


NTTコミュニケーション科学基礎研究所
 上席特別研究員

永田 昌明 Masaaki Nagata

進化・変化が著しい機械翻訳で、 より精度が高く使いやすい システムをめざす



生成AI（人工知能）ChatGPTが、急速に世の中で注目されてきています。ChatGPTは、入力された文や単語を基に新たな文を生成して出力するもので、特に指定がない限り入力文の言語で文が出力されます。ところが、出力の言語を指示すると翻訳された文が出力される、機械翻訳としての利用が可能です。20年以上にわたり自然言語処理とその応用である機械翻訳の研究に取り組んできた、NTT コミュニケーション科学基礎研究所 永田昌明上席特別研究員に、商用化を抑えた日英特許対訳コーパスによる機械翻訳、大規模言語モデル（LLM）による翻訳の動向や特徴、そして研究プロセス・アイデアは「出会い」という思いについて伺いました。

 **機械翻訳は統計的機械翻訳からニューラル翻訳,そしてLLMを活用した翻訳へ**
 現在、手掛けていらっしゃる研究について教えていただけますでしょうか。

機械翻訳における、日英対訳コーパスと単語対応に関する研究を行っています。

前回（2021年4月号）は、統計的機械翻訳からニューラル翻訳に変わる過渡期的話をさせていただきました。この数年間で急速

に普及が進んできたニューラル翻訳では、対訳データを収集することが重要です。そこで、WebをクローリングしてJParaCrawlという2000万文対を超える日英対訳コーパスをつくった経験に基づいて、特許の対訳データをつくってみました。特許は公文書で、日本でも米国等の海外でも特許出願の文書が全部公開されています。私たちには大規模対訳データをつくるノウハウがあったので、それを活用したところ3億文対を超える対訳データを日英特許対訳コーパスとしてつくることができました（図1）。

