解説 02

機械翻訳技術の研究と動向

後藤功雄

国際社会における日本への理解を促進するために，外国語による情報発信の強化が必要とされている。また，外国人観光客の増加で，訪日旅行者向けの翻訳へのニーズも増えている。近年，機械翻訳の翻訳品質が向上し，さまざまな場面での活用が期待されている。本稿では，当所におけるこれまでの機械翻訳研究への取り組みを紹介するとともに，機械翻訳技術の研究動向を解説する。特に，この2,3年の間に急速に発展し，翻訳品質の向上に寄与しているニューラル機械翻訳の原理と動向を中心に，主な機械翻訳技術について解説する。

1．はじめに

グローバル化が進む中で，国際社会の日本への理解を促進するために，外国語による情報発信の重要性が増してきている。また，2020年に向けて，海外から日本への関心も高まっている。NHKは，日本の観光，文化，科学技術をはじめ，日本の魅力や姿を，世界に対して外国語で積極的に発信している[1)]。また，外国人観光客の増加に伴い，訪日旅行者向けの外国語でのサービスや情報提供，外国人とのコミュニケーションのニーズも増加している。このような外国語での情報発信を容易にし，言葉の壁を越えるための技術として機械翻訳技術がある。

機械翻訳技術を大きく分類すると，規則ベース方式とコーパス*1ベース方式に分けられる。規則ベース方式は，翻訳知識を規則として人手で作成し，それらを組み合わせて訳文を生成する方式であり，機械翻訳技術初期の1950年代から研究されたものである。パソコン用パッケージの翻訳ソフトとして現在市販されているものは主にこの方式である。

一方，コーパスベース方式は，大量の対訳文から翻訳知識を自動獲得して利用する方式であり，現在盛んに研究が進められている。この方式が提案された背景には，人手で作成した翻訳知識の規則ではさまざまな言語現象を網羅することが困難であるということがある。数十万文から数億文という大量の対訳文があれば，その中にさまざまな言語現象が含まれていることが期待できる。また，文レベルの文脈の情報も活用することができる。

コーパスベース方式は「用例翻訳」，「統計的機械翻訳（SMT：Statistical Machine Translation）」，「ニューラル機械翻訳（NMT：Neural Machine Translation）」の3つに分類できる。用例翻訳は，翻訳したい入力文と似た文を大量の対訳文から検索して，対訳文の中で入力文と一致しない部分に対応する翻訳先言語（目的言語）の表現を書き換える方式であり，定型的な表現の翻訳に適用できる。SMTとNMTはコーパスの統計情報から構築したモデルを用いる方式であり，NMTはモデルにニューラルネットワー

*1 テキストを集めてデータベース化した言語資料。

ク*2を用いるものである。SMTとNMTは適用範囲が広く，2000年頃からSMTが盛んに研究されてきたが，2014年にNMTの翻訳品質がSMTに追いついてからは，NMTが急速に発展して現在の機械翻訳の研究の中心になっている。

本稿では，２章でこれまでの当所における機械翻訳技術の研究への取り組みを紹介し，３章でSMTの原理を概説し，４章でNMTの原理と動向について解説する。

*2 脳内の神経回路網の処理の一部を計算機上のシミュレーションによって表現したもの。

2. 当所における機械翻訳技術の研究

海外との効率的な番組交換や放送の多様化への貢献を目的として，当所では1986年から放送における自動翻訳の可能性に関する検討を開始した[2)]。当時，日本では日英・英日の機械翻訳がコンピューターメーカーを中心に盛んに研究開発されていたが，それらの対象はマニュアルなどの分野を限定した書き言葉であり，話し言葉で分野が限定されていないニュースの翻訳にそのまま利用するのは困難であった。当所では1988年から，分野を限定しないシステムの実現を目指した英日機械翻訳の研究を本格的に開始した。研究の早急な立ち上げを図るために，規則ベース方式に基づいた国内ソフトメーカーの英日機械翻訳システムを導入し，それを核とした基本システムを，同メーカーとの共同研究により作成した[2)]。

