改良に伴い、音声認識精度を表す単語誤り率も半分程度に改善されました。

■ 音声コーパス

音声コーパスは、1言語当たり1,000～2,000時間の音声データに対して言いよどみ、フィラー（間投詞など）、同型語の読み分けまで含めた正確な文字起こしを人手によって行ったものです。音声

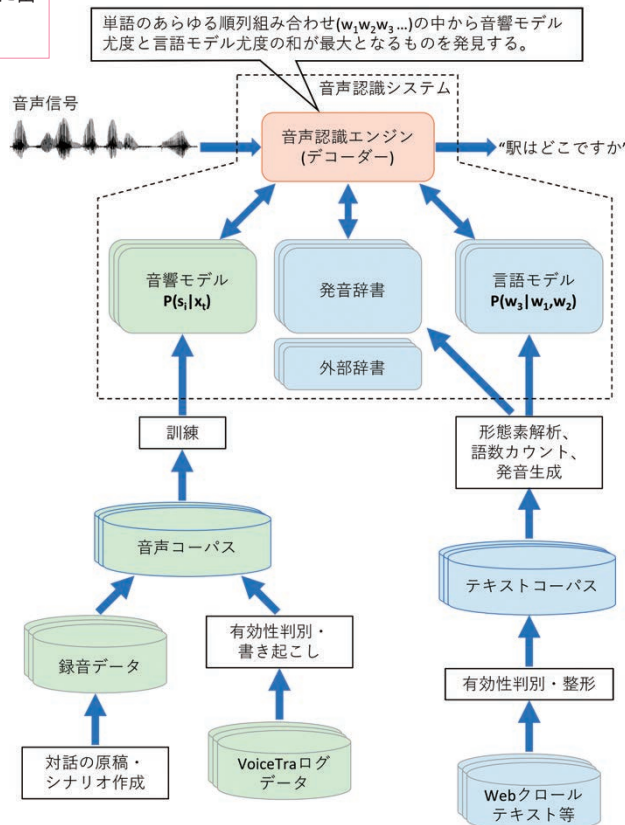


図1 音声認識システムとコーパス

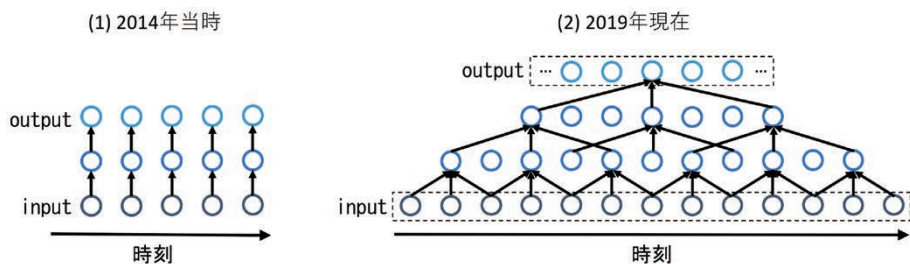


図2 DNN音響モデルの構造の進化

データは、NICTが公開中の多言語音声翻訳実証実験アプリ“VoiceTra”^{ボイストラ}の利用件数が多い日本語、英語、ミャンマー語等では、ログデータを中心に使用しますが、それ以外の言語では、話者を募集し、会議室、屋外等で模擬会話を行うなどして録音した音声を使用します。文長が5～10単語の会話アプリに音声認識を使用する場合は、VoiceTra ログデータを音響モデルの訓練データの主体とした方が良好な結果が得られます。ログデータが少ない言語については、ログデータがある程度たまった段階で訓練データに追加する→音声認識精度改善によりVoiceTraの利用を促進する→ログデータが増加、というサイクルを回す必要があります。

ログデータを使用する場合は、原則として非母語話者の発話を人手で排除します。話者数は、録音データの場合は、1時間当たり6名程度です。文字起こしの誤りは、音声認識精度に悪影響を与えるため、各言語のネイティブ話者によって厳格に品質管理を行い、不良発話数を10%以下に抑えています。

■言語モデル

言語モデルは、発話中に2つの単語の並び w_1, w_2 があった時、その次の単語が w_3 である確率 $P(w_3/w_1, w_2)$ を与えます。図3は、言語モデルの一例です。言語モデルは、テキストコーパスに出現する1～3単語の連鎖を数え上げることによって作成します。テキストコーパス中に事例が存在しない単語の3つ組に対しては、単語2つ組と孤立単語の出現確率を合成

して $P(w_3/w_1, w_2)$ を推定します。日本語のように単語境界を明示しない言語では、形態素解析という処理を行って単語境界を推定する必要があります。

テキストコーパスは、音声コーパスの文字起こしデータだけでは量が不十分なので、Webからクローリングによって自動収集したテキストデータなどを併用しますが、音声認識を適用する分野の文表現(言い回し)を含むテキストを大量に集めることが重要です。Webクローリングコーパスを使用する場合は、HTMLタグ除去などのクリーニング処理に加えて、言語を正しく識別する必要があります。

■発音辞書と外部辞書

言語モデルに含まれるすべての単語を集めた辞書が音声認識可能な単語の範囲となり、新たな単語を音声認識できるようにしたい場合は、その単語を含む文例多数をテキストコーパスに追加し、言語モデルを再作成する必要があります。しかしながら、固有名詞は数が膨大であり、日々新しい単語が作られているため、単語追加が簡単にできないと実用上不都合です。このため、言語モデルに含まれる単語を変数化することにより、単語追加を可能にします。具体的には、図3で黄色マーカーを付した3つの単語を{新幹線}という1つの変数にまとめ、変数のメンバーである「のぞみ」、「つばめ」、「さくら」を外部辞書に登録しておきます。メンバーの確率値は、基本的に均等ですが、ウェイトをつけることも可能です。新たな新幹線の名称を追加する場合は、