この技術の社会における応用として、海外への特許出願時等における関連文書の作成・チェックが考えられます。海外への特許

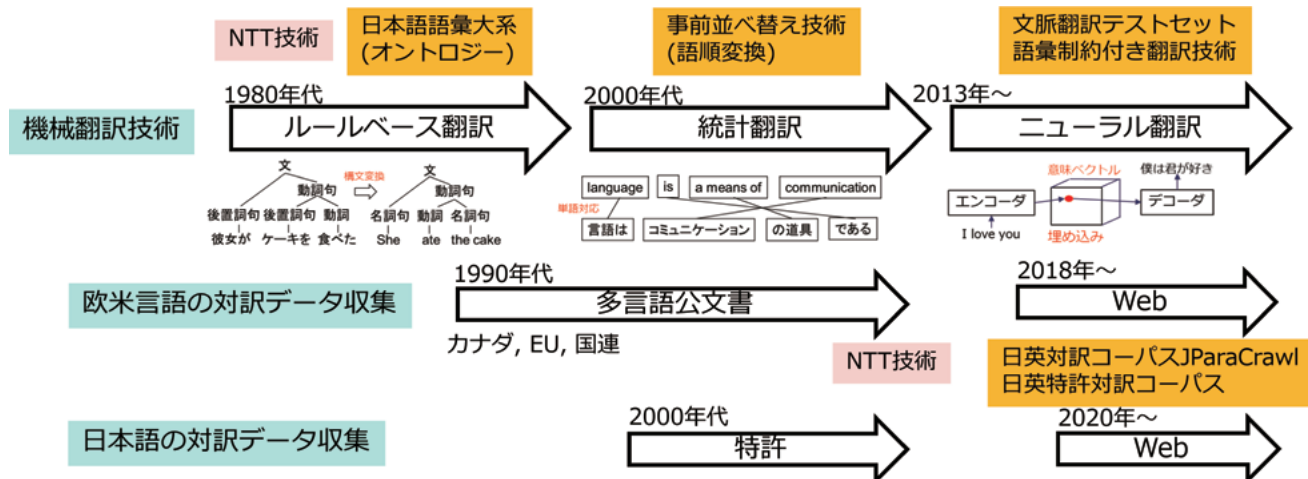


図1 機械翻訳の歴史とNTTの研究開発

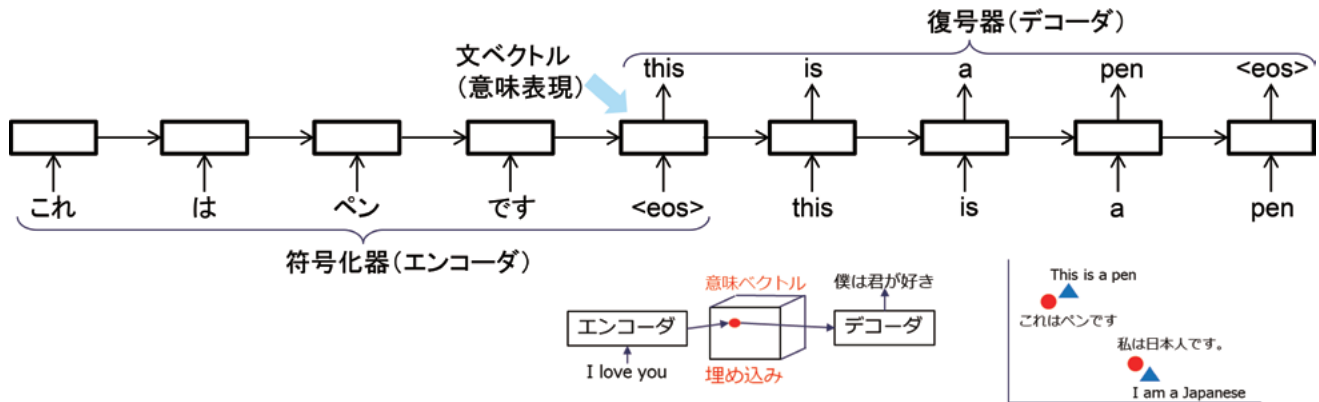


図2 ニューラル機械翻訳 (RNN エンコーダデコーダモデル)

出願時に作成する特許明細書等の特許関連文書の翻訳においては、必要な権利を担保するために修飾・被修飾の関係の保持や適切な技術用語の選定が重要であり、一般的な翻訳で重視される流暢性のある程度犠牲にしてでも、論理的な正確さや厳密さが要求されます。こうした課題への対応を評価するために、三菱重工の日英の特許明細書（請求項含む）を評価データとして、この日英特許対訳文コーパスを用いて構築した特許専用翻訳エンジン、汎用翻訳エンジン、および既存の特許専用翻訳エンジンの3種類のAI（人工知能）自動翻訳エンジンによる翻訳結果について、三菱重工・NTT両社の知的財産部門の特許出願業務を行う社員によって翻訳品質の順位付けを行いました。評価結果として、平均順位はそれぞれ1.5位、2.0位、1.9位であり、この日英特許対訳文コーパスを用いて構築した特許専用翻訳エンジンの翻訳結果がもっとも高い順位でした。また、プロの翻訳家によって事前に作成された正解翻訳との類似性を用いた自動評価（100点満点）についても、汎用翻訳38.6点、既存特許専用翻訳44.0点であったのに対し、この特許専用翻訳は57.5点と大幅な向上がみられました。今後は、この日英特許対訳文コーパスを用いて構築した特許専用翻訳エンジンが、NTTグループ企業である株式会社みらい翻訳からサービス提供される予定です。

さて、翻訳精度はかなり高くなったとはいえ、特許請求項のような長文になると翻訳精度が下がる傾向にあります。そこで、私たちが考案した世界最高精度の単語対応の技術を用いて、長文の翻訳精度を上げることができないか、あるいは誤訳もしくは誤訳のおそれがある部分を自動的に指摘する仕組みをつくれなにかというのが直近の研究テーマです。また特許以外の分野での翻訳精度向上については、ここ1年くらいの間に急激に広まってきた、大規模言語モデル（LLM: Large Language Model）を活用することを考えています。