この規則ベース方式の基本システムは，当時の主流となっていた構文トランスファー方式に基づく翻訳システムであった。構文トランスファー方式は，以下のような処理を行う。

①まず入力文を構文解析して構文構造を得る。

（例：“I（主語）saw（述語）a car（目的語）”）

②次に英語の構文構造を日本語の構文構造に変換して語順を並べ替える。

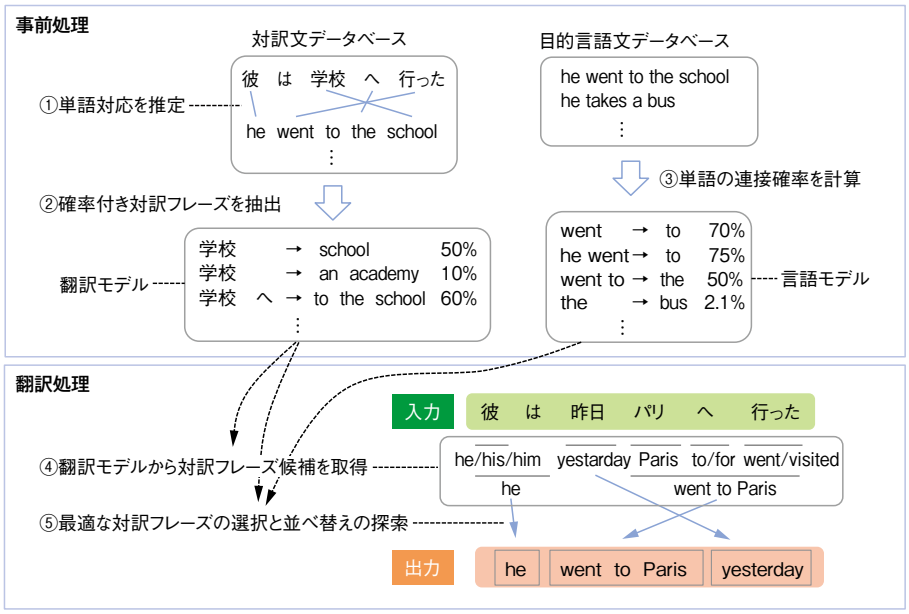
（例：“I（主語）a car（目的語）saw（述語）”）

③最後に各英単語の日本語の訳語を，辞書を使って出力する。

（例：“私は　一つの　車を　見た”）

これらの処理にはさまざまな規則を使うが，この規則を言語学の知見や人の知識に頼って人手で作成していた。基本システムは，約２万語の辞書と約３千の文法規則から成るものであったが，辞書については収録語数が根本的に不足しており，またニュースの文体を扱うための規則が未整備であった。このため，辞書と規則の拡充，整備を進めていった。言語の翻訳には，さまざまなあいまい性がある。前記の例でも“saw”は「ノコギリ」という名詞，「ノコギリで切る」という動詞，動詞“see”の過去形の可能性がある。また動詞“see”だとすると，これには「見える，会う，気づく…」と多くの日本語訳語がある。このようなあいまいな候補から正しい解を認定するには，単語の周りの文脈を考えた規則の精密化が求められる。これは極めて難しい作業であった。このような中で，1989年８月，開始間もない衛星放送における斬新な演出や翻訳システムの実用化の加速を狙って，英語ニュースの日本語テロップ*3の作成において，このシステムの試用が始められた[3)]。開始当初は，研究中のシステムを搭載した研究者用のワークステーションを翻訳現場に持ち込んだが，このシステムは日本語テロップ作成の多数の工程の中で翻訳だけしかできず，また操作が複雑であった。そのため，辞書や文法の拡充，アルゴリズムの研究に加えて，テロップの作成に適したシステムを開発して現場に導入した[4)]。この英日機械翻訳の研究開発は，1997年まで実施した。

*3 常時表示される字幕（オープンキャプション）。



1図　SMT の概略

英日機械翻訳の研究と並行して，国際放送のための翻訳を支援することを目的として1991年から日英機械翻訳の研究を開始した。規則ベース方式に加えてコーパスベース方式の１つである用例翻訳の研究開発を進め，2000年から規則ベース方式の研究開発は共同研究先の（株）国際電気通信基礎技術研究所（ATR：Advanced Telecommunications Research Institute International）に移行した。用例翻訳の研究においては，経済ニュースの対訳文データベースを構築し，データベースに含まれる対訳文と一致する定型的な経済ニュース文を翻訳できるようになった。

2017年から，NHKの国際放送局を中心とした外国語での情報発信の強化を目指したプロジェクトに当所も参加し，外国語による放送のための，言語間の機械翻訳の研究開発を開始した。現在当所では，ニュースの日英翻訳，および番組とニュースの英語から多言語への機械翻訳の研究開発に取り組んでいる。当所で開発しているシステムはNMTとSMTのハイブリッド方式であり，NMTを主として，NMTが不得意な部分や問題点をSMTで補っている。以下の章では，これらのSMTとNMTについて説明する。

3. 統計的機械翻訳（SMT）

SMTは1990年頃に提案された翻訳方式である[5]。初期の手法は，単語を単位として翻訳するモデルを用いており，文脈の情報を活用しにくいという課題があった。そこで，文脈の情報を活用しやすいように部分単語列（フレーズ）を単位として翻訳する，フレーズベースSMTが提案された[6]。現在のSMTは，このフレーズを単位として翻訳するフレーズベースSMTが主流である。

3．1　処理の流れ

SMTにはいくつかの手法があるが，ここでは代表的な手法であるフレーズベースSMT[6]を説明する。**1図**にSMTの概略を示す。SMTには，事前にモデルを構築する処理（事前処理）と入力文を翻訳する処理（翻訳処理）の２種類の処理がある。各処理

を１図中の番号を用いて説明する。

事前処理では，翻訳モデルと言語モデルを次のように構築する。

①共起*4などの統計情報を用いて，訓練データ*5の対訳文対における単語対応を推定する。例えば１図では「学校」と「school」を対応付けている。

②単語対応を用いて，対訳文対から対訳の部分単語列（フレーズ）を抽出し，抽出した対訳のフレーズから翻訳モデルを構築する。翻訳モデルは，入力側言語（原言語）のフレーズとそれに対応する出力側言語（目的言語）のフレーズ候補，およびその確率を保持する。確率は，訓練データ中の統計情報（出現頻度）から求める。