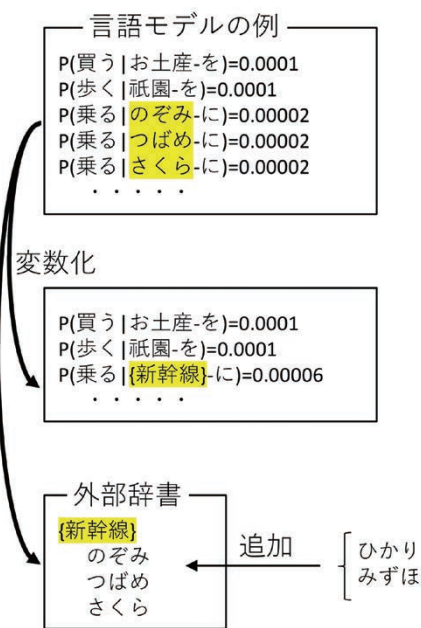


図3 n-gram言語モデルと変数化の例

外部辞書に追加するだけでよく、言語モデルを更新する必要はありません。

言語モデルの辞書には、単語の表記に加えて、単語の発音を音素記号で記述します。これにより音響モデルを介して言語モデルを入力音声と突き合わせる事が可能になります。発音は、人手で正確な表記を指定するのが理想的ですが、単語数が多い場合は、G2P (Grapheme-To-Phoneme) という処理によって自動生成します。複数通りの発音がある場合は、すべて列挙しておけば、入力音声に即した候補が自動的に参照されます。

■社会の様々な場所で利用中

当研究室では、既に前記10言語についてSNR (信号雑音比) が5dB以上の環境での使用に適した音声認識システムを開発し、民間企業等にライセンスを供与しており、様々な音声翻訳サービス・アプリで使用されています。さらに、日本で需要の多いブラジルポルトガル語、フィリピン語、ネパール語の音声認識システムの開発も進めています。

* 10言語：日本語、英語、中国語、韓国語、タイ語、インドネシア語、ベトナム語、ミャンマー語、スペイン語、フランス語

グローバルコミュニケーション計画の実現に向けた多言語翻訳技術の開発



隅田 英一郎

(すみた えいいちろう)

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

大学院修士課程修了後、日本IBM、国際電気通信基礎技術研究所（ATR）を経て、2007年NICT入所、多言語音声翻訳、e-Learningの研究に従事。博士（工学）。

先 進的翻訳技術研究室では、①2020年東京オリンピック・パラリンピック競技大会での社会実装に向けた10言語の実用的な自動翻訳技術の開発及び2021年以降の世界を見据えた同技術の高度化として、②対訳依存性を最小化する研究と③同時通訳の基礎研究を行っています。本稿では分かりやすい話題である①と③について紹介します。

■自動翻訳の仕組み

最先端の自動翻訳は、対訳データ（入力言語文と翻訳である出力言語文を対にしたもの）を大規模に用意し、そこから翻訳知識を学習します。最近では、AIの基本技術である深層学習によって、翻訳知識をニューラルネットで表現します（この技術はニューラル翻訳（NMT）と呼ばれています）。対訳データから学習する方法は1980年代に遡り、世界中で多数の研究者が様々なアルゴリズムを研究し精度も改善を続けていたのですが、最近は改善速度が落ちていました（この

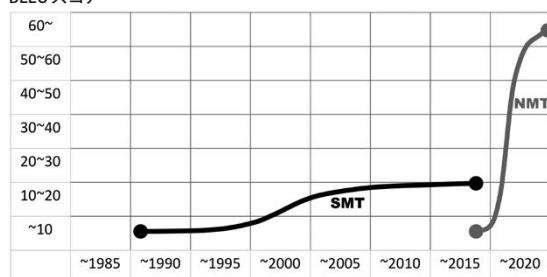
技術は統計的翻訳（以下SMT）と呼ばれています）。図1に示したようにSMTの停滞後に登場したNMTは瞬く間にSMTの性能を追い抜き、現在も継続的に性能改善中です。

■公開サイト「みんなの自動翻訳」

NICTで研究開発している自動翻訳は、2020年東京オリンピック・パラリンピック競技大会での社会実装に向けた10言語の実用的なシステムとして、多言語音声翻訳の無償トライアルアプリ VoiceTra^{ボイストラ}*2として公開され、さらに、広く技術移転先企業各社のアプリや専用機に搭載され活用されています。

また、図2のみんなの自動翻訳^{テキストトラ}@TexTra^{テキストトラ}*3では、ログインのアカウントをお作りいただくだけでNICTの開発したNMTの実力をご評価いただけます。当サイトは、非商用のトライアル専用としており無料で利用可能です。また、NICTの技術移転先企業では、当サイトと同等性能やUIに各社の付加価値を付

BLEUスコア



* 同じ対訳データにおける翻訳精度 BLEU^{スコア}*1 の変化（BLEUが高いほど高性能）

図1 学習方法変更による劇的性能向上～SMTからNMTへ移行～



図2 みんなの翻訳@TexTraでNICTのNMTをご体験ください。
<https://mt-auto-minhon-mlt.ucrri.jgn-x.jp/>

けて商用サービスを提供しています。

当サイトは、日本語・英語・中国語・韓国語の4言語を中心に多言語に対応しており、一般的なWeb翻訳と同様に、どなたでも簡単にお使いいただけます。

さらに、翻訳支援エディタをインストールなしでご利用いただけるほか、ワードやエクセルなどのオフィスソフトウェアのアドインが利用可能です。また、各種翻訳支援ツール（Trados等）から自動翻訳エンジン呼び出すことができますし、自分のプログラムから直接WebAPIを呼ぶこともできます。

このように、みんなの自動翻訳は、自動翻訳に関する一式を便利に提供しています。

■公開サイト「翻訳バンク」

実用化には、NMTだけでなく、「翻訳バンク」*4が重要です。「翻訳バンク」とは、総務省とNICTが連携しながら、NICT*5に翻訳データを集積し自動翻訳の多分野化・高精度化を進める取組です。図3に示すように、ファイナンス、医薬、製造等の分野ごとに対訳データを集め、分野専用のNMTを作ることによ