LLMを活用した翻訳はどのように行われるのでしょうか。

LLMを使った機械翻訳は、ひと言で例えると、英語で多くの事柄を学んだ多言語話者に、ある言語の文を別の言語に言い換えてもらうようなものです。LLMによる翻訳と従来のニューラル翻訳の違いを理解するためには、ニューラル翻訳からTransformerを経てLLMが登場するまでの経緯を知る必要があります。ニューラル翻訳では、文の意味を1000次元程度の実数ベクトルで表現し、ある言語の入力文を数値ベクトルに変換するエンコーダ（符号化器）と、この数値ベクトルを別の言語の出力文に変換するデコーダ（復号器）を連結して翻訳を実現します。初期のニューラル翻訳では、単語を1つずつ入力して内部状態を更新し、内部状態に基づいて単語を1つずつ出力する再帰ニューラルネットワーク（RNN）を使ってエンコーダとデコーダを実現していました（図2）。入力文の終わり（end of sentence）を表す特殊記号<eos>が入力された際のRNNの内部状態は、入力文のすべての単語の情報を持っているので、これを入力文の意味表現（意味ベクトル）とみなすことができます。たくさんの対訳データを使ってこのRNNエンコーダデコーダモデルを訓練すると、例えば、互いに翻訳になっている文の対である日本語の「これはペンです」と英語の“this is a pen”の意味ベクトルが距離的に近い位置に配置されるようになります。つまり文の類似度がベクトル空間の距離に対応するようになるのです。このように2つの言語が1つのベクトル空間を共有することで翻訳を実現するのがニューラル翻訳です。

RNNを用いたエンコーダとデコーダは、入力文を固定長ベクトルで表現するので長い文は表現能力が低い、単語を1つずつ入力し1つずつ出力するので並列化が難しいといった課題があります。そこで、注意機構（attention mechanism）と呼ばれる複数の対象から情報を選択的に収集する仕組みを用いて、エンコーダは入力文のすべての単語から情報を取捨選択して入力文の各単語