③訓練データの目的言語文から，言語モデルを構築する。言語モデルは，目的言語の単語の連接確率を保持する。連接確率も，訓練データ中の統計情報（出現頻度）から求める。

翻訳処理では，翻訳モデルと言語モデルを用いて次のように翻訳する。

④入力文中で，翻訳モデルに含まれるフレーズと一致するものを探索し，その対訳の目的言語のフレーズ候補を取得する。例えば１図では，「彼」のフレーズに対して「he」，「his」，「him」の３つが目的言語のフレーズ候補として取得され，「彼 は」のフレーズに対して「he」が目的言語のフレーズ候補として取得されている。

⑤目的言語のフレーズ候補を並べ替えてつなぎ合わせて得られる目的言語文の候補から，翻訳モデルの確率と言語モデルの確率を用いて計算するスコア*6が最大になるものを探索して出力する。このとき，入力文と過不足なく一致するフレーズを選択する。例えば１図では，「彼 は」，「昨日」，「パリ へ 行った」という３つのフレーズや，「彼 は」，「昨日」，「パリ」，「へ」，「行った」という５つのフレーズは入力文と過不足なく一致する。これらのフレーズの訳を並べ替えてつなぎ合わせたものが目的言語文の候補となる。

*4
ここでは，ある原言語単語とある目的言語単語が対訳文対に出現すること。

*5
入力と出力のペアの事例で，その出力を正解として機械学習での学習に用いるデータ。

*6
各対訳フレーズの翻訳モデルの確率の対数の和と，各目的言語単語の言語モデルの連接確率の対数の和を重み付けして加算した値。

３．２　SMTの特徴

訓練データの中に，入力文に似た文がたくさんあれば，翻訳処理で長いフレーズを活用できる可能性が高くなり，翻訳品質が高くなる。低頻度語でも，対訳の単語対応が正しく推定できれば，訳語選択の誤りは少ない。翻訳処理では，入力文中の各部分がすべて一度だけ翻訳されるように制御しているため，翻訳すべき内容が出力に含まれない「訳抜け」や，入力文中の同じ部分を２回以上訳出してしまう「訳出の繰り返し」はほとんど発生しない。

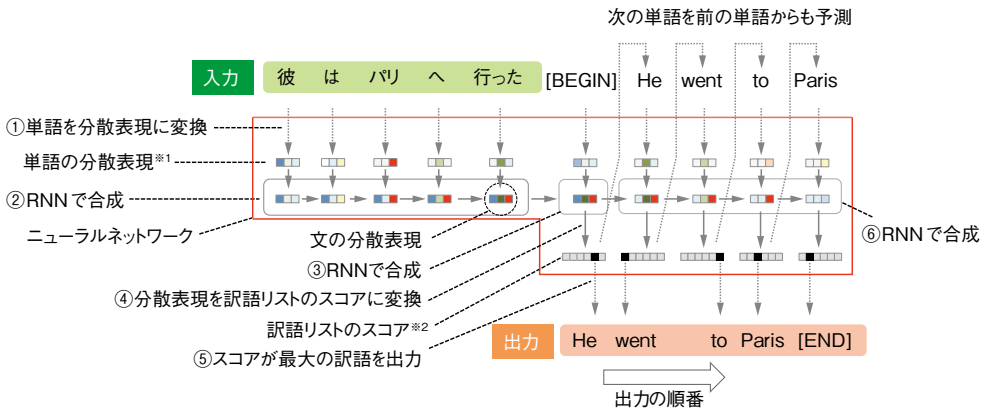
SMTの課題としては，次の点が挙げられる。翻訳モデルで活用できる文脈の情報は，フレーズ内の単語列の情報に限られる*7。言語モデルで活用できる文脈も，直前の数単語（通常は４単語程度）までに限られる。訳語選択は局所的な情報であいまい性を解消できることが多いが*8，語順の並べ替えはより長い文脈の情報を必要とする。このためフレーズベースSMTは，訳語選択の性能は高い一方で，語順の並べ替えの性能は高くない。

*7
例えば，１図の対訳フレーズ「へ → to」の確率は「へ」の周囲の語が何であるかを考慮しない。一方で，「パリ へ 行った → went to Paris」の確率は，「へ」の訳が「to」なのかどうかについて周囲の単語（「パリ」，「Paris」，「行った」，「went」）を考慮したものとなっている。しかし，この場合でも，フレーズの外側の文脈の情報（例えば「彼 は 昨日」）の情報は考慮していない。

*8
例えば，「bank」の訳が「銀行」なのか「土手」なのかは，「bank」の周囲の数単語の文脈からどちらなのか分かることが多い。

３．３　SMTの動向

フレーズベースSMTは，語順の並べ替えに課題がある。そのため，日英など語順が大きく異なる言語間の翻訳性能は低かった。そこで，任意のフレーズに置換可能な変数



※1　例として，3つの数値の並び（3次元）から成る分散表現を色で表している。
※2　訳語候補の各スコアを並べたイメージ。ここでは，黒が最も高いスコアを示す。

2図　基本的なNMTの翻訳処理

をフレーズ内に導入することで翻訳モデルに並べ替えルールを埋め込んだ手法[7]や，並べ替えの確率を計算するモデル[8)9)]，構文構造を用いた手法[10)11)]などが提案され，語順が大きく異なる言語間の翻訳の性能も向上した。