て高精度化を実現します。

NMTは優れた自動翻訳アルゴリズムではありますが、学習（訓練）データである対訳データがなければ自動翻訳エンジンはできません。NICTでは、以前から、特許庁などと協力して、対訳データの収集に努めてきました。翻訳バンクはこの取組を加速するものです。

まず、NMTの訓練自体は、非常に大まかに言えば、現時点のNMTで原文を翻訳してみて、その結果が対訳の出力側と異なる度合いに応じて、ニューラルネットのパラメータを調整することです。この過程を大量の対訳文に対して何回か繰り返すことにより、共通基盤となるNICT汎用システムが構築されます。

次に、上記パラメータ調整が完了した汎用システムに対して、翻訳対象分野の対訳文を利用して、同様なパラメータ調整を追加実施するのがアダプテーション（適応）です。このように、アダプテーションでは、既に学習済みの汎用NMTパラメータを翻訳分野に応じて追加調整するので、比較的少量の対訳データで当該分野に対する高精度な翻訳エンジンを構築できます。

表1に医薬のアダプテーションの例を

示しました。圧倒的な訳質の向上をご確認いただけると思います。

■同時通訳にむけた翻訳の漸次化

NICTは、いずれ深層学習による同時通訳も可能になると考えています。既に、そのプロトタイプを稼働させています。プロトタイプでは英語話者が話すいくつかの単語で翻訳可能な塊にまとまった時に、日本語の訳として出力されていきます。例えば、「I'd like to have some tea」という入力が入力が翻訳可能かどうかはもう少し先を見ないと判定できません。続く言葉が「and cake」なら切れませんし、「how much」なら切れるでしょう。このように、入力されてくる単語の関係や文法あるいは無音区間等様々な情報源を統合して、訳出すべき単位を決めます。

最後まで聞いて翻訳すると時間がかかってしまうので、途中でポンポンと出していくのが同時通訳の特徴です。これからの性能を高めていきたいと考えていますが、英語を一生懸命聞き取っていくよりも、時々訳誤があっても日本語を五月雨的に理解する方が、日本人の大半にとってははるかに楽ですね。

この同時通訳技術を応用すれば、テレビ会議システムとつなげて東京の本社と外国の支社との間で、それぞれ母語でコミュニケーションすることもできます。

自動翻訳が進化すれば、外国人と一緒に仕事できる領域が広がると思います。

■おわりに

NICTでは、最新のNMT研究成果を展開すると同時に、翻訳バンクにより集積された対訳データを活用することにより、実用的で高精度なNMTを社会還元することを目標としています。皆様のご協力をよろしくお願いいたします。

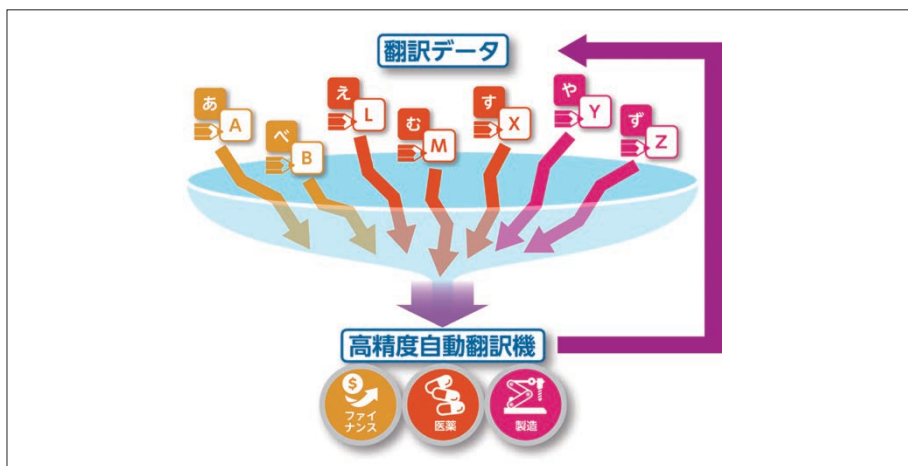


図3 翻訳バンクの概念図

表1 アダプテーションの例（医薬）

原文	A majority of subjects had AEs that were mild or moderate in severity and not related to investigational product (IP).
汎用NMT	大多数の被験者は、重症度が軽度または中等度のAESを有し、治験薬 (IP) とは無関係であった。
アダプテーション後 NMT	大多数の被験者では、軽度又は中等度のAEが認められ、治験薬との因果関係は否定された。

*1 BLEU:自動翻訳の性能を評価する自動尺度の一つ。BLEUでは、自動翻訳の結果の訳文と、参照訳としての手による訳文との類似度が高いほど、高い数値となる。

*2 <https://voicetra.nict.go.jp/>

*3 <https://mt-auto-minhon-mlt.ucrijn-x.jp/>

*4 <https://h-bank.nict.go.jp/>

*5 NICT内では先進的音声翻訳研究開発推進センター (ASTREC) と知能科学融合研究開発推進センター (AIS) が協力して運営している。

研究成果を社会実装するための ソフトウェア開発とサービス運用



葦苅 豊

(あしかり ゆたか)

先進的音声翻訳研究開発推進センター
統合システム開発室
室長

大学院博士前期課程修了後、数社のソフトウェア開発企業を経て、2001年より株式会社国際電気通信基礎技術研究所（ATR）、2006年NICT入所。音声翻訳システムの研究開発に従事。

統合システム開発室では、先進的音声翻訳研究開発推進センターの研究成果である音声認識、音声合成、言語翻訳技術などの社会実装を進めるために、音声翻訳システム等の統合システムを開発し、すぐに商用利用できる品質にするとともに、広く世間に技術を周知するためにスマートフォン用の多言語音声翻訳アプリ（VoiceTra）として一般公開しています。また、各種技術の応用例として様々な音声翻訳アプリケーションを開発し、展示会等で提案し続けています。