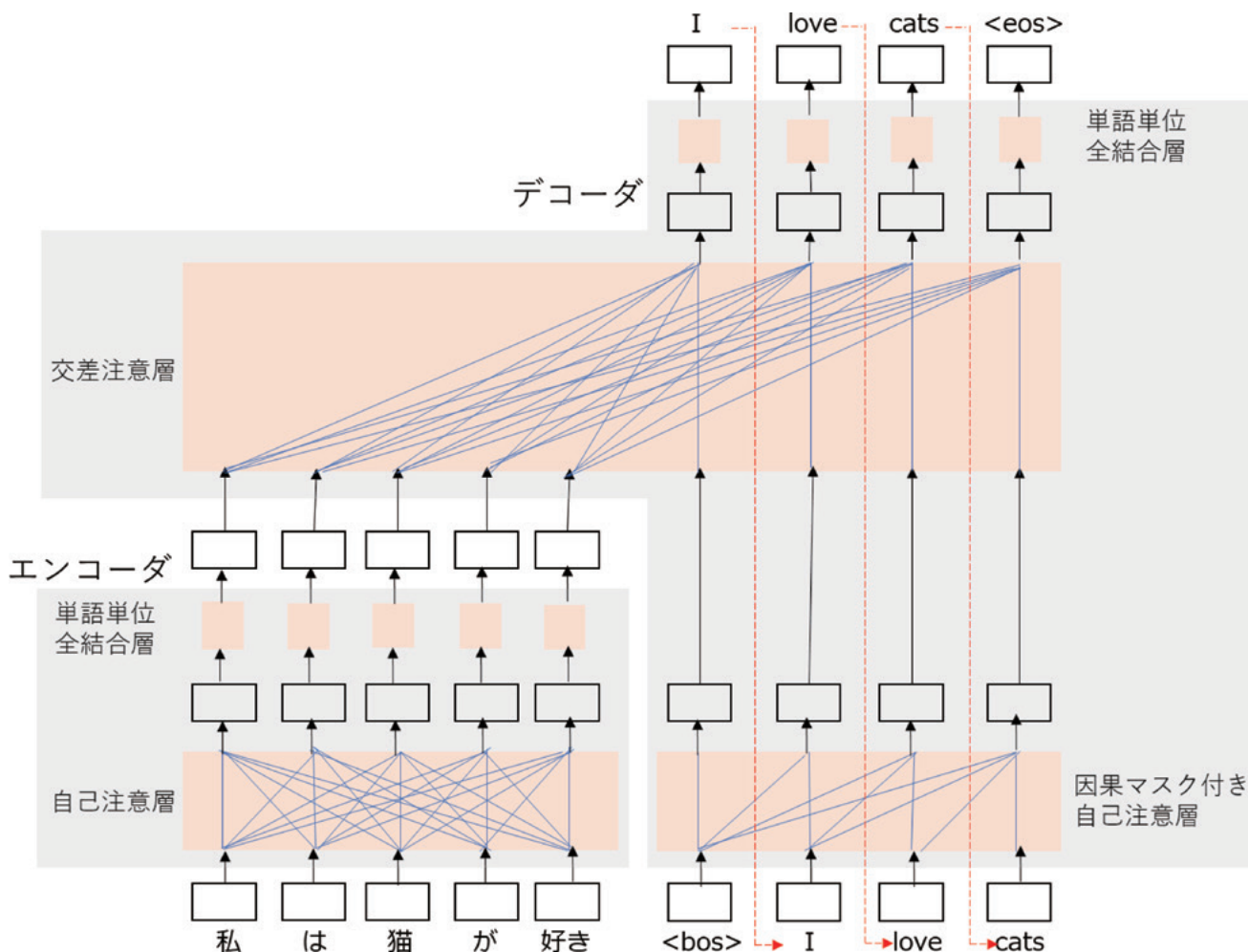


図3 ニューラル機械翻訳 (Transformer)

ベクトルを再構成し、デコーダは再構成された入力文の各単語ベクトルと直前までに出力したすべての単語ベクトルから情報を取捨選択して次に出力する単語を決める等の工夫を施した、Transformerというニューラルネットが考案されました (図3)。TransformerはRNNエンコーダデコーダに比べて精度が高いだけでなく、注意機構によって並列化が可能になったため、大量の対訳データを使って翻訳モデルを訓練したり、翻訳モデル自体を大規模化することが可能になりました。前述の日英特許対訳文コーパスもこのTransformerを利用して翻訳モデルを作成しています。

さて、生成AI、ChatGPTのベースとなっているGPT (Generative Pre-trained Transformer) -3というLLMが注目されています。これは、その名のとおりTransformerからデコーダだけを取り出し、モデルを大きくして大量のテキストデータで訓練したものです。Transformerデコーダだけのモデルは、直前までに出力したすべての単語ベクトルから注意機構を用いて情報を取捨選択して次に出力する単語を決めるニューラルネットワークであり、言語生成モデルまたは言語モデルと呼ばれます。GPT-3の場合は96層、175 Billion(B)パラメータのモデルを0.3 Tera(T)トークンのテキストで訓練しています。それ以前の翻訳モデルや言語モデ

ルに比べて約1000倍大きいのでLLMと呼ばれます。GPT-3をベースとして作成された対話システムChatGPTは日本語で違和感なく利用可能ですし、なぜか翻訳もできてしまいます。しかし、GPT-3の訓練データはほとんどが英語で、日本語は全体の0.2%程度しかありません。ChatGPTが英語以外の言語でも違和感なく使える理由まだ説明されていませんが、おそらくGPT-3の訓練データに含まれるすべての言語が、単語のような細かな粒度で1つの意味ベクトル空間を共有しているためと思われます (図4)。LLMによる翻訳は、例えば日英バイリンガルな人が日本語で聞いたことを反射的に英語で話しているようなものといえるでしょう。

GPTが急速に進化していますが、機械翻訳という観点から今後どのように変わっていくのでしょうか。

LLMによる機械翻訳について、現段階において翻訳精度は従来のニューラル翻訳と同等ですが、パラメータ数はニューラル翻訳の1000倍もあるため、明らかに処理は遅くなります。さらに、規模が大きいゆえに作成・運用が大変だという課題もあります。したがって、従来型のニューラル翻訳はなくなるのではないかと