4. ニューラル機械翻訳（NMT）

現在のNMTの基本的な方式であるリカレントニューラルネットワーク*9（RNN：Recurrent Neural Network）を用いた翻訳方式は1997年に発表されている[12)13)]。しかし当時は，訓練データ量，演算能力，パラメーターの最適化の技術などが十分ではなかった。NMTがSMTに比べて同等以上の性能を達成し[14)]，機械翻訳の研究の中心となったのは2015年ごろからである。この基本的なNMTは長い文の翻訳に課題があり，この課題への対策としてアテンション機構と呼ばれる機能の導入が提案された[15)]。現在，NMTの手法は，このアテンション機構を備えたNMTが主流になっている。

本章では，まず基本的なNMTについて概説し，次にアテンション機構を備えたNMTについて説明する。さらに，NMTの特徴と動向を説明する。

4．1　基本的なNMTとその処理の流れ

NMTにおいてもSMTと同様に，事前にモデルを構築する処理（事前処理）と入力文を翻訳する処理（翻訳処理）の２種類の処理がある。本節では，はじめに**2図**に示す基本的なNMTの翻訳処理について説明し，次に事前処理について説明する。

翻訳処理では，まず入力文全体の情報を１つの分散表現*10にまとめ，次にその分散表現から訳語を推定して，文頭から１単語ずつ出力していくことで訳文を生成する。以下，２図中の番号を用いて処理の内容を順番に説明する。

①入力文の単語を分散表現と呼ばれる数値の並びに変換する。

②RNNを用いて入力文中のすべての単語の分散表現を合成し*11，文の分散表現を得る。

文の分散表現は入力文の情報を保持している。この合成は具体的には，次のように計算する。j番目の単語の分散表現を縦ベクトル$\boldsymbol{x}_j$とし，直前のRNNの出力の分散表現を縦ベクトル$\boldsymbol{h}_{j-1}$とし，重み行列$\boldsymbol{W}_{\mathrm{xh}}$と$\boldsymbol{W}_{\mathrm{hh}}$を用いて，$\boldsymbol{x}_j$をRNNに入力して合成した出力$\boldsymbol{h}_j$を

*9
ニューラルネットワークのうち，再帰構造を持つもの。再帰構造とは，時系列データをネットワークに順番に入力した際に，新たな入力と同時にその直前のネットワークで保持していた値も入力する構造である。

*10
単語や文などの特徴を数百～千程度の数値の並びで表現したもの。単語の場合は，主にその周囲の単語（文脈）がその単語の特徴として用いられる。似た意味の単語（例えば，“子供”と“少年”）は，似た分散表現になる。

*11
RNNでは，入力を繰り返すと以前に入力した情報が減衰してしまう。そこで実際には，入力を繰り返しても以前に入力した情報が減衰しにくいように機能を拡張したRNNであるLSTM（Long Short-Term Memory）[16)]またはGRU（Gated Recurrent Unit）[17)]を利用する。

$$\boldsymbol{h}_j=\tanh\left(\boldsymbol{W}_{\mathrm{xh}}\boldsymbol{x}_j+\boldsymbol{W}_{\mathrm{hh}}\boldsymbol{h}_{j-1}\right) \quad (1)$$

で計算する*12。ここでtanhはニューラルネットワークにおける活性化関数で，ベクトルの各要素に対して非線形*13な変換を行う関数である。重み行列$\boldsymbol{W}_{\mathrm{xh}}$と$\boldsymbol{W}_{\mathrm{hh}}$の値は，後で説明する事前処理であらかじめ設定しておく。

③訳出の開始記号（２図では［BEGIN］）の分散表現をRNNに入力し，文の分散表現と合成する。合成して得られた分散表現は，入力文の情報と，次に出力するのは文頭の語であるという情報を保持している。

④合成して得られた分散表現を，目的言語の語彙数*14と等しいサイズを持つ，訳語リストのスコアに変換する。この変換は具体的には，“訳語リストの次元数×分散表現の次元数”の重み行列と，分散表現の縦ベクトルとの積で計算する。

⑤訳語リストのスコアの中で最も高いスコアの訳語を出力する。

⑥出力した訳語の分散表現をRNNで合成する。これによって，何を出力したかをニューラルネットワークに入力する。合成して得られた分散表現は，入力文の情報とこれまでに出力した単語の系列の情報を保持している。これにより，次の単語を予測するときに，入力文の情報に加えて，それまでに出力した単語の系列の情報も使って予測することができる。

⑦さらに手順の④と⑤を実施して次の訳語を出力する。これら⑥，④，⑤の３つの手順を，文末を表す単語（［END］）が出力されるまで繰り返す。

次にNMTの事前処理について説明する。事前処理においては，訓練データの対訳文の入力側言語（原言語）の文を入力して，翻訳処理と同じ処理を行い，訳語のスコアを計算する。そして，対訳文中の目的言語単語を正解として，正解の単語のスコアが高くなるように，誤差逆伝播法*15によりニューラルネットワークで用いられているパラメーター*16を最適化する。パラメーターには最初にランダムな値を付与し，パラメーターを最適化することで，ネットワークが出力する訳語スコアが所望のスコアに近づく。このパラメーターの最適化がNMTの事前処理であり，モデルの学習とも呼ばれる。