■多言語音声翻訳アプリ VoiceTra ボイストラ

VoiceTra（図1）は、話した文を他の言語に翻訳して読み上げる多言語音声翻訳アプリです。例えば、日本語で話すとそれを英語に翻訳して英語の音声として読み上げます。2010年7月に世界初のクラウド型音声翻訳アプリとして一般公開しました。その後、いったんは民間企業にサービスの運営を移管しましたが、2015年5月からは、再びNICTによる開発と運用を再開しました。翻訳できる言語はアジア言語を中心に31言語で、そのうち、音声入力できる言語は18言語、音声出力できる言語は16言語です*。また、2019年10月からは、入力された音声は何語かを自動的に判別する言語識別機能を追加しました。これにより相手の言語が分からない場合でも会話することができるようになりました。2019年12月時点で識別できる言語は8言語ですが、今後増やしていく予定です。

このようにVoiceTraには、研究成果を世間に知っていただくためのショーケースという一面がありますが、もうひ

とつの重要な役割としてデータ収集という一面もあります。VoiceTraアプリは、2019年11月30日時点で約560万ダウンロードされており、利用数も累計で2億発話（音声入力された数）を超えています。この大量の音声データをAI学習に利用することにより性能向上を図っています。

■音声翻訳サーバシステムの開発と運用

クラウド型の音声翻訳システムを稼働させるためには、先進的音声翻訳研究開発推進センターの各研究室で開発された音声認識エンジン、音声合成エンジン、言語翻訳エンジン、言語識別エンジン等



図1 VoiceTraの画面

* <https://voicetra.nict.go.jp/>

をサーバに配置するだけでなく、スマートフォン用アプリとインターネット経由で通信し、アプリからのリクエスト（例えば、日本語から韓国語に翻訳し、女性の声で読み上げるという命令）に従って適切なエンジンを利用して処理を行い、リクエスト元のアプリに処理結果を返す仕組みが必要です。これを実行するのが音声翻訳サーバソフトウェアです。

統合システム開発室では、すぐに商用利用可能な品質水準のサーバソフトウェアを開発するとともに、各研究室のエンジンの品質テストも行い、音声翻訳システム全体の品質を担保することによって、研究成果の社会実装推進に寄与しています。

■応用システムの提案

VoiceTraは、研究成果を利用したひとつの実現形ですが、音声翻訳技術を利用してほかにも様々なシステムを構築することが可能です。当室では、継続的に様々なデモンストレーションシステムを開発し、展示会等で提案することにより、サービス提供者やシステム開発者の興味を刺激し続けています。これまでに開発

したシステムの例を以下に示します。

(1) イヤホン型音声翻訳システム

お互いにイヤホン型ヘッドセットを装着するだけで、お互いが母国語で話した内容が相手の言語に翻訳されて、それが相手のイヤホンで聞こえるシステムです。周りで見ている人には、例えば、日本人が日本語で話し、アメリカ人が英語で話して、会話が成立しているように見えます（図2）。現在は、それぞれが携帯するスマートフォンとイヤホン型ヘッドセットが繋がって音声翻訳を実現していますが、近い将来、イヤホンを装着するだけで音声翻訳できるデバイスの開発が期待されます。

(2) リアルタイム字幕付与システム

音声認識結果、または、それを翻訳した結果を字幕として表示するシステムを各種開発しています。図3は、パワーポイントのスライド上に日本語で話した説明の内容を英語に翻訳した結果を画面下部にリアルタイムに表示したものです。日本語話者と外国語話者が混在している国際会議などでの利用を想定しています。また、翻訳せずに、例えば、日本語

の音声認識結果のみを字幕として表示することもでき、聴覚障害者の情報保障として十分実用になるレベルに到達していると考えています。今後は、デモンストレーションだけではなく、ろう学校などと協力して実証実験を行っていく予定です。

■今後の展望

これまでの音声翻訳システムは、比較的短い一文単位の翻訳でしたが、今後はスピーチなどの長い話を同時通訳的に翻訳するステージに突入します。そのためのエンジンの研究開発は各研究室に委ねますが、同時通訳システム用のサーバソフトウェアも新たに必要となります。同時通訳システムでは、文脈の利用やマルチモーダル情報の利用などで音声認識や翻訳の精度の向上を図ることが検討されており、今後は様々な情報を普遍的に扱える枠組みが必要となります。現時点では、エンジン側が要求する情報やその保管方法等が不明ですが、エンジンの性能を120%生かすような拡張が容易で柔軟なプラットフォームを開発していきたいと思っています。



図2 イヤホン型音声翻訳システム



図3 リアルタイム字幕付与システム

グローバルコミュニケーション計画の実現に向けた多言語音声翻訳技術の社会展開



内元 清貴

(うちもと きよたか)

先進的音声翻訳研究開発推進センター
企画室
室長

1996年大学院修士課程修了。同年、郵政省通信総合研究所(現NICT)に入所。以来、自然言語処理、音声翻訳に係る研究開発・社会実証及び研究成果の社会還元に従事。2015年から現職。博士(情報学)。