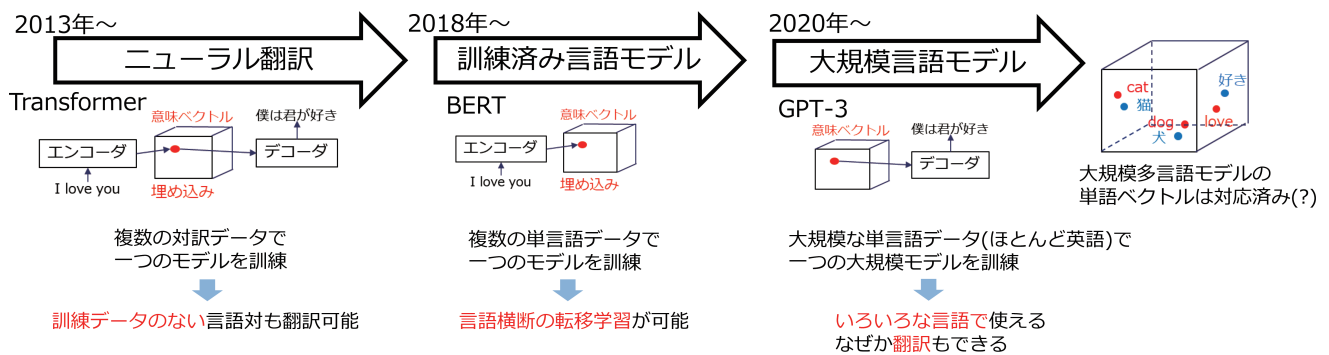


図4 言語モデルの多言語性 (multilinguality)

と予測します。とはいえ、現実としてChatGPTによる翻訳がユーザに支持されているのも事実です。これは、長い文脈の考慮、文体や用語の指定、入出力フォーマット変換等、何でもできるところと、それらを自然言語でコントロールできることにありと思っています。

こうした状況において、今後の機械翻訳研究としては、LLMを用いた機械翻訳の品質評価、誤りの検出と訂正、事実性の確認、ジェンダーバイアス等のバイアスの検出等が課題になると考えられます。さらに、現在の巨大で全能な知能をつくる流れは限界がありどこかで止まることが予想されるので、プロの翻訳者の使い方を反映した指示チューニング用データを作成し、機械翻訳とその周辺機能を自然言語で指示できるようカスタマイズされた、10B前後の小さめのLLMによる「翻訳アシスタント」を開発することが今後の課題になると考えています。

研究プロセスやアイデアは「出会い」だ。 試行錯誤で出会うチャンスを増やす

研究者として心掛けていることを教えてください。

これまでの経験の中で、あらかじめ落としどころを狙って進めてきた研究はうまくいかなかったことが多かったように感じます。試しにやってみたら新しい結果につながったようなこと、例えば、前述のLLMによる翻訳において、普通に対訳データでLLMをファインチューンしても上手くいかず、試行錯誤していく中で、上手くいくパターンが見つかりました。このようないいプロセスやいいアイデア等はある種の出会いなのではないかと思います。

そこで、「出会う」機会を多く持つために、共同研究を含めて人との出会い、そして試行錯誤を大切にしています。共同研究においては、極力重なる部分がないように、そして先入観を持たずにいろいろな試行錯誤を繰り返すのですが、各人が自分の思うことを勝手にやることもあり、逆にこれがいい結果につながることも

もあります。また、こうした「出会い」は、技術の変革期に多くあるのではないかと思います。変革期であるがゆえに、確立されたものがほとんどない状態なので、必然的に試行錯誤が多くなるからです。現在はLLMの登場でまさに変革期の中にあり、いいチャンスなので、翻訳の切り口から試行錯誤に注力していきたいと思っています。

後進の研究者へのメッセージをお願いします。

私は元々自然言語処理をテーマにしていたのですが、ここ20年くらいは自然言語処理の1つである翻訳に関するテーマに取り組んでいます。私は外国語を学ぶことが趣味なので、この分野は私が興味を持っている分野なのです。基礎研究は一般的に息の長い研究なので、興味を持ち続けることができれば、長期にわたって取り組むことはできません。また、長く同じ分野にいと、研究を取り巻く外的要因は時々刻々と変化していくので、それについていくための勉強もしなければならぬのですが、1本筋の通ったものを持っていることで、自分を見失うことなく前進することができます。

さて、今回は「人と違ったことをしていい意味で物議をかもし」ようなことを言いましたが、異なることを言っているわけではありません。研究者である以上、何か新しいことをやる、新しい知見を得る、新しい技術をつくることとは常に意識しなければならず、それは同じ分野の中であっても必要なことで、まさに人と違ったことをすることなのです。そうすると、新しいことに取り組むためには、今を知らなければならず、最先端の技術を勉強したうえで、試行錯誤をしなければなりません。その試行錯誤は1人でできないことが多いので、いろいろな人とさまざまな方向を試していく、それを行うためには議論が必要になってくるのです。こうした試行錯誤のプロセスを繰り返す先に解が見つかるのではないのでしょうか。