４．２　アテンション機構を備えたNMTの処理の流れ

前節で説明した基本的なNMTでは，入力文の情報を１つの分散表現にまとめているため，文が長くなると，入力文中の情報を正しく保持することが難しくなるという課題がある。この課題への対策として，入力文の情報を１つの分散表現にまとめるのではなく，分散表現の系列として入力文の情報を保持し，訳語の系列を出力していく際に，入力文中で次に訳出すべき部分を推定して，その部分の情報を用いて訳語を選択する手法が提案された[15]。この手法が，アテンション機構を備えたNMTである。アテンション機構とは，次に訳出すべき入力文中の部分を推定するもので，語順の並べ替え機能になっている。

アテンション機構を備えたNMTにおいても，事前処理と翻訳処理がある。事前処理の内容は前節と同じである。本節では，**３図**に示すアテンション機構を備えたNMTの翻訳処理を説明する。前節の基本的なNMTと同様に，翻訳したい文を入力し，目的言語の単語を文頭から１単語ずつ出力していくことで訳文を生成する。

以下，３図中の番号を用いて翻訳処理の内容を順番に説明する（前節と同じ処理の説

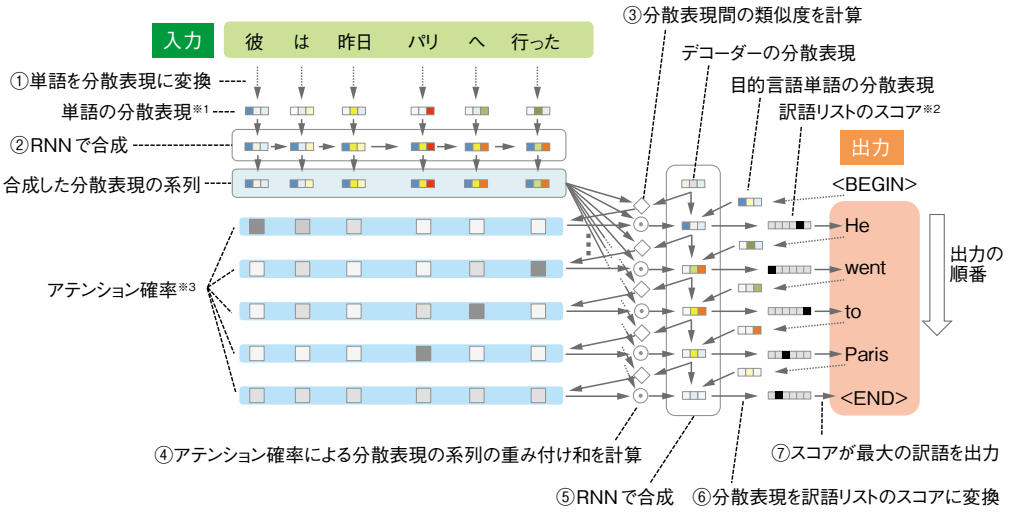
＊12
最初の$\boldsymbol{h}_1$の計算に必要な$\boldsymbol{h}_0$は，要素がすべてゼロのベクトルを用いればよい。

＊13
１次式（$ax+b$で表せる式）では表せない関係。

＊14
NMTで扱う語彙はあらかじめ決めておく。あらかじめ決めた語彙に含まれない語（未知語）は，未知語を表す語（［unk］）に変換して扱う。

＊15
ニューラルネットワークの出力と正解との誤差を，ネットワークの出力側から入力側へ伝播させることで，ネットワーク全体のパラメーターを効率的に更新するアルゴリズム。

＊16
パラメーターの具体的な内容は，次のとおりである。各原言語単語の分散表現，②③⑥のRNNでの合成に用いる重み行列（$\boldsymbol{W}_{\mathrm{xh}}$，$\boldsymbol{W}_{\mathrm{hh}}$），訳語リストのスコアへの変換に用いる重み行列，各目的言語単語の分散表現。



※1　例として，3つの数値の並び（3次元）から成る分散表現を色で表している。
※2　訳語候補の各スコアを並べたイメージ。ここでは，黒が最も高いスコアを示す。
※3　ここでは，色が濃いものが高い確率を示す。

3図　アテンション機構を備えたNMTの翻訳処理

明は簡略化する）。

①入力文の単語を分散表現に変換する。

②単語の組み合わせの情報も扱えるようにRNNを用いて分散表現の系列を合成する。合成して得られた分散表現の系列は入力文の情報を保持している。

③分散表現の系列において，次に訳出すべき箇所を推測する。具体的には，合成した分散表現の系列と，出力を生成する機構（デコーダー）の分散表現との類似度を計算し，アテンション確率を得る。ここでデコーダーの分散表現は，その時点までの翻訳で既に使用した入力の情報と，それまでに出力した目的言語単語列の情報を保持するものである。なお，最初はまだ何も出力していないことを表す初期値*17となる。また，類似度は次のようにして計算する。出力単語の位置をiとし，デコーダーのi番目の分散表現を縦ベクトル$\boldsymbol{s}_i$とすると，各入力文位置jに対応する$\boldsymbol{h}_j$（入力文の情報を保持する分散表現の系列中で位置jの分散表現）との類似スコアe_{ij}を

$$e_{ij}=\boldsymbol{v}^{\mathrm{T}}\tanh\left(\boldsymbol{W}_{\mathrm{se}}\boldsymbol{s}_{i-1}+\boldsymbol{W}_{\mathrm{he}}\boldsymbol{h}_j\right) \tag{2}$$