- *1 https://www.soumu.go.jp/main_content/000285578.pdf
- *2 <https://voicetra.nict.go.jp/>
- *3 <https://gcp.nict.go.jp/>
- *4 <https://www.nict.go.jp/press/2015/10/26-1.html>
- *5 <https://news.kddi.com/kddi/corporate/newsrelease/2019/11/12/4134.html>
- *6 <https://pr.fujitsu.com/jp/news/2017/09/19.html>
- *7 https://www.hitachi-solutions-tech.co.jp/iot/solution/voice/Ruby_Concierge/railway_index.html
- *8 https://www.keikyu.co.jp/company/news/2017/20180328HP_17271TS.html
- *9 <http://www.madoguchi-honyaku.jp/>
- *10 https://gcp.nict.go.jp/news/products_and_services_GCP.pdf
- *11 <https://www.toppa.co.jp/news/2018/05/newsrelease1805141.html>
- *12 <https://iamili.com/ja/>
- *13 <https://panasonic.biz/cns/inv/taimenhonyaku/>
- *14 <https://pocketalk.jp/>
- *15 <https://www.konicaminolta.jp/melon/>
- *16 <https://www.toppa.co.jp/news/2018/04/newsrelease1804252.html>
- *17 https://miraitranslate.com/uploads/2019/04/MiraiTranslate_MultilingualPlatform_pressrelease_20190426.pdf
- *18 https://www.fairydevices.jp/mimi_about/mimi_price.html
- *19 https://jpn.nec.com/cloud/service/platform_service/multilingual/index.html
- *20 <https://h-bank.nict.go.jp/>

近年、訪日外客数は増え続けており、2020年には4,000万人を超えられていると言われています。このような状況の下、総務省は2014年4月に、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}」を発表しました。この目標を達成すべく、NICTでは民間企業と共にオールジャパン体制で多言語音声翻訳技術の性能向上と対応可能な言語・分野の拡大及び実証実験・社会実装を進めています。この技術を世の中に普及させることにより、言葉で困らない社会の実現を目指します。

■ 音声翻訳の仕組み

NICTの多言語音声翻訳技術は、ネットワーク型の音声翻訳アプリ「VoiceTra^{ボイストラ}®^{*2}」に実装しApp StoreやGoogle Playで公開しています。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ます。スマートフォンから入力された音声は、図1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ます。サーバ内の各処理は、いずれも、言葉のデータベース(コーパス)から統計的に学習し処理する機構を採用しています。音声翻訳の精度は基盤となるコーパスの量と質に大きく依存します。NICTでは、対象分野を絞り、VoiceTraの利用ログ情報も活用し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ました。これまで2020年を目途に、旅行会話から生活、災害、医療の分野へ対応範囲を広げてきました。

■ 産学官連携の取組

グローバルコミュニケーション計画の推進に資することを目的として、2014年12月に設立されたグローバルコミュニケーション開発推進協議会^{*3}、2015年10月に設立された総務省委託研究開発・多言語音声翻訳技術推進コン

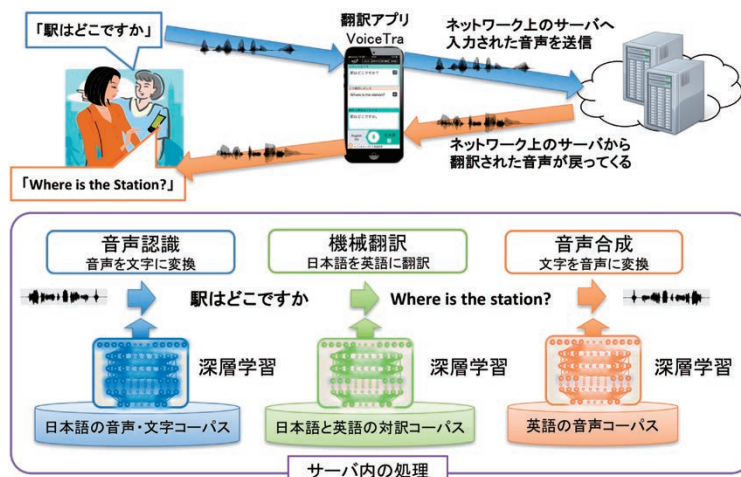


図1 ネットワーク型多言語音声翻訳の仕組み

ソーシアム^{*4}では、NICTと民間企業が協力してハードとソフトの両面で社会実装に向けた研究開発や実証実験に取り組み、様々なシーンで、様々なユーザインタフェース (UI) により技術検証を行いました。例えば、タクシー分野では、KDDIが開発した運転座席・後部座席連動型の音声翻訳機を活用し、鳥取市、東京都、那覇市などで実証実験を実施しました^{*5}。医療分野では、富士通研究所が開発したIDカード型ハンズフリー音声翻訳端末^{*6}などを活用し、全国各地の病院や介護施設で臨床試験や実証実験を行いました。鉄道分野では、京浜急行電鉄、ブリックス、日立グループとの共同研究の成果を活用し、日立ソリューションズ・テクノロジーが新たな鉄道向け多機能翻訳アプリ“駅コンシェル[®]*7”を開発・商用化し、2018年7月、京浜急行電鉄の全駅（泉岳寺駅を除く）に本格導入しました^{*8}。ここでは、鉄道分野でよく用いるフレーズへの対応、UIの工夫、電話通訳との組み合わせなどにより、鉄道分野に適したアプリを実現しました。

消防庁の消防研究センターとの共同研究では、VoiceTraに定型文機能を追加したアプリ“救急ボイストラ”を開発しました。このアプリは2019年10月1日時点で、47都道府県における726消防本部中476本部（65.6%）で導入済みです。警察関連では、29都道府県の警察でVoiceTraが試験利用されており、岡山県警のように独自のアプリとサーバによる運用を始める県警も増えています。

日本に長期滞在する外国人も多く、自治体での多言語対応も求められています。NICTでは、委託研究「自治体向け音声翻訳システムに関する研究開発^{*9}」により、自治体窓口業務を対象としたコーパスの構築や音声翻訳サーバ、アプリの試作を行い、前橋市、板橋区、綾瀬市などの自治体窓口で実証実験を行いました。