で計算する。ここで$\boldsymbol{v}$は重みの縦ベクトル，$\boldsymbol{W}_{\mathrm{se}}$，$\boldsymbol{W}_{\mathrm{he}}$は重み行列である。類似度すなわちアテンション確率$a_{ij}$は各$e_{ij}$を指数関数で変換して正規化した値として計算し，

$$a_{ij}=\exp(e_{ij})/\Sigma_k\exp(e_{ik}) \tag{3}$$

となる。アテンション確率は，次に翻訳する部分の予測，すなわち，語順の並べ替えの推定になっている。例えば，3図では，最初の訳語（He）を出力するための，最初のアテンション確率は「彼」に対応する部分の確率が高くなっている。これは最初の時点で入力文中の「彼」の部分を翻訳すべきであると推定したことになる。

④アテンション確率で重み付けして分散表現の系列の重み付け和を計算し，1つの分散表現を得る。得られた分散表現は，アテンション確率が高い位置の入力の情報を多く含んだものになっている。3図では，最初の訳語（He）を出力する直前に，この処

＊17
初期値の設定には，いくつかのやり方がある。文献[15]では，入力文の分散表現を変換した値に設定している。文献[15]の著者の実装DL4MT（https://github.com/nyu-dl/dl4mt-tutorial）では，入力文の分散表現の系列の平均値を変換した値に設定している。文献[15]の実装の1つであるKyoto-NMT（https://github.com/fabiencro/knmt）[18]では，初期値をパラメーターとして扱い，誤差逆伝播法により最適化している。

理で得る分散表現は「彼」の情報を多く含んでいる。

⑤直前に出力した目的言語単語の分散表現，④で得た分散表現，デコーダーの直前の分散表現の３つをRNNで合成する。

⑥合成して得られた分散表現を，目的言語の語彙数のサイズである訳語リストのスコアに変換する。

⑦訳語リストのスコアで最も高いスコアの訳語を出力する。

この③から⑦の手順を，文末を表す単語（[END]）が出力されるまで繰り返す。

４．３　NMTの特徴

翻訳したい文と分野が一致する訓練データの量が多い場合（例えば数百万文対以上），NMTはSMTより翻訳品質が高い傾向がある。そして流暢（りゅうちょう）な訳文を出力する傾向がある。また比較的短い文（例えば20単語程度まで）では，語順の並べ替え精度も高い。その理由として次のことが挙げられる。SMTの翻訳モデルがフレーズ内の文脈しか扱えないのに対して，NMTでは入力文全体の情報を文脈として扱える。また，SMTでは意味が似ていても異なる単語は統計量（出現頻度）を共有して利用することができないが，NMTでは分散表現を用いることで，単語が異なっていても似た意味の単語の統計量を共有して利用することができる。一般に訓練データ中の出現頻度が高いほどその統計量の信頼性は高くなるので，統計量を共有できるという特徴はNMTに有利であると考えられる。

NMTは出力の計算において，入力文全体の他に，出力の先頭から直前までの目的言語単語列もすべて活用できる。さらに，アテンション機構による語順の学習能力も，短い文であれば高い。

一方で，以下に述べるような課題もある。

NMTで高い翻訳品質を達成するためには，パラメーターの最適化に大量の訓練データを必要とする。訓練データの量が少ない場合（例えば数万文対程度），SMTも翻訳品質は低くなるが，NMTはパラメーターを適切に最適化することが難しく，SMTよりも翻訳品質が低くなる傾向がある。また，NMTは低頻度語の訳語選択の精度が低い。

NMTでは翻訳処理中に，入力文で訳された部分と訳されていない部分を区別しないため，翻訳すべき内容が出力に含まれない「訳抜け」や，入力文中の同じ部分を２回以上訳出してしまう「訳出の繰り返し」がしばしば発生する。また，短い文の翻訳品質は高いが，長い文の翻訳は難しい。

NMTは入力と出力のデータ対から学習し，アテンション確率以外の内部状態は，解釈が難しいブラックボックスになっている。そのため，ネットワークの内部に特定の目的のためのコントロール処理を導入することが難しい。

４．４　NMTの動向

訓練データ量が少ない場合の対策として，NMTで目的言語の文を原言語に翻訳することで対訳データを生成することや[19) 20)]，他の分野で多くの対訳データが存在する場合に，分野が一致する少量の対訳データに加えて他の分野の対訳データも利用することが試みられている[21) 22)]。この際，学習の最後に分野が一致する訓練データのみを用いたり[21)]，分野を表すタグ*18をデータに追加したりする[22)]工夫が行われている。また，訓練デー

*18
分野などの付加情報を表す文字列で，本文と区別がつく形式のもの。

タの文および低頻度語の出現回数を増やすために，訓練データの対訳文中の単語を低頻度語に置換して新たな対訳文を構築する研究もある[23]。さらに，翻訳したい言語対とは異なる言語対の対訳データを活用する試みもある[24]。

低頻度語の扱いに関しては，低頻度語は特別な語彙（[unk]）に置き換えて翻訳し，後処理で対訳辞書を用いてその単語を翻訳する方法がある[25]。そのほか，低頻度語もNMTで直接扱えるようにするために，単語を部分文字列に分割したり[26]，文字単位に分割したりする[27)~29)] ことで，処理の単位となる表現の種類数（文字単位であれば文字の種類数，単語単位であれば語彙数）を減らし，低頻度の表現を減らす取り組みがある。データ量が少ない場合の対策としても紹介したが，訓練データの対訳文中の単語を低頻度語へ置換して訓練データを増やすことにより，低頻度語の頻度を高めることができる[23]。

NMT特有の問題である訳抜けと訳出の繰り返しについては，これらを軽減するための分散表現の導入[30) 31)] や，訳抜けの軽減のために出力側の情報から入力文に逆翻訳する確率の利用[32) 33)] が提案されている。訳抜けした出力候補を逆翻訳して入力文を強制的に生成すると，そのときの入力文中の各単語の生成確率のうち，訳抜けしている内容を表す入力単語の生成確率（逆翻訳後の，その単語のスコア）が，訳抜けしていない出力候補を逆翻訳して入力文を生成した場合に比べて小さくなることを利用して，訳抜けの少ない出力候補を選択することができる。また，目的言語単語列に構文構造を表す開閉括弧を挿入することで，訳出の繰り返しが減ったことが確認されている[34]。

また，構文構造の利用は，長い文の語順並べ替えに有効であると考えられるが，本稿で概説したNMTは構文構造を利用していない。構文構造を括弧などで表現して原言語文または目的言語文に追加することで，NMTで構文構造を利用することができる[35) 36)]。構文構造を利用する他の方法も複数提案されており，例えば，入力文の句単位の分散表現の導入が試みられている[37]。

ブラックボックスとなっているNMTの動作の可視化・内部状態の解析についての研究もある。アテンション確率を見ることで，それぞれの出力単語への各入力単語の影響の強さは分かるが，それ以外に，既出力の各目的言語単語の影響の強さも計算する方法がある[38]。また，ネットワークの内部状態を他のタスク（単語の品詞推定など）に転用して，転用先のタスクでの精度（例えば品詞の推定精度）を測ることで，どのような情報が内部状態に保存されているかを調べる取り組みもある[39]。

5. おわりに

本稿では，当所におけるこれまでの機械翻訳研究への取り組みを紹介するとともに，この２，３年で急速に発展し，最近の翻訳品質の向上に寄与しているニューラル機械翻訳を中心に，コーパスベース方式の機械翻訳の原理と動向を解説した。コーパスベース方式の機械翻訳技術で高い翻訳品質を達成するために重要なことは，翻訳したいテキストと似た内容の対訳文を大量に訓練データとして利用することである。そのため，機械翻訳が幅広く社会で役立つようになるには，翻訳手法の改善とともに，さまざまな分野や言語対の対訳データの収集・構築，および効果的な構築方法の開発が重要な課題である。

外国語での情報発信の強化や，外国人観光客の増加により，さまざまな場面で機械翻

訳技術の活用が期待されている。今後，機械翻訳技術の導入・活用が進み，言葉の壁を越えた情報発信やコミュニケーションが容易になることが期待される。当所では，2020年に向けて高まる日本への関心に言葉の壁を越えて応えられるよう，機械翻訳の研究開発を進めていく。

参考文献

1) NHK経営計画 2015-2017年度，http://www.nhk.or.jp/pr/keiei/plan/index.html

2) 日本放送協会放送技術研究所：研究史'80 ～ '89（1991）

3) 相沢，浦谷，田中："放送ニュースへの機械翻訳システムの適用，" 信学技報，NLC-91-20, pp.31-38（1991）

4) 住吉，田中，畑田，江原："字幕作成のための翻訳ワークベンチ，" 信学技報，NLC-93-61, pp.53-58（1993）

5) P. E. Brown，S. A. D. Pietra，V. J. D. Pietra and R. L. Mercer：The Mathematics of Statistical Machine Translation: Parameter Estimation, Computational Linguistics，Vol.19, No.2（1993）

6) P. Koehn，F. J. Och and D. Marcu："Statistical Phrase-Based Translation，" Proc. HLT-NAACL（2003）

7) D. Chiang："A Hierarchical Phrase-Based Model for Statistical Machine Translation，" Proc. ACL（2005）

8) P. Koehn, A. Axelrod, A. B. Mayne, C. Callison-Burch, M. Osborne and D. Talbot："Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation，" Proc. IWSLT, pp.68-75（2005）

9) I. Goto，M. Utiyama，E. Sumita，A. Tamura and S. Kurohashi："Distortion Model Considering Rich Context for Statistical Machine Translation，" Proc. ACL（2013）

10) A. Zollmann and A. Venugopal："Syntax Augmented Machine Translation via Chart Parsing，" Proc. WMT（2006）

11) H. Isozaki，K. Sudoh，H. Tsukada and K. Duh："Head Finalization: a Simple Reordering Rule for SOV Languages，" Proc. WMT，pp.244-251（2010）