■多言語音声翻訳技術の社会実装

分野や使われるシーンに合わせた商用製品・サービス^{*10}も複数生まれています（図2参照）。例えば、凸版印刷の音声翻訳アプリ“VoiceBiz[®]”^{*11}、ログバーのオフライン音声翻訳機“山^{イリー}”^{*12}、日本電気の多言語音声翻訳サービス、パナソニックの多言語音声翻訳サービス“対面ホンヤク[®]*13”、ソースネクストのクラウド型音声通訳機“POCKETALK[®] W/POCKETALK[®] S”^{*14}、コニカミノルタの医療通訳タブレット“MELON[®]*15”などです。凸版印刷の音声翻訳アプリは、日本郵便「郵便局窓口音声翻訳^{*16}」として、全国約20,000の郵便局（簡易郵便局は除く）に導入されました。自治体窓口での実証実験を踏まえ、自治体による採用も広がりつつあります。これらの商用製品・サービスは、NICTの多言語音声翻訳技術のライセンスを受けて実現されています。2019年4月には、みらい翻訳により、多言語音声翻訳プラットフォームのサービス提供と音声翻訳ソフトウェアのライセンス事業が開始されました^{*17}。音声翻訳のAPIサービスはフェアリーデバイス^{*18}や日本電気^{*19}からも提供されています。

■今後の展開

多言語音声翻訳をいつでもどこでも使える技術とするためには、更なる性能向上と共に使い方の工夫が必要です。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・構築です。NICTでは、低コスト化や多分野化、技術の高精度化の加速のため、産学官の協力により対訳コーパスをNICTに集積する“翻訳バンク[®]*20”の運用を開始し、拡充しています。多言語音声翻訳技術の使い方の工夫としては、状況によって、地図などの併用や、電話通訳サービスとのシームレスな接続など、使われるシーンに合わせてベストミックスとなる組み合わせを見つけることが重要です。

グローバル化が加速する中、ビジネス・国際会議など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も広がると予想しています。総務省では次期グローバルコミュニケーション計画の策定が進んでいます。今後、大阪・関西万博が開催される2025年までに、文脈などから多様な情報を取り込み、高精度と低遅延を両立する実用レベルの同時通訳の実現を目指すとともに、その技術の普及に努めます。

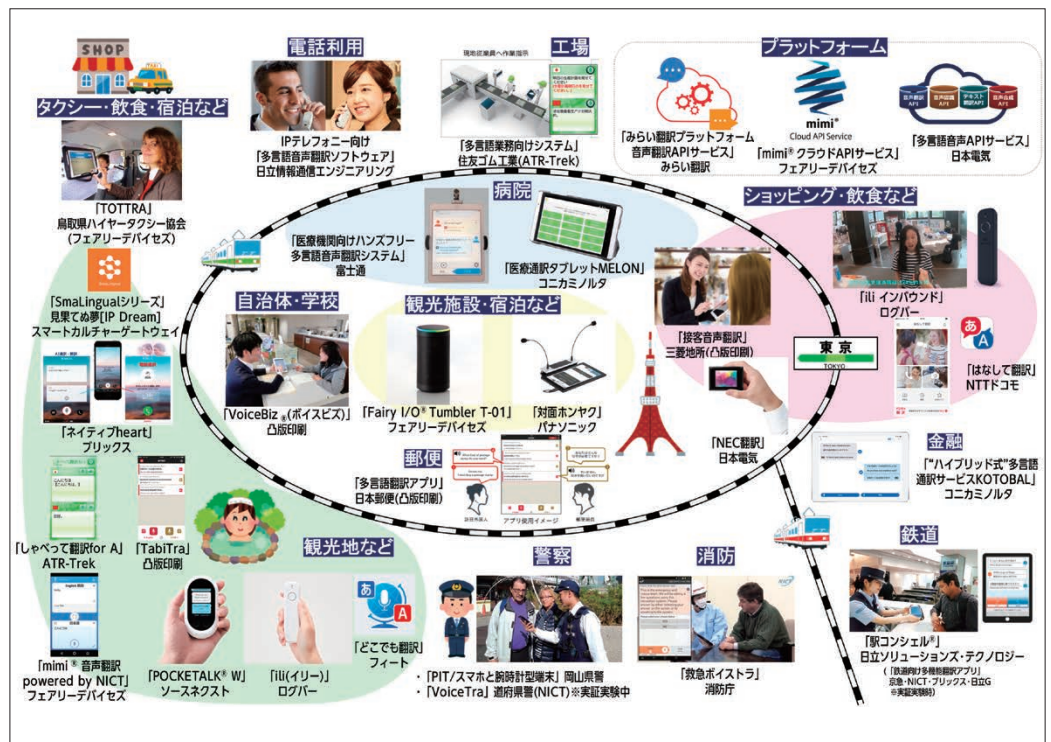


図2 NICT多言語音声翻訳技術の社会展開例

光出力500mWを超える極めて高出力な 265nm帯深紫外LEDを実現

未来ICT研究所
深紫外光ICTデバイス先端開発センター
センター長 井上 振一郎

光とは様々な波長をもつ電磁波の一種ですが、波長範囲としておおむね200~300 nmの深紫外 (Deep-Ultraviolet: DUV) 光は、空气中を伝搬できる光の中で最も波長の短い光に分類されます。この深紫外光の中でも特に、波長280 nm未満 (UV-C領域) の光は、オゾン層ですべて吸収されるため、自然界 (地表) の太陽光の中には含まれておらず、極めて特異な性質を有しています。例えば、深紫外光が降り注がない環境下で進化してきたウィルスや細菌は、深紫外領域の中に強い吸収構造を持つため、深紫外光を照射することで、塩素などの薬剤を用いずに、ウィルスや細菌を極めて効果的かつクリーンに殺菌・無害化できます。また自然界に存在しない深紫外光は、太陽光の背景ノイズの影響を受けない通信・センシングや、大気中の高い散乱係数を利用した見通し外 (Non Line of Sight: NLOS) 光通信などへの応用が期待されています。このような応用以外にも、空气中を伝搬できる光の中で最も波長の短い深紫外光は、光加工や3Dプリンタの高精細化、樹脂の硬化、印刷、環境汚染物質の分解、分光分析、医療応用など、多様な技術領域において今後画期的な役割を果たしていくものと期待されています。