12) R. P. Neco and M. L. Forcada："Asynchronous Translations with Recurrent Neural Nets，" International Conference on Neural Networks，pp.2535-2540（1997）

13) A. Castano and F. Casacuberta："A Connectionist Approach to Machine Translation，" Proc. EUROSPEECH，pp.91-94（1997）

14) I. Sutskever，O. Vinyals and Q. V. Le："Sequence to Sequence Learning with Neural Networks，" Proc. NIPS，pp.3104-3112（2014）

15) D. Bahdanau，K. Cho and Y. Bengio："Neural Translation by Jointly Learning to Align and Translate，" Proc. ICLR（2015）

16) S. Hochreiter and J. Schmidhuber："Long Short-Term Memory，" Neural Computation, Vol.9，Issue 8，pp.1735-1780（1997）

17) J. Chung，C. Gulcehre，K. Cho and Y. Bengio："Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling，" Proc. NIPS 2014 Workshop on Deep Learning（2014）

18) F. Cromieres："Kyoto-NMT: a Neural Machine Translation Implementation in Chainer，" Proc. COLING，pp.307-311（2016）

19) R. Sennrich，B. Haddow and A. Birch："Improving Neural Machine Translation Models with Monolingual Data，" Proc. ACL，pp.86-96（2016）

20) D. He，Y. Xia，T. Qin，L. Wang，N. Yu，T. Liu and W.-Y. Ma："Dual Learning for Machine Translation，" Proc. NIPS，pp.820-828（2016）

21) M.-T. Luong and C. D. Manning："Stanford Neural Machine Translation Systems for Spoken Language Domain，" Proc. IWSLT（2015）

22) C. Chu, R. Dabre and S. Kurohashi："An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation，" Proc. ACL, pp.385-391（2017）

23) M. Fadaee，A. Bisazza and C. Monz："Data Augmentation for Low-Resource Neural Machine Translation，" Proc. ACL，pp.567-573（2017）

24) M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes and J. Dean：“Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation,” Transactions of the Association of Computational Linguistics – Vol.5, Issue 1, pp.339-351（2017）

25) T. Luong, I. Sutskever, Q. Le, O. Vinyals and W. Zaremba：“Addressing the Rare Word Problem in Neural Machine Translation,” Proc. ACL, pp.11-19（2015）

26) R. Sennrich, B. Haddow and A. Birch：“Neural Machine Translation of Rare Words with Subword Units,” Proc. ACL, pp.1715-1725（2016）

27) M.-T. Luong and C. D. Manning：“Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models,” Proc. ACL, pp.1054-1063（2016）

28) J. Chung, K. Cho and Y. Bengio：“A Character-level Decoder Without Explicit Segmentation for Neural Machine Translation,” Proc. ACL, pp.1693-1703（2016）

29) M. R. Costa-jussà and J. A. R. Fonollosa：“Character-based Neural Machine Translation,” Proc. ACL, pp.357-361（2016）

30) Z. Tu, Z. Lu, Y. Liu, X. Liu and H. Li：“Modeling Coverage for Neural Machine Translation,” Proc. ACL, pp.76-85（2016）

31) H. Mi, B. Sankaran, Z. Wang and A. Ittycheriah：“Coverage Embedding Models for Neural Machine Translation,” Proc. EMNLP, pp.955-960（2016）

32) Z. Tu, Y. Liu, L. Shang, X. Liu and H. Li：“Neural Machine Translation with Reconstruction,” Proc. AAAI（2017）

33) I. Goto and H. Tanaka：“Detecting Untranslated Content for Neural Machine Translation,” Proc. the First Workshop on Neural Machine Translation, pp.47-55（2017）

34) A. N. Le, A. Martinez, A. Yoshimoto and Y. Matsumoto：“Improving Sequence to Sequence Neural Machine Translation by Utilizing Syntactic Dependency Information,” Proc. IJCNLP, pp.21-29（2017）

35) J. Li, D. Xiong, Z. Tu, M. Zhu, M. Zhang and G. Zhou：“Modeling Source Syntax for Neural Machine Translation,” Proc. ACL, pp.688-697（2017）

36) R. Aharoni and Y. Goldberg：“Towards String-To-Tree Neural Machine Translation,” Proc. ACL, pp.132-140（2017）

37) A. Eriguchi, K. Hashimoto and Y. Tsuruoka：“Tree-to-Sequence Attentional Neural Machine Translation,” Proc. ACL, pp. 823-833（2016）

38) Y. Ding, Y. Liu, H. Luan and M. Sun：“Visualizing and Understanding Neural Machine Translation,” Proc. ACL, pp.1150-1159（2017）

39) Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad and J. Glass：“What Do Neural Machine Translation Models Learn About Morphology?,” Proc. ACL, pp.861-872（2017）

後藤（ごとう） 功雄（いさお）

1997年入局。仙台放送局を経て，1999年から放送技術研究所において，自然言語処理の研究に従事。2004年から2006年まで（株）国際電気通信基礎技術研究所（ATR）に出向。2008年から2013年まで（独）情報通信研究機構に出向。現在，放送技術研究所ヒューマンインターフェース研究部に所属。博士（情報学）。