従来、深紫外光を照射する光源として、産業的には主に水銀ランプが用いられてきました。しかし、水銀ランプは人体や環境に有害な水銀を含み、大型で環境負荷の高い製品であるため、水銀廃絶に向けた「水銀に関する水俣条約」(2017年8月16日発効) などにおいて、現在その代替が強く求められています。このため、深紫外発光ダイオード (Light-Emitting Diode: LED) の研究開発が、世界中の研究機関において活発に進められています。しかしながら、深紫外LEDの光出力は水銀ランプと比べると極めて微弱で、その高出力化が課題でした。

NICT未来ICT研究所 深紫外光ICTデバイス先端開発センターでは、ナノ光構造技術を基盤とした深紫外LEDの高出力化に取り組んできました。内部光吸収や光出力飽和現象 (効率ドループ) を抑制する、ナノ光取出し構造や半導体デバイ

ス・チップ電極構造・パッケージ、実装技術等の総合的な開発を進め、単チップ (室温・連続駆動) において、世界最高出力となる520 mW超の265 nm帯深紫外LEDの実証に成功しました (図1)。

本深紫外LEDは、環境にやさしく、小型・ポータブル・高出力で、水銀ランプの置き換えだけではなく、持ち運び可能なウィルス殺菌システムやポイントオブケア (臨床現場即時) 型医療、家電搭載など、小型・低電圧駆動・ウォームアップ不要なDUV固体光源の特色を活かした、これまでにない様々な新しい応用分野の開拓が期待されます。現在、社会実装への取組として、民間企業との産官連携も積極的に展開しています。今後は、早期社会普及、新規産業創出への貢献に向けて、性能や信頼性の更なる向上を図っていくとともに、NLOS光通信やDUVセンシングなどの画期的な光ICT応用への拡張も視野に入れ、DUV光制御素子開発等を含めた深紫外光デバイスの研究開発を継続していきます。

なお、今回の成果は、応用物理学会誌「応用物理」(2019年10月号) に掲載され、その表紙に取上げられました (図2)。



図2 本成果が飾った「応用物理」2019年10月号の表紙写真。開発した深紫外LEDの表面に形成したナノ光構造の電子顕微鏡写真及び、筆者の言葉が掲載。

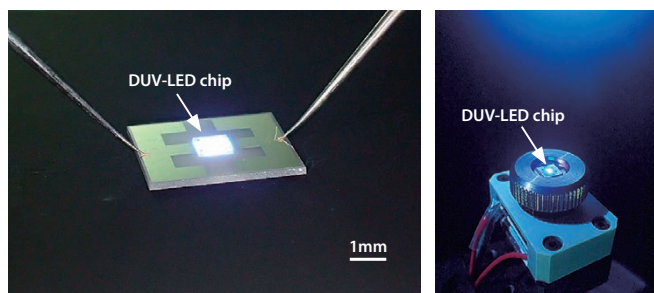


図1 電流を印加し発光中の深紫外LEDの外観写真 (左: サブマウント上、右: 実装後)

井上 振一郎 (いのうえ しんいちろう)

大学院博士課程修了後、理化学研究所基礎科学特別研究員、九州大学助教を経て2010年NICT入所。神戸大学大学院工学研究科客員教授。06年光科学技術研究振興財団研究表彰、14年応用物理学会論文賞、18年独創性を拓く先端技術大賞フジサンケイビジネスアイ賞など。博士 (工学)。



計算言語学・自然言語処理の アジアを中心とした多言語化



● 経歴

- 1986年 中国 山東省 済南市にて誕生
- 2009年 中国 山東大学 数学統計学院 卒業
- 2012年 筑波大学大学院 博士前期課程 修了
- 2015年 筑波大学大学院 博士後期課程 修了後、NICT入所
- 2018年 現職

● 受賞歴等

Best Paper Award of PACLING 2017

丁 塵辰

(テイ ジンシン)

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
テニューアトラック研究員
博士 (工学)

一問一答

Q 生まれ変わったら？

A 複数の母語を持ちたいです。世界に対する見方・考え方が広まるかもしれません。言語類型論の視点から典型的な孤立語（中国語等）・膠着語（日本語等）・屈折語（ドイツ語・ロシア語等）をそれぞれ1つ身につけたいです。

Q 最近ハマっていること

A クラリネット練習（2018年から13年ぶり）昨年9月からウェーバーの小協奏曲（作品26）に挑戦し始めています。

Q 研究者志望の学生さんにひとこと

A 君子不器

世 界の言語数は6,000に上ります。ところが、自然言語処理技術研究の対象言語はヨーロッパ言語及びごく少数のアジア言語に偏しており、他の数多くの言語に関する研究は未熟・未開拓です。また、各言語のデータ整備も不足しがちです。広く世界の言語間で翻訳するという究極の目的を実現するためには、未熟・未開拓・未整備などの課題を克服しなければなりません。

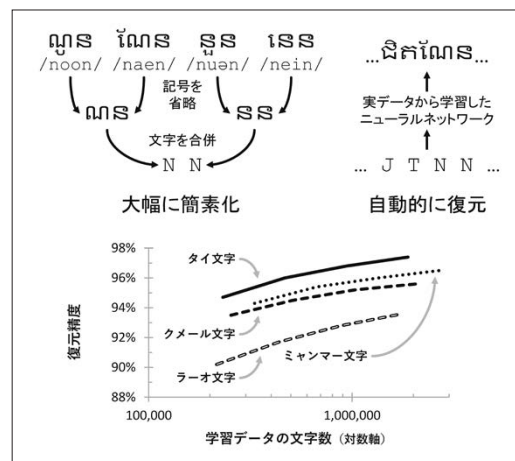
しかしながら、現実的には世界の全言語を対象にはできないので、研究の波及効果及び社会経済的な影響を考慮すると、NICTではアジア諸言語を優先すべきだと判断できます。私は、アジア諸言語を主たる対象とした言語データの整備及び汎用的言語解析技術創出を目指しています。今まで70年以上にわたって発展してきた言語処理技術をアジア諸言語に創造的に応用することにより、アジア諸言語の文字・音声における自動処理の水準を英語・日本語・中国語と同等の高

い水準にまで引き上げたいと思っています。

私は、これまでの研究の成果のひとつとして、多くのアジア言語で使われているアブギダ系文字の効率的な入力プログラム『AKKHARA』を開発しました。文字の入力は、IT化や言語処理の基盤となるものですが、各国において必ずしも十分な技術ができておらず、混乱が生じたり、時間を浪費したりしています。この

問題解決のため、現在、このAKKHARAをミャンマー、カンボジアをはじめとするアジア諸国のPC利用者へ普及させる活動を行っています。

本研究の実行により、計算言語学・自然言語処理など研究分野に貢献でき、さらに、アジア地域における情報通信技術の発展、日本の企業のアジアへの進出、最終的にはアジア全体の繁栄に寄与できると信じています。



複雑な用字系において効率的な入力を実現するため、文字体系にある冗長性を低減し、最先端のニューラルネットワーク技術による自動的・高精度の復元ができる。図の上部は、カンボジアのクメール文字を例として、簡素化・復元のイメージを示している。下部は、タイ文字・ミャンマー文字・クメール文字・ラーオ文字における復元精度を示している。利用される実データ量の増加に従い、復元性能が向上する。

パーマネント研究職・総合職 採用2021

NICTは、情報通信分野を専門とした我が国唯一の研究機関です。研究者と総合職が一体になることで、高いパフォーマンスを発揮し、情報通信の分野で『安心・安全で豊かな社会の実現』を目指して仕事をしています。このNICTで、是非一緒に様々なことに挑戦していきませんか？

研究職・研究技術職・テニュアトラック研究員

- 募集職種 パーマネント研究職員、パーマネント研究技術職員及びテニュアトラック研究員
- 採用時期 2021年4月1日
(場合により早期採用の可能性有り)
- 応募方法 弊機構採用情報のwebページからのエントリー
(研究職) https://www.nict.go.jp/employment/research_staff.html
(研究技術職) https://www.nict.go.jp/employment/technical_staff.html
- 応募締切 2020年4月10日(金) 17:00
- お問い合わせ
総務部人事室人事グループ/経営企画部 研究職 採用担当
MAIL: jinji-r@ml.nict.go.jp TEL: 042-327-7304

総合職

- 仕事内容 研究開発の推進及び研究開発成果の社会還元のため、経営企画、人事、財務、法務、広報等の組織マネジメント、産学官・地域連携、国際連携、知的財産管理、技術移転などの業務に従事します。
- 応募資格 4年制大学(海外の大学を含む)以上を2021年3月に卒業・修了する見込みの方。あるいはそれらを卒業・修了した30歳以下(1990年4月1日生まれまで)の方。
※学部や専攻は問いません。
- エントリー方法
下記URLからマイナビ2021に登録後、エントリーをされた方へ順次ご案内いたします。
<https://www2.nict.go.jp/employment/clerk/clerk.html>
※マイナビにエントリーできない方は、下記までお問い合わせください。
- お問い合わせ
総務部人事室人事グループ 総合職 採用担当
MAIL: jinji@ml.nict.go.jp TEL: 042-327-7304

2020年3月1日

エントリースタート!

その他、詳細については、採用情報のURLをご覧ください。

<http://www.nict.go.jp/employment/index-top.html>

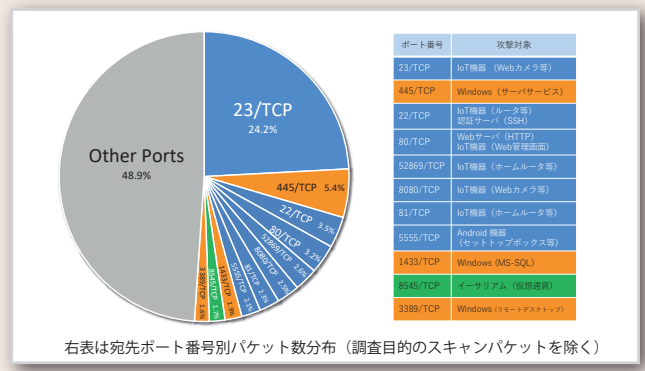
Interop Tokyo 2020

2020年4月13日(月)～15日(水)
会場：幕張メッセ(国際展示場)

展示内容

- インシデント分析センターNICTER
- ナショナルサイバートレーニングセンターの3事業
- NICTER スピンオフ技術 社会展開最新事例

NICTER 観測レポート2019の公開



NICTは、2月10日、サイバーセキュリティ研究所において、NICTER観測レポート2019を公開しました*。NICTERプロジェクトの大規模サイバー攻撃観測網で2019年に観測されたサイバー攻撃関連通信は、2018年の約1.5倍と著しい増加をみせましたが、その内訳をみると、主に海外からの調査目的とみられるスキャンが観測パケット全体の53%を占め(2018年は全体の35%)、マルウェアによるとみられる通信はほぼ横ばいでした。詳細は以下をご覧ください。

https://www.nict.go.jp/cyber/report/NICTER_report_2019.pdf

* NICTサイバーセキュリティ研究所では、NICTERプロジェクトにおいて大規模サイバー攻撃観測網(ダークネット観測網)を構築し、2005年からサイバー攻撃関連通信の観測を続けてきました。

NICTオープンハウス2020

in 小金井

6月12日(金)・13日(土)開催

- 「セカイを変えるICT」をテーマに、幅広い年代層の方々にお楽しみいただける企画を準備しています。
- 皆様のご来場をお待ちしています。
- 詳細は、3月中旬ごろHPに掲載